



HAL
open science

Analyse comparative des tests de permutations en régression multiple et application à l'analyse de tableaux de distances.

Ali Shadrokh

► **To cite this version:**

Ali Shadrokh. Analyse comparative des tests de permutations en régression multiple et application à l'analyse de tableaux de distances.. Mathématiques [math]. Université Joseph-Fourier - Grenoble I; Université Pierre Mendès-France - Grenoble II, 2007. Français. NNT: . tel-00201481

HAL Id: tel-00201481

<https://theses.hal.science/tel-00201481>

Submitted on 30 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

Présentée par

Ali SHADROKH

Pour obtenir le grade de

Docteur de l'Université Joseph Fourier - Grenoble I

(Arrêtés ministériels du 5 Juillet et du Mars 1992)

**en mathématiques appliquées
Spécialité : Statistique**

Titre

**Analyse comparative des tests de permutations en
régression multiple et application à l'analyse de tableaux
de distances.**

Soutenue le 20 juillet 2006 devant la commission d'examen

Rapporteur	Farid Beninel
Rapporteur	Pierre Cazes
Examineur	Serge Dégerine
Examineur	Alain Latour
Directeur	Gérard d'Aubigny

Table des matières

1	Quelques méthodologies de régression linéaire simple	17
1.1	Introduction	17
1.2	Rappels et notations	18
1.2.1	Moments d'une loi bivariée	18
1.2.2	Moments empiriques d'une loi bivariée	20
1.3	Régression et modèle de prévision	21
1.3.1	La meilleure fonction de prévision	21
1.3.2	Meilleure prévision linéaire	23
1.4	Le modèle de régression linéaire simple	25
1.4.1	Le modèle de régression linéaire classique	25
1.4.2	Le modèle de régression linéaire gaussien	26
1.4.3	Estimation au sens du maximum de vraisemblance conditionnelle aux $\{x_i\}$	27
1.4.4	Estimation au sens des moindres carrés ordinaires	29
1.5	Inférence statistique en régression linéaire classique	31
1.5.1	Tests usuels portant sur le coefficient de corrélation	31
1.5.2	Tests d'hypothèse simple sur les paramètres de coefficient de régression	32
1.5.3	La relation entre tests du coefficient corrélation et du coefficient de régression	33
1.6	Inférence de plan en régression	34
1.6.1	Statistique approchée	34
1.6.2	Test de permutation	35
1.6.3	Les méthodes de Jackknife et de Bootstrap	52
1.7	La régression linéaire multiple	59
1.7.1	Le modèle classique de régression linéaire multiple	59
1.7.2	Le modèle de régression linéaire gaussien	61
1.7.3	Estimation au sens des moindres carrés ordinaires	64
1.8	Inférence statistique en régression linéaire multiple	65
1.8.1	Test de nullité d'un coefficient de régression linéaire	65
1.8.2	Test de nullité d'un coefficient de corrélation partielle	67

2	Résultat des simulations	77
2.1	Introduction	77
2.2	Description des méthodes de permutation	80
2.2.1	Permutation des résidus du modèle contraint : Freedman-Lane	80
2.2.2	Permutation des observations de la variable Y : méthode de Manly	83
2.2.3	Permutation des résidus du modèle général : ter Braak	84
2.2.4	Permutation des résidus du modèle contraint : Kennedy	85
2.3	Comparaison des méthodes	86
2.3.1	La méthode de Freedman et Lane	86
2.3.2	La méthode de Manly	88
2.3.3	La méthode de ter Braak	89
2.3.4	Méthode de Kennedy	89
2.4	Simulation d'Anderson et Legendre	90
2.4.1	Analyse du risque d'erreur de première espèce du test	90
2.4.2	Analyse de la puissance du test	93
2.4.3	Analyse d'effet d'un point extrême	95
2.5	Notre simulation	99
2.5.1	Analyse du risque d'erreur de première espèce du test	99
2.5.2	Analyse de la puissance du test	113
2.5.3	Étude la puissance de la méthode de Kennedy	122
2.5.4	Effet d'un point extrême	133
2.5.5	Étude d'effet d'un point extrême pour la méthode de Kennedy	141
3	Résultats théoriques	157
3.1	Introduction	157
3.2	Estimateurs MCO de $\beta_{2,\pi}^F$ et $\beta_{2,\pi}^K$	158
3.2.1	Cas du modèle de Freedman et Lane	158
3.2.2	Cas du modèle de Kennedy	162
3.3	Estimateurs MCO de la Variance de $\hat{\beta}_{2,\pi}$	164
3.3.1	Cas du modèle de Freedman et Lane	164
3.3.2	Cas du modèle de Kennedy	166
3.4	Étude comparative des deux méthodes	170
3.4.1	Estimation de $\text{Var}(\hat{\beta}_{2,\pi})$	171
3.4.2	Comparaison des estimateurs de $\text{Var}(\hat{\beta}_{2,\pi})$	174
3.4.3	Méthode de Kennedy et degrés de liberté	179
3.4.4	Influence de la normalisation	184
3.5	Espérance permutationnelle des variables	187
3.5.1	Cas univarié	187
3.5.2	Cas bivarié	188
3.6	Moments permutationnels de $\hat{\beta}_{2,\pi}$	190
3.6.1	Espérance permutationnelle de $\hat{\beta}_{2,\pi}$	191
3.6.2	La Variance permutationnelles de $\hat{\beta}_{2,\pi}$	192

3.7	Espérances permutacionnelle de $\widehat{\text{Var}}(\hat{\beta}_{2,\pi})$	195
3.7.1	Cas de la méthode de Freedman-Lane	196
3.7.2	Cas de la méthode de Kennedy	199
3.8	Étude comparative pour p variables explicatives	203
3.8.1	Cas de la méthode de Freedman-Lane	205
3.8.2	Cas de la méthode de Kennedy	209
3.8.3	Influence de la normalisation	214
4	Tests d'association entre matrices de distance	223
4.1	Introduction	223
4.2	Le test de Mantel	224
4.2.1	Simulation de Legendre	225
4.3	Notre simulation	227
4.3.1	Analyse du risque d'erreur de première espèce du test	228
4.3.2	Analyse de la puissance du test	244

Introduction

L'objet premier du travail présenté dans cette thèse était l'étude des dépendances partielles entre tableaux de distances, ou plus généralement de tableaux de dissimilarités.

La situation qui est au centre de cette étude est caractérisée par la donnée de trois matrices de distances : \mathbf{D}^1 , \mathbf{D}^2 et \mathbf{D}^3 , observées sur un même échantillon de n objets et la question posée consiste à évaluer le degré de liaison entre \mathbf{D}^1 et \mathbf{D}^2 sachant \mathbf{D}^3 et d'autre part, à tester la significativité statistique de cette liaison partielle.

Cette question peut s'interpréter en termes de régression (linéaire). Smouse *et al.* [53] proposèrent une extension du test de Mantel [39] à cette question. Ils élaborent un test de corrélation à la Mantel, reposant sur une stratégie permutacionnelle équivalente à la méthode proposée plus tard par Kennedy [33] dans le cas de données d'échantillon¹. Cette approche n'est pas la seule possible. Ainsi, Wolfe [62, 63] montre qu'on peut poser le problème en termes de mesure de la corrélation simple entre \mathbf{D}^1 et $(\mathbf{D}^2 - \mathbf{D}^3)$. Mais notre travail privilégie les approches directement exprimables en termes de régression multiple.

On est alors conduit à développer des outils d'inférence particularisés par trois spécificités :

S1 : L'interprétation des \mathbf{D}^j en termes de distances interpoints rend naturelle l'idée de spécifier un modèle de régression linéaire sans terme constant, du fait du sens particulier accordé au zéro et donc à la nullité d'une distance : $d_{il} = 0$;

S2 : Il est peu pertinent de traiter la loi de répartition de données distancielles dans un cadre gaussien, et il n'existe pas de modèle probabiliste paramétrique privilégié et simple pour traiter de tels problèmes d'inférence.

S3 : Les matrices de distances sont en général des constructions déduites de répartitions multivariées $\mathbf{D}^j = D(\mathbf{X}^j)$ qui interdisent de traiter leurs éléments $\{D_{kl}^j; k, l = 1, 2, \dots, n, k \neq l\}$ comme mutuellement indépendants. Il en résulte des comportements des estimateurs de la régression plus complexes que dans le cas

¹Par convention, nous appelons *données d'échantillon* les observations x_i attachées à chaque unité statistique appartenant à un échantillon. Par opposition, la donnée d_{il} attachée à une paire (i, l) d'individus de l'échantillon est dite *dyadique*.

usuel. Les problèmes d'inférence posés par l'analyse des liaisons entre matrices de dissimilarités ne relèvent donc pas du cadre classique d'application du modèle linéaire général sur données d'échantillon.

D'un point de vue méthodologique, l'analyse de la variance et la régression linéaire offrent les deux exemples les plus célèbres de cette classe de méthodes : elles sont fondées sur des modèles probabilistes ou mathématiques plus ou moins sophistiqués, censés intégrer les caractéristiques principales des données. Par suite, elles ne peuvent pas prétendre répondre de façon optimale à tous les problèmes d'inférence posés par l'utilisateur. Le domaine de validité de ces méthodes est limité par l'ensemble des postulats qui définissent le modèle. Par exemple la pertinence d'une analyse de la variance dépend de postulats d'additivité des effets et d'homogénéité des variances des erreurs. Le recours à un test de Student suppose de plus un tirage aléatoire simple garantissant des échantillons indépendants.

Le domaine d'application du test de Student est en fait plus large, parce que la loi théorique de cette statistique de test est réputée donner une approximation raisonnable de la loi réelle pour une classe de modèles probabilistes plus large que le modèle gaussien pour lequel elle a été établie. On qualifie de robuste une telle méthode, qui reste opérante pour plusieurs modèles, même si elle n'est optimale pour aucun ou seulement pour quelques uns.

Le test de Student est l'un des outils les plus souvent utilisés en régression linéaire gaussienne, pour éprouver la significativité de la corrélation partielle entre la variable réponse Y et l'un des prédicteurs présents dans le modèle. La grande diversité des contextes d'utilisation de ce modèle, en économie, en biologie, en météorologie, etc. confronte de façon courante les méthodes d'inférence statistique, telles que le test de Student à des situations ne respectant pas les postulats justifiant leur emploi.

Autrement dit, lorsque le processus de génération des données ne respecte pas les postulats fondant l'analyse statistique du modèle de régression linéaire classique, les statistiques de test d'hypothèse usuelles, portant par exemple sur les coefficients de régression partielle, ne suivent plus nécessairement la loi d'échantillonnage que leur assigne la théorie et en général on ne connaît plus la loi exacte de la statistique de test retenue. On doit alors recourir à d'autres méthodologies d'inférence statistique.

Lorsque la taille de l'échantillon est grande, on peut souvent approcher la loi d'échantillonnage de la statistique de test par une loi limite, par exemple une loi déduite de la loi de Gauss en application du théorème central limite. Il n'est alors pas nécessaire de connaître exactement la loi de répartition des résidus, ou de supposer *ab initio* qu'ils suivent une loi de Gauss. En pratique, cette approximation n'est pas toujours justifiée car on se confronte assez souvent à des situations où la taille d'échantillon est faible.

Lorsque la taille de l'échantillon est petite, les résultats asymptotiques ne sont plus utilisables et l'on doit recourir à des modes d'analyse orientés données, c'est-à-dire à des méthodes qui utilisent l'information contenue dans les données pour réduire sensiblement les postulats que doivent satisfaire les modèles potentiellement utiles. On recourt alors de façon privilégiée à la classe des méthodes de ré-

échantillonnage, bien qu'elles nécessitent une forte capacité de calculs. Cette classe inclue en particulier la méthode du Jackknife, présentée la première fois par Quénouille [51] et reprise par Tukey [60]). Une description plus théorique et plus générale de cette méthode est détaillée par Miller [43] et Efron [15]. Parmi les méthodes de ré-échantillonnage on peut citer aussi la méthode du Bootstrap présentée la première fois par Efron [14], et les méthodes permutationnelles.

Les tests de permutation sont au coeur de notre travail. Ils offrent une méthode non-paramétrique de test d'hypothèses. Si l'on trouve les premiers écrits sur ce sujet dans les travaux de Pearson (1900) sur l'analyse statistique des tables de contingence, la première application de cette méthode pour le modèle linéaire est en général attribuée à Fisher [18] et à Pitman [48, 49, 50]. Cette méthode comme le Bootstrap est fondée sur des postulats moins astreignants que l'approche paramétrique classique. Elle est en effet libre, au sens (distribution-free) où sa validité ne dépend pas de la loi de répartition des observations. Cette généralité est cependant obtenue au prix d'une contrepartie : sa mise en oeuvre nécessite des moyens de calcul et de simulation de données importants. Cette caractéristique explique d'ailleurs pourquoi elle n'a pas retenue l'attention des chercheurs avant l'apparition d'ordinateurs puissants.

Nous avons centré l'objectif principal de cette recherche plus particulièrement sur l'étude comparative des tests de permutation adaptés aux hypothèses de nullité d'un coefficient de régression partielle, dans le cadre de modèles linéaires en les paramètres.

Considérons tout d'abord un modèle de régression linéaire simple déterminé par les variables ordinaires Y , X^1 et ϵ , qui s'écrit :

$$Y = \beta_0 + \beta_1 x^1 + \epsilon \quad (1)$$

où la loi des résidus ϵ est inconnue. Nous avons ici une seule variable explicative (X^1) et dans ce cas, la méthode de test de permutation à retenir et présentée dans la suite, fait consensus. Elle porte sur le coefficient de régression (β_1) :

$$H_0 : \beta_1 = 0 \text{ contre } H_1 : \beta_1 \neq 0 \quad (2)$$

Dans ce cas, on dispose en effet d'une propriété de symétrie de la loi des erreurs sous l'hypothèse nulle, puisqu'alors les observations vérifient une propriété d'échangeabilité, imposée au terme erreur, qui permet de proposer un test exact (*cf.* [37] p. 248-253, [12] p. 195-216). Par contre, le problème devient difficile dès qu'interviennent au moins deux variables explicatives dans un modèle de régression linéaire multiple, car on perd les propriétés d'échangeabilité assurant l'existence d'un test de permutation exact. Plusieurs méthodes de tests de permutation approchés existent pour mettre à l'épreuve l'hypothèse de nullité d'un coefficient de régression partielle dans les modèles de régression multiple conditionnellement à l'observation d'un échantillon. Puisqu'on a perdu les propriétés d'échangeabilité, il existe plusieurs façons de les restaurer par transformation du problème. Chaque méthode repose sur des caractéristiques distinctes du modèle classique de régression linéaire multiple, conduisant

à la proposition de plusieurs logiques de test nées dans divers contextes d'application : Freedman et Lane [19]; Smouse *et al.* [53]; Oja, [47]; Collins, [6]; Gail, Tan and Piantadosi, [20]; Welch, [61]; ter Braak, [59]; Kennedy, [33]; Kennedy et Cade, [34]). Un domaine important d'utilisation de ces tests de permutation est l'analyse canonique des données multivariées dans les applications à l'écologie, la biologie et l'agriculture (ter Braak, [57, 58]; Legendre et Legendre [36]).

Notre mémoire de thèse considère souvent le modèle de régression linéaire double, pour simplifier la présentation :

$$Y = \beta_0 + \beta_1 x^1 + \beta_2 x^2 + \epsilon \quad (3)$$

Nous avons donc $p = 2$ variables explicatives dans le modèle, et l'objectif est de construire un test d'hypothèse portant sur le coefficient de régression partielle β_2 :

$$H_0 : \beta_2 = 0 \text{ contre } H_1 : \beta_2 \neq 0 \quad (4)$$

Nous montrons dans la suite à diverses reprises qu'une présentation des méthodes de test de permutation limitée au cas du modèle de régression linéaire double ne conduit pas à une perte substantielle de généralité et que l'extension des méthodes au cas d'une régression linéaire multiple avec $p > 2$ va en général de soi.

Dans le premier chapitre, nous rappelons tout d'abord les notions et les concepts fondamentaux. Ensuite, nous présentons le modèle classique de régression linéaire simple adapté au cadre gaussien et nous rappelons les postulats nécessaires pour valider l'inférence statistique. Ensuite nous donnons l'estimateur du maximum de vraisemblance (MV) et la statistique du test d'hypothèse usuelle portant sur le coefficient de régression simple. En particulier, nous montrons que le test d'hypothèse de nullité du coefficient de régression linéaire est identique au test de nullité du coefficient de corrélation linéaire. Ensuite, nous présentons les estimations de moindres carrés ordinaires (MCO) de façon à rappeler que dans un cadre gaussien les deux estimateurs MCO et MV sont identiques. L'approche reçoit alors une interprétation géométrique indépendante des hypothèses de loi de probabilité. On n'a en particulier pas besoin de l'hypothèse de loi de Gauss pour justifier le recours aux estimateurs MCO du modèle.

Le cas non gaussien, correspondant à des grands échantillons est alors traité en considérant la situation où une approximation gaussienne de la loi d'échantillonnage des estimateurs des paramètres de la régression se justifie, et nous en expliquons les termes, avant de nous intéresser au cas plus commun où l'emploi d'une telle approximation gaussienne n'est pas justifié, en particulier lorsque la taille n de l'échantillon est petite.

Nous rappelons ensuite le principe général qui fonde la méthode de test de permutation. Les postulats définissant le cadre de justification et d'interprétation des résultats sont présentés en nous appuyant sur le cas de la comparaison des moyennes de deux populations ainsi qu'à l'aide d'un exemple qui concerne le test de nullité du coefficient de régression dans le modèle linéaire simple.

Nous présentons ensuite les éléments d'analyse du modèle de régression linéaire multiple - en particulier le cas double : $p = 2$ - utiles à la présentation de notre travail ainsi que les postulats imposés dans le cas gaussien. Nous présentons en particulier l'estimateur du maximum de vraisemblance, la statistique de test usuelle pour tester la nullité d'un coefficient de régression partielle, dans le cas particulier $p = 2$. Nous montrons en particulier que dans le cas de la régression linéaire multiple en univers gaussien, réaliser le test d'hypothèse de nullité du coefficient de régression partielle équivaut à tester la nullité du coefficient de corrélation partielle.

Le chapitre suivant présente les différentes méthodes de test de permutation proposées dans la littérature pour l'hypothèse de nullité d'un coefficient de régression partielle. La discussion est le plus souvent limitée au cas de deux prédicteurs pour rester aussi proche que possible du cadre de simulations retenu. Il faut pourtant noter que l'essentiel des résultats présentés s'étendent sans difficulté au cas de deux groupes de prédicteurs X^k de dimension p_k , avec $k = 1, 2$ et $p_1 + p_2 = p \geq 2$, comme conséquence du célèbre théorème de Frisch et Waugh, *cf.* par exemple Greene [23]. On peut de même étendre ces résultats avec un peu plus de technicité, au cas de la régression multivariée.

Nous distinguons deux catégories d'approches :

- 1) La première regroupe les méthodes qui réalisent le test d'hypothèse (4), en permutant seulement les valeurs observées de la variable explicative X^2 . Cela revient à dire qu'on considère que les valeurs observées de la variable X^2 sont des unités échangeables sous l'hypothèse $H_0 : \beta_2 = 0$.
- 2) La deuxième regroupe les méthodes qui s'interprètent comme effectuant des permutations des valeurs prises sur l'échantillon par le terme d'erreur ϵ . Cela revient à considérer que les observations de ϵ sont échangeables.

La différence essentielle qui oppose ces deux catégories de méthodes tient au fait que le premier groupe (p.ex., Oja, [47]) ignore au cours des permutations, la corrélation qui existe en général entre les variables explicatives X^1 et X^2 . Certains auteurs, (*cf.* Anderson et Legendre [1]) tels que Welch, [61] ou ter Braak, [59] présentent cette caractéristique comme un désavantage. D'autres auteurs, en particulier Hall et Titterington [24] et Hall et Wilson [25] insistent sur l'importance qu'il y a à utiliser une statistique pivotale dans le contexte très voisin des tests par simulation et de ceux par le Bootstrap.

Nous distinguerons quatre approches parmi les méthodes de la première catégorie, toutes fondées sur des principes distincts : les approches de Freedman et Lane (1983), de Manly (1991), de ter Braak (1992) et celle de Kennedy (1995) respectivement. Du point de vue de la statistique, un test est jugé bon lorsque son risque d'erreur de première espèce est petit et lorsque conjointement sa puissance est élevée. Mais la plupart des comparaisons de tests de permutation relèvent de la simulation. Ainsi, en 1991, Manly [40] a réalisé une simulation limitée pour évaluer la qualité de sa méthode. : les résultats de cette simulation l'ont conduit à écrire que sa méthode fonctionnait bien. Pourtant, Kennedy [34] a critiqué la méthode de Manly et a également fait une simulation limitée qui montre que sa méthode est meilleure que celle

de Manly, au sens précédent. Kennedy démontre également que l'estimateur MCO de β_2 issu de sa méthode et l'estimateur MCO de β_2 a une expression analytique identique à celle qui est proposée par Freedman et Lane. Il en déduit que les deux méthodes étaient équivalentes. Notre étude analytique de ces deux méthodes montre pourtant que la situation est plus complexe. En 1992, ter Braak proposa une autre méthode de test qui calcule la valeur de la statistique de test associée à chaque permutation sous l'hypothèse alternative $H_1 : \beta_2 = b_2$. Dans ce chapitre nous comparons ces diverses approches et nous essayons également d'expliquer les avantages et les inconvénients de chacune. À notre connaissance, aucune comparaison n'a été faite avant 1999. Cette année là, Anderson et Legendre [1] ont étudié les quatre méthodes précédentes à l'aide d'une simulation large et riche, qui permet d'analyser les effets des facteurs de variabilité suivants sur le comportement du test :

- a) la taille de l'échantillon ;
- b) le degré de corrélation entre les variables explicatives $\rho(X^1, X^2)$;
- c) la valeur du paramètre β_1 associé à la covariable X^1 ;
- d) la loi de probabilité du terme d'erreur aléatoire ϵ_i .

Plus précisément, trois caractéristiques sont retenues pour évaluer l'influence de ces facteurs :

1. le risque d'erreur de première espèce
2. la puissance du test
3. l'effet d'un point extrême dans l'ensemble des valeurs données de la covariable X^1 , sur le risque d'erreur de première espèce.

Anderson et Legendre déduisent de leur simulation que la méthode de Freedman et Lane présente l'avantage sur les autres méthodes de conserver le niveau d'erreur de première espèce considéré. Le niveau de couverture du test varie très faiblement autour du niveau nominal d'erreur de première espèce. Ces auteurs ont également constaté que les trois méthodes de test Freedman-Lane, Manly et ter Braak ont des puissances sensiblement identiques. Ils ont de plus montré par simulation qu'en présence d'un point extrême introduit dans l'ensemble des valeurs de la covariable X^1 , la méthode de Freedman-Lane conservait mieux que les autres méthodes, le niveau du risque d'erreur de première espèce considéré

Nous avons repris l'ensemble de ces simulations avec une déclinaison plus systématique que celle d'Anderson et Legendre, et un choix plus réaliste de lois de répartition des variables explicatives. L'apport le plus important de ces simulations réside dans la mise en évidence du caractère erroné des affirmations de Kennedy lorsque cet auteur affirme (*cf.* Kennedy [33]) que sa méthode est équivalente à celle de Freedman et Lane. Nos simulations, présentées au chapitre 2, confirment les observations d'Anderson et Legendre : les deux méthodes sont approximativement sans biais, même sur des échantillons de petite taille. Mais ces simulations illustrent de plus le fait que l'erreur de première espèce encourue par la méthode de Kennedy est toujours supérieure ou égale à celle qui résulte de l'emploi du test de Freedman

et Lane. Ce constat nous a convaincu d'approfondir la question suivante : pourquoi ces méthodes, bien qu'utilisant le même estimateur du paramètre β_2 , sont-elles différentes en pratique ?

Nous avons complété notre étude par simulation des méthodes de Freedman et Lane et de Kennedy en veillant à respecter le cadre de travail d'Anderson et Legendre, mais en menant une étude plus complète et plus riche en information. Nous avons donc pu dégager des résultats non mis en lumière par Anderson et Legendre. D'une part, en autorisant un plus large spectre de valeurs possibles pour les prédicteurs, nous avons rencontré des situations où l'erreur de première espèce de la méthode de Kennedy était inférieure à celle de Freedman-Lane. D'autre part, les résultats de notre recherche confirment que la puissance du test fondé sur la méthode de Kennedy est plus grande que celle de Freedman et Lane. Les résultats de simulations obtenues dans ce chapitre nous ont convaincu de l'intérêt d'une étude théorique de ces méthodes de test de permutation et les résultats de ce travail sont présentés dans le chapitre suivant.

Dans ce troisième chapitre, nous comparons les propriétés formelles de ces deux méthodes de test de permutation. Nous démontrons en particulier que l'estimation MCO du coefficient de régression partielle β_2 associé à chaque permutation π des observations par la méthode de Freedman et Lane (notée $\hat{\beta}_{2,\pi}^F$) est égale à celle qui lui est associée par la méthode de Kennedy (notée $\hat{\beta}_{2,\pi}^K$). Si l'on note $\hat{\beta}_{2,\pi}$ cet estimateur commun, nous établissons les résultats complémentaires suivants :

- a) Notons $\text{Var}(\hat{\beta}_{2,\pi})$ la variance de $\hat{\beta}_{2,\pi}$ associée à une permutation quelconque, notée π , des éléments de l'échantillon observé. Notons de plus $S^2(\hat{\beta}_{2,\pi})_F$ l'estimateur de $\text{Var}(\hat{\beta}_{2,\pi})$ fourni par la méthode de Freedman et Lane, et $S^2(\hat{\beta}_{2,\pi})_K$ celui qui est obtenu par la méthode de Kennedy. Alors, $S^2(\hat{\beta}_{2,\pi})_F \leq S^2(\hat{\beta}_{2,\pi})_K$. Par suite, la valeur absolue de la statistique de test de Student associée à chaque permutation par la méthode de Freedman et Lane est toujours plus grande que celle qui lui est associée par la méthode de Kennedy, c'est-à-dire :

$$S^2(\hat{\beta}_{2,\pi})_K \geq S^2(\hat{\beta}_{2,\pi})_F \implies |T_{2,\pi}^K| = \frac{|\hat{\beta}_{2,\pi}|}{S(\hat{\beta}_{2,\pi})_K} \leq \frac{|\hat{\beta}_{2,\pi}|}{S(\hat{\beta}_{2,\pi})_F} = |T_{2,\pi}^F| \quad (5)$$

On en déduit que la fonction de répartition de $|T_{2,\pi}^F|$ est toujours supérieure ou égale à celle de $|T_{2,\pi}^K|$ et par suite, la fonction puissance du test permutationnel de Freedman et Lane est toujours supérieure ou égale à celle de la méthode de Kennedy. Sous l'hypothèse $H_0 : \beta_2 = 0$, ce résultat dit que l'erreur de première espèce de la méthode de Freedman et Lane est plus petite que celle de Kennedy. Inversement, sous l'hypothèse alternative $H_1 : \beta_2 \neq 0$, la puissance de la méthode de Kennedy est plus grande que celle de Freedman et Lane.

- b) Nous avons de plus calculé la variance permutationnelle de $\hat{\beta}_{2,\pi}$. Elle est notée $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ et quantifie la variabilité de $\hat{\beta}_{2,\pi}$ pour sa loi permutationnelle. Nous avons de plus adopté la notation \mathcal{E} , pour désigner la notion d'espérance

permutationnelle dans toute notre thèse.

Si nous désignons par $S^2(\hat{\beta}_{2,\pi})_{K^*}$ l'estimateur de la variance de $\hat{\beta}_{2,\pi}$, obtenu lorsque la somme des carrés des résidus est normalisée par la quantité $n - 2$ où n est le nombre observations. Nous montrons alors que :

$$\mathcal{E}(S^2(\hat{\beta}_{2,\pi})_F) = \mathcal{V}ar(\hat{\beta}_{2,\pi}).$$

Cette écriture signifie que l'estimateur de $\mathbb{V}ar(\hat{\beta}_{2,\pi})$ obtenu par la méthode de Freedman et Lane est un estimateur sans biais de la variance permutationnelle de $\hat{\beta}_2$, au sens de la loi permutationnelle.

c) On montre de façon analogue que :

$$\mathcal{E}(S^2(\hat{\beta}_{2,\pi})_K) \neq \mathcal{V}ar(\hat{\beta}_{2,\pi}),$$

c'est-à-dire que l'estimateur de $\mathbb{V}ar(\hat{\beta}_{2,\pi})$ obtenu par la méthode de Kennedy est biaisé pour la loi permutationnelle de $\hat{\beta}_2$.

d) Le choix du terme de normalisation est important ici, puisque contrairement au cas (a), nous avons aussi montré que l'estimateur de $\mathbb{V}ar(\hat{\beta}_{2,\pi})$ obtenu par la méthode de Kennedy défini en retenant maintenant $(n - 2)$ degrés de liberté et noté par $S^2(\hat{\beta}_{2,\pi})_{K^*}$, se comporte différemment de l'estimateur précédent normalisé en $(n - k)$ degrés de liberté : dans certaines situations, il est plus grand que celui de Freedman et Lane ($S^2(\hat{\beta}_{2,\pi})_{K^*} - S^2(\hat{\beta}_{2,\pi})_F > 0$) et dans d'autres situations, il est plus petit ($S^2(\hat{\beta}_{2,\pi})_{K^*} - S^2(\hat{\beta}_{2,\pi})_F < 0$).

Le choix de la normalisation de la statistique de test utilisée par Kennedy par $(n - 2)$ degrés de liberté au lieu des $(n - 3)$ degrés de liberté (ici, $k = 3$) retenus par cet auteur, modifie la situation traitée en (a) : i) on ne peut plus affirmer que le risque d'erreur de première espèce de la méthode de Freedman et Lane est plus petit que celui de Kennedy normalisé par $(n - 2)$ degrés de liberté ; ii) la puissance de la méthode de Freedman et Lane n'est plus systématiquement inférieure à celle de la méthode de Kennedy normalisée par $(n - 2)$ degrés de liberté. Il existe des cas où le risque d'erreur de première espèce de la méthode de Freedman et Lane est plus grand que celui de la méthode de Kennedy avec $(n - 2)$ degrés de liberté et la puissance de la méthode de Freedman et Lane est plus grande que celle de Kennedy avec $(n - 2)$ degrés de liberté. Néanmoins, les résultats de nos simulations montrent que l'erreur de première espèce résultant de la méthode de Freedman et Lane est plus petite que celle de Kennedy dans la plupart des cas et que la puissance de la méthode de Freedman et Lane est plus petite que celle de Kennedy dans la plupart des cas.

e) L'estimateur de $\mathbb{V}ar(\hat{\beta}_{2,\pi})$ dans la méthode de Kennedy corrigée en considérant $(n - 2)$ degrés de liberté est toujours plus petit que celui qui est obtenu par la méthode de Kennedy normalisé par $(n - 3)$ degrés de liberté.

$$S^2(\hat{\beta}_{2,\pi})_K \geq S^2(\hat{\beta}_{2,\pi})_{K^*}.$$

On en déduit que :

$$|T_{2,\pi}^K| = \frac{|\hat{\beta}_{2,\pi}|}{S(\hat{\beta}_{2,\pi})_K} \leq \frac{|\hat{\beta}_{2,\pi}|}{S(\hat{\beta}_{2,\pi})_{K^*}} = |T_{2,\pi}^{K^*}|.$$

La fonction de répartition de $|T_{2,\pi}^K|$ est donc toujours supérieure ou égale à celle de $|T_{2,\pi}^{K^*}|$. En conséquence, la fonction puissance du test permutational de Kennedy corrigé est toujours supérieure ou égale à celle de la méthode de Kennedy. Sous l'hypothèse $H_0 : \beta_2 = 0$, ce résultat dit que l'erreur de première espèce de la méthode de Kennedy est plus petite que celle de Kennedy corrigée. Inversement, sous l'hypothèse alternative $H_1 : \beta_2 \neq 0$, la puissance de la méthode de Kennedy corrigée est plus grande que du test original de Kennedy.

- f) L'estimateur de $\text{Var}(\beta_{2,\pi})$ obtenu par la méthode de Kennedy corrigée est un estimateur sans biais pour la variance permutational de $\hat{\beta}_2$, au sens de la loi permutational :

$$\mathcal{E}(\widehat{\text{Var}}(\hat{\beta}_{2,\pi})_{K^*}) = \text{Var}(\hat{\beta}_{2,\pi})$$

Nous étendons de plus ces résultats établis dans le cas du modèle de régression double, au cas du modèle de régression multiple.

Le chapitre suivant est consacré au cas de données dyadiques. Supposons par exemple qu'on étudie un ensemble de n individus (ou objets) où chaque individu est mesuré à l'aide d'un vecteur aléatoire de dimension p . Le choix d'une métrique sur l'espace engendré par ces p variables permet de résumer ces mesures multivariées en associant une distance à chaque paire d'individus (ou dyade) (i, l) . Il en résulte une matrice de distances \mathbf{D} , carrée d'ordre n , à éléments positifs, symétrique et à diagonale nulle, qui a pour élément (i, l) la distance entre les individus i et l :

$$\mathbf{D} = \begin{pmatrix} 0 & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{12} & 1 & d_{23} & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{1n} & d_{2n} & \cdots & d_{(n-1)n} & 1 \end{pmatrix}.$$

Par exemple, lorsqu'on calcule les distances entre chaque paire d'individus par la formule :

$$d_{ij} = \left(\sum_{k=1}^p (x_k^i - x_k^j)^2 \right)^{1/2} \quad (6)$$

on obtient une matrice de distances euclidiennes. Cette situation est particulière puisque nous avons supposé que l'on disposait des mesures initiales des p variables. Une situation plus courante en pratique correspond au cas où l'on mesure directement la dissimilarité d_{il} entre éléments de la dyade (i, l) . Les matrices de distances ont beaucoup d'applications en économie, en biologie et génomique, en écologie, en

analyse des réseaux sociaux, etc. Il existe des cas où les chercheurs veulent étudier les dépendances entre matrices de distances, par exemple à l'aide d'un modèle de régression sur matrices de distances. Le chapitre 4 est donc consacré à l'étude du comportement des quatre méthodes de test de permutation présentées au chapitre 2 dans le cas d'un modèle de régression linéaire sur matrices de distances. Les critères de comparaisons retenus sont ici encore :

1. le risque d'erreur de première espèce ;
2. la puissance du test.

Cette étude se justifie, du fait des spécificités de ce type de données et de la difficulté inhérente à la définition d'une notion utile de corrélation partielle entre matrices de distances. Ainsi, Legendre ([35]) a étudié par simulation l'application des quatre méthodes de tests de permutation présentés plus haut au cas de matrices de distances. Dans ce cas, les variables X^1 , X^2 et Y , définissant un modèle de régression, représentent chacun les éléments de la partie triangulaire inférieure d'une matrice de distances.

Ce chapitre explique tout d'abord brièvement la méthode de simulation adoptée et les résultats obtenus par Legendre : la statistique de test utilisée par cet auteur est le coefficient de corrélation partielle. Nous avons tenté de compléter son étude en retenant un plan de simulation plus large et plus riche, susceptible de lever certaines ambiguïtés de sa présentation. Puis nous comparons nos résultats à ceux obtenus dans la simulation effectuée par Legendre. Notre simulation est composée de deux parties :

1. Dans la première partie, les données générées pour les variables X^1 , X^2 vérifient les propriétés d'une matrice de distances euclidiennes alors que les données générées pour le terme d'erreur ϵ , ne vérifient pas nécessairement ces propriétés. Ensuite les données de la variable Y sont calculées sous la forme : $\mathbf{Y} = \mathbf{D}_R \boldsymbol{\beta} + \boldsymbol{\epsilon}$ où $\mathbf{D}_R = [\mathbf{X}^1 | \mathbf{X}^2] \mathbf{R}^{1/2}$, $\mathbf{R}^{1/2}$ est la matrice de Cholesky induite par la matrice de corrélation considérée entre X^1 et X^2 et notée \mathbf{R} , ${}^t\boldsymbol{\beta} = (\beta_0, \beta_1)$ et $\boldsymbol{\epsilon}$ est le terme erreur (ou bruit). Remarquons que les valeurs calculées pour Y comme les valeurs générées pour ϵ définissent une matrice de dissimilarités, mais ne vérifient pas nécessairement les propriétés d'une matrice de distances euclidiennes.
2. Les simulations présentées dans la deuxième partie de ce chapitre se veulent plus contraignantes et plus complètes que celles qui sont présentées dans la première, afin d'étudier les effets de transformations de distances euclidiennes sur le risque d'erreur de première espèce et sur la puissance du test. Nous avons donc généré des données de trois manières différentes, de façon à étudier l'influence de la répartition relative des grandes et petites distances sur le comportement du test.
 - I) les données générées pour les variables X^1 , X^2 vérifient les propriétés d'une matrice de distances euclidiennes ;

- II) les données générées pour les variables X^1, X^2 sont les valeurs obtenues en (I) élevées au carré.
- III) les données générées pour les variables X^1, X^2 sont les transformées logarithmiques des valeurs obtenues en (I), qui vérifient les propriétés d'une matrice de distances euclidiennes.

Dans les trois cas, les valeurs des variables Y et ϵ sont générées de façons à intégrer deux caractéristiques distinctes. Tout d'abord, nous avons imposé une structure de corrélation entre prédicteurs. Pour cela, nous avons formulé le modèle comme suit : $\mathbf{Y} = \mathbf{D}_R \boldsymbol{\beta} + \boldsymbol{\epsilon}$ où \mathbf{R} est la matrice de corrélation retenue entre X^1 et X^2 , on définit : $\mathbf{D}_R \boldsymbol{\beta} = [\mathbf{X}^1 | \mathbf{X}^2] \mathbf{R}^{1/2} \boldsymbol{\beta}$ comme le signal dans un modèle de type signal+bruit du modèle : $\mathbf{Y} = \mathbf{D}_R \boldsymbol{\beta} + \boldsymbol{\epsilon}$ où $\boldsymbol{\epsilon}$ est le terme erreur (ou bruit).

- a) les données générées pour ϵ vérifient les propriétés d'une matrice de distances euclidiennes telle que
 - 1) la variance des données générées pour le terme de signal (\mathbf{W}_R) est au moins quatre fois plus grande que la variance des données générées pour le terme de bruit (ϵ);
 - 2) les valeurs de Y qui sont calculées par : $\mathbf{Y} = \mathbf{D}_R \boldsymbol{\beta} + \boldsymbol{\epsilon}$ doivent vérifier les propriétés d'une matrice de distances euclidiennes.
- b) les données générées pour ϵ vérifient les propriétés d'une matrice de distances euclidiennes telle que
 - 1) la variance des données générées pour le terme de signal (\mathbf{D}_R) est au moins quatre fois plus petite que la variance des données générées pour le terme de bruit (ϵ);
 - 2) les valeurs de Y qui sont calculées par $\mathbf{Y} = \mathbf{D}_R \boldsymbol{\beta} + \boldsymbol{\epsilon}$ doivent vérifier les propriétés d'une matrice de distances.

Le choix des statistiques de test utilisées en régression linéaire est largement motivé par leurs propriétés d'optimalité démontrées dans le cas d'échantillons gaussiens. Notre étude n'échappe pas à cette pratique alors même que nous avons utilisé une méthode de test de permutation, reposant sur une approche nonparamétrique, donc non liée à un postulat de processus de génération de données gaussiennes. De plus, les caractéristiques probabilistes propres aux mesures de dissimilarités rendent moins naturelle la référence aux propriétés de la loi de Gauss pour l'analyse des dépendances entre matrices de dissimilarités. Nous avons donc souhaité évaluer l'influence du choix d'une statistique de test sur la performance comparée des méthodes de test de permutation discutées dans ce mémoire. Pour chaque permutation d'un ensemble de données générées et pour chaque méthode de test permutationnel, nous avons calculé trois statistiques dont l'équivalence en régression est démontrée dans le cadre classique : le rapport T , le $\hat{\rho}_{Y2,1}$ (utilisé par Legendre) et le $\hat{\rho}_{Y2,1}^2$.

Nous avons de plus complété cette comparaison en traitant le cas d'une hypothèse alternative unilatérale $H_0 : \beta_0 \leq 0$ contre $H_1 : \beta_2 > 0$ et celui d'une hypothèse bilatérale $H_0 : \beta_0 = 0$ contre $H_1 : \beta_2 \neq 0$. Pour 3000 ensembles de données gé-

nées et les deux types d'hypothèses alternatives, nous avons calculé les p -valeurs adaptées aux deux types de tests d'hypothèses, le risque d'erreur de première espèce (empirique) et la puissance (empirique).

Les erreurs de première espèce et les puissances obtenues par les méthodes sur les données simulées dans la première et la deuxième partie sont assez différentes. Nous allons en discuter les raisons, expliquer en détails dans le chapitre 5.

Nous présentons les résultats de ces simulations dans la deuxième partie. Nous montrons en particulier que l'étude des quatre méthodes de tests de permutation discutées plus haut conduit aux observations suivantes :

- a) lorsque la variance du terme de bruit est plus grande que celle du terme de signal, la puissance du test diminue fortement ;
- b) les trois statistiques de test habituellement utilisées : le T de Student, le coefficient de corrélation partielle et le carré du coefficient de corrélation partielle produisent approximativement le même risque d'erreur de premier espèce et la même puissance.
- c) la puissance du test obtenue sur des données générées dans le cas (II) (où les valeurs des variables X^1 et X^2 sont des carrés de distances interpoints) est plus grande que les puissances de tests obtenues dans les cas (I) (où les valeurs des variables X^1 et X^2 sont des distances interpoints) et (III) (où les valeurs des variables X^1 et X^2 sont les logarithmes des distances interpoints).

Chapitre 1

Quelques méthodologies de régression linéaire simple

1.1 Introduction

Dans de nombreuses applications du modèle linéaire, la connaissance du processus de génération des données ou bien une étude des résidus montre que l'utilisation des méthodes standard d'inférence statistique portant sur les paramètres du modèle est contestable ou même trompeuse, lorsque le terme d'erreur du modèle ne vérifie pas les postulats classiques justifiant les méthodes de test paramétrique pour des échantillons indépendants, identiquement distribués (i.i.d.) et gaussiens. Dans le cas i.i.d. non gaussien, une solution souvent adoptée consiste à recourir à une méthode non paramétrique. Une telle solution consiste à utiliser une méthode de ré-échantillonnage, par exemple les méthodes de Jackknife ou de Bootstrap. Une autre méthode non paramétrique est fournie par la classe des tests de permutation. Les premières descriptions des tests de permutation adaptés aux modèles linéaires furent présentées dans la première moitié du 20^{ème} siècle par Fisher(1953) et Pitman(1937a, b, 1938).

Dans ce chapitre, nous rappelons tout d'abord, les notions et les concepts fondamentaux de la régression linéaire simple. Ensuite, nous présentons le modèle de régression linéaire simple classique sous hypothèse gaussienne ainsi que les postulats nécessaires à la construction d'inférences statistiques. Puis, nous construisons l'estimateur du maximum de vraisemblance (MV) des paramètres et la statistique du test d'hypothèse portant sur le coefficient de régression simple sous hypothèse gaussienne. Nous montrons en particulier que le test d'hypothèse de nullité du coefficient de régression linéaire et celui de nullité du coefficient de corrélation linéaire sont identiques. Ensuite, nous reprenons la présentation des estimateurs des paramètres du modèle de régression simple au sens des moindres carrés ordinaires (MCO) sans référence à l'hypothèse gaussienne et nous rappelons leurs caractéristiques importantes. Puis, nous rappelons que les estimateurs MCO et MV sont identiques dans le cas gaussien.

Dans le cas non gaussien, on doit recourir à une méthode non paramétrique pour effectuer des inférences statistiques. Une méthode naturelle s'impose dans le cas de grands échantillons. C'est la méthode d'approximation gaussienne que nous exposons. Mais l'emploi de cette approximation gaussienne n'est pas toujours possible puisqu'elle suppose vérifiés les postulats du théorème central limite et il n'est pas justifié lorsque la taille n de l'échantillon est petite.

Puis, nous rappelons la méthode des tests de permutation, ses hypothèses justificatives et ses caractéristiques, en mentionnant l'exemple d'application de cette méthode à la comparaison des moyennes de deux populations ainsi qu'un exemple qui concerne le test de nullité du coefficient de régression dans le modèle de régression linéaire simple. Ensuite, nous présentons le modèle de régression linéaire multiple (en particulier double) et précisons les postulats justifiant la méthode classique de test d'hypothèse dans le cas gaussien. Notre étude concerne principalement la statistique de test utilisée pour réaliser le test de nullité du coefficient de régression partielle (en particulier dans le cas double). Ainsi, nous montrons que dans le cas d'une régression linéaire multiple, réaliser le test portant sur l'hypothèse de nullité du coefficient de régression partielle et le test portant sur le coefficient de corrélation partielle sont équivalents.

1.2 Rappels et notations

1.2.1 Moments d'une loi bivariée

Soit (Ω, \mathcal{F}, P) un espace probabilisé où Ω est un ensemble quelconque, \mathcal{F} est une tribu de parties de Ω et P est une mesure de probabilité définie sur \mathcal{F} . Une variable aléatoire réelle X est une fonction mesurable $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$ où \mathcal{B} est la tribu des boréliens de \mathbb{R} . On désigne par $x = X(\omega)$ la valeur prise par X en $\omega \in \Omega$.

définition 1.2.1. *Considérons la variable aléatoire réelle X et un entier positif k . Lorsqu'elle existe, la valeur :*

$$\mu_k = \mathbb{E}(X^k) \tag{1.1}$$

s'appelle moment simple d'ordre k . De même, si elle existe, la valeur :

$$\mu'_k = \mathbb{E}((X - \mu)^k) \tag{1.2}$$

s'appelle moment centré d'ordre k .

D'après (1.1) μ_1 est la moyenne de X que l'on note plus simplement μ ou μ_X s'il y a plusieurs variables aléatoires à distinguer. De plus, d'après (1.2) pour $k = 1$ on a $\mathbb{E}(X - \mu) = 0$, ce qui caractérise le centrage de X . Pour $k = 2$ on obtient la variance de X qui est notée :

$$\text{Var}(X) = \mathbb{E}(X - \mu)^2 = \sigma_X^2 \tag{1.3}$$

Dans la suite, nous considérons un couple de variables aléatoires X et Y réelles définies sur (Ω, \mathcal{F}, P) et de carré intégrable, i.e. $\mathbb{E}(X^2) < \infty$ et $\mathbb{E}(Y^2) < \infty$.

définition 1.2.2. *Considérons un couple de variables aléatoires réelles (X, Y) , lorsque p et q sont deux entiers positifs, la valeur (si elle existe) :*

$$\mu_{pq} = \mathbb{E}(X^p Y^q)$$

s'appelle moment croisé d'ordre p et q . De même, la valeur (si elle existe) :

$$\mu'_{pq} = \mathbb{E}((X - \mu_X)^p (Y - \mu_Y)^q)$$

s'appelle moment croisé centré d'ordre p et q .

Pour $p = q = 1$, on obtient la covariance entre deux variables X et Y , qui est donnée par :

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \quad (1.4)$$

Si $\text{Cov}(X, Y) = 0$ on dit que X et Y sont non corrélées et l'on note $X \perp\!\!\!\perp Y$.

Soient X et Y deux variables aléatoires définies sur (Ω, \mathcal{F}, P) et de carré intégrable. Alors, on vérifie classiquement les propriétés suivantes :

Propriété 1.2.1.

1. $\text{Cov}(X, X) = \text{Var}(X)$;
2. $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y), \forall a, b \in \mathfrak{R}$ (propriété de bilinéarité) ;
3. $\text{Cov}(X + a, Y) = \text{Cov}(X, Y), \forall a \in \mathfrak{R}$ (invariance par translation) ;
4. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ (symétrie) ;
5. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$;
6. Si X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$.

définition 1.2.3. *Le coefficient de corrélation de Pearson*

Soient X et Y deux variables aléatoires réelles définies sur (Ω, \mathcal{F}, P) , de carré intégrable et telles que $\sigma_X^2 > 0$ et $\sigma_Y^2 > 0$. Le coefficient de corrélation linéaire entre X et Y est alors donné par la formule suivante :

$$\rho(X, Y) \equiv \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}. \quad (1.5)$$

Propriété 1.2.2. *Sous les hypothèses de la définition (1.2.3),*

1. *Pour tous nombres réels $a > 0, c > 0, b$ et d , le coefficient de corrélation linéaire vérifie :*

$$\rho(aX + b, cY + d) = \rho(X, Y);$$

2. $\rho(X, Y)$ est un cosinus :

$$-1 \leq \rho(X, Y) \leq 1;$$

3. Si les variables X et Y vérifient la relation linéaire $aX + bY + c = 0$, où a , b et c sont des réels alors

$$\rho(X, Y) = \begin{cases} 1 & \text{si } ab < 0 \\ -1 & \text{si } ab > 0 \end{cases}$$

4. Si X et Y sont indépendantes, alors $\rho(X, Y) = 0$.

1.2.2 Moments empiriques d'une loi bivariée

Soit X_1, X_2, \dots, X_n , un échantillon de taille n , issu de la loi de X . La variable aléatoire $T_n = f(X_1, X_2, \dots, X_n)$, fonction de X_1, X_2, \dots, X_n , est une statistique d'échantillon. Les exemples les plus courants de statistiques d'intérêt sont les suivants :

définition 1.2.4. *Considérons un échantillon de taille n , de la variable aléatoire X . La statistique*

$$M_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

où k est un entier positif, s'appelle moment empirique d'ordre k .

Pour $k = 1$ on parle de moyenne empirique. Remarquons que M_K est une moyenne arithmétique et donc si la loi mère admet un moment μ_k d'ordre k , M_K est un estimateur sans biais de μ_k : $\mathbb{E}(M_k) = \mu_k$.

définition 1.2.5. *On appelle moment empirique centré d'ordre k de la variable aléatoire X , la statistique*

$$M'_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

où k est un entier positif.

Pour $k = 2$ on obtient la variance empirique :

$$M'_2 = \tilde{S}^2(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (1.6)$$

$\tilde{S}^2(X)$ est un estimateur biaisé de σ_X^2 lorsque le centrage est effectué par rapport à la moyenne de l'échantillon \bar{X} et non pas avec la moyenne théorique μ de sa loi. On notera dans la suite

$$S^2(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (1.7)$$

la variance corrigée de X . On démontre alors (cf. Dodge et Rousson [9]) que $S^2(X)$ est un estimateur sans biais de σ^2 .

Considérons maintenant un échantillon de taille n d'un couple de variables aléatoires (X, Y) . Les moments croisés empiriques d'ordres p et q et leurs correspondants centrés sont respectivement :

$$\frac{1}{n} \sum_{i=1}^n X_i^p Y_i^q \quad (1.8)$$

et

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^p (Y_i - \bar{Y})^q. \quad (1.9)$$

Dans le cas $p = q = 1$, le moment centré est appelé la covariance empirique :

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}). \quad (1.10)$$

Par analogie à la définition de la corrélation linéaire rappelée 'a l'équation (1.5.1), on peut définir la corrélation linéaire empirique en divisant la covariance empirique par le produit des écarts types empiriques des variables aléatoires X et Y .

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1.11)$$

1.3 Régression et modèle de prévision

1.3.1 La meilleure fonction de prévision

De nombreux modèles de prévision décrivent les valeurs prises par une variable aléatoire comme fonction des valeurs d'une ou de plusieurs autres variables aléatoires, appelée fonction de prévision ou encore fonction de régression.

Dans le cas simple où deux variables aléatoires X et Y sont indépendantes, la connaissance des valeurs prises par X n'apporte aucune information sur les variations de Y et la fonction de prévision est une fonction constante en X . À l'opposé, lorsque X et Y sont dépendantes, la connaissance d'une réalisation de X apporte une information sur Y et permet donc de construire une fonction de prévision.

Le problème de spécification de la meilleure fonction de prévision de Y sachant X se pose alors de la façon suivante : on considère un couple de variables aléatoires réelles (X, Y) défini sur un espace probabilisé (Ω, \mathcal{F}, P) :

$$(X, Y) : \Omega \longrightarrow \mathbb{R}^2$$

dont la loi de probabilité conjointe $f_{X,Y}(x, y)$ peut se calculer à l'aide de la densité de probabilité marginale de X : $f_X(x)$ et de la densité conditionnelle de Y sachant X : $f_{Y|X}(y|x)$. On note de même $f_Y(y)$: la densité de probabilité marginale de Y .

Prévoir Y à l'aide d'une réalisation x de X nécessite la connaissance d'une fonction de x notée $g(x)$ qui associe à tout x une prévision $g(x)$ de la valeur de Y . À toute prévision $g(x)$ de Y est de plus attachée une erreur de prévision, définie comme l'écart associé : $(Y - g(x))$

Une méthode simple et classique d'estimation de la fonction g , consiste à minimiser une fonction critère telle que l'erreur quadratique moyenne. Dans ce cas, $g(x)$ doit vérifier :

$$\mathbb{E}((Y - g(X))^2) = \min_{h(X) \in H_X} \mathbb{E}((Y - h(X))^2) \quad (1.12)$$

où H_X est l'espace de Hilbert formé par les fonctions réelles, définies sur l'espace l'image de X , noté $Im(X)$ et de carré intégrable :

$$H = \{h(x) | \mathbb{E}(h(X))^2 < \infty\}$$

et H_X est muni de la norme $\|X\| = \sqrt{\mathbb{E}(X)^2}$.

Théorème 1.3.1. *Soit (X, Y) un couple de variables aléatoires et $f(Y)$ est une fonction borélienne de Y telle que $\mathbb{E}(f(Y)) < \infty$, alors*

$$\mathbb{E}[f(Y)] = E_X[E_Y[f(Y)|X]]$$

En particulier :

$$E[Y] = E_X[E_Y[Y|X]]$$

Preuve : (cf. Mood et al. [45])

En utilisant le théorème 1.3.1 on écrit :

$$\mathbb{E}_{X,Y}(Y - g(X))^2 = \mathbb{E}_X[\mathbb{E}_{Y|X}(|Y - g(X)|^2)|X]$$

l'espérance conditionnelle de Y sachant $\{X = x\}$ réalise le minimum de l'erreur quadratique moyenne et donc :

$$g(x) = \mathbb{E}(Y|X = x). \quad (1.13)$$

Exemple 1.3.1. *Soit un couple (X, Y) de densité conjointe :*

$$f_{X,Y}(x, y) = \begin{cases} x + y, & \text{si } 0 < x, y < 1; \\ 0, & \text{sinon.} \end{cases}$$

Nous voulons trouver la meilleure fonction, c'est-à-dire :

$$\mathbb{E}(Y|X = x).$$

il faut donc trouver d'abord la densité marginale de X . Nous obtenons :

$$f_X(x) = \begin{cases} x + 1/2, & \text{si } 0 < x < 1; \\ 0, & \text{sinon.} \end{cases}$$

On en déduit immédiatement la densité conditionnelle de Y sachant X :

$$f_{Y|X}(y|x) = \begin{cases} \frac{2x+2y}{2x+1} & \text{si } 0 < x, y < 1; \\ 0 & \text{sinon.} \end{cases}$$

En utilisant la proposition (1.3.1) nous obtenons :

$$g(x) = \mathbb{E}(Y|X = x) = \frac{3x + 2}{6x + 3}$$

qui est une fonction non linéaire de x . Par exemple, si $x = \frac{1}{3}$, la valeur prévue pour Y est :

$$y = \frac{3\frac{1}{3} + 2}{6\frac{1}{3} + 3} = \frac{3}{5}$$

définition 1.3.1. Soient X et Y deux variables aléatoires t.q. $E(|Y|) < \infty$. La fonction $g(x) : \mathfrak{R} \rightarrow \mathfrak{R}$ définie par

$$g(x) = \mathbb{E}(Y|X = x) \tag{1.14}$$

s'appelle la fonction de régression de Y en X .

1.3.2 Meilleure prévision linéaire

La minimisation de l'erreur quadratique moyenne nécessite de connaître la densité conditionnelle de Y sachant $X = x$. On suppose souvent que celle-ci appartient à une classe paramétrée de fonctions. Le cas le plus simple est donné par une relation fonctionnelle entre X et Y de forme linéaire, la fonction de prévision s'écrit

$$g(x) = \beta_0 + \beta_1 x$$

Cette classe de fonctions est paramétrée par $\theta = \{\beta_0, \beta_1\} \in \Theta$, un ouvert de \mathbb{R}^2 . Dans ce cas, la fonction linéaire optimale au sens de l'erreur quadratique moyenne est connue lorsque les valeurs β_0 et β_1 qui réalisent le minimum de :

$$\mathbb{E}((Y - \beta_0 - \beta_1 X)^2 | X = x)$$

sont connues. On les note $\hat{\beta}_0$ et $\hat{\beta}_1$. En utilisant la formule de décomposition de la variance, on obtient pour la loi conjointe de (X, Y) :

$$\begin{aligned} \mathbb{E}(Y - \hat{\beta}_0 - \hat{\beta}_1 X)^2 &= \text{Var}(Y - \hat{\beta}_0 - \hat{\beta}_1 X) + \mathbb{E}^2(Y - \hat{\beta}_0 - \hat{\beta}_1 X) \\ &= \sigma_Y^2 + \hat{\beta}_1^2 \sigma_X^2 - 2\hat{\beta}_1 \sigma_{XY} + (\mu_Y - \hat{\beta}_0 - \hat{\beta}_1 \mu_X)^2 \\ &= \sigma_Y^2 + \sigma_X^2 \left(\hat{\beta}_1 - \frac{\sigma_{XY}}{\sigma_X^2}\right)^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} + (\mu_Y - \hat{\beta}_0 - \hat{\beta}_1 \mu_X)^2 \end{aligned} \tag{1.15}$$

de sorte que la solution est obtenue pour ;

$$\hat{\beta}_1 = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{\rho \sigma_X \sigma_Y}{\sigma_X^2} = \rho \frac{\sigma_Y}{\sigma_X} \tag{1.16}$$

$$\hat{\beta}_0 = \mu_Y - b\mu_X = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X \quad (1.17)$$

c'est-à-dire pour la fonction (linéaire en x) :

$$g(x) = \mu_X + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_x) = \beta_0 + \beta_1 x \quad (1.18)$$

On peut retrouver la fonction $g(x)$ figurant dans l'expression précédente par d'autres voies.

Proposition 1.3.1. *Soit (X, Y) un couple aléatoire de carré intégrable (non nécessairement gaussien), de moyenne, variances et corrélation linéaire notés respectivement $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ et ρ . Si $\mathbb{E}(Y|X = x)$ est une fonction linéaire de x , cette fonction est de la forme :*

$$\mathbb{E}(Y|X = x) = \mu_X + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_x) = \beta_0 + \beta_1 x$$

Preuve : Nous savons que :

$$\mathbb{E}(Y|X = x) = \beta_0 + \beta_1 x$$

Par suite :

$$\begin{aligned} \mu_Y &= \mathbb{E}(Y) = \mathbb{E}_X(\mathbb{E}(Y|X)) \\ &= \mathbb{E}_X(\beta_0 + \beta_1 X) \\ &= \beta_0 + \beta_1 \mu_X \end{aligned} \quad (1.19)$$

Opérons de façon analogue après avoir multipliés par x :

$$\begin{aligned} \mu_{XY} &= \mathbb{E}(XY) = \mathbb{E}_X(X\mathbb{E}(Y|X)) \\ &= \mathbb{E}_X(X(\beta_0 + \beta_1 X)) \\ &= \beta_0 \mu_X + \beta_1 \mathbb{E}(X^2) \end{aligned} \quad (1.20)$$

Les équations 1.19 et 1.20 ont pour conséquences :

$$\beta_0 = \mu_Y - \beta_1 \mu_X = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X \quad (1.21)$$

$$\beta_1 = \frac{\sigma_{XY}}{\sigma_X^2} = \rho \frac{\sigma_Y}{\sigma_X} \quad (1.22)$$

par suite :

$$g(x) = \mathbb{E}(Y|X = x) = \mu_X + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_x) \quad (1.23)$$

Remarque 1.3.1. *Lorsque les variables X et Y suivent une loi conjointe gaussienne, la loi conditionnelle de Y sachant X est une loi de Gauss et alors $\mathbb{E}(Y|X)$ est une forme linéaire et $\text{Var}(Y|X)$ ne dépend pas de X , justifiant ainsi le choix d'une fonction de régression linéaire $g(x)$.*

1.4 Le modèle de régression linéaire simple

Dans cette section nous rappelons le modèle de régression linéaire et les postulats nécessaires à la construction d'inférences statistiques relatives aux paramètres du modèle.

1.4.1 Le modèle de régression linéaire classique

Considérons un couple de variables (X, Y) défini sur (Ω, \mathcal{F}, P)

$$(X, Y) : (\Omega, \mathcal{F}, P) \longrightarrow (\mathcal{R}^2, \mathcal{B}_{\mathcal{R}^2})$$

où $\mathcal{B}_{\mathcal{R}^2}$ est la tribu des boréliens de \mathcal{R}^2 . Nous voulons prévoir les valeurs de la variable Y à l'aide d'une fonction linéaire de la réalisation de la variable X . La fonction de régression s'écrit sous cette contrainte :

$$g(x) = \mathbb{E}(Y|X = x) = \beta_0 + \beta_1 x,$$

de sorte que la valeur moyenne attendue de Y en x , notée $\mathbb{E}(Y|X)$, diffère de la valeur réelle Y par un terme d'erreur ϵ . On obtient ainsi le modèle classique de régression linéaire simple :

$$Y = \beta_0 + \beta_1 x + \epsilon \tag{1.24}$$

On parle alors de variable expliquée pour Y , de variable explicative ou prédictive pour X et d'erreur de prévision de Y ou résidu pour ϵ .

Quand on observe n répliquions indépendantes du modèle, on écrit :

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, n \tag{1.25}$$

et la loi de probabilité des ϵ_i décrit les erreurs du modèle retenu. La suite de cette thèse étudie les variations de Y conditionnellement aux valeurs prises par X . Les valeurs x_i observées définissent alors un plan d'expérience et peuvent être traitées comme des contraintes fixées par le plan et l'on retrouve ainsi le cadre classique d'analyse de la régression sur données expérimentales. Les méthodes classiques de test de permutation traitent sans modification du cas plus général et plus fréquent en sciences sociales où l'on traite des données de simple observation, caractérisées par l'impossibilité d'expérience, donc de maîtriser les valeurs de X . La méthodologie des tests de rééchantillonnage prend alors toute sa valeur du fait de sa capacité à intégrer d'éventuelles dépendances entre observations. Afin d'estimer les paramètres du modèle classique de régression linéaire simple et de préciser les propriétés d'optimalité statistique des estimateurs retenus, on doit formuler certains postulats concernant la distribution des erreurs. Nous commençons par donner trois postulats essentiels concernant cette loi de probabilité (*cf.* Hayashi, [26]). Puisque la loi de Y sachant $X = x$ se déduit de la loi conditionnelle de ϵ sachant $X = x$, on formule les postulats sur le couple (X, ϵ) plutôt que sur le couple (X, Y) .

P1. (Exogénéité forte) : L'espérance conditionnelle de ϵ_i sachant toutes les valeurs x_i prises par la variables explicatives X sur les éléments i de l'échantillon est nulle. Formellement, en notant \mathbf{X} l'ensemble des valeurs des prédicteurs prises sur l'ensemble de l'échantillon, et pas uniquement sur le i -ème individu,

$$\mathbb{E}(\epsilon_i|\mathbf{X}) = 0, \quad \forall i, i = 1, 2, \dots, n$$

P2. Propriétés d'homoscédasticité

$$\mathbb{E}(\epsilon_i^2|\mathbf{X}) = \sigma^2 > 0 \quad \forall i = 1, 2, \dots, n$$

P3. Propriétés de non corrélation des erreurs.

$$\mathbb{E}(\epsilon_i, \epsilon_j|\mathbf{X}) = 0 \quad i \neq j = 1, 2, \dots, n$$

On déduit les propriétés suivantes du postulat P1 pour tout $i = 1, 2, \dots, n$,

a) $\mathbb{E}(\epsilon_i) = \mathbb{E}[\mathbb{E}(\epsilon_i|\mathbf{X})] = 0$ ¹

b) $\mathbb{E}(Y_i|x_i) = \beta_0 + \beta_1 x_i$

et pour tout couple d'observations $i, j = 1, 2, \dots, n \quad i \neq j$:

c) $\mathbb{E}(X_j \epsilon_i) = 0$ ²

d) $Cov(X_j, \epsilon_i) = 0$ ³

On déduit aussi les propriétés suivantes des postulats P1 et P2 pour tout $i = 1, 2, \dots, n$,

a) $Var(\epsilon_i|\mathbf{X}) = \mathbb{E}(\epsilon_i^2|\mathbf{X}) - (\mathbb{E}(\epsilon_i|\mathbf{X}))^2 = \mathbb{E}(\epsilon_i^2|\mathbf{X}) = \sigma^2$

b) $Var(Y_i|\mathbf{X}) = Var(\epsilon_i|\mathbf{X}) = \sigma^2$

et pour tout couple $(i, j) \in \{1, 2, \dots, n\}^2 \quad i \neq j$

c) $Cov(\epsilon_i, \epsilon_j|\mathbf{X}) = \mathbb{E}(\epsilon_i \epsilon_j|\mathbf{X}) - \mathbb{E}(\epsilon_i|\mathbf{X})\mathbb{E}(\epsilon_j|\mathbf{X}) = \mathbb{E}(\epsilon_i \epsilon_j|\mathbf{X}) = 0$

d) $Cor(Y_i, Y_j|\mathbf{X}) = Cov(\epsilon_i, \epsilon_j|\mathbf{X}) = 0$

1.4.2 Le modèle de régression linéaire gaussien

Nous reprenons dans cette section le modèle 1.25 :

$$Y_i|\mathbf{X} = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, n$$

en supposant satisfait un quatrième postulat sur la loi de ϵ

P4. $\epsilon_i|\mathbf{X} \sim LG(0, \sigma^2) \quad \forall i = 0, 1, \dots, n$

¹ $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X))$

² $\mathbb{E}(X_j \epsilon_i) = \mathbb{E}[\mathbb{E}(X_j \epsilon_i|X_j)] = \mathbb{E}[X_j \mathbb{E}(\epsilon_i|X_j)] = 0$

³ $Cov(X_j, \epsilon_i) = \mathbb{E}(X_j \epsilon_i) - \mathbb{E}(X_j)\mathbb{E}(\epsilon_i) = \mathbb{E}(X_j \epsilon_i) = 0$

Nous supposons donc que les ϵ_i sont identiquement distribués, de loi gaussienne et indépendants conditionnellement à \mathbf{X} . Dans ce cas, $Y_i|X = x_i \sim LG(\beta_0 + \beta_1 x_i, \sigma^2)$ et l'on écrit pour tout i :

$$f_{Y|X}(y_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right\}$$

Les paramètres du modèle sont $\theta = (\beta_0, \beta_1, \sigma^2)$.

1.4.3 Estimation au sens du maximum de vraisemblance conditionnelle aux $\{x_i\}$

Les quatre postulats, $P1 - P4$ permettent de connaître la loi de probabilité conjointe de l'échantillon, c'est-à-dire la distribution des variables Y_1, Y_2, \dots, Y_n étant données les valeurs x_1, x_2, \dots, x_n , et donc d'estimer les paramètres $(\beta_0, \beta_1, \sigma^2)$ par la méthode du maximum de vraisemblance, qui consiste à retenir comme estimateurs des paramètres inconnus les valeurs de ces derniers qui maximisent la densité conjointe de l'échantillon observé.

Rappelons que dans le cas de la régression linéaire simple sur échantillon indépendant, on cherche alors à maximiser la fonction de vraisemblance :

$$L(\beta_0, \beta_1, \sigma^2; y_1, \dots, y_n) = \prod_{i=1}^n f(Y_i, \beta_0, \beta_1, \sigma^2) \quad (1.26)$$

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(Y_i - \beta_0 - \beta_1 x_i)^2\right\} \quad (1.27)$$

ou de façon équivalente la fonction de log-vraisemblance, définie comme :

$$\begin{aligned} l(\beta_0, \beta_1, \sigma^2; y_1, \dots, y_n) &= \ln L(\beta_0, \beta_1, \sigma^2; y_1, \dots, y_n) \\ &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2 \end{aligned} \quad (1.28)$$

où \ln désigne le logarithme népérien. En annulant les dérivées partielles on obtient :

$$\frac{\partial}{\partial \beta_0} l(\beta_0, \beta_1, \sigma^2; y_1, \dots, y_n) = \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i) = 0 \quad (1.29)$$

$$\frac{\partial}{\partial \beta_1} l(\beta_0, \beta_1, \sigma^2; y_1, \dots, y_n) = \frac{1}{2\sigma^2} \sum_{i=1}^n x_i (Y_i - \beta_0 - \beta_1 x_i) = 0 \quad (1.30)$$

$$\frac{\partial}{\partial \sigma^2} l(\beta_0, \beta_1, \sigma^2; y_1, \dots, y_n) = \frac{n}{2\sigma^2} - \frac{1}{2\sigma^4} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad (1.31)$$

et la résolution de ce système de trois équations à trois inconnues donne :

$$\hat{\beta}_1 = \frac{S(x, Y)}{S^2(x)} = \frac{\sum (Y_i - \bar{Y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i Y_i - n\bar{x}\bar{Y}}{\sum x_i^2 - n\bar{x}^2} \quad (1.32)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}, \quad (1.33)$$

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \quad (1.34)$$

On construit ainsi le modèle de prévision estimé :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (1.35)$$

Remarque 1.4.1. $(\hat{\beta}_0, \hat{\beta}_1)$ est solution des deux équations 1.32 et 1.33 indépendamment de la solution de 1.34, de sorte que cet estimateur, minimise aussi la somme des carrés des écarts :

$$Q = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$$

On retrouve ainsi la formulation du problème de régression au sens des MCO.

Remarque 1.4.2. Un calcul élémentaire (cf. Neter et al. [46]) montre que $\tilde{\sigma}^2$ est un estimateur biaisé de σ^2 . On obtient un estimateur sans biais en corrigeant son espérance par le coefficient $n/(n-2)$, qui conduit à la statistique :

$$\hat{\sigma}^2 = \frac{n}{n-2} \tilde{\sigma}^2 \quad (1.36)$$

$\hat{\sigma}^2$ est un estimateur sans biais, de variance $2\sigma^4/(n-2)$ (cf. Neter et al. [46]).

Les résultats suivants sont fondamentaux pour connaître la loi d'échantillonnage des estimateurs dans le cas gaussien.

Proposition 1.4.1. Sous l'hypothèse gaussienne (P.4) on a (cf. Neter et al. [46]) :

- a) $\hat{\beta}_0 \sim LG(\beta_0, \frac{\sigma^2}{n} \times \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2})$
- b) $\hat{\beta}_1 \sim LG(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2})$
- c) $\frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$
- d) $\hat{\sigma}^2$ est indépendant de \bar{Y} , $\hat{\beta}_0$ et $\hat{\beta}_1$

Remarque 1.4.3. Puisque $\hat{\sigma}^2$ est un estimateur sans biais pour σ^2 , alors :

$$S^2(\hat{\beta}_0) = \frac{\hat{\sigma}^2}{n} \times \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.37)$$

$$S^2(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.38)$$

sont des estimateurs sans biais respectivement pour $\text{Var}(\hat{\beta}_1)$ et $\text{Var}(\hat{\beta}_1)$.

1.4.4 Estimation au sens des moindres carrés ordinaires

La méthode des moindres carrés (MCO), due à Bernoulli, fut élaborée de façon indépendante par Gauss (1777-1855) et Legendre (1752-1833), en termes d'ajustement mathématique d'abord, puis dans un cadre d'estimation en univers gaussien, afin de comparer des données expérimentales, généralement entachées d'erreurs de mesure à un modèle mathématique censé décrire ces données. Son utilisation en régression a pour objet de construire une estimation des paramètres du modèle de régression optimale au sens où elle minimise la somme des carrés des résidus.

Cette formulation décrit un problème d'ajustement susceptible de justifications sans qu'il soit nécessaire d'imposer des contraintes de loi de probabilité aux erreurs ϵ pour construire une estimation. Rappelons cependant qu'avec l'ajout du postulat P4, nous avons montré la proposition suivante en section 1.4.3.

Proposition 1.4.2. *L'estimateur des moindres carrés ordinaires de $(\beta_0$ et $\beta_1)$ coïncide avec l'estimateur du maximum de vraisemblance de $(\beta_0$ et $\beta_1)$ dans le modèle de régression linéaire simple gaussien.*

Preuve : Afin de trouver les estimateurs des moindres carrés ordinaires des paramètres β_0 et β_1 , nous avons vu plus haut qu'il faut minimiser :

$$Q = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$$

par rapport à β_0 et β_1 . Par ailleurs, d'après la remarque 1.4.1, pour estimer (β_0, β_1) par la méthode du maximum de vraisemblance, on doit minimiser la même expression (par rapport à β_0 et β_1), donc l'estimateur MCO de (β_0, β_1) est $(\hat{\beta}_0, \hat{\beta}_1)$.

Propriétés des estimateurs MCO

Nous rappelons ici les propriétés des estimateurs MCO des paramètres d'une régression linéaire dans un cadre non gaussien. En effet, de nombreux utilisateurs continuent d'utiliser des estimateurs MCO, alors même qu'on doit s'interroger sur les propriétés d'optimalité qu'ils conservent dans un cadre aussi général, non contraint par un postulat de loi de Gauss.

A-Propriétés à horizon fini

Proposition 1.4.3. *Sous les postulats P1-P3 les estimateurs MCO de β_0 et β_1 sont des estimateurs sans biais et vérifient :*

$$\text{Var}(\hat{\beta}_0|\mathbf{X}) = \frac{\sigma^2}{n} \times \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.39)$$

$$\text{Var}(\hat{\beta}_1|\mathbf{X}) = \frac{\sigma^2}{n} \times \frac{n}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.40)$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1|\mathbf{X}) = -\frac{\sigma^2}{n} \times \frac{n\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.41)$$

$$\text{Cov}(\hat{\beta}_1, \bar{Y}|\mathbf{X}) = 0 \quad (1.42)$$

Greene [23]

Propriétés d'efficacité

La classe des estimateurs sans biais des paramètres d'une régression est en général non vide. Mais lorsqu'un paramètre admet plusieurs estimateurs sans biais, on doit définir un critère de comparaison pour déterminer un meilleur parmi eux. Un critère d'optimalité d'un estimateur souvent privilégié est la notion d'efficacité relative. Celle-ci interprète la variance d'un estimateur comme une mesure de sa précision. On compare donc les estimateurs sans biais d'un paramètre donné pour déterminer celui dont la précision est la meilleure, au sens où sa variance est la plus petite. On parle alors d'estimateur de variance minimum. La comparaison concerne donc la classe des estimateurs de carré intégrable dans laquelle on a, lorsque la borne de Fréchet-Cramer-Rao existe :

définition 1.4.1. *Lorsqu'on compare deux estimateurs sans biais d'un paramètre θ , on dit que T^* est plus efficace que T si $\text{Var}(T^*) < \text{Var}(T)$. Donc, dans la classe des estimateurs sans biais de θ de carré intégrable, un estimateur de variance minimale⁴ est efficace, s'il atteint la borne de Fréchet-Cramer-Rao.*

Théorème 1.4.1. (Gauss-Markov)

Sous les postulats P1-P3, parmi tous les estimateurs sans biais de (β_0, β_1) qui dépendent linéairement des observations de Y , les estimateurs MCO $(\hat{\beta}_0, \hat{\beta}_1)$ sont de variance minimale, On parle alors d'estimateur BLUE (Best Linear Unbiased Estimator).

B-Propriétés asymptotiques

De façon classique, les notations doivent être alourdis pour parler des propriétés asymptotiques d'un estimateur. Il faut considérer que l'on s'intéresse en réalité à une suite d'estimateurs $\{\beta^n; n \in \mathbb{N}\}$ construite en prenant des échantillons e_n de taille

⁴La précision obtenue pour T^* ne peut donc pas descendre en deçà d'une certaine limite, dite la borne de Fréchet-Cramer-Rao (cf. [38])

n croissante et à la limite de cette suite d'estimateurs quand n tend vers l'infini. Reprenons l'expression des moments d'ordre deux des estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$ donnés dans la proposition 1.4.3. On observe aussi que la condition $\ll (\sum_{i=1}^n x_i^2/n)$ admet une limite finie quand n tend vers l'infini \gg est suffisante pour assurer la nullité des termes suivants :

$$\mathbb{V}ar(\hat{\beta}_0^n | \mathbf{X}) = \frac{\sigma^2}{n} \times \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \xrightarrow{n \rightarrow \infty} 0 \quad (1.43)$$

$$\mathbb{V}ar(\hat{\beta}_1^n | \mathbf{X}) = \frac{\sigma^2}{n} \times \frac{n}{\sum_{i=1}^n (x_i - \bar{x})^2} \xrightarrow{n \rightarrow \infty} 0 \quad (1.44)$$

$$Cov(\hat{\beta}_0^n, \hat{\beta}_1^n | \mathbf{X}) = -\frac{\sigma^2}{n} \times \frac{n\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \xrightarrow{n \rightarrow \infty} 0 \quad (1.45)$$

On montre aussi que cette condition assure la convergence des estimateurs MCO respectifs de β_0 et β_1 , ainsi que leur non corrélation asymptotique, indépendamment de toute hypothèse de loi d'échantillonnage.

1.5 Inférence statistique en régression linéaire classique

1.5.1 Tests usuels portant sur le coefficient de corrélation

A) La loi d'échantillonnage de R

Même dans le cas d'un couple aléatoire gaussien, l'expression analytique de la loi d'échantillonnage de R est inconnue. Fisher [18] propose de contourner cette difficulté en utilisant la loi approchée de la statistique transformée :

$$Z = \frac{1}{2} \ln \frac{1+R}{1-R} = \arctan h(R) \quad (1.46)$$

Fisher [18] a en effet montré que lorsque $n \geq 25$, Z suit approximativement une loi de Gauss de moyenne et variance asymptotiques connues :

$$\mu_Z = \mathbb{E}(Z) \approx \frac{1}{2} \ln \frac{1+\rho}{1-\rho} = \arctan h(\rho) \quad (1.47)$$

$$\sigma_Z^2 = \mathbb{V}ar(Z) \approx \frac{1}{n-3} \quad (1.48)$$

La transformation de Fisher, de R en Z , conduit à une méthodologie de test simple, puisque Z est une fonction monotone de R de loi approchée

$$\frac{Z - \mu_Z}{\sigma_Z} \sim LG(0, 1). \quad (1.49)$$

On peut ainsi réaliser un test asymptotique ou construire un intervalle de confiance asymptotique pour ρ . Par exemple, le test d'hypothèse

$$H_0 : \rho = \rho_0 \quad \text{contre} \quad H_1 : \rho \neq \rho_0 \quad (1.50)$$

peut être mis en oeuvre à l'aide de la statistique de test :

$$T = \frac{\arctan h(r) - \arctan h(\rho_0)}{\frac{1}{\sqrt{n-3}}} \quad (1.51)$$

La consultation d'une table de la loi de Gauss centrée réduite conduit à rejeter H_0 si $|T| > z_{\frac{\alpha}{2}}$ où α est le risque de première espèce. De façon liée, T permet aussi de construire un intervalle de confiance pour ρ (*cf.* Dodge [9]).

B)- Étude du cas gaussien

Soit (X, Y) un couple de variables aléatoires réelles suivant une loi conjointe gaussienne et intéressons nous au cas où X et Y sont indépendantes, $\rho_0 = 0$. Dans ce cas, nous pouvons précisément calculer la densité de probabilité de la statistique R comme suit (*cf.* Hogg et Craig [28]) :

$$f_R(r) = \frac{\Gamma(\frac{n-1}{2})(1-r^2)^{\frac{n-4}{2}}}{\Gamma(\frac{1}{2})\Gamma(\frac{n-2}{2})}, \quad -1 < r < 1$$

Pour réaliser le test d'hypothèse $H_0 : \rho = \rho_0$ contre $H_1 : \rho \neq \rho_0$ dans ce cas particulier, on utilise comme statistique de test :

$$T = \frac{R \sqrt{n-2}}{\sqrt{1-R^2}} \sim t(n-2) \quad (1.52)$$

qui admet une loi d'échantillonnage connue, la loi de Student à $(n-2)$ degrés de liberté. Dit autrement, T est la statistique *pivotal* sur laquelle repose le test de Student classique. Comme Z , la statistique T est une transformation monotone croissante du coefficient de corrélation empirique R .

1.5.2 Tests d'hypothèse simple sur les paramètres de coefficient de régression

La proposition 1.4.1 montre de plus que le test d'hypothèse portant sur chacun des cas particuliers des paramètres β_0 et β_1 (tests de Student) et sur σ^2 (tests du khi-deux) sont des cas particuliers d'application de la théorie classique des tests d'hypothèse établis dans le cas d'une loi de Gauss. Nous nous intéressons plus particulièrement dans la suite aux généralisations du test d'hypothèse classique :

$$H_0 : \beta_1 = 0 \quad \text{contre} \quad H_1 : \beta_1 \neq 0 \quad (1.53)$$

utile pour vérifier si Y dépend significativement (de façon linéaire) de la variable explicative X . Considérons le cas d'un échantillon i.i.d. de taille n . Si le vecteur des résidus ϵ vérifie la condition $\epsilon \sim LG(0, \mathbf{I}_n \sigma^2)$, alors

$$\hat{\beta} = \mathbf{X}^* \mathbf{Y} = \mathbf{X}^* (\mathbf{X} \hat{\beta} + \epsilon) = \hat{\beta} + \mathbf{X}^* \epsilon \quad (1.54)$$

où ${}^t \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ et \mathbf{X}^* est l'adjointe de $\mathbf{X} = (\mathbf{1}, \mathbf{x})$ - où $\mathbf{1}$ est le vecteur constant de coordonnées 1 et \mathbf{x} est la matrice d'ordre $n \times p$ des valeurs des variables observées sur l'échantillon - définie comme $\mathbf{X}^* \equiv ({}^t \mathbf{X} \mathbf{X})^{-1} {}^t \mathbf{X}$. Donc, $\hat{\beta}$ est un vecteur gaussien, de sorte que lorsque σ^2 est connue sous $H_0 : \beta_1 = 0$, le rapport

$$Z = \frac{\hat{\beta}_1}{\frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}} \quad (1.55)$$

définit une variable centrée réduite, qui suit une loi de probabilité gaussienne $LG(0, 1)$. Le test de niveau α de hypothèse H_0 conduit à consulter une table de cette loi et à rejeter H_0 si $|Z| > z_{\alpha/2}$, où $z_{\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$.

Lorsque σ^2 est inconnue, on estime $\text{Var}(\hat{\beta}_1)$ par l'estimateur sans biais $S^2(\hat{\beta}_1)$ défini dans la remarque (1.4.3). Alors, sous $H_0 : \beta_1 = 0$, la statistique :

$$T = \hat{Z} = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)} \quad (1.56)$$

suit une loi de Student à $n - 2$ degrés de libertés. Le test de hypothèse H_0 de niveau α conduit à consulter une table de Student et à rejeter H_0 si $|Z| > z_{\alpha/2}$, où $z_{\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ d'une loi de Student à $n - 2$ degrés de liberté.

1.5.3 La relation entre tests du coefficient corrélation et du coefficient de régression

Dans le cas d'une régression simple en univers gaussien, on montre de plus que les tests de signification fondés sur le coefficient de corrélation linéaire, présentés en section(1.5.1), sont équivalents au test rappelé ci-dessus. En effet, si (X, Y) est un couple aléatoire défini sur (Ω, \mathcal{F}, P) de moyennes, variances et corrélation linéaire $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ et ρ , si de plus la fonction de régression est linéaire, alors :

$$g(x) = \mathbb{E}(Y|X = x) = \mu_X + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$$

et donc $\beta_1 = \rho \frac{\sigma_Y}{\sigma_X}$. Une relation analogue lie les estimateurs $\hat{\beta}_1$ et R :

$$\hat{\beta}_1 = R \frac{S(Y)}{S(X)} \quad (1.57)$$

où R est le coefficient de corrélation linéaire empirique, $S^2(X)$ et $S^2(Y)$ sont les variances échantillonnelles de X et Y . Lorsque σ_Y et σ_X sont strictement positifs, $\beta_1 = 0$ si et seulement si $\rho = 0$, de sorte que les hypothèses $H_0 : \beta_1 = 0$ et $H_0 : \rho = 0$ sont équivalentes.

1.6 Inférence de plan en régression

Cette section étudie le test d'hypothèse $H_0 : \beta_1 = 0$ contre $H_1 : \beta_1 \neq 0$, lorsque l'erreur ϵ ne vérifie plus le postulat P4, c'est-à-dire dans le cas non gaussien. La statistique $T = \hat{\beta}_1 / S(\hat{\beta}_1)$ ne suit plus nécessairement une loi de Student et on ne connaît pas sa loi exacte en général. On doit donc recourir à d'autres méthodologies d'inférence statistique.

1.6.1 Statistique approchée

Dans le cas général, non nécessairement gaussien, il est possible de trouver la loi de la statistique T sous l'hypothèse où les résidus ϵ_i sont i.i.d. et sous certaines hypothèses complémentaires (cf. Antoniadis et al. [2]) à condition de se placer dans le cas où n est suffisamment grand. Reprenons le modèle (1.25) sous forme matricielle :

$$\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \epsilon_n \quad (1.58)$$

où \mathbf{X}_n est une matrice d'ordre $n \times (p + 1)$ s'il existe p prédicteurs et de première colonne le vecteur constant $\mathbf{1}$. Nous savons (cf. par exemple Draper et Smith. [11]) que l'estimateur MCO de $\boldsymbol{\beta}$ est :

$$\hat{\boldsymbol{\beta}}^n = (\mathbf{X}_n \mathbf{X}_n)^{-1} \mathbf{X}_n \mathbf{Y}_n$$

et que $\mathbb{E}(\hat{\boldsymbol{\beta}}^n) = \boldsymbol{\beta}$ et $\text{Var}(\hat{\boldsymbol{\beta}}^n) = \sigma^2 (\mathbf{X}_n \mathbf{X}_n)^{-1}$. On voit donc qu'une condition suffisante pour que la suite $(\hat{\boldsymbol{\beta}}^n)$ converge vers $\boldsymbol{\beta}$ en moyenne quadratique (et donc en probabilité) est que $V_n = (\mathbf{X}_n \mathbf{X}_n)^{-1}$ converge vers 0 (cf. Antoniadis et al. [2]).

Proposition 1.6.1. *Soit m un entier positif. Considérons le modèle $\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \epsilon_n$ et supposons que les ϵ_i sont i.i.d. Si :*

- a) *la matrice \mathbf{X}_n est de plein rang pour tout $n \geq m$;*
- b) *tout élément diagonal v_{jj}^n de la matrice $V_n = (\mathbf{X}_n \mathbf{X}_n)^{-1}$ définit une suite convergente vers 0 ;*
- c) *la suite des matrices $\frac{1}{n}(\mathbf{X}_n \mathbf{X}_n)$ converge vers une matrice définie positive Q quand n tend vers l'infini,*

alors la suite de vecteurs aléatoires $\sqrt{n}(\hat{\boldsymbol{\beta}}^n - \boldsymbol{\beta})$ converge en loi vers une loi de Gauss $LG(0, \sigma^2 Q^{-1})$.

Preuve : cf. Antoniadis et al. [2]

On en déduit que $\sqrt{n}(\hat{\beta}_j^n - \beta_j) \sim LG(0, \sigma^2 (Q^{-1})_{jj})$ où $(Q^{-1})_{jj}$ est le terme diagonal (j, j) de Q^{-1} . Alors, la statistique de test de l'hypothèse $H_0 : \beta_1 = 0$ contre $H_1 : \beta_1 \neq 0$:

$$T = \frac{\sqrt{n} \hat{\beta}_j^n}{\sqrt{\sigma^2 (Q^{-1})_{jj}}} \quad (1.59)$$

suit approximativement une loi $LG(0, 1)$.

Les résultats de ce type permettent d'effectuer un test de H_0 optimal dans le cas de grands échantillons, mais ne justifient pas l'emploi de cette approximation gaussienne lorsque n est petit. Une façon de procéder dans ce cas consiste à recourir à une méthode non paramétrique.

Les méthodes non paramétriques d'inférence statistique adaptées au cas de la régression linéaire se sont développées plus tardivement parce que leur mise en oeuvre nécessite le recours à des moyens de calcul intensif, donc guère utilisables avant l'apparition d'ordinateurs assez puissants. Deux grandes sous-classes de méthodes partagent actuellement les faveurs des utilisateurs. La première s'est développée dans une optique d'estimation et de maîtrise de la variabilité des estimateurs autour des idées de Jackknife et de Bootstrap (*cf.* Efron [16], Manly [41]). Plus tardivement des travaux ont été menés pour construire des méthodologies de tests d'hypothèse et de construction d'intervalles de confiances cohérents avec cette approche. La section suivante fait quelques brefs rappels à ce sujet, à des fins de comparaison.

La deuxième approche est au centre des préoccupations de cette thèse. Elle s'est développée dans le but de construire des tests d'hypothèse et reste peu adaptée au traitement de questions d'estimation. On parle alors de méthodologie des *tests de permutation* ou des *tests de randomisation* ou encore des tests de re-randomisation.

Les deux approches partagent les mêmes hypothèses de loi des erreurs qui les différencient donc de façon identique par rapport au cadre classique de la régression linéaire comme plus généralement de l'inférence statistique paramétrique classique. Toutes deux développent des méthodes de tests conditionnel aux données fournies par l'échantillon observé, qui n'impose aucune contrainte à la procédure de tirage de l'échantillon.

1.6.2 Test de permutation

Les tests de permutation offrent sans doute l'un des exemples les plus clairs de la démarche statistique pour tester une hypothèse de symétrie des observations. Ils constituent aussi une démarche très générale qu'on peut appliquer à une grande variété de situations. Elles ne prétendent cependant pas à l'universalité car il existe des problèmes d'inférence qui n'admettent pas de solution en termes de permutations.

La validité statistique des tests de permutation est fondée sur la notion d'échangeabilité des observations, due à Fréchet (*cf.* [22]), qui est un postulat (de symétrie des observations) fondamental en théorie statistique, vérifiable en pratique et plus intuitif que la notion voisine d'indépendance. Dans un cadre général, on souhaite tester une hypothèse H_0 contre l'alternative H_1 au vu d'un échantillon $\mathbf{e} = \{e_1, e_2, \dots, e_n\}$. On suppose de plus que sous H_0 l'échantillon $\mathbf{e} = \{e_1, e_2, \dots, e_n\}$ est échangeable. Autrement dit, si l'hypothèse nulle est vraie, alors toutes les $n!$ permutations possibles des données ont une probabilité égale d'apparaître.

On peut alors réaliser un test de permutation. On choisit une statistique de

test T , dont on calcule d'abord la valeur observée $t_{obs} = T(\mathbf{e})$ sur l'échantillon originel (non permuté). t_{obs} est considérée comme une valeur de référence. Ensuite, on calcule la statistique de test T sur tous les échantillons permutés \mathbf{e}^π obtenus par une permutation π des valeurs de \mathbf{e} . Ces valeurs sont notés $t_\pi = T(\mathbf{e}^\pi)$. On dénombre $n!$ permutations possibles. A partir des valeurs obtenues t_π pour la statistique de test T sur ces $n!$ permutations possibles, on construit la distribution permutationnelle de T . La proportion des valeurs de t_π qui sont supérieures ou égales à t_{obs} dans la distribution permutationnelle s'appelle la p -valeur permutationnelle de la statistique de test T . Elle est notée p_π . On rejette H_0 lorsque p_π est inférieure ou égale au niveau nominal α retenu *a priori* pour ce test.

Par comparaison aux méthodes statistiques standard, la méthode des tests de permutation bénéficie de deux avantages principaux : (1) elle construit des inférences de plan, de portée limitée, mais sans imposer aux échantillons d'être issus d'une procédure de sondage aléatoire (2) elle est capable de prendre en compte les particularités de la situation (*cf.* Manly [41]).

Les tests de permutation sont souvent décrits comme des tests conditionnels à l'échantillon. Autrement dit, ils étudient la variabilité des statistiques d'intérêt, conditionnellement aux données. Ce faisant, ils utilisent toute l'information disponible contenue dans l'échantillon observé et seulement celle-ci. Certains regardent cette limitation comme un désavantage de la méthode des tests de permutation par rapport à ses concurrents relevant de la démarche classique. Cela a pour conséquence de ne rien dire sur la possibilité de généraliser les conclusions d'un test de permutation à une population visée dont l'échantillon serait une extraction représentative. Le test de permutation ignore la phase de sélection d'un échantillon dans une population et répond à une question de portée tout à fait particulière aux données disponibles.

Certains statisticiens affirment que le manque de théorie pour généraliser les résultats des tests de permutation à une super population signifie que ces tests ont peu de valeur en comparaison avec les tests standards classiques. D'autres affirment que dans la réalité, les échantillons ne sont pas tout le temps construits selon une procédure assurant leur caractère aléatoire mais plutôt, qu'ils constituent simplement des observations directement disponibles. La généralisation des résultats à la population Ω , pose une autre question et la validité de la réponse est en particulier assurée sans la condition que l'échantillon obtenu soit aléatoire. Dans ce cas, comme pour les tests classiques, on peut difficilement généraliser toute inférence obtenue en utilisant un test de permutation à la population globale Ω , car on ne maîtrise pas le risque d'erreur.

Les premières utilisations d'une méthode de test de permutation adaptée au modèle linéaire furent présentées dans la première moitié du 20ème siècle par Fisher, [18] et par Pitman, [48, 49, 50]). Leur utilisation se heurtait cependant à une lourdeur de calculs notoire par rapport au cadre méthodologie classique de tests d'hypothèse. Selon Diaconis (1988), on peut même en trouver l'application plus tôt et sous une forme non explicite dans les écrits de Pearson.

A) Permutation des n valeurs observées d'une variable aléatoire

Une présentation précise de la méthodologie des tests de permutation repose sur un certain nombre de définitions. On rappelle tout d'abord la définition d'une permutation des valeurs observées d'une variable aléatoire univariée sur un échantillon de taille n .

définition 1.6.1. Soient n une taille d'échantillon fixée a priori et $\mathbf{I}_n = \{1, 2, \dots, n\}$ l'ensemble ordonné indiquant les éléments de l'échantillon observé. Soit de plus π une permutation quelconque définie sur \mathbf{I}_n . π associe au n -uplet ordonné $\mathbf{i}_n = (1, 2, \dots, n)$ le n -uplet ordonné $\mathbf{i}_n^\pi = (\pi(1), \pi(2), \dots, \pi(n))$, noté aussi $\mathbf{i}_n^\pi = (\pi_1, \pi_2, \dots, \pi_n)$. Toute permutation π définie sur \mathbf{I}_n induit une permutation des coordonnées de tout vecteur \mathbf{x} de \mathbb{R}^n indicé par \mathbf{I}_n que l'on note encore π :

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \xrightarrow{\pi} \mathbf{x}^\pi = (x^{\pi_1}, x^{\pi_2}, \dots, x^{\pi_n}) \quad (1.60)$$

Soit \mathcal{P} l'ensemble des permutations de \mathbf{I}_n dans lui-même, muni de la loi uniforme. Alors chaque permutation π peut être considérée comme la réalisation d'une variable aléatoire $\mathbf{\Pi} = (\Pi_1, \Pi_2, \dots, \Pi_n)$ à valeurs dans \mathcal{P} , et de loi uniforme. La variable aléatoire $\mathbf{\Pi}$ est appelée permutation aléatoire sur \mathbf{I}_n .

Soit maintenant \mathbf{x} un vecteur de \mathbb{R}^n indicé par \mathbf{I}_n . On lui associe la variable aléatoire $\mathbf{x}^\mathbf{\Pi}$ induite par la permutation aléatoire $\mathbf{\Pi}$ à l'aide de l'application suivante, que l'on note encore (par abus) $\mathbf{\Pi}$:

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \xrightarrow{\mathbf{\Pi}} \mathbf{x}^\mathbf{\Pi} = (x^{\Pi_1}, x^{\Pi_2}, \dots, x^{\Pi_n}) \quad (1.61)$$

On peut étendre ces définitions, en s'intéressant maintenant au cas des permutations d'indices appliquées à un vecteur bivarié, $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ de $\mathbb{R}^n \times \mathbb{R}^n$ résultat de l'observation du vecteur aléatoire $\mathbf{z} = (\mathbf{X}, \mathbf{Y})$ sur un échantillon de n répliques d'un phénomène.

$$\{Z_1, Z_2, \dots, Z_n\} = \{(X_1, Y_1), (X_2, Y_2) \dots (X_n, Y_n)\} \quad (1.62)$$

Toute permutation π définie sur \mathbf{I}_n , induit une permutation des coordonnées du vecteur quelconque \mathbf{z} indicé par \mathbf{I}_n que l'on note encore π :

$$\mathbf{z} = (z_1, z_2, \dots, z_n) \xrightarrow{\pi} \mathbf{z}^\pi = (z^{\pi_1}, z^{\pi_2}, \dots, z^{\pi_n}) \quad (1.63)$$

Ici, $z^{\pi_i} = (x^{\pi_i}, y^{\pi_i})$. Autrement dit, π associe à $\mathbf{z} = ((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$, le vecteur $\mathbf{z}^\pi = ((x^{\pi_1}, y^{\pi_1}), (x^{\pi_2}, y^{\pi_2}), \dots, (x^{\pi_n}, y^{\pi_n}))$. Lorsque la permutation π est interprétée comme une réalisation d'une permutation aléatoire $\mathbf{\Pi}$ définie comme précédemment, on peut associer à tout vecteur \mathbf{z} de \mathbb{R}^n indicé par \mathbf{I}_n , le vecteur aléatoire bivarié $\mathbf{Z}^\mathbf{\Pi} = (\mathbf{x}^\mathbf{\Pi}, \mathbf{y}^\mathbf{\Pi})$ image de \mathbf{z} par l'application aléatoire $\mathbf{\Pi}$. On note encore (par abus) $\mathbf{\Pi}$ cette application.

Soit maintenant Π une permutation aléatoire définie comme précédemment sur \mathbf{I}_n et soit \mathbf{z} un vecteur de $\mathbb{R}^n \times \mathbb{R}^n$, indicé par \mathbf{I}_n . On lui associe la variable aléatoire bivariée \mathbf{z}^Π par application du procédé présenté ci-dessus que l'on note encore Π :

$$\mathbf{z} = (z_1, z_2, \dots, z_n) \xrightarrow{\Pi} \mathbf{z}^\Pi = (z^{\Pi_1}, z^{\Pi_2}, \dots, z^{\Pi_n}) \quad (1.64)$$

où $z^{\Pi_i} = (x^{\Pi_i}, y^{\Pi_i})$. Au bilan, on construit l'élément aléatoire \mathbf{z}^π suivant :

$$\mathbf{z} = ((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)) \xrightarrow{\Pi} \mathbf{z}^\Pi = ((x^{\Pi_1}, y^{\Pi_1}), (x^{\Pi_2}, y^{\Pi_2}), \dots, (x^{\Pi_n}, y^{\Pi_n})) \quad (1.65)$$

B) Test exact et test sans biais

Soit X une variable aléatoire dont la loi de probabilité est supposée appartenir à une famille paramétrée de distributions caractérisées par leur densité de probabilité $\{f_\theta : \theta \in \Theta\}$. On souhaite tester l'hypothèse :

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1$$

Soit $T(\mathbf{x}) = \{x_1, x_2, \dots, x_n\}$ une fonction de l'échantillon. On définit la fonction de test

$$\phi(T(\mathbf{x})) = \phi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) \in C; \\ 0, & T(\mathbf{x}) \in \bar{C}; \end{cases}$$

où la région critique C définit une règle de décision pour un échantillon observé \mathbf{x} telle que pour $T(\mathbf{x})$ appartenant à C , on a $\phi(\mathbf{x}) = 1$, et on rejette H_0 , alors que pour un échantillon de valeurs observé \mathbf{x} avec $T(\mathbf{x})$ appartenant à son complémentaire \bar{C} , $\phi(\mathbf{x}) = 0$, et on accepte H_0 . On suppose de plus ici que $\mathcal{X} = C \cup \bar{C}$ recouvre l'ensemble de tous les échantillons possibles. Autrement dit, une fonction de test ϕ partage l'ensemble \mathcal{X} en deux sous ensembles C et \bar{C} tels que

$$C = \{\mathbf{x} : \phi(\mathbf{x}) = 1\} \subseteq \mathcal{X} \text{ et } \bar{C} = \{\mathbf{x} : \phi(\mathbf{x}) = 0\} \subseteq \mathcal{X}$$

et le risque d'erreur de première espèce α s'écrit :

$$\alpha = P_{H_0}(\mathbf{x} \in C) = \int_C f_{\theta_0}(x) d\mu(x) = \int \phi(\mathbf{x}) f_{\theta_0}(x) d\mu(x) = \mathbb{E}_{\Theta_0}[\phi(\mathbf{x})]$$

De façon analogue, β l'erreur de deuxième espèce est définie par

$$\beta = P_{H_1}(\mathbf{x} \in \bar{C}) = \int_{\bar{C}} f_{\theta_1}(x) d\mu(x) = \int [1 - \phi(\mathbf{x})] f_{\theta_1}(x) d\mu(x) = \mathbb{E}_{\Theta_1}[1 - \phi(\mathbf{x})]$$

et la puissance du test, notée η , vérifie par définition :

$$\begin{aligned} \eta &= 1 - \beta \\ &= 1 - \mathbb{E}_{\Theta_1}[1 - \phi(\mathbf{x})] = \mathbb{E}_{\Theta_1}[\phi(\mathbf{x})] = \int_C f_{\theta_1} d\mu(x) = P_{H_1}(\mathbf{x} \in C). \end{aligned}$$

Dans la théorie fishérienne des tests, les conclusions du chercheur sont pilotées en fixant un niveau de risque α , dont on déduit la forme de la région critique $C \equiv C_\alpha$. Pour simplifier les notations, nous ignorons l'indice α dans la suite. Soit $\mathcal{P}(\mathbf{x})$ l'ensemble des $n!$ permutations possibles du vecteur \mathbf{x} de dimension n .

Lemme 1.6.1. *Soit \mathbf{x} un échantillon de taille n de la variable X . Un test de permutation au niveau α consiste en la donnée d'une région critique C , telle que pour tout \mathbf{z}^π , $\phi(\mathbf{z}^\pi) = 1$ si et seulement si*

$$\sum_{\mathbf{x}^\pi \in \mathcal{P}(\mathbf{x})} \phi(\mathbf{x}^\pi) \leq \alpha n!$$

Preuve : La définition du risque d'erreur de première espèce encouru par le test de permutation s'écrit :

$$\alpha = \mathbb{E}_{H_0}(\phi(\mathbf{x}^\pi)) = \int \phi(\mathbf{x}) f_{\theta_0}(x) d\mu(x) = \sum_{\mathbf{x}^\pi \in \mathcal{P}(\mathbf{x})} P_{\theta_0}(\mathbf{x}) \phi(\mathbf{x}^\pi)$$

mais la probabilité associée à chacune des $n!$ permutations possibles de \mathbf{x} sous H_0 est $1/n!$. Par suite, $P_{\theta_0} = \{\mathbf{x}^\pi = \mathbf{x}\} = 1/n!$ et l'on a pour un test de niveau α :

$$\mathbb{E}_{H_0}(\phi(\mathbf{x}^\pi)) = \sum_{\mathbf{x}^\pi \in \mathcal{P}(\mathbf{x})} \frac{1}{n!} \phi(\mathbf{x}^\pi) \leq \alpha$$

si et seulement si

$$\sum_{\mathbf{x}^\pi \in \mathcal{P}(\mathbf{x})} \phi(\mathbf{x}^\pi) \leq \alpha n! \quad \square \quad (1.66)$$

définition 1.6.2. *Une hypothèse nulle est dite composite lorsqu'elle est la réunion de plusieurs hypothèses simples (cf. Good [22]).*

Par exemple, considérons une variable aléatoire X qui suit une loi de densité de probabilité $f_X(x)$. L'hypothèse $H_0 : \mu = 0$ inclut les hypothèses $H_{0,1} : X \sim LG(\mu = 0, \sigma^2 = 1)$, $H_{0,2} : X \sim LG(\mu = 0, \sigma^2 = 1.6)$ et bien d'autres telles que $H_{0,3} X \sim Beta(0, 4)$. Remarquons que la moyenne de la loi Beta de paramètres (a, b) est $a/(a+b)$, etc.

Soit \mathbb{H}_0 une hypothèse composite. Elle s'écrit $\mathbb{H}_0 = \{H_0^j, j \in J\}$ où H_0^j est une hypothèse simple et J est un ensemble éventuellement non dénombrable.

définition 1.6.3. *Un test $\phi(X)$ de l'hypothèse H_0 est dit exact, si :*

$$\mathbb{E}_{H_0^j}(\phi(X)) = \alpha, \quad \forall j \in J \quad (1.67)$$

D'autre part, le test $\phi(X)$ de \mathbb{H}_0 est dit conservateur si :

$$\mathbb{E}_{H_0^j}(\phi(X)) \leq \alpha, \quad \forall j \in J. \quad (1.68)$$

Évidemment, un test exact est conservateur, mais la réciproque n'est pas toujours vraie. L'importance pratique des tests exacts tient à ce que leur niveau, qui est un majorant du risque d'erreur de première espèce encouru pour chaque H_0^j , ne peut pas être surestimé par ϕ . Cette propriété est remarquable en particulier lorsqu'un test est exact quelle que soit la loi du processus de génération des données. À l'inverse, si un test de niveau nominal α est en fait de niveau réel γ , son utilisation peut se révéler délicate. Si $\gamma > \alpha$, le risque d'erreur de première espèce est plus grand qu'il n'est envisagé. Si $\gamma < \alpha$, le test est sous-optimal, et l'on peut l'améliorer en agrandissant sa région critique. Pour plus d'informations sur la question, le lecteur est renvoyé à Good, [22], chapitre 11.

L'exactitude d'un test réfère à la maîtrise des risques de première espèce. De façon complémentaire, la notion de biais d'un test permet de définir des optimalités en termes de puissance d'un test. Nous présentons cette notion en identifiant l'hypothèse \mathbb{H}_0 à une partie non vide de l'espace Θ de définition des paramètres d'un modèle. On ne perd rien en généralité si l'on admet que Θ peut éventuellement être de dimension inférieure. On écrit alors $\Theta = \Theta_0 \cup \Theta_1$. On dit que θ satisfait à l'hypothèse nulle, si $\theta \in \Theta_0$ et que θ satisfait à l'hypothèse alternative sinon.

définition 1.6.4. *Test sans biais*

Un test est dit sans biais de niveau α , si sa fonction puissance, notée η , est conservative, c'est-à-dire si η satisfait aux deux conditions suivantes :

$$\begin{cases} \eta(\theta) \leq \alpha & \text{pour tout } \theta \text{ satisfaisant à l'hypothèse nulle} \\ \eta(\theta) \geq \alpha & \text{pour tout } \theta \text{ satisfaisant à une hypothèse alternative} \end{cases}$$

Comme en estimation, l'absence de biais d'un test d'hypothèse est considéré comme une propriété souhaitable par de nombreux auteurs. Ainsi, Good y voit un principe normal et souhaitable. Pourtant ce jugement reste l'objet de discussions et de critiques, cf. par exemple, Barra [3], Ch. V-1 et Suissa et Shuster [56].

C) Échangeabilité

La recherche de conditions suffisantes garantissant l'exactitude et l'absence de biais d'un test de permutation a fait l'objet de nombreuses publications (Lehman, [37]). Le cadre généralement retenu repose sur la notion d'échangeabilité.

définition 1.6.5. *Soit (X_1, X_2, \dots, X_n) un vecteur aléatoire de taille n et de loi conjointe $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ donnée. Les X_i sont dits échangeables si la densité conjointe des observations est invariante par permutation des indices. C'est-à-dire si :*

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_{X_1^\pi, X_2^\pi, \dots, X_n^\pi}(x_1^\pi, x_2^\pi, \dots, x_n^\pi), \forall \pi \in \mathcal{P}$$

Les deux exemples les plus classiques d'observations échangeables sont fournis d'une part, par le cas de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) et d'autre part par un ensemble de variables aléatoires de loi conjointe

gaussienne et de covariances identiques (Good 2000 [22]). D'autres exemples sont présentés par Galambos, [21] ou Draper et Smith, [10]. Pour plus d'information, voir par exemple Good [22] chapitre 14.

Une condition suffisante garantissant l'exactitude et l'absence de biais d'un test de permutation est l'échangeabilité des observations formant l'échantillon, *cf.* Lehman [37].

Reprenons ici le cadre classique de comparaison de deux populations. Soit donc un échantillon de taille n X_1, \dots, X_n , i.i.d. de la loi $F(x)$ et Y_1, \dots, Y_m un échantillon de taille m i.i.d. de la loi $F(x-\Delta)$. Nous notons en conséquence $\mathbf{z} = \{x_1, \dots, x_n, y_1, \dots, y_m\}$ l'ensemble des valeurs observées. On suppose ainsi qu'on a observé m éléments de la population 1 et n éléments de la population 2. On définit $\mathcal{P}(\mathbf{z})$ comme l'ensemble des $(m+n)!$ permutations possibles de l'échantillon de valeurs observées \mathbf{z} . De plus, on suppose que F admet une densité et l'on note f la densité de probabilité associée à la fonction de répartition F .

Théorème 1.6.1. *Soit \mathcal{F} la famille de toutes les densités f de $(n+m)$ dimensions qui sont intégrables et symétriques en leurs arguments. On souhaite tester l'hypothèse $H_0 : \Delta = 0$ contre l'alternative $f(x_1, \dots, x_n, y_1 - \Delta, \dots, y_m - \Delta)$ avec $\Delta \neq 0$. Un test ϕ pour toutes les $f \in \mathcal{F}$ est dit sans biais si seulement si*

$$\sum_{\mathbf{z}^\pi \in \mathcal{P}(\mathbf{z})} \phi(\mathbf{z}^\pi) = \alpha(n+m)! \quad (1.69)$$

Preuve *cf.* Lehman [37].

Remarquons que le test de permutation qui vérifie (1.69), mis en place pour tester cette hypothèse, encoure un risque d'erreur de première espèce donné par :

$$\mathbb{E}_{H_0}(\phi(\mathbf{z}^\pi)) = \frac{1}{(n+m)!} \sum_{\mathbf{z}^\pi \in \mathcal{P}(\mathbf{z})} \phi(\mathbf{z}^\pi) = \alpha \quad (1.70)$$

Pour toutes les fonctions $f \in \mathcal{F}$ et donc, le test de permutation est sans biais pour toute hypothèse H_0^j formant \mathbb{H}_0 . C'est donc un test exact de \mathbb{H}_0 .

Théorème 1.6.2. *Sous les hypothèses du théorème ci-dessus, les tests de permutation sont sans biais.*

Preuve : C'est une conséquence directe de la définition (1.6.4) et du théorème ci-dessus, *cf.* Good [22].

Considérons maintenant deux tests $\phi(x)$ et $\phi^*(x)$ de niveau α de l'hypothèse simple $H_0 : \theta = \theta_0$ contre l'hypothèse simple $H_1 : \theta = \theta_1$, c'est-à-dire tels que :

$$\mathbb{E}_{\theta_0}[\phi(X)] \leq \alpha \text{ et } \mathbb{E}_{\theta_0}[\phi^*(X)] \leq \alpha$$

on dit que le test $\phi^*(X)$ est plus puissant que $\phi(X)$, si :

$$\mathbb{E}_{\theta_1}[\phi^*(X)] \geq \mathbb{E}_{\theta_1}[\phi(X)]$$

définition 1.6.6. *Considérons l'hypothèse $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$. Un test $\phi^*(X)$ est dite uniformément le plus puissant au niveau α si :*

$$a) \mathbb{E}_{\theta_0}[\phi^*(X)] \leq \alpha$$

$$b) \forall \phi \text{ qui vérifie (a), } \mathbb{E}_{\theta_1}[\phi^*(X)] \geq \mathbb{E}_{\theta_1}[\phi(X)]$$

Théorème 1.6.3. *Les tests de permutation sont uniformément les plus puissants pour tester l'hypothèse $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ où $\theta - \theta_1 < 0$.*

Preuve : cf. Good [22]

C) Mise en oeuvre d'un test de permutation

Good [22] a décrit les étapes d'une procédure de construction d'un test de permutation. Cette façon d'opérer se décompose en cinq étapes, que nous illustrons sur l'exemple présenté à de nombreuses reprises par cet auteur. Il s'agit d'un problème de comparaison de deux sous-populations en termes de moyenne, donc analogue au problème initialement traité par Student dans un cadre paramétrique et qui conduisit à la proposition du test qui porte son nom.

Étape 1- Analyse du problème

L'analyse du problème consiste tout d'abord à identifier l'hypothèse nulle et l'alternative d'intérêt. Supposons que l'on veuille évaluer les propriétés de prolongation de la vie d'une cellule par un nouveau traitement expérimental. Pour ce faire, on a divisé des cultures de cellules en deux groupes : un groupe développé dans un milieu standard et un groupe développé en présence d'une vitamine (le traitement) E. À l'issue de l'expérience et après élimination de plusieurs cultures contaminées, chaque groupe a été divisé en trois cultures traitées indépendamment. On appelle les trois cultures du premier groupe (groupe traité en présence vitamine E) X_1, X_2 et X_3 et de façon analogue, on distingue dans le deuxième groupe Y_1, Y_2 et Y_3 .

L'hypothèse nulle dit que le potentiel de croissance d'une culture n'est pas affecté par la présence de la vitamine E. L'alternative d'intérêt affirme que les cellules développées en présence de la vitamine E sont capables de beaucoup plus de divisions cellulaires. Si on note la moyenne du premier groupe par μ_X et celle du deuxième groupe par μ_Y , alors les hypothèses s'écrivent :

$$H_0 : \mu_X = \mu_Y \text{ contre } H_1 : \mu_X \neq \mu_Y$$

Sous l'hypothèse nulle, le fait d'être traité ou non ne fournit aucune information au sujet des résultats, car on s'attend à ce que les observations aient plus ou moins les mêmes valeurs dans chacun des deux groupes expérimentaux. On est donc libre d'échanger (de permuter) les étiquettes sans changer les résultats pour autant.

Étape 2- Choix d'une statistique de test et distribution sous l'hypothèse nulle et l'alternative

L'étape suivante de la méthode consiste à choisir une statistique de test qui discrimine entre l'hypothèse nulle et l'alternative. Lorsqu'on soumet les données à une permutation des étiquettes des observations. Dans l'exemple présenté, la statistique considérée est la somme des comptages dans le groupe été traité avec de la vitamine E :

$$T = \sum_{i=1}^3 X_i$$

Si l'alternative est vraie, cette somme doit être plus grande que la somme calculée dans le groupe non traité. Si l'hypothèse nulle est vraie, alors les sommes calculées dans les deux groupes d'observations devraient être identiques aux fluctuations d'échantillonnage près. Une somme peut bien sûr être plus petite ou plus grande que l'autre par le seul jeu du hasard, mais les deux ne devraient pas être trop différentes. On s'interroge donc sur la significativité du décalage Δ de la position centrale de la répartition, décrite formellement plus haut. Sous cette forme, on retrouve la formulation d'une méthode de test de permutation étudiée par Fisher (1935), puis Pitman (1935b).

Étape 3-Calcul de la statistique de test pour les données originelles

La troisième étape dans la méthode consiste à calculer la statistique de test sur les observations originelles (réellement observées). De fait, pour les données du premier groupe les trois premières observations, $x_1 = 121, x_2 = 118$ et $x_3 = 110$, appartiennent aux cultures qui avaient reçu la vitamine E. La valeur de la statistique de test est :

$$t_{obs} = \sum_{i=1}^3 x_i = 121 + 118 + 110 = 349$$

et t_{obs} , joue le rôle de valeur de référence pour T .

Étape 4- Permutation des observations

On permute maintenant les observations. Pour cela, on attribue les six étiquettes aléatoirement, aux six observations, en veillant à respecter la structure du plan d'expérience. On distingue donc dans chaque permutation trois cultures traitées et trois non traitées. Soit $\mathbf{z} = (x_1, x_2, x_3, y_1, y_2, y_3)$ les six observations traitées et non traitées et π une permutation de \mathbf{Z} , on obtient :

$$\mathbf{z} = (x_1, x_2, x_3, y_1, y_2, y_3) \xrightarrow{\pi} \mathbf{z}^\pi = (x_1^\pi, x_2^\pi, x_3^\pi, y_1^\pi, y_2^\pi, y_3^\pi) \quad (1.71)$$

Par exemple, les observations permutées qui se répartissent en traitées : $x_1^\pi = 121, x_2^\pi = 118$ et $x_3^\pi = 34$, et non traitées : $y_1^\pi = 110, y_2^\pi = 22$ et $y_3^\pi = 12$ (voir

le tableau (1.6.2)). Pour cette permutation π la somme des observations dans le premier groupe fournit la statistique de test et vaut :

$$t_\pi = \sum_{i=1}^3 x_i^\pi = 121 + 118 + 34 = 273.$$

Lorsqu'on répète cette étape pour la totalité des $\binom{6}{3} = 20$ permutations possibles, on décrit de façon parfaite la loi de répartition permutationnelle de T , de sorte qu'on obtient un test exact de l'hypothèse H_0 , puisque les variables $(X^1, X^2, X^3, Y^1, Y^2, Y^3)$ sont échangeables sous H_0 .

	Premier groupe			Somme	Deuxième groupe		
	x_1	x_2	x_3		t_{obs}	y_1	y_2
1	121	118	110	349	34	22	12
	x_1^π	x_2^π	x_3^π	t^π	y_1^π	y_2^π	y_3^π
2	121	118	34	273	110	22	12
3	121	110	34	265	118	22	12
4	118	110	34	262	121	22	12
5	121	118	22	261	110	34	12
6	121	110	22	253	118	34	12
7	121	118	12	251	110	34	22
8	118	110	22	250	121	34	12
9	121	110	12	243	118	34	22
10	118	110	12	240	121	34	22
11	121	34	22	177	118	110	12
12	118	34	22	174	121	110	12
13	121	34	12	167	118	110	22
14	110	34	22	166	121	118	12
15	118	34	12	164	121	110	22
16	110	34	12	156	121	118	22
17	121	22	12	155	118	110	34
18	118	22	12	152	121	110	34
19	110	22	12	144	121	118	34
20	34	22	12	68	121	118	110

TAB. 1.1 – Toutes les 20 permutations possibles

Étape 5- Conclusion du test

Parmi les 20 permutations distinctes possibles, la somme (t_π) des observations dans le groupe traité par la vitamine E est une fois seulement supérieure ou égale à $t_{obs} = 349$. Par suite, la p -valeur permutationnelle du test est :

$$p_\pi = \frac{\#(|t_\pi| \geq |t_{obs}|)}{20} = \frac{1}{20} = 0.05$$

On rejette donc l'hypothèse nulle au niveau $\alpha = 5\%$ et l'on conclut que le traitement est efficace.

F) Test de permutation en régression linéaire simple

La démarche précédente s'adapte de façon directe au cas d'un modèle de régression linéaire. Pour illustrer la chose dans un cas simple, reprenons le modèle 1.24

$$Y = \beta_0 + \beta_1 x + \epsilon$$

L'objet de cette application de la méthodologie des tests de permutation consiste à construire un test de l'hypothèse $H_0 : \beta_1 = 0$ contre l'alternative $H_1 : \beta_1 \neq 0$. Dans le modèle classique de régression linéaire simple sous l'hypothèse gaussienne, le terme d'erreur ϵ est postulé i.i.d. sachant \mathbf{X} et distribué selon une loi de Gauss $LG(0, \sigma^2)$. Dans le cadre permutatif, les valeurs observées de ϵ , notées $\hat{\epsilon}$ ou encore e , sont échangeables (cf. [48], [49] et [50]), parce que supposées i.i.d.

Les étapes du test

On dispose d'un échantillon de n couples de valeurs observées du vecteur aléatoire (X, Y) . Comme dans le cas classique, on veut savoir s'il existe une liaison linéaire entre les variables X et Y .

Étape 1- Analyse du problème

Le coefficient β_1 du modèle (1.24) est nul quand il n'existe pas de liaison linéaire entre les variables X et Y . Par suite, l'hypothèse nulle et l'alternative sont respectivement :

$$H_0 : \beta_1 = 0 \text{ contre } H_1 : \beta_1 \neq 0$$

On réalise tout d'abord une régression linéaire de la variable Y en x , au sens des MCO, en utilisant le modèle (1.24). La loi de Y dépend du paramètre $\theta = (\beta_0, \beta_1, \sigma^2)$. On obtient alors l'estimateur MCO $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)$ et la prévision des résidus $\hat{\epsilon} = e$.

Étape 2- La statistique de test

Soit $S(X)$ une fonction de l'échantillon aléatoire (x_1, x_2, \dots, x_n) . x_i est la i -ème valeur observée d'une variable aléatoire X dont la loi de probabilité est spécifiée par la fonction densité $f(x, \theta)$ où f est de forme connue mais dépend d'un paramètre θ appartenant à un ouvert de \mathbb{R}^k . Ici, $k = 3$ et $\Theta = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$.

définition 1.6.7. (cf. Mood et al. [45]) Une statistique $Q(\theta, S(X))$ qui est monotone en θ et dont la distribution de probabilité ne dépend pas de θ est dite pivotale.

Remarque 1.6.1. Les études récentes, menées sur les tests de permutation et dans le contexte proche des tests simulés (Monte Carlo tests) et par le Bootstrap, montrent

l'importance que revêt le choix d'une statistique pivotale pour maîtriser le niveau des tests de permutation, lorsque l'on travaille sur des échantillons de taille limitée, cf. Hall et Titterington [24] et Hall et Wilson [25]. Le dernier chapitre de notre thèse tente d'évaluer l'influence de cette qualité sur le comportement des tests de permutation dans le cas où l'on traite des matrices de dissimilarités. Dans ce cas, la loi du processus de génération des données n'est pas connue et son étude reste un problème ouvert. Nous ignorons donc si les statistiques de test retenues sont pivotales. Nous avons donc dû sonder quelques transformations monotones des données, pour en étudier l'influence sur la taille et la puissance des tests effectués.

La deuxième étape dans l'application de la méthodologie des tests de permutation consiste à choisir une statistique de test qui distingue l'hypothèse nulle et l'alternative. Ici, on retient la statistique de test de Student :

$$T = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)}$$

où $S^2(\hat{\beta}_1)$ est défini dans la remarque 1.4.3 et désigne l'estimateur MCO de $\text{Var}(\hat{\beta}_1)$.

En effet si $S(X) = \frac{\hat{\beta}_2}{S(\hat{\beta}_2)}$ et $\theta = \beta_2$ alors,

1. $Q(\theta, S(X)) = \frac{\hat{\beta}_2 - \beta_2}{S(\hat{\beta}_2)} = T$ et on obtient $Q(\theta, S(X)) = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)} = T$ sous $H_0 : \beta_2 = 0$;
2. T est une statistique pivotale car elle suit une loi de Student à $(n - 2)$ degrés de liberté, quelque soit $\theta \in \Theta_0$.

Étape 3- Le calcul de la statistique de test

La troisième étape consiste à calculer la statistique T sur les observations d'origine. La valeur observée de cette statistique T est notée :

$$t_{obs} = \frac{b_1}{s(\hat{\beta}_1)} = \frac{b_1}{s(B_1)}$$

où, b_1 est la valeur observée de $\hat{\beta}_1 = B_1$ sur l'échantillon, *i.e.* l'estimation de β_1 , et $s^2(\hat{\beta}_1)$ (la valeur observée de $S^2(\hat{\beta}_1)$) est l'estimation MCO associée de la variance de cet estimateur sur cet échantillon.

Remarque 1.6.2. *Comme déjà dit, le cadre permutational suppose que les valeurs observées de ϵ , notées e , sont échangeables. Par ailleurs, $H_0 : \beta_1 = 0$ dit qu'il n'existe pas de liaison linéaire entre les variables X et Y , donc les valeurs prises par X n'ont pas d'effet sur celles de Y . Mieux, $Y_i = \beta_0 + \epsilon_i$, et puisque β_0 est constante, on en déduit que les valeurs observées de Y comme celles de ϵ sont échangeables. On peut donc permuter les valeurs observées y_i au lieu des e_i .*

Étape 4- La randomisation des étiquettes des observations

Cette étape permute les observations originelles $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ par des permutations aléatoires π . On obtient ainsi pour chaque permutation π :

$$\mathbf{y} = (y_1, y_2, \dots, y_n) \xrightarrow{\pi} \mathbf{y}^\pi = (y^{\pi_1}, y^{\pi_2}, \dots, y^{\pi_n})$$

Ensuite, on recalcule la valeur de la statistique T pour chaque permutation π .

$$t_\pi = \frac{b_{1,\pi}}{s(\hat{\beta}_{1,\pi})} = \frac{b_{1,\pi}}{s_\pi(B_1)}$$

où $b_{1,\pi}$ et $s^2(\hat{\beta}_{1,\pi}) = s_\pi^2(\hat{\beta}_1) = s_\pi^2(B_1)$ sont les estimations MCO respectives de β_1 et de $S^2(\hat{\beta}_1) = S^2(B_1)$ calculées sur les observations permutées par π . Rappelons que $S^2(\hat{\beta}_1) = S^2(B_1)$ est l'estimateur de la variance de l'estimateur MCO de β_1 , notée $\sigma^2(\hat{\beta}_1) = \sigma^2(B_1)$. On obtient ainsi la distribution permutationnelle exacte de la statistique t_π , en calculant cette statistique pour les $n!$ permutations possibles de $\mathbf{y} = (y_1, y_2, \dots, y_n)$. En pratique, dès que n est grand, on doit limiter la masse de calculs induite, en approchant la loi exacte de T par le tirage d'un échantillon de de permutations indépendantes π de l'échantillon observé de taille assez grande pour approcher de façon satisfaisante la loi permutationnelle de T .

Étape 5-Conclusion du test

Pour effectuer un test de permutation bilatéral à partir des valeurs obtenues t_π , on construit la distribution des valeurs $|t_\pi|$ rangées par ordre croissant. Puis, on place la valeur $|t_{obs}|$ parmi les statistiques d'ordre de cette distribution. La p -valeur permutationnelle exacte (notée p_π) est calculée comme la proportion de valeurs $|t_\pi|$ qui sont supérieures ou égales à la valeur $|t_{obs}|$ (cf. Hope [29]) pour la distribution où probabilité définie comme la loi permutationnelle de $|T|$ définie sur l'échantillon observé $\{(x_i, y_i); i = 1, 2, \dots, n\}$, c'est-à-dire :

$$p_\pi = Pr\{|T| \geq |t_{obs}|\} = \frac{\#\{|t_\pi| \geq |t_{obs}|\}}{n!}$$

Cette étape conduit à décider si l'on accepte H_0 ou si l'on rejette cette hypothèse. Dans ce cas, on décide d'agir comme si l'alternative était vraie. On fixe pour cela, avant l'observation des données, un niveau nominal α , par exemple $\alpha = 0.05$. Si $p_\pi < 0.05$ on rejette H_0 .

Exemple 1.6.1. (cf. Dodge [9])

Nous illustrons la méthodologie décrite dans la section précédente sur un corpus de données pédagogiques de faible taille. Les données proviennent d'une enquête visant à comparer selon certains critères, différents aliments vendus dans un restaurant rapide et les auteurs ont retenu les informations se trouvant dans le tableau ci-dessous :

Observation i	Aliment	Poids(g) x_i	Prix (CHF) y_i
1	Sandwich végétarien	150	3,90
2	Sandwich parisien au jambon	92	3,40
3	Big Mac	193	5,70
4	Hamburger	90	2,90
5	Hot Dog	135	3,80
6	Fallafel	241	6,00
7	Quiche lorraine	169	3,3
8	Pizza tomates et jambona	165	3,3

TAB. 1.2 – Prix de différents aliments

Ils ont appliqué le modèle 1.24 à cette étude :

$$Y = \beta_0 + \beta_1 x + \epsilon$$

avec Y ici le prix de l'aliment, x le poids de cet aliment. $\theta = (\beta_0, \beta_1, \sigma)$ est le paramètre du modèle, ϵ est un terme d'erreur. Il est supposé réparti suivant une loi inconnue mais de variance finie σ^2 . On suppose de plus les ϵ_i , $i = 1, 2, \dots, 8$ échangeables sous H_0 . La méthode de test de permutation permet alors de réaliser un test exact (c'est-à-dire fondé sur $8!$ permutations possibles) de l'hypothèse : $H_0 : \beta_1 = 0$ contre $H_1 : \beta_1 \neq 0$.

On effectue d'abord une régression de la variable Y en x , au sens MCO. On obtient alors $b_1 = 0.0188$ et $s(\hat{\beta}_1) = 0.0054$ et donc la valeur de la statistique de test sur les observations est

$$t_{obs} = \frac{b_1}{s(\hat{\beta}_1)} = 3.429.$$

Ensuite, on permute les observations de la variable Y et pour les $8! = 40320$ permutations possibles, on calcule la statistique t_π sur les données permutées. La distribution permutationnelle de la statistique $|T^\pi|$ ainsi obtenue est représentée dans la figure (1.1). On calcule ensuite la p-valeur permutationnelle :

$$p_\pi = \frac{\#(|t_\pi| \geq |t_{obs}|)}{n!} = \frac{\#(|t_\pi| \geq 3.429)}{8!} = 0.0143$$

Par suite, on rejette $H_0 : \beta_1 = 0$ au niveau de signification $\alpha = 0.05$. Le prix de cet aliment dépend de son poids.

Remarque 1.6.3. Puisque le nombre de permutations possibles de n objets est égal à $n!$, lorsque la taille de l'échantillon (n) croît, les calculs nécessaires pour réaliser un test de permutation deviennent vite énormes. Par suite, on se contente en pratique d'une estimation de p_π obtenue en tirant un échantillon aléatoire de M permutations parmi ces $n!$ on obtient alors :

$$\hat{p}_\pi = \frac{\#(|t_{1,\pi}| \geq |t_{obs}|)}{M}$$

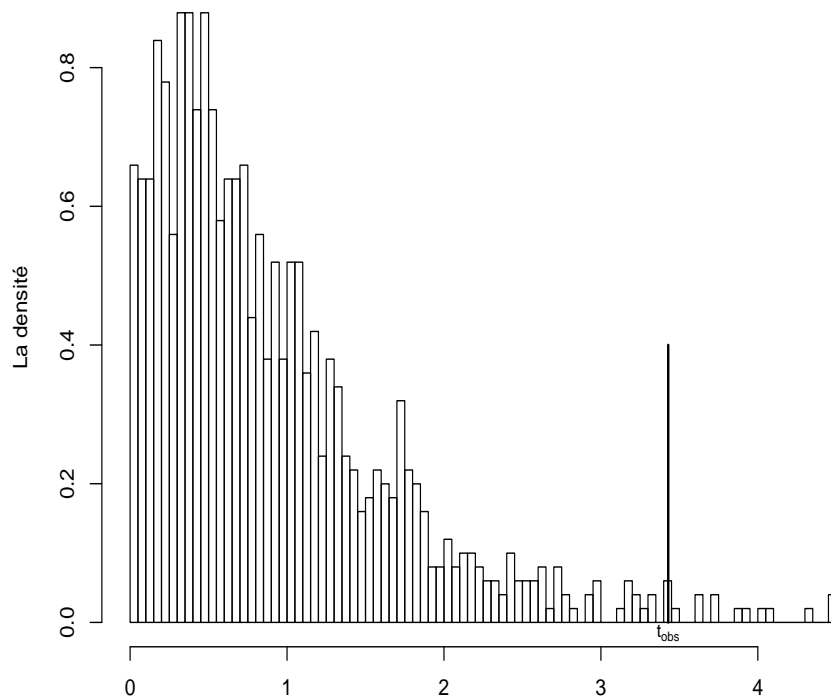


FIG. 1.1 – La distribution permutationnelle de la statistique de test $|T_\pi|$ sur $8!$ permutations possibles. La valeurs observée t_{obs} est distinguée sur la figure.

qui est une estimation d'autant meilleure de P_π que M est grand. (cf. Freedman et Lane [19]).

Par ailleurs, s'inspirant des résultats établis par Marriott [42] dans le cas de test de Monte Carlo, Manly[41] montra qu'il faut effectuer au moins 1000 permutations, pour effectuer un test au niveau 0.05 sur un ensemble de données et qu'au moins 5000 permutations sont nécessaires pour un test de niveau 0.01. Ces permutations forment un échantillon tiré avec remise dans l'ensemble des permutations possibles. On parle en conséquence parfois de méthode de rééchantillonnage ou de randomisation.

Exemple 1.6.2. Reprenons l'exemple 1.6.1, nous voulons réaliser le même test de permutationnel en limitant M par 1000 permutations aléatoires. La distribution permutationnelle de T_π obtenue dans ce cas est fournie par la figure (1.2) qui fournit une répartition approchée de la répartition exacte fournie dans la figure (1.1). De plus, la p-valeur permutationnelle approchée obtenue est :

$$\hat{p}_\pi = \frac{\#(|t_\pi| \geq |t_{obs}|)}{M} = \frac{\#(|t_\pi| \geq 3.429)}{1000} = 0.0139$$

qui est très proche de la p -valeur permutationnelle obtenue sur $8!$ permutations (0.0143). Ici, encore le test approché de niveau de signification $\alpha = 0.05$ rejette $H_0 : \beta_1 = 0$.

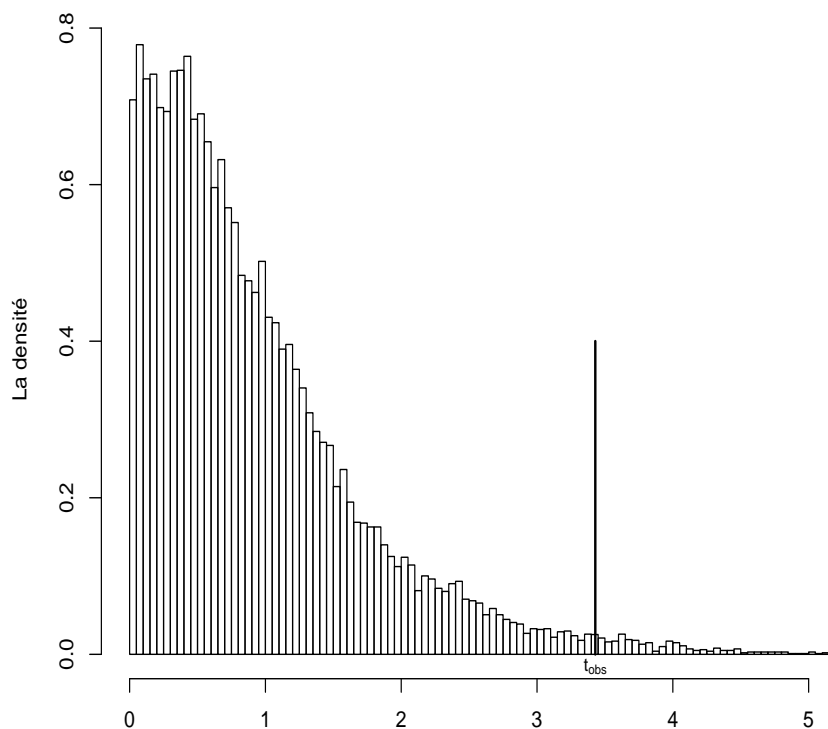


FIG. 1.2 – La distribution permutationnelle de la statistique de test $|T_\pi|$ sur 1000 permutations aléatoires

Pour compléter cette comparaison des deux distributions, on peut tracer un graphique quantile-quantile, ou QQ-plot, fondé sur les deux ensembles d'observations de la statistique de test obtenus par chacune des méthodes permutationnelles : le test exact et le test sur échantillon de 1000 permutations.

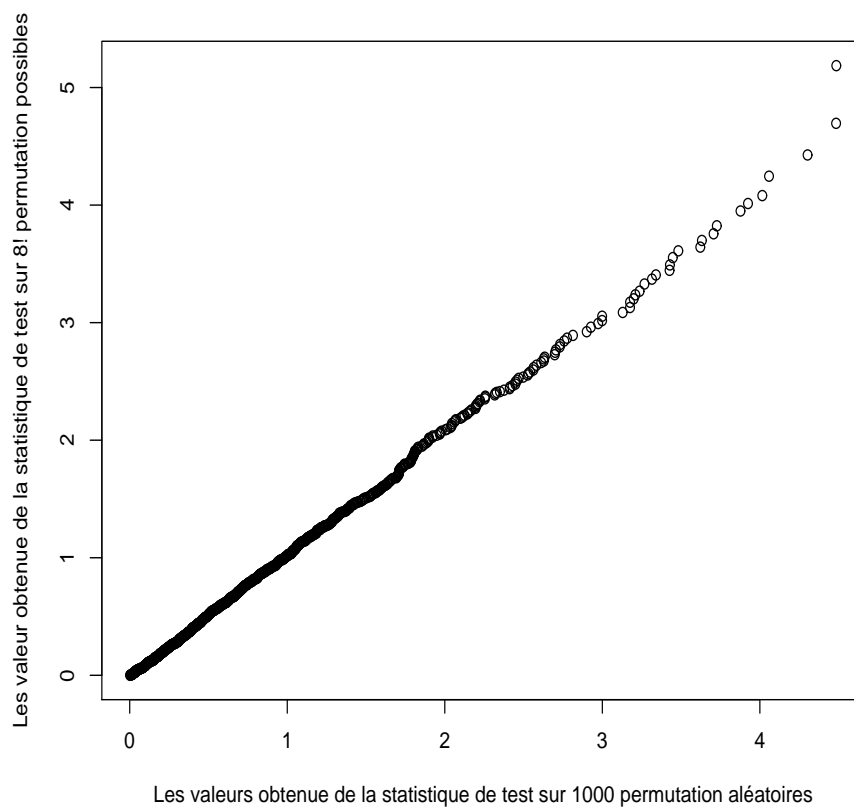


FIG. 1.3 – La figure de QQ-plot sur les deux ensembles d'observations de la statistique de test

Ce graphique montre que les deux ensembles des valeurs de la statistique de test fournissent des distributions très proches, avec une légère dégradation de cet accord pour les grandes valeurs de la statistique de test.

G) Équivalence des statistiques de test

Lorsqu'on développe l'expression analytique de la statistique $T = \hat{\beta}_1 / S(\hat{\beta}_1)$, nous vérifions que :

$$S(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{(n-2) \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2 - (\hat{\beta}_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2}{(n-2) \sum_{i=1}^n (x_i - \bar{x})^2}$$

on obtient ainsi :

$$T = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)} = \frac{\hat{\beta}_1 \sqrt{(n-2) \sum_{i=1}^n (x_i - \bar{x})^2}}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2 - \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}}$$

Cette écriture montre que pour toute valeur constante du nombre des observations n , T est une fonction monotone croissante de $\hat{\beta}_1$. Par suite, $\hat{\beta}_1$ est une statistique de test équivalente à T dans l'optique d'un test de permutation. Par ailleurs, les statistiques $\hat{\beta}_1$ et R sont équivalentes d'après l'équation (1.57). D'autre part, la statistique :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i Y_i - n\bar{x}\bar{Y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

à pour seul terme non invariant par permutation, la somme des produits $\sum x_i Y_i$. Les autres termes intervenant dans cette expression \bar{x} , \bar{Y} et $\sum x_i^2$ sont invariants par permutation des Y_i . Cette propriété est partagée par les composantes de la statistique R , c'est-à-dire :

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{\sum_{i=1}^n x_i Y_i - n\bar{x}\bar{Y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(\sum_{i=1}^n Y_i^2 - n\bar{Y}^2)}}$$

$\sum x_i Y_i$ est le seul terme qui n'est pas invariant par permutation des Y_i . On en déduit que dans un cadre permutationnel, le test d'hypothèse $H_0 : \beta_1 = 0$ contre

$H_1 : \beta_1 \neq 0$ est équivalent au test d'hypothèse $H_0 : \rho = 0$ contre $H_1 : \rho \neq 0$, au sens où les statistiques T , R et $\hat{\beta}_1$ sont des statistiques de test équivalentes à la statistique

$$M_{xY} = \sum_{i=1}^n x_i Y_i$$

Ce produit croisé joue un rôle privilégié dans le développement de la théorie, puisque son étude fut à l'origine des travaux de Pitman, [48, 49, 50], fixant le cadre théorique des tests de permutation, après les travaux pionniers de Fisher, [18]. Souvent redécouverte ensuite, cette statistique de test de permutation est généralement connue sous le nom de statistique de Mantel [39] depuis son introduction en biométrie par cet auteur. Le cadre de la présentation de Mantel est en fait plus élémentaire que celui d'un plan d'expérience à 1 facteur de variabilité et $M > 1$ modalités qui fut adopté par Pitman. En pratique, la statistique de Mantel conduit à des calculs plus simples et moins coûteux que les autres statistiques telles que R , β_1 et T . Elle présente cependant l'inconvénient d'une interprétabilité concrète peu évidente.

Remarque 1.6.4. *Afin de tester d'hypothèse $H_0 : \beta_1 = 0$ contre $H_0 : \beta_1 \neq 0$ dans le cas permutational, pour calculer la statistique de test $M_{xY} = \sum_{i=1}^n x_i Y_i$ ou d'autres statistiques équivalentes, la permutation des observations de Y et des observations de X sont équivalentes dans le cas d'une régression linéaire simple.*

Les tests de permutation n'offrent pas la seule méthode de construction de tests non-paramétriques, conditionnels aux données observées. Nous rappelons dans cette section, le cadre d'emploi de deux méthodes de rééchantillonnage concurrentes utiles pour évaluer la variabilité d'observations faites sur un échantillon dont la forme de la loi de répartition n'est pas connue.

1.6.3 Les méthodes de Jackknife et de Bootstrap

Soit une population Ω de taille N finie ou non, pour laquelle on veut estimer un paramètre, par exemple la moyenne μ . Pour cela, on observe un échantillon aléatoire de n individus de cette population et on mesure la caractéristique d'intérêt sur chaque individu formant cet échantillon. Nous notons $x_i, i = 1, \dots, n$ la valeur de cette caractéristique sur l'individu i . Un estimateur classique de μ est la moyenne arithmétique :

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i. \quad (1.72)$$

La question centrale traitée par les méthodes de Bootstrap et de Jackknife est celle de la qualité de cette estimation. Une réponse classique consiste à calculer la variance échantillonnale $\text{Var}(\hat{\mu})$ de l'estimateur $\hat{\mu}$, vu comme une variable aléatoire définie sur l'espace des échantillons tirés dans Ω , suivant le même mode d'échantillonnage que celui qui a produit l'échantillon observé. Calculer la variance de l'estimateur

$\hat{\mu}$, impose d'étudier sa distribution d'échantillonnage. Les méthodes de Jackknife et de Bootstrap ont pour finalité d'en proposer une approximation fondée sur les seules données observées.

Comme toutes les méthodes non paramétriques abordées dans cette thèse, les méthodes de Jackknife et de Bootstrap étudient la variabilité des statistiques d'intérêt, conditionnellement aux données. Cela signifie que faute de connaître la loi d'échantillonnage de cette statistique, on s'intéresse aux variations induites sur ses valeurs par une technique de rééchantillonnage, dont la pertinence est fixée par une liste de postulats.

Ainsi, la méthode du Bootstrap a pour principe de prélever une série d'échantillons aléatoires simples avec remise parmi les observations de l'échantillon initial, considéré comme une population. Elle permet ainsi d'évaluer la variabilité d'échantillonnage, et par suite d'estimer les variations de chaque «paramètre» du modèle lorsqu'on tire des échantillons i.i.d., dits échantillons Bootstrap, dans l'échantillon initial. Ce «paramètre» de la loi conditionnelle à l'échantillon observé est en fait une estimation de la valeur du paramètre attaché à la loi dans Ω . On peut donc estimer la loi d'échantillonnage de ces estimations. Sous certaines conditions de régularité (*cf.* Beran [4]), la théorie permet d'établir des résultats de comportement asymptotique des estimateurs sous-jacents. Elle précise dans quelles conditions la distribution de cette suite d'estimations converge vers la distribution d'échantillonnage de ces estimateurs dans Ω . Ces méthodes utilisent en particulier le postulat que les observations sont i.i.d. .

A) Méthode du Jackknife

Cette méthode, fut proposée par Quenouille (1949)[51] et reprise par Tukey (1958[60]). Le terme de Jackknife est dû à Tukey (1958) qui entendait décrire une approche d'inférence statistique générale, utilisable pour construire des tests d'hypothèse et des intervalles de confiance dans la situation où aucune méthode paramétrique n'est facilement utilisable. Un aspect plus théorique et général de la méthode est détaillé par Miller ([43]) et Efron ([15]).

Une façon de justifier l'idée du Jackknife consiste à réfléchir à ce qui est fait pour estimer la valeur d'une moyenne, d'un point de vue un peu inhabituel. Supposons, en effet qu'on dispose d'un échantillon de taille n aléatoire x_1, x_2, \dots, x_n représentatif. La moyenne empirique est classiquement :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Cette formule fournit une estimation de la moyenne de l'échantillon. Pour chaque élément x_j de l'échantillon, on peut calculer la moyenne empirique en enlevant la j -ème observation de l'échantillon, on obtient ainsi les :

$$\bar{x}_{-j} = \frac{1}{n-1} \left(\sum_{i=1}^n x_i - x_j \right) \quad (1.73)$$

À partir des deux dernières équations on peut ré-écrire x_j :

$$x_j = n\bar{x} - (n-1)\bar{x}_{-j}$$

Ceci montre qu'il est possible de déterminer la valeur x_j à partir de la moyenne globale sur toutes observations (\bar{x}) et de la moyenne lorsqu'on a enlevé la valeur x_j (\bar{x}_{-j}). Ce résultat n'est pas utile si les valeurs d'échantillon sont connues en premier lieu. Cependant, il est potentiellement utile dans des situations plus complexes où le paramètre à estimer est autre chose qu'une moyenne. Ainsi, dans la situation générale où un paramètre θ est estimé par une fonction des valeurs de l'échantillon de taille n observé. Cette estimation est notée par $s = s(x_1, \dots, x_n)$. En enlevant la j -ème observation, on obtient une estimation partielle, notée s_{-j} . Par analogie avec l'équation (1.73) on construit alors un ensemble des pseudo-valeurs pour $j = 1, \dots, n$:

$$s_j^* = ns - (n-1)s_{-j} \quad (1.74)$$

Ces pseudo-valeurs jouent un rôle analogue aux valeurs observées x_j sur un échantillon. On peut donc se poser dans ce cadre des questions statistiques telles l'estimation d'une moyenne et la moyenne des pseudo-valeurs

$$s^* = \frac{1}{n} \sum_{j=1}^n s_j^*. \quad (1.75)$$

Cela définit la méthode d'estimation de θ au sens du Jackknife. L'utilisation des pseudo-valeurs en tant qu'un échantillon aléatoire (d'estimations indépendantes) suggère alors que l'on peut estimer la variance de cette estimation du Jackknife par S^2/n où S^2 est la variance empirique des pseudo-valeurs. De plus, on peut construire un intervalle de confiance pour θ au niveau $100(1-\alpha)\%$ par : $S^* \pm t_{\alpha/2, n-1} s/\sqrt{n}$.

B) La méthode du Jackknife en régression linéaire simple

Reprenons le modèle (1.25) $Y_i = \beta_0 + \beta_1 x_i$, $i = 1, 2, \dots, n$. On peut décrire ce modèle sous forme matricielle par

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

où $\mathbf{X} = (\mathbf{1}, \mathbf{x})$, ${}^t\mathbf{x} = (x_1, x_2, \dots, x_n)$ et ${}^t\boldsymbol{\beta} = (\beta_0, \beta_1)$. L'estimateur MCO de $\boldsymbol{\beta}$ est (cf. Draper et Smith [11]) classiquement :

$$\hat{\boldsymbol{\beta}} = ({}^t\mathbf{X}\mathbf{X})^{-1} {}^t\mathbf{X}\mathbf{Y} \quad (1.76)$$

On suppose que $\mathbb{E}(\boldsymbol{\epsilon}|\mathbf{X}) = \mathbf{0}$, $\mathbb{E}(\epsilon_i \epsilon_j | \mathbf{X}) = 0$, $i \neq j = 1, \dots, n$, et $\text{Var}(\epsilon_i | \mathbf{X}) = \sigma^2$, $i = 1, \dots, n$

L'estimateur du Jackknife pour ce modèle linéaire est donné comme suit. Soit $\hat{\boldsymbol{\beta}}_j$ l'estimateur MCO de $\boldsymbol{\beta}$ obtenu en recalculant $\boldsymbol{\beta}$ dans (1.76) en supprimant le

j -ème couple (x_j, y_j) de l'échantillon. Pour estimer β on définit des pseudo-valeurs $\mathbf{p}_j = n\hat{\beta} - (n-1)\hat{\beta}_j$. L'estimateur ponctuel du Jackknife pour β est alors donné par :

$$\tilde{\beta} = \frac{1}{n} \sum_{j=1}^n \mathbf{p}_j \quad (1.77)$$

et l'estimateur du Jackknife pour $\text{Var}(\tilde{\beta})$ s'en déduit :

$$\widehat{\text{Var}(\tilde{\beta})}_J = \frac{1}{n(n-1)} \sum_{i=1}^n (\mathbf{p}_i - \tilde{\beta})^t (\mathbf{p}_i - \tilde{\beta}) \quad (1.78)$$

Les propriétés asymptotiques de $\tilde{\beta}$ et de $\widehat{\text{Var}(\tilde{\beta})}_J$ ont été étudiées par Miller (1974b [44]) sous certaines conditions fortes. D'autre part, Hinkley [27] a précisé deux imperfections de la méthode :

- a) $\tilde{\beta}$ est un estimateur sans biais mais sa variance est généralement plus grande que celle de l'estimateur MCO $\hat{\beta}$;
- b) $\widehat{\text{Var}(\tilde{\beta})}_J$ est un estimateur biaisé de $\text{Var}(\hat{\beta})$ et de $\text{Var}(\tilde{\beta})$.

Ces imperfections sont illustrées dans les simulations faites par Wu ([64]). Cet auteur montre en particulier que les problèmes mentionnés proviennent juste de la nature équilibrée de la méthode du Jackknife qui néglige la nature non équilibrée des données généralement modélisées en régression. Hinkley ([27]) a proposé une modification pour remédier à ces problèmes.

C) La méthode du Bootstrap

La motivation première de la méthode du Bootstrap est d'approcher par simulation la distribution d'un estimateur lorsque l'on ne connaît pas la loi de l'échantillon. L'idée de base consiste à remplacer des hypothèses probabilistes pas toujours vérifiées ou même invérifiables par des simulations et donc une connaissance empirique acquise après des calculs effectués sur les données observées. Le principe fondamental de cette technique de rééchantillonnage est de substituer à la distribution de probabilité inconnue F , dont est issu l'échantillon d'apprentissage, la distribution empirique \hat{F}_n qui donne un poids $1/n$ à chaque réalisation. Ainsi, on obtient un échantillon de taille n , dit échantillon Bootstrap, distribué selon la distribution empirique \hat{F}_n puisqu'il résulte de n tirages aléatoires avec remise. Ainsi en effet, dans chaque tirage toutes les observations de l'échantillon ont la même probabilité parmi les n observations initiales qui sont alors considérées comme formant une population.

Lorsqu'on dispose de moyens de calculs, il est facile de construire un grand nombre d'échantillons Bootstrap sur lesquels on peut calculer l'estimateur concerné. Sous des hypothèses peu contraignantes, assurant que l'échantillon est indépendant de même loi et que l'estimateur est indépendant de l'ordre des observations, la loi

simulée de cet estimateur est une approximation asymptotiquement convergente de la loi de l'estimateur (*cf.* Efron [16]).

Reprenons l'équation (1.72) où $\hat{\mu}$ est un estimateur de μ . Étant donnée une caractéristique d'intérêt, suivant une distribution F de moyenne inconnue μ sur les individus de la population Ω , afin d'étudier la distribution d'échantillonnage de $\hat{\mu}$ et celle de la variance de $\hat{\mu}$, une approximation pourrait se construire de la manière suivante.

- 1) On tire un nombre B d'échantillons de taille ns de cette population et on calcule B estimations $\hat{\mu}_1, \dots, \hat{\mu}_B$ de la moyenne inconnue μ .
- 2) La distribution empirique de ces estimations $\hat{\mu}_1, \dots, \hat{\mu}_B$ est une approximation de la distribution d'échantillonnage de l'estimateur $\hat{\mu}$.

De plus un estimateur de la variance $\text{Var}(\hat{\mu})$ s'en déduit :

$$\widehat{\text{Var}}(\hat{\mu}) = \frac{1}{B-1} \sum_{k=1}^B (\hat{\mu}_k - \bar{\hat{\mu}})^2$$

où $\bar{\hat{\mu}} = \frac{1}{B} \sum \hat{\mu}_k$ est la moyenne des B estimations de μ . Remarquons que pour connaître la véritable variance $\text{Var}(\hat{\mu})$, on doit évaluer tous les échantillons de taille ns possibles tirés selon la même procédure dans la population Ω , soit $\tilde{k} = \binom{n}{K}$ si Ω est de cardinal K . Cela suppose un très grand nombre de calculs. De plus, cette démarche n'est pas réalisable car elle nécessite plusieurs échantillons alors que la procédure réelle n'en produit qu'un seul. La méthode du Bootstrap permet de remédier à ce problème grâce à la simulation d'échantillons. Dans un cadre plus général, considérons un paramètre θ et la statistique (ou estimateur) S d'intérêt, ayant une distribution d'échantillonnage notée F_S . La procédure suivante construit une distribution empirique qui fournit une approximation de la distribution d'échantillonnage de (F_S) .

1. On tire B échantillons de taille n aléatoires simples avec remise parmi les observations de l'échantillon initial.
2. Pour tous ces échantillons, on calcule les estimations $\hat{s}_1^*, \hat{s}_2^*, \dots, \hat{s}_B^*$
3. A partir de $s_1^*, s_2^*, \dots, s_B^*$, on construit la distribution empirique, dite distribution Bootstrap, qui est une approximation de F_S .

Remarquons qu'on peut calculer une approximation de $\text{Var}(S)$ à l'aide de $s_1^*, s_2^*, \dots, s_B^*$:

$$\widehat{\text{Var}}(S)^* = \frac{1}{B-1} \sum_{k=1}^B (s_k^* - \bar{s}^*)^2$$

où $\bar{s}^* = \sum_{i=1}^B s_i^*$. A l'aide de cette distribution on peut réaliser un test d'hypothèse ou construire un intervalle de confiance pour le paramètre θ .

D) Le test du Bootstrap en régression linéaire simple

L'objet de cette section est d'étudier la liaison entre les variables Y et X . Soit un échantillon de taille n indépendant du couple (X, Y) , noté $\{(x_i, y_i), i = 1, 2, \dots, n\}$. Afin de comparer cette méthode avec la méthode des tests de permutation, nous la décomposons en cinq étapes.

Étape 1- Analyse du problème

Reprenons le modèle 1.24 : $Y = \beta_0 + \beta_1 x + \epsilon$. Comme dans le cas de la méthode permutationnelle, on veut tester l'hypothèse $H_0 : \beta_1 = 0$ contre $H_1 : \beta_1 \neq 0$. On réalise une régression linéaire de la variable Y en x pour le modèle 1.24. La connaissance des estimations respectives b_0 et b_1 de β_0 et β_1 , permet alors de calculer les prévisions :

$$\hat{y}_i = b_0 + b_1 x_i$$

et les résidus observés associés e_i comme suit :

$$e_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_i$$

Étape 2-La statistique de test

Contrairement au cas permutationnel, dans cette étape la statistique de test est spécifiquement un estimateur du paramètre β_1 . Par exemple :

$$S^B = \hat{\beta}_1$$

Étape 3-Le calcul de la statistique

Ici, comme dans la méthode de test de permutation, on calcule la statistique de test sur les observations d'origine. La valeur observée de cette statistique qui est déjà calculée dans la première étape est notée :

$$b_{1,obs} = b_1$$

Nous nous référons désormais à cette valeur à l'aide de la notation $\hat{\beta}_1$.

Étape 4-Le tirage avec remise des observations

Permuter les observations de Y peut être interprété comme un tirage sans remise d'un échantillon de taille n , dans une population de taille n , elle aussi. La méthode permutationnelle est donc interprétable comme une méthode de rééchantillonnage.

Le Bootstrap s'en différencie par le mode de tirage adopté. On effectue maintenant un tirage de B échantillons aléatoires simples de taille n avec remise dans l'ensemble des observations initiales, où B est fixé - grand - par l'utilisateur. Dans le cas de la régression simple, Il existe deux façons différentes de procéder au tirage. Dans la première, on tire un échantillon de taille n simple avec remise parmi les

résidus calculés à partir des observations initiales (dans l'étape 1). Cela correspond à une définition du modèle de la régression de la forme 1.24, où la loi des résidus est première dans l'analyse statistique. Dans la deuxième, on tire un échantillon de taille n aléatoire simple avec remise parmi des paires observées (x_i, y_i) initiales. Cela correspond à une spécification conditionnelle du modèle, de la forme :

$$\mathbb{E}(Y|X = x) = \beta_0 + \beta_1 x$$

où les résidus se déduisent de la loi conjointe de (X, Y) .

1) Tirage dans l'ensemble des résidus : Cette première façon de procéder opère par rééchantillonnage des résidus de la régression effectuée sur les données d'origine :

- a) Pour $k = 1$, on tire un échantillon de taille n aléatoire simple avec remise des résidus observés e_i , que l'on note :

$$\mathbf{e}_k^* = (e_1^*, e_2^*, \dots, e_n^*)$$

et l'on construit alors les nouvelles pseudo-valeurs de Y comme suit :

$$Y_i^* = b_0 + b_1 x_i + e_i^*$$

- b) On réalise alors une régression linéaire de Y^* en x d'après le modèle 1.24 appliqué au pseudo-échantillon $\{(x_i, Y_i^*); i = 1, 2, \dots, n\}$ obtenu et on en déduit une nouvelle estimation des paramètres β_0 et β_1 , notée b_0^{*k} et b_1^{*k} .
- c) On reproduit les étapes a) et b) un grand nombre de fois $k = 1, 2, \dots, B$ de sorte qu'on obtient en particulier la distribution de rééchantillonnage du paramètre β_1 , en répétant B fois les parties (a) et (b). Alors on obtient un vecteur des valeurs b_1^* , noté par $\mathbf{B}_1^* = (b_1^{*1}, b_1^{*2}, \dots, b_1^{*B})$ associées à $k = 1, 2, \dots, B$.

2) Le tirage dans l'ensemble des couples observés :

Cette deuxième façon de procéder traite directement les couples (X, Y) :

- a) Pour $k = 1$, on tire un échantillon de taille n aléatoire simple avec remise des couples observés (x_i, y_i) , $i = 1, \dots, n$, que l'on note :

$$(\mathbf{x}^*, \mathbf{y}^*)_k = ((x_1^*, y_1^*), (x_2^*, y_2^*), \dots, (x_n^*, y_n^*))$$

alors on obtient les nouvelles pseudo-valeurs de Y et de x notées $\mathbf{Y}^* = \{y_1^*, \dots, y_n^*\}$ et $\mathbf{x}^* = \{x_1^*, \dots, x_n^*\}$.

- b) On réalise alors une régression linéaire de \mathbf{Y}^* en \mathbf{x}^* utilisant le modèle 1.24 et on en déduit une nouvelle estimation des paramètres β_0 et β_1 notée b_0^{*k} et b_1^{*k} .
- c) Cette étape est la même que l'étape (c) décrite pour le cas du tirage des résidus.

Étape 5- Conclusion du test

Cette étape est analogue à celle qui est adoptée par la méthode de test de permutation. Pour effectuer un test de Bootstrap bilatéral à partir des valeurs obtenues b_1^* , on construit la distribution des valeurs $|b_1^*|$ rangées par ordre croissant. Puis, la p -valeur Bootstrap, notée (p_1^B) , est calculée comme la proportion des valeurs b_1^{*k} qui sont supérieures ou égales à la valeur observée sur l'échantillon dans cette distribution, c'est-à-dire :

$$\hat{p}^B = \frac{\#(|b_1^{*k}| \geq |b_1^{obs}|)}{B}$$

Si la valeur obtenue \hat{p}_1^B est inférieure au niveau α considéré, on rejette l'hypothèse $H_0 : \beta_1 = 0$ au profit de l'alternative $H_1 : \beta_1 \neq 0$.

En résumé, la comparaison des méthodes de test de permutation et des test de Bootstrap montre deux différences. La première apparaît dans l'étape 2 de choix de la statistique de test. En effet, pour la méthode de permutation on considère une statistique pivotale ou une statistique équivalente (cf. 1.6.2) alors que pour la méthode de Bootstrap, la statistique est un estimateur du paramètre considéré.

La deuxième différence tient à l'étape 4 puisque la méthode de permutation permute des observations y_i alors que la méthode de Bootstrap tire un échantillon aléatoire simple avec remise parmi des couples (x_i, y_i) observées ou bien parmi les résidus observés calculé à partir des observations initiales.

1.7 La régression linéaire multiple

Le modèle de régression linéaire simple suppose que la variable à expliquer Y ne dépend que d'un seul prédicteur X . Nous abordons maintenant le cas de modèles de régression prenant en compte la possibilité d'effet de plusieurs prédicteurs X^1, X^2, \dots, X^p , $p > 1$.

1.7.1 Le modèle classique de régression linéaire multiple

On considère un vecteur aléatoire donné par le k -uplet $(X^1, X^2, \dots, X^p, Y)$ et défini sur un espace probabilisé (Ω, \mathcal{F}, P) :

$$(X^1, X^2, \dots, X^p, Y) : \Omega \longrightarrow \mathfrak{R}^k$$

Ici, $k = p + 1$. Le modèle de régression multiple permet de prévoir les variations de la variable Y en fonction d'une réalisation de (X^1, X^2, \dots, X^p) . La méthode revient à trouver la meilleure approximation de cette variable aléatoire par une combinaison linéaire des variables explicatives, au sens de la moyenne des carrés des écarts.

Soit $g(x^1, x^2, \dots, x^p)$ une fonction de (x^1, x^2, \dots, x^p) , qui associe à tout (x^1, x^2, \dots, x^p) une prévision $\hat{y} = g(x^1, x^2, \dots, x^p)$ de la valeur de Y . Lorsque la relation fonctionnelle qui lie Y aux X^1, X^2, \dots, X^p est linéaire, la fonction de prévision s'écrit

$$g(x) = \beta_0 + \beta_1 x^1 + \beta_2 x^2 + \dots + \beta_p x^p$$

Nous avons rappelé dans la section (1.3.2) que l'interprétation probabiliste de la fonction de régression exprime le modèle linéaire sous la forme :

$$g(x) = \mathbb{E}(Y|X^1 = x^1, X^2 = x^2, \dots, X^p = x^p) = \beta_0 + \beta_1 x^1 + \beta_2 x^2 + \dots + \beta_p x^p$$

Cette écriture dit d'une part que la valeur moyenne attendue de Y correspondant à x^1, x^2, \dots, x^p , notée $\mathbb{E}(Y|X^1 = x^1, X^2 = x^2, \dots, X^p = x^p)$ diffère de la valeur observée Y par un terme d'erreur que l'on note ϵ . Le modèle classique de régression linéaire multiple dit de plus que l'espérance conditionnelle $g(x)$ est une fonction linéaire des prédicteurs et donc :

$$Y = \beta_0 + \beta_1 x^1 + \beta_2 x^2 + \dots + \beta_p x^p + \epsilon \quad (1.79)$$

Remarquons qu'ici comme dans la section (1.4.1), on s'intéresse plus particulièrement au cas où les prédicteurs $X^j, j = 1, \dots, p$ sont fixés et on note x^1, x^2, \dots, x^p leurs valeurs observées respectives, supposées connues ou prédéterminées.

Lorsqu'on dispose de n réalisations du vecteur $(X^1, X^2, \dots, X^p, Y)$, le modèle s'écrit de façon analytique :

$$Y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \dots + \beta_p x_i^p + \epsilon_i, \quad i = 1, \dots, n \quad (1.80)$$

ou de façon équivalente, comme suit en utilisant les notations matricielles :

définition 1.7.1. *Un modèle de régression linéaire est défini par une équation de la forme :*

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (1.81)$$

$(n \times 1) \quad (n \times k)_{(k \times 1)} \quad (n \times 1)$

où $k = p + 1$ et

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \in \mathbb{R}^n, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \in \mathbb{R}^n$$

$$\mathbf{X} = \begin{bmatrix} {}^t \mathbf{x}_1 \\ {}^t \mathbf{x}_2 \\ \vdots \\ {}^t \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} 1 & x_1^1 & \dots & x_1^p \\ 1 & x_2^1 & \dots & x_2^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^1 & \dots & x_n^p \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \in \mathbb{R}^k$$

Ces diverses formes montrent que la valeur observée du vecteur aléatoire \mathbf{Y} est la somme d'une composante déterministe $\mathbf{X}\boldsymbol{\beta}$ et d'une composante aléatoire $\boldsymbol{\epsilon}$ qui modélise les variations non prises en compte par le modèle.

Afin d'estimer les paramètres $(\boldsymbol{\beta}, \sigma)$ du modèle classique de régression linéaire multiple et de préciser les propriétés statistiques des estimateurs retenus, on doit formuler certains postulats caractérisant la distribution du terme d'erreur. Nous reprenons ici la formulation retenue par Hayashi [26]) :

P1. (Exogénéité forte) : L'espérance conditionnelle de ϵ_i sachant toutes les observations x_i^k des variables explicatives est nulle.

$$\mathbb{E}(\epsilon_i|\mathbf{X}) = \mathbb{E}(\epsilon_i|x_1, x_2, \dots, x_n) = 0, \quad \forall i, i = 1, 2, \dots, n$$

Pour tout triplet (i,j,k) tel que $i, j = 1, 2, \dots, n$ et $k = 1, 2, \dots, p$, on en déduit les relations

- a) $\mathbb{E}(\epsilon_i) = \mathbb{E}[\mathbb{E}(\epsilon_i|\mathbf{X})] = 0$ ⁵
- b) $\mathbb{E}(Y_i|x_i^1, x_i^2, \dots, x_i^p) = \beta_0 + \beta_1 x_i^1 + \dots + \beta_p x_i^p$
- c) $\mathbb{E}(X_j^k \epsilon_i) = 0$ ⁶
- d) $Cov(X_j^k, \epsilon_i) = 0$ ⁷

P2. La matrice \mathbf{X} est de rang plein, c'est-à-dire que la matrice carrée ${}^t\mathbf{X}\mathbf{X}$ d'ordre k est inversible.

P3. propriétés d'homoscédasticité :

$$\mathbb{E}(\epsilon_i^2|\mathbf{X}) = \sigma^2 > 0 \quad i, j = 1, 2, \dots, n$$

P4. propriétés de non corrélation des erreurs :

$$\mathbb{E}(\epsilon_i, \epsilon_j|\mathbf{X}) = 0 \quad i \neq j = 1, 2, \dots, n$$

On déduit des postulats P1, P2, P3 et P4, les propriétés suivantes, vraies pour tous $(i, j) \in \{1, 2, \dots, n\}^2$:

- a) $Var(\epsilon_i|\mathbf{X}) = \mathbb{E}(\epsilon_i^2|\mathbf{X}) - (\mathbb{E}(\epsilon_i|\mathbf{X}))^2 = \mathbb{E}(\epsilon_i^2|\mathbf{X}) = \sigma^2$
- b) $Cov(\epsilon_i, \epsilon_j|\mathbf{X}) = \mathbb{E}(\epsilon_i \epsilon_j|\mathbf{X}) - \mathbb{E}(\epsilon_i|\mathbf{X})\mathbb{E}(\epsilon_j|\mathbf{X}) = \mathbb{E}(\epsilon_i \epsilon_j|\mathbf{X}) = 0$
- c) $Var(Y_i|\mathbf{X}) = Var(\epsilon_i|\mathbf{X}) = \sigma^2$
- d) $Cor(Y_i, Y_j|\mathbf{X}^1, \mathbf{X}^2) = 0, \quad i \neq j$

1.7.2 Le modèle de régression linéaire gaussien

Complétons notre étude du modèle 1.81 : $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ en supposant satisfait un cinquième postulat :

P5. La loi conditionnelle de $\boldsymbol{\epsilon}$ sachant \mathbf{X} est gaussienne :

$$\boldsymbol{\epsilon}|\mathbf{X} \sim LG(\mathbf{0}, \sigma^2\mathbf{I})$$

Lorsque les ϵ_i sont i.i.d. de loi gaussienne conditionnellement à \mathbf{X} , \mathbf{Y} est alors distribuée selon une loi de Gauss, conditionnellement à \mathbf{X} :

$$\mathbf{Y}|\mathbf{X} \sim LG(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \tag{1.82}$$

⁵ $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|\mathbf{X}))$

⁶ $\mathbb{E}(X_j^k \epsilon_i) = \mathbb{E}[\mathbb{E}(X_j^k \epsilon_i | X_j^k)] = \mathbb{E}[X_j^k \mathbb{E}(\epsilon_i | X_j^k)] = 0$

⁷ $Cov(X_j^k, \epsilon_i) = \mathbb{E}(X_j^k \epsilon_i) - \mathbb{E}(X_j^k)\mathbb{E}(\epsilon_i) = \mathbb{E}(X_j^k \epsilon_i) = 0$

A) Estimation au sens du maximum de vraisemblance

Les postulats $P1 - P5$ permettent de connaître la loi de probabilité conjointe de l'échantillon, c'est-à-dire la distribution du vecteur \mathbf{Y} étant donné la valeur de \mathbf{X} et donc d'estimer le paramètre $(\boldsymbol{\beta}, \sigma^2)$, par la méthode du maximum de vraisemblance. En effet, la fonction de vraisemblance du vecteur gaussien \mathbf{Y} est

$$L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = \frac{-1}{(2\pi\sigma^2)^{n/2}} \exp\left\{\frac{1}{2\sigma^2} {}^t(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

et le logarithme népérien de la fonction vraisemblance s'écrit :

$$l(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = \ln(L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y})) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}Q \quad (1.83)$$

où $Q = {}^t(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ est la somme des carrés des erreurs. Par suite, quelle que soit la valeur de σ^2 , maximiser $L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y})$ par rapport à $\boldsymbol{\beta}$ revient à minimiser Q . On annule donc la dérivée partielle Q par rapport à $\boldsymbol{\beta}$

$$\frac{\partial}{\partial \boldsymbol{\beta}} Q = -2 {}^t \mathbf{X} \mathbf{Y} + 2 {}^t \mathbf{X} \mathbf{X} \boldsymbol{\beta} = 0$$

on obtient ainsi l'estimateur des moindres carrés ordinaires (MCO) et de maximum de vraisemblance (MV) de $\boldsymbol{\beta}$ dans un cadre gaussien :

$$\hat{\boldsymbol{\beta}} = ({}^t \mathbf{X} \mathbf{X})^{-1} {}^t \mathbf{X} \mathbf{Y} \quad (1.84)$$

Remarque 1.7.1. *Sous le postulat (P5) (cas gaussien), l'estimateur $\hat{\boldsymbol{\beta}}$ est de variance minimum parmi les estimateurs sans biais (non nécessairement linéaires) de $\boldsymbol{\beta}$ (c.f. Rao [52])*

On définit $Q(\hat{\boldsymbol{\beta}}) = {}^t(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ comme la somme des carrés des erreurs empiriques. De plus, si l'on annule la dérivée partielle $l(\boldsymbol{\beta}, \sigma^2; \mathbf{Y})$ par rapport à σ^2

$$\frac{\partial}{\partial \sigma^2} l(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} Q = 0 \quad (1.85)$$

On obtient donc l'estimateur de maximum de vraisemblance de σ^2 :

$$\tilde{\sigma}^2 = \frac{1}{n} Q(\hat{\boldsymbol{\beta}}) = \frac{1}{n} {}^t(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (1.86)$$

$\tilde{\sigma}^2$ est la moyenne empirique des carrés des erreurs :

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - {}^t \mathbf{x}_i \hat{\boldsymbol{\beta}})^2$$

Remarque 1.7.2. $\tilde{\sigma}^2$ est un estimateur biaisé de σ^2 . On obtient un estimateur sans biais en corrigeant son espérance par le coefficient multiplicatif $n/(n-k)$ (cf. Antoniadis et al. [2]). On obtient ainsi la statistique :

$$\hat{\sigma}^2 = \frac{n}{n-k} \tilde{\sigma}^2 = \frac{1}{n-k} {}^t(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (1.87)$$

On montre alors que $\hat{\sigma}^2$ est bien un estimateur sans biais de σ^2 et qu'il s'interprète comme la moyenne des carrés des erreurs, à la correction près du biais.

Proposition 1.7.1. Sous l'hypothèse gaussienne (P5)

a) L'estimateur $\hat{\boldsymbol{\beta}}$ est un vecteur gaussien :

$$\hat{\boldsymbol{\beta}} \sim LG(\boldsymbol{\beta}, \sigma^2({}^t\mathbf{X}\mathbf{X})^{-1})$$

b) L'estimateur $\hat{\sigma}^2$ vérifie :

$$\frac{(n-k)\hat{\sigma}^2}{\sigma^2} = \frac{{}^t(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{\sigma^2} \sim \chi_{n-k}^2$$

c) Les estimateurs $\hat{\boldsymbol{\beta}}$ et $\hat{\sigma}^2$ sont indépendants

Preuve : cf. Antoniadis et al. [2] ou d'Aubigny [8]

Remarque 1.7.3. Puisque $\hat{\sigma}^2$ est un estimateur sans biais de σ^2 , alors un estimateur sans biais de $\text{Var}(\hat{\boldsymbol{\beta}})$ est donné par :

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2({}^t\mathbf{X}\mathbf{X})^{-1} \quad (1.88)$$

B) Le modèle de régression linéaire double

Nous particularisons le modèle (1.81) au cas de deux variables explicatives afin d'introduire les expressions analytiques utilisées dans la suite. Le modèle s'écrit alors de façon analytique comme :

$$Y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \epsilon_i \quad (1.89)$$

Dans ce cas, l'équation (1.84) dit que les estimateurs de maximum de vraisemblance de β_0, β_1 et β_2 sont (cf. Johnston et al. [30]) :

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}^1 - \hat{\beta}_2 \bar{x}^2 \quad (1.90)$$

$$\hat{\beta}_1 = \frac{S^2(X^2)S(X^1, Y) - S(X^2, Y)S(X^1, X^2)}{S^2(X^1)S^2(X^2) - [S(X^1, X^2)]^2} \quad (1.91)$$

$$\hat{\beta}_2 = \frac{S^2(X^1)S(X^2, Y) - S(X^1, Y)S(X^1, X^2)}{S^2(X^1)S^2(X^2) - [S(X^1, X^2)]^2} \quad (1.92)$$

$$(1.93)$$

où

1. $S(X^1, X^2) = \frac{1}{n} \sum_i^n (x_i^1 - \bar{x}^1)(x_i^2 - \bar{x}^2)$
2. $S^2(X^j) = S(X^j, X^j) = \frac{1}{n} \sum_i^n (x_i^j - \bar{x}^j)^2, \quad j = 1, 2$
3. $S(X^j, Y) = \frac{1}{n} \sum_i^n (x_i^j - \bar{x}^j)(Y_i - \bar{Y}), \quad j = 1, 2$

De plus, les variances de $\hat{\beta}_0, \hat{\beta}_1$ et $\hat{\beta}_2$ s'écrivent :

1. $\text{Var}(\hat{\beta}_0) = \frac{\sigma^2}{n} \times \frac{1}{S^2(X^1)S^2(X^2) - [S(X^1, X^2)]^2}$
2. $\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{n} \times \frac{S^2(X^2)}{S^2(X^1)S^2(X^2) - [S(X^1, X^2)]^2}$
3. $\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{n} \times \frac{S^2(X^1)}{S^2(X^1)S^2(X^2) - [S(X^1, X^2)]^2}$

1.7.3 Estimation au sens des moindres carrés ordinaires

Les calculs précédents montrent que la méthode des moindres carrés ordinaires se généralise sans problème au cas de la régression linéaire multiple. Il faut alors chercher le vecteur $\hat{\beta} \in \mathbb{R}^k$ (où $k = p + 1$) tel que $\mathbf{X}\hat{\beta}$ soit le plus proche possible de \mathbf{Y} au sens de la norme euclidienne classique. On est donc amené à minimiser la norme euclidienne standard du vecteur résiduel $\epsilon = \mathbf{Y} - \mathbf{X}\hat{\beta}$ évalué en $\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$ choisi parmi tous les vecteurs qui s'écrivent $\mathbf{X}\beta$. Trouver l'estimateur de β au sens MCO conduit à $\hat{\beta}$ qui réalise le minimum de l'expression :

$$Q = {}^t(\mathbf{Y} - \mathbf{X}\beta)(\mathbf{Y} - \mathbf{X}\beta) = \|\mathbf{Y} - \mathbf{X}\beta\|^2$$

c'est la quantité que nous avons minimisée dans le cas gaussien pour trouver l'estimateur du maximum de vraisemblance. Dans le cas d'une régression linéaire multiple, les estimateurs de β optimaux au sens du maximum de vraisemblance et des MCO sont identiques.

Propriétés des estimateurs MCO

On rappelle maintenant les principales propriétés d'optimalité statistique des estimateurs MCO. Ceux-ci étendent au cadre de la régression multiple les résultats obtenus dans la section (1.4.4) pour le cas d'un seul prédicteur. La distribution d'échantillonnage de β vérifie la proposition suivante :

Proposition 1.7.2. *Notons encore ici $\hat{\beta}$ l'estimateur MCO de β*

1. *Sous les postulats P1-P2, $\hat{\beta}$ est un estimateur sans biais : $\mathbb{E}(\hat{\beta}|\mathbf{X}) = 0$*

2. Sous les postulats P1-P4 : $\text{Var}(\hat{\beta}|\mathbf{X}) = \sigma^2 ({}^t\mathbf{X}\mathbf{X})^{-1}$

3. Sous les postulats P1-P4 : $\text{Cov}(\hat{\beta}|\mathbf{X}) = \mathbf{0}$

Preuve : (cf. Hayashi [26])

On rappelle de plus la propriété fondamentale d'optimalité des estimateurs MCO de β .

Théorème 1.7.1. (Gauss-Markov)

Sous les postulats P1 – P4, l'estimateur $\hat{\beta}$ est efficace dans la classe des estimateurs linéaires et sans biais. Cela veut dire que pour tout estimateur $\tilde{\beta}$ de β qui est linéaire en Y , on a $\text{Var}(\tilde{\beta}) \geq \text{Var}(\hat{\beta})$ au sens matriciel.

Preuve : (cf. Hayashi [26])

1.8 Inférence statistique en régression linéaire multiple

Le test classique de l'hypothèse de nullité des coefficients de régression, fait l'objet de cette section.

1.8.1 Test de nullité d'un coefficient de régression linéaire

A) Cas de la régression linéaire multiple avec $p > 2$ prédicteurs

Reprenons le modèle de régression linéaire dans un cadre gaussien (1.81)

$$\underset{(n \times 1)}{\mathbf{Y}} = \underset{(n \times k)}{\mathbf{X}} \underset{(k \times 1)}{\beta} + \underset{(n \times 1)}{\epsilon}, \quad \epsilon \sim LG(\mathbf{0}, \sigma^2 \mathbf{I})$$

et considérons tout d'abord une hypothèse de nullité du j -ème coefficient de régression

$$H_0 : \beta_j = 0 \text{ contre } H_0 : \beta_j \neq 0$$

nous voulons soumettre cette hypothèse à un test de signification. La proposition ((1.7.1)) assure que :

$$(\hat{\beta} - \beta) \sim LG(\mathbf{0}, \sigma^2 ({}^t\mathbf{X}\mathbf{X})^{-1}) \quad (1.94)$$

On en déduit la loi de probabilité de l'estimateur β_j

$$(\hat{\beta}_j - \beta_j) \sim LG(0, \sigma^2 ({}^t\mathbf{X}\mathbf{X})_{jj}^{-1})$$

où $({}^t\mathbf{X}\mathbf{X})_{jj}^{-1}$ est le terme diagonal en position (j, j) dans $({}^t\mathbf{X}\mathbf{X})^{-1}$. Par suite, lorsque σ^2 est connue, la statistique :

$$Z_j = \frac{\hat{\beta}_j}{\sqrt{\sigma^2 ({}^t\mathbf{X}\mathbf{X})_{jj}^{-1}}} \quad (1.95)$$

suit une loi de probabilité gaussienne $LG(0, 1)$ sous l'hypothèse $H_0 : \beta_j = 0$. La consultation d'une table de cette loi conduit à rejeter H_0 si $|Z_j| > z_{\frac{\alpha}{2}}$ où α est le niveau du test, où $z_{\frac{\alpha}{2}}$ est le fractile de la loi de Z correspondant au niveau de probabilité $1 - \alpha/2$. Lorsque σ^2 est inconnue, $\text{Var}(\hat{\beta}_j)$ est elle aussi inconnue, et pour estimer $\text{Var}(\hat{\beta}_j)$ on doit estimer σ^2 . A partir de la remarque (1.7.2), on vérifie qu'un estimateur sans biais de σ^2 est :

$$\hat{\sigma}^2 = \frac{1}{n-k} {}^t(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

et alors,

$$S^2(\hat{\beta}_j) = \widehat{\text{Var}(\hat{\beta}_j)} = \hat{\sigma}^2 ({}^t\mathbf{X}\mathbf{X})_{jj}^{-1}$$

est l'estimateur souhaité de $\text{Var}(\hat{\beta}_j)$. Par suite, on montre de façon classique que la statistique :

$$T = \hat{Z}_j = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 ({}^t\mathbf{X}\mathbf{X})_{jj}^{-1}}} \quad (1.96)$$

suit une loi de probabilité de Student à $(n-k)$ degrés de liberté. Le test de l'hypothèse H_0 de niveau α conduit donc à consulter une table de loi de Student à $(n-k)$ degrés de liberté.

B) Cas de la régression linéaire double

Pour ce cas particulier de la section précédente le modèle de régression linéaire double gaussien (1.89) s'écrit :

$$Y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \epsilon_i$$

Pour effectuer un test d'hypothèse de nullité du coefficient de régression partielle : $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$, lorsque σ^2 est connue, on doit calculer

$$\text{Var}(\hat{\beta}_2) = \sigma^2 ({}^t\mathbf{X}\mathbf{X})_{22}^{-1}$$

et sous l'hypothèse $H_0 : \beta_2 = 0$, la statistique

$$Z_2 = \frac{\hat{\beta}_2}{\sqrt{\sigma^2 ({}^t\mathbf{X}\mathbf{X})_{22}^{-1}}} \quad (1.97)$$

suit une loi de Gauss $LG(0, 1)$. Le test de l'hypothèse H_0 de niveau α conduit donc à consulter une table de la loi de Gauss centrée réduite. Mais σ^2 est en général inconnue, de sorte que $\text{Var}(\hat{\beta}_j)$ l'est aussi. Un estimateur sans biais de $\text{Var}(\hat{\beta}_j)$ est :

$$S^2(\hat{\beta}_2) = \widehat{\text{Var}(\hat{\beta}_2)} = \hat{\sigma}^2 ({}^t\mathbf{X}\mathbf{X})_{22}^{-1}$$

où l'on déduit de (1.7.2) que

$$\hat{\sigma}^2 = \frac{1}{n-3} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i^1 - \beta_2 x_i^2)^2$$

est un estimateur sans biais de σ^2 . Sous $H_0 : \beta_2 = 0$, la statistique de test

$$T = \hat{Z}_2 = \frac{\hat{\beta}_2}{S(\hat{\beta}_2)} \quad (1.98)$$

suit la loi probabilité de Student à $(n-3)$ degrés de liberté (*cf.* Johnston et al. [30]) où $\hat{\beta}_2$ est l'estimateur MCO de β_2 et $S^2(\hat{\beta}_2)$ est l'estimateur MCO de $\text{Var}(\hat{\beta}_2)$. La consultation d'une table de loi de Student à $(n-3)$ degrés de liberté permet de rejeter H_0 si $|T| > t_{(n-3, \frac{\alpha}{2})}$ où α est le niveau du test.

1.8.2 Test de nullité d'un coefficient de corrélation partielle

Les méthodes présentées dans la section précédente ont pour point commun d'étudier l'apport spécifique d'un prédicteur X^j au modèle à l'aide de la valeur estimée du coefficient de régression partielle β_j . Nous montrons ici qu'une façon équivalente de présenter le problème conduit à accorder aux corrélations partielles un rôle fondamental dans la pratique de la régression. Ici, nous rappelons les aspects formels de cette caractéristique importante.

A) La corrélation partielle : cas de régression double

Considérons un modèle de régression linéaire multiple qui fait intervenir deux variables explicatives seulement :

$$Y = \beta_0 + \beta_1 x^1 + \beta_2 x^2 + \epsilon$$

Le coefficient de corrélation partielle de Y et X^2 sachant X^1 mesure l'information apportée par X^2 sur Y , une fois retirée l'influence (linéaire) de X^1 sur ces deux variables. Autrement dit, le coefficient de corrélation partielle mesure la liaison linéaire existant entre deux variables lorsque l'influence d'une troisième est contrôlée. Par suite, le coefficient de corrélation partielle est très utile en régression multiple, parce qu'il permet d'évaluer directement la part de la variation de Y non expliquée par X^1 qui devient expliquée par ajout de la variable X^2 dans le modèle.

Afin d'enlever l'effet de la variable X^1 sur les variables X^2 et Y on ajuste deux modèles de régression linéaire simple :

$$Y = \alpha_0 + \alpha_1 x^1 + \epsilon^{Y,1} \quad (1.99)$$

$$X^2 = \gamma_0 + \gamma_1 x^1 + \epsilon^{2,1} \quad (1.100)$$

On est alors conduit à prévoir les résidus $\epsilon^{Y.1}$ et $\epsilon^{2.1}$

$$R^{Y.1} = Y - \hat{\alpha}_0 - \hat{\alpha}_1 x^1 \quad (1.101)$$

$$R^{2.1} = X^2 - \hat{\gamma}_0 - \hat{\gamma}_1 x^1 \quad (1.102)$$

Ces modèles de régression linéaire destinés à séparer l'effet imputable aux variations de X^1 au travers de relations linéaires avec Y et X^2 respectivement et les résidus observés $R^{Y.1}$ et $R^{2.1}$ qui traduisent respectivement la part de variation de Y et X^2 qui n'est pas corrélée linéairement à X^1 . On ajuste alors un modèle de régression linéaire décrivant la liaison linéaire entre les résidus observés $R^{Y.1}$ et les résidus observés $R^{2.1}$

$$R^{Y.1} = \delta_0 + \delta_1 R^{2.1} + \eta \quad (1.103)$$

de façon à décrire la part de covariation de Y expliquée par X^2 qui est non corrélée à l'effet linéaire de X^1 , donc laissée inexpliquée (linéairement) par ce prédicteur. La corrélation partielle théorique de Y et X^2 à X^1 fixée, peut être définie directement à partir du coefficient de corrélation linéaire simple calculé entre deux variables $\epsilon^{Y.1}$ en $\epsilon^{2.1}$. Notons que ces deux variables sont centrées par construction et donc :

$$\rho_{Y2.1} = \rho(\epsilon^{Y.1}, \epsilon^{2.1}) = \frac{\mathbb{E}(\epsilon^{Y.1} \epsilon^{2.1})}{\sqrt{\mathbb{E}(\epsilon^{Y.1})^2 \mathbb{E}(\epsilon^{2.1})^2}} \quad (1.104)$$

de plus, la corrélation partielle empirique fournit un estimateur convergeant de ce paramètre défini comme :

$$r_{Y2.1} = \text{Corr}(R^{Y.1}, R^{2.1}) = \frac{\sum R_i^{Y.1} R_i^{2.1}}{\sqrt{\sum (R_i^{Y.1})^2 \sum (R_i^{2.1})^2}} \quad (1.105)$$

On peut aussi définir la corrélation partielle à partir des corrélations simples (ordinaires) entre les couples de variables expliquée et explicatives considérées (*cf.* Kendall et al. ([32])). Considérons en effet la matrice de corrélation liant les variables Y, X^1, X^2 et notée :

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{Y1} & \rho_{Y2} \\ \rho_{Y1} & 1 & \rho_{12} \\ \rho_{Y2} & \rho_{12} & 1 \end{pmatrix}$$

le coefficient de corrélation partielle entre Y et X^2 sachant la variable X^1 est noté $\rho_{Y2.1}$,

$$\rho_{Y2.1} = -\frac{C_{13}}{\sqrt{C_{11}C_{33}}} = \frac{\rho_{Y2} - \rho_{Y1}\rho_{12}}{\sqrt{(1 - \rho_{Y1}^2)(1 - \rho_{12}^2)}} \quad (1.106)$$

$$= \frac{\sigma_1^2 \sigma_{Y2} - \sigma_{Y1} \sigma_{12}}{\sqrt{\sigma_Y^2 \sigma_1^2 - \sigma_{Y1}^2} \sqrt{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2}} \quad (1.107)$$

Ainsi, le coefficient de corrélation partielle entre Y et X^1 sachant la variable X^2 , noté $\rho_{Y1.2}$, s'écrit :

$$\rho_{Y1.2} = -\frac{C_{12}}{\sqrt{C_{11}C_{22}}} = \frac{\rho_{Y1} - \rho_{Y2}\rho_{12}}{\sqrt{(1 - \rho_{Y2}^2)(1 - \rho_{12}^2)}} \quad (1.108)$$

$$= \frac{\sigma_2^2\sigma_{Y1} - \sigma_{Y2}\sigma_{12}}{\sqrt{\sigma_Y^2\sigma_2^2 - \sigma_{Y2}^2}\sqrt{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}} \quad (1.109)$$

où C_{mn} désigne le cofacteur du terme en (m, n) -ème position dans la matrice symétrique \mathbf{R} , et $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$, où $\sigma_{ij} = Cov(X_i, X_j)$, $\sigma_j^2 = Var(X^j)$, pour $\{i, j=1, 2\}$ et $\sigma_Y^2 = Var(Y)$.

Proposition 1.8.1. *Les équations (1.104) et (1.108) sont égales.*

Preuve : L'équation (1.104) a pour conséquence :

$$\begin{aligned} \rho_{Y2.1} &= \rho(\epsilon^{Y.1}, \epsilon^{2.1}) = \frac{\mathbb{E}(\epsilon^{Y.1}\epsilon^{2.1})}{\sqrt{\mathbb{E}(\epsilon^{Y.1})^2\mathbb{E}(\epsilon^{2.1})^2}} \\ &= \frac{\mathbb{E}(Y - \alpha_0 - \alpha_1 X^1)(X^2 - \gamma_0 - \gamma_1 X^1)}{\sqrt{\mathbb{E}(Y - \alpha_0 - \alpha_1 X^1)^2\mathbb{E}(X^2 - \gamma_0 - \gamma_1 X^1)^2}} \\ &= \frac{\mathbb{E}((Y - \mu_Y) - \alpha_1(X^1 - \mu_{X^1}))((X^2 - \mu_{X^2}) - \gamma_1(X^1 - \mu_{X^1}))}{\sqrt{\mathbb{E}((Y - \mu_Y) - \alpha_1(X^1 - \mu_{X^1}))^2\mathbb{E}((X^2 - \mu_{X^2}) - \gamma_1(X^1 - \mu_{X^1}))^2}} \\ &= \frac{\sigma_{y2} - \gamma_1\sigma_{y1} - \alpha_1\sigma_{12} + \alpha_1\gamma_1\sigma_1^2}{\sqrt{(\sigma_Y^2 + \alpha_1^2\sigma_1^2 - 2\alpha_1\sigma_{Y1})(\sigma_2^2 + \gamma_1^2\sigma_1^2 - 2\gamma_1\sigma_{12})}} \end{aligned}$$

D'après la proposition (1.3.1) on a de plus

$$\begin{aligned} \alpha_1 &= \rho_{Y1} \frac{\sigma_Y}{\sigma_{X^1}} \quad \text{et} \quad \alpha_0 = \mu_Y - \beta_1\mu_{X^1} = \mu_Y - \rho_{Y1} \frac{\sigma_Y}{\sigma_{X^1}} \mu_{X^1} \\ \gamma_1 &= \rho_{12} \frac{\sigma_{X^2}}{\sigma_{X^1}} \quad \text{et} \quad \gamma_0 = \mu_{X^2} - \gamma_1\mu_{X^1} = \mu_{X^2} - \rho_{12} \frac{\sigma_{X^2}}{\sigma_{X^1}} \mu_{X^1} \end{aligned}$$

alors, on obtient par substitution l'équation :

$$\rho_{Y2.1} = \frac{\sigma_1^2\sigma_{Y2} - \sigma_{Y1}\sigma_{12}}{\sqrt{\sigma_Y^2\sigma_1^2 - \sigma_{Y1}^2}\sqrt{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}} \quad (1.110)$$

$$= \frac{\rho_{Y2} - \rho_{Y2}\rho_{12}}{\sqrt{(1 - \rho_{Y2}^2)(1 - \rho_{12}^2)}} \quad (1.111)$$

Proposition 1.8.2. *Soit (Y, X^1, X^2) un triplet aléatoire (non nécessairement gaussien) de moyenne, variance et corrélations linéaires $\mu_Y, \mu_{X^1}, \mu_{X^2}, \sigma_Y^2, \sigma_{X^1}^2, \sigma_{X^2}^2, \rho_{Y1}, \rho_{Y2}$ et ρ_{12} . Si $\mathbb{E}(Y|X = x)$ est une fonction linéaire de x^1 et x^2 de la forme :*

$$\mathbb{E}(Y|X^1 = x^1, X^2 = x^2) = \beta_0 + \beta_1 x^1 + \beta_2 x^2$$

les paramètres β_0, β_1 et β_2 sont :

$$a) \beta_0 = \mu_Y - \beta_1 \mu_{X^1} - \beta_2 \mu_{X^2}$$

$$b) \beta_1 = \frac{\sigma_2^2 \sigma_{Y1} - \sigma_{Y2} \sigma_{12}}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} = \frac{\sigma_Y (\rho_{Y1} - \rho_{Y2} \rho_{12})}{\sigma_1 (1 - \rho_{12}^2)}$$

$$c) \beta_2 = \frac{\sigma_1^2 \sigma_{Y2} - \sigma_{Y1} \sigma_{12}}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} = \frac{\sigma_Y (\rho_{Y2} - \rho_{Y1} \rho_{12})}{\sigma_2 (1 - \rho_{12}^2)}$$

Preuve :

Elle repose sur deux propriétés. Tout d'abord, pour $j = 1, 2$, on a si $\mathbf{X} = (X^1, X^2)$:

$$\mu_{X^j} = E(X^j) = \mathbb{E}_{\mathbf{X}}(\mathbb{E}(X^j | \mathbf{X}))$$

D'autre part,

$$\mu_Z = \mathbb{E}(Z) = \mathbb{E}_{\mathbf{X}}(\mathbb{E}(Z | \mathbf{X}))$$

Par conséquent, le modèle s'écrit :

$$\mathbb{E}(Y | X^1, X^2) = \beta_0 + \beta_1 x^1 + \beta_2 x^2$$

et donc :

$$\begin{aligned} \mu_Y &= \mathbb{E}_{\mathbf{X}}(\mathbb{E}(Y | \mathbf{X})) \\ &= \mathbb{E}_{\mathbf{X}}(\mathbb{E}(\beta_0 + \beta_1 X^1 + \beta_2 X^2 | \mathbf{X})) \\ &= \beta_0 + \beta_1 \mathbb{E}_{\mathbf{X}}(\mathbb{E}(X^1 | \mathbf{X})) + \beta_2 \mathbb{E}_{\mathbf{X}}(\mathbb{E}(X^2 | \mathbf{X})) \\ &= \beta_0 + \beta_1 \mathbb{E}(X^1) + \beta_2 \mathbb{E}(X^2) \end{aligned}$$

c'est-à-dire

$$\mu_Y = \beta_0 + \beta_1 \mu_{X^1} + \beta_2 \mu_{X^2} \quad (1.112)$$

Opérons de façon analogue pour $Z = Y X^1$:

$$\begin{aligned} \mu_Z = \mathbb{E}(X^1 Y) &= \mathbb{E}_{\mathbf{X}}(\mathbb{E}(Y X^1 | \mathbf{X})) \\ &= \mathbb{E}_{\mathbf{X}}(\mathbb{E}((\beta_0 + \beta_1 X^1 + \beta_2 X^2) X^1 | \mathbf{X})) \\ &= \mathbb{E}_{\mathbf{X}}(\mathbb{E}(\beta_0 X^1 + \beta_1 (X^1)^2 + \beta_2 (X^1 X^2) | \mathbf{X})) \end{aligned}$$

Opérons comme ci-dessus, il vient :

$$\mathbb{E}(X^1 Y) = \beta_0 \mu_{X^1} + \beta_1 \mathbb{E}((X^1)^2) + \beta_2 \mathbb{E}(X^1 X^2) \quad (1.113)$$

Le cas de $Z = Y X^2$ est totalement symétrique et donc :

$$\mathbb{E}(X^2 Y) = \beta_0 \mu_{X^2} + \beta_1 \mathbb{E}(X^1 X^2) + \beta_2 \mathbb{E}((X^2)^2) \quad (1.114)$$

À partir des équations (3.2.1), (3.2.2) et (3.3.1) on peut alors écrire les trois équations :

$$\begin{aligned} \mu_Y &= \beta_0 + \beta_1 \mu_{X^1} + \beta_2 \mu_{X^2} \\ \mathbb{E}(X^1 Y) &= \beta_0 \mu_{X^1} + \beta_1 \mathbb{E}((X^1)^2) + \beta_2 \mathbb{E}(X^1 X^2) \\ \mathbb{E}(X^2 Y) &= \beta_0 \mu_{X^2} + \beta_1 \mathbb{E}(X^1 X^2) + \beta_2 \mathbb{E}((X^2)^2) \end{aligned}$$

qui se synthétisent en une équation matricielle $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$:

$$\begin{bmatrix} 1 & \mu_{X^1} & \mu_{X^2} \\ \mu_{X^1} & \mathbb{E}((x^1)^2) & \mathbb{E}(X^1 X^2) \\ \mu_{X^2} & \mathbb{E}(X^1 X^2) & \mathbb{E}((X^2)^2) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \mu_Y \\ \mathbb{E}(X^1 Y) \\ \mathbb{E}(X^2 Y) \end{bmatrix}$$

La résolution du système par la méthode de Cramer permet d'obtenir :

$$\beta_0 = \frac{\det[\mathbf{A}_1(\mathbf{c})]}{\det(\mathbf{A})}, \quad \beta_1 = \frac{\det[\mathbf{A}_2(\mathbf{c})]}{\det(\mathbf{A})} \quad \text{et} \quad \beta_2 = \frac{\det[\mathbf{A}_3(\mathbf{c})]}{\det(\mathbf{A})} \quad (1.115)$$

où $\mathbf{A}_i(\mathbf{c})$ est construite à partir de la matrice \mathbf{A} dans laquelle la colonne \mathbf{a}_i est remplacée par \mathbf{c} . Par exemple, on écrit pour β_2 :

$$\beta_2 = \frac{\begin{vmatrix} 1 & \mu_{X^1} & \mu_Y \\ \mu_{X^1} & \mathbb{E}((x^1)^2) & \mathbb{E}(X^1 Y) \\ \mu_{X^2} & \mathbb{E}(X^1 X^2) & \mathbb{E}(X^2 Y) \end{vmatrix}}{\begin{vmatrix} 1 & \mu_{X^1} & \mu_{X^2} \\ \mu_{X^1} & \mathbb{E}((X^1)^2) & \mathbb{E}(X^1 X^2) \\ \mu_{X^2} & \mathbb{E}(X^1 X^2) & \mathbb{E}((X^2)^2) \end{vmatrix}}$$

et l'on obtient après simplification :

$$= \frac{-\mathbb{E}(X^1 X^2)(\mathbb{E}(X^1 X^2) - \mu_{X^1} \mu_{X^2}) + \mathbb{E}((X^1)^2)(\mathbb{E}((X^2)^2) - \mu_{X^2}^2) + \mu_Y(\mu_{X^1} \mathbb{E}((X^1 X^2) - \mu_{X^2}^2) \mathbb{E}((X^1)^2))}{\mathbb{E}(X^1 Y)(\mathbb{E}((X^2)^2) - \mu_{X^2}^2) - \mathbb{E}(X^1 X^2)(\mathbb{E}(X^2 Y) - \mu_Y \mu_{X^2}) - \mu_Y \mu_{X^1} \mathbb{E}((X^2)^2) + \mu_{X^1} \mu_{X^2} \mathbb{E}(X^2 Y)}$$

En remplaçant $\mathbb{E}(X^i X^j)$ par $\sigma_{ij} + \mu_{X^i} \mu_{X^j}$ ainsi que $\mathbb{E}(X^i Y)$ par $\sigma_{Yi} + \mu_{X^i} \mu_Y$ ($i, j = 1, 2$), on obtient après simplification :

$$\beta_2 = \frac{\sigma_1^2 \sigma_{Y2} - \sigma_{Y1} \sigma_{12}}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} = \frac{\sigma_Y}{\sigma_2} \frac{(\rho_{Y2} - \rho_{Y1} \rho_{12})}{(1 - \rho_{12}^2)} \quad (1.116)$$

De la même façon, on obtient pour β_0 et β_1 :

$$\beta_1 = \frac{\sigma_2^2 \sigma_{Y1} - \sigma_{Y2} \sigma_{12}}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} = \frac{\sigma_Y}{\sigma_1} \frac{(\rho_{Y1} - \rho_{Y2} \rho_{12})}{(1 - \rho_{12}^2)} \quad (1.117)$$

$$\beta_0 = \mu_Y - \beta_1 \mu_{X^1} - \beta_2 \mu_{X^2} \quad \square \quad (1.118)$$

Remarque 1.8.1. On obtient à partir de la proposition précédente :

$$\beta_j = -\frac{\sigma_Y}{\sigma_j} \frac{C_{1k}}{C_{11}}, \quad j = 1, 2$$

où $k = j + 1$ et C_{mn} est le cofacteur du (m, n) -ème élément de la matrice symétrique \mathbf{R} ,

Remarque 1.8.2. On obtient ainsi à partir des équations (3.3.2) et (3.4.1) :

$$\beta_2 = \rho_{Y2.1} \frac{\sqrt{\sigma_Y^2 \sigma_1^2 - \sigma_{Y1}^2}}{\sqrt{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2}} = \rho_{Y2.1} \frac{\sigma_Y \sqrt{(1 - \rho_{Y1}^2)}}{\sigma_1 \sqrt{(1 - \rho_{12}^2)}} \quad (1.119)$$

si $(1 - \rho_{Y2}^2)$ et $(1 - \rho_{12}^2)$ sont strictement positifs (c'est-à-dire $(\rho_{Y1}^2 < 1)$ et $(\rho_{12}^2 < 1)$), on en déduit que $\beta_2 = 0$ si et seulement si $\rho_{Y2.1} = 0$, de sorte que les hypothèses $H_0 : \beta_2 = 0$ et $H_0 : \rho_{Y2.1} = 0$ sont équivalentes. On peut montrer de la même façon que les hypothèses $H_0 : \beta_1 = 0$ et $H_0 : \rho_{Y1.2} = 0$ sont équivalentes.

B) Corrélation partielle dans le cas général

Décomposons maintenant la matrice \mathbf{X} formée par les p variables prédictes en deux parties, avec $p_1 = p - 1$:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^1 & : & \mathbf{x}^j \end{bmatrix} \quad (1.120)$$

$(n \times p)$ $(n \times p_1)$ $(n \times 1)$

Le coefficient de corrélation partielle de Y et X^j sachant \mathbf{X}^1 mesure alors l'information sur Y , apportée par X^j complémentirement à l'information sur Y apportée par \mathbf{X}^1 . Dit autrement, le coefficient de corrélation partielle mesure la liaison entre deux variables lorsque l'influence des autres sont contrôlés.

Comme dans le cas précédent, on réalise les régressions linéaires de la variable Y en \mathbf{X}^1 et aussi de X^j en \mathbf{X}^1 au sens des MCO et les équations

$$Y = \alpha_0 + \alpha_1 x^1 + \alpha_2 x^2 + \dots + \alpha_{j-1} x^{j-1} + \alpha_{j+1} x^{j+1} + \dots + \alpha_p x^p + \epsilon^Y \quad (1.121)$$

$$X^j = \gamma_0 + \gamma_1 x^1 + \gamma_2 x^2 + \dots + \gamma_{j-1} x^{j-1} + \gamma_{j+1} x^{j+1} + \dots + \gamma_p x^p + \epsilon^{j.1} \quad (1.122)$$

La corrélation partielle théorique de Y et X^j à \mathbf{X}^1 fixée, se définit directement à partir du coefficient de corrélation linéaire simple calculé entre $\epsilon^{Y.1}$ et $\epsilon^{j.1}$. En considérant que ces deux variables sont centrées par construction, on obtient :

$$\rho_{Yj.1} = \rho(\epsilon^{Y.1}, \epsilon^{j.1}) = \frac{\mathbb{E}(\epsilon^{Y.1} \epsilon^{j.1})}{\sqrt{\mathbb{E}(\epsilon^{Y.1})^2 \mathbb{E}(\epsilon^{j.1})^2}} \quad (1.123)$$

De même, comme dans le cas précédent, on peut définir le coefficient de corrélation partielle à partir des corrélation simples entre les couples de variables considérées (cf. Kendall *et al.* ([32])). Soit \mathbf{R} la matrice de corrélation de $(p + 1)$ variables Y, X^1, X^2, \dots, X^p :

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{Y1} & \rho_{Y2} & \dots & \rho_{Yp} \\ \rho_{Y1} & 1 & \rho_{12} & \dots & \rho_{1p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{Yp} & \rho_{1p} & \rho_{2p} & \dots & 1 \end{pmatrix}$$

Le coefficient de corrélation partielle entre Y et X^j sachant les autres variables est noté :

$$\rho_{Yj.12,\dots,j-1,j+1,\dots,p} = \rho(Y, X^j | X^1, X^2, \dots, X^{j-1}, X^{j+1}, \dots, X^p) = \rho_{Yj.q_j} \quad (1.124)$$

où $q_j = \{1, 2, \dots, p\} - \{j\}$ est défini (cf. Kendall et al. ([32]) comme :

$$\rho_{Yj.q_j} = \frac{C_{1j}}{\sqrt{C_{11}C_{jj}}} \quad (1.125)$$

et où C_{mn} est le cofacteur du (m, n) -ème élément de la matrice symétrique \mathbf{R} .

Remarque 1.8.3. Comme dans le cas précédent, on peut montrer que les équations (3.4.2) et (1.125) sont identiques.

Proposition 1.8.3. Soit (Y, X^1, \dots, X^p) un $(p+1)$ -uplet aléatoire (non nécessairement gaussien) de moyenne $\mu_Y, \mu_{X^1}, \dots, \mu_{X^p}$, de matrice de corrélation \mathbf{R} et de la matrice de variances-covariances :

$$\mathbf{S} = \begin{pmatrix} \sigma_Y^2 & \sigma_{Y1} & \sigma_{Y2} & \dots & \sigma_{Yp} \\ \sigma_{Y1} & \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{Yp} & \sigma_{1p} & \sigma_{2p} & \dots & \sigma_p^2 \end{pmatrix}$$

où $\sigma_{ij} = \text{Cov}(X^i, X^j)$ et $\sigma_{Yj} = \text{Cov}(Y, X^j)$, $i, j = 1, \dots, p$.

Si $\mathbb{E}(Y|X = x)$ est une fonction linéaire de x^1, x^2, \dots, x^p de la forme :

$$\mathbb{E}(Y|X^1 = x^1, \dots, X^p = x^p) = \beta_0 + \beta_1 x^1 + \dots + \beta_p x^p$$

alors, les paramètres $\beta_0, \beta_j, j = 1, \dots, p$ sont :

$$\begin{aligned} \text{a) } \beta_j &= -\frac{\sigma_Y C_{1k}}{\sigma_j C_{11}}, & (k = j + 1) \\ \text{b) } \beta_0 &= \mu_Y - \beta_1 \mu_{X^1} - \dots - \beta_p \mu_{X^p} \end{aligned}$$

où C_{mn} est le cofacteur du (m, n) -ème élément de la matrice symétrique \mathbf{R} .

Preuve :

Elle utilise les mêmes arguments que la démonstration de la proposition 1.8.2, puisque seuls interviennent les moments croisés d'ordre 2.

a) Comme dans le cas de la régression linéaire double, on écrit le modèle :

$$\mathbb{E}(Y|X^1 = x^1, \dots, X^p = x^p) = \beta_0 + \beta_1 x^1 + \dots + \beta_p x^p$$

et on obtient le résultat voulu, du fait des propriétés de l'espérance mathématique. En particulier, on montre comme plus haut que :

$$\mu_Y = \beta_0 + \beta_1 \mu_{X^1} + \dots + \beta_p \mu_{X^p} \quad (1.126)$$

$$\mathbb{E}(X^1 Y) = \beta_0 \mu_{X^1} + \beta_1 \mathbb{E}((X^1)^2) + \dots + \beta_p \mathbb{E}(X^1 X^p) \quad (1.127)$$

On itère le procédé, de façon analogue après avoir multiplié par X^2, X^3, \dots, X^p . Chaque terme multiplicateur X^j fournit une équation :

$$\mathbb{E}(X^j Y) = \beta_0 \mu_{X^j} + \beta_1 \mathbb{E}(X^1 X^j) + \dots + \beta_j \mathbb{E}((X^j)^2) + \dots + \beta_p \mathbb{E}(X^j X^p) \quad (1.128)$$

On obtient ainsi un système de $k = p + 1$ équations à $p + 1$ inconnues que l'on peut résumer sous la forme matricielle $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$. Ainsi pour $p = 2$, on écrit :

$$\begin{aligned} \mu_Y &= \beta_0 + \beta_1 \mu_{X^1} + \beta_2 \mu_{X^2} \\ \mathbb{E}(X^1 Y) &= \beta_0 \mu_{X^1} + \beta_1 \mathbb{E}((X^1)^2) + \beta_2 \mathbb{E}(X^1 X^2) \\ \mathbb{E}(X^2 Y) &= \beta_0 \mu_{X^2} + \beta_1 \mathbb{E}(X^1 X^2) + \beta_2 \mathbb{E}((X^2)^2) \end{aligned}$$

qui particularise l'équation matricielle $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$; dans laquelle

$$\begin{bmatrix} 1 & \mu_{X^1} & \mu_{X^2} & \dots & \mu_{X^p} \\ \mu_{X^1} & \mathbb{E}((X^1)^2) & \mathbb{E}(X^1 X^2) & \dots & \mathbb{E}(X^1 X^p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_{X^p} & \mathbb{E}(X^1 X^p) & \mathbb{E}(X^2 X^p) & \dots & \mathbb{E}((X^p)^2) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \mu_Y \\ \mathbb{E}(X^1 Y) \\ \mathbb{E}(X^2 Y) \\ \vdots \\ \mathbb{E}(X^p Y) \end{bmatrix}$$

Comme dans le cas précédent, on déduit β_j par la méthode de Cramer :

$$\beta_j = \frac{\det[\mathbf{A}_{j+1}(\mathbf{c})]}{\det(\mathbf{A})}, \quad j = 0, 1, \dots, p \quad (1.129)$$

où $\mathbf{A}_i(\mathbf{c})$ est déduit de la matrice \mathbf{A} en remplaçant la colonne \mathbf{a}_i par \mathbf{c} . Par exemple, pour β_j on obtient après simplification :

$$\beta_j = -\frac{\sigma_Y C_{1k}}{\sigma_j C_{11}}, \quad (k = j + 1) \quad (1.130)$$

De plus,

$$\beta_0 = \mu_Y - \beta_1 \mu_{X^1} - \dots - \beta_p \mu_{X^p} \quad \square \quad (1.131)$$

définition 1.8.1. (cf. Kendall et al. ([32])

On appelle variance conditionnelle de la variable X^j sachant les $\{Y, X^l, l \neq j\}$, la quantité notée $\sigma_{j,1,2,\dots,j-1,j+1,\dots,p,y}^2$ ou plus simplement $\sigma_{j,q}^2$, définie par :

$$\sigma_{j,q}^2 = \sigma_j^2 \frac{|\mathbf{R}|}{C_{kk}}$$

et variance conditionnelle de la variable Y sachant les $\{X^l, l = 1, \dots, p\}$, la quantité notée $\sigma_{Y,1,\dots,p}^2$ ou plus simplement $\sigma_{Y,q}^2$ définie par :

$$\sigma_{Y,q}^2 = \sigma_Y^2 \frac{|\mathbf{R}|}{C_{11}}$$

où $k = j + 1$ et C_{mn} est le cofacteur du (m, n) -ème élément de la matrice \mathbf{R} ,

Remarque 1.8.4. A partir des équations (1.125) et (3.5.1) on obtient la relation liant la corrélation partielle et le coefficient de régression partielle :

$$\beta_j = \rho_{Yj.qj} \frac{\sigma_Y \sqrt{C_{kk}}}{\sigma_k \sqrt{C_{11}}} = \rho_{Yj.qj} \frac{\sigma_{Y.q}}{\sigma_{j.q}}$$

où $\sigma_{j.q}^2 = \sigma_j^2 \frac{|\mathbf{R}|}{C_{jj}}$ et $\sigma_{Y.q}^2 = \sigma_Y^2 \frac{|\mathbf{R}|}{C_{11}}$ sont des variances conditionnelles. Lorsque $\sigma_{Y.q}$ et $\sigma_{j.q}$ sont strictement positifs, on en déduit que $\beta_j = 0$ si et seulement si $\rho_{Yj.qj} = 0$, de sorte que les hypothèses $H_0 : \beta_1 = 0$ et $H_0 : \rho = 0$ sont équivalentes. Par suite, réaliser le test sur le coefficient de régression partielle et la corrélation de partielle sont deux démarches équivalentes.

On peut aussi traiter formellement une situation plus générale. Pour cela, il suffit de décomposer la matrice des prédicteurs en deux blocs :

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^1 & : & \mathbf{X}^2 \end{bmatrix} \quad (1.132)$$

$(n \times p)$ $(n \times p_1)$ $(n \times p_2)$

où $p_2 = p - p_1 > 1$. On décompose aussi de façon coordonnée le vecteur des paramètres β , et l'on écrit le modèle sous la forme :

$$\mathbf{Y} = \mathbf{X}\beta = [\mathbf{X}^1 : \mathbf{X}^2] \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \epsilon = \mathbf{X}^1\beta_1 + \mathbf{X}^2\beta_2 + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \quad (1.133)$$

Notons de plus, $\mathbf{H}_k \equiv \mathbf{X}^k ({}^t\mathbf{X}^k \mathbf{X}^k)^{-1} \mathbf{X}^k$ le projecteur orthogonal sur le sous espace vectoriel F_{X^k} engendré par les colonnes prédicteurs du groupe $k = 1, 2$; $\mathbf{P}_k \equiv (\mathbf{I}_n - \mathbf{H}_k)$ est l'annulateur associé; $\mathbf{R}^{Y.1} = \mathbf{P}_1 \mathbf{Y}$ est alors le vecteur des résidus de \mathbf{Y} dans la régression MCO de cette variable sur les prédicteurs formant les colonnes de \mathbf{X}^1 et $\mathbf{R}^{2.1} = \mathbf{P}_1 \mathbf{X}^2$ est la matrice dont les colonnes sont les résidus de la régression MCO des variables donnant les colonnes de \mathbf{X}^2 sur les prédicteurs formant les colonnes de \mathbf{X}^1 . Nous rappelons maintenant un théorème très important, attribué à Frisch-Waugh (1993) qui dit comment opérer de façon équivalente soit par une régression globale, soit à l'aide de régressions intermédiaires pour construire l'estimateur MCO des paramètres d'intérêt, dans un cadre général de régression multiple.

Théorème 1.8.1. (Frisch-Waugh)

Sous les postulats P1-P4, il est équivalent de construire l'estimateur MCO $\hat{\beta}_2$ de β_2 par estimation MCO globale de β ou par une régression MCO de $\mathbf{R}^{Y.1}$ en $\mathbf{R}^{2.1}$. Si \mathbf{U} est le vecteur des résidus de la régression de \mathbf{Y} en \mathbf{X} . on vérifie dans les deux cas,

1. $\hat{\beta}_2 = ({}^t\mathbf{R}^{2.1} \mathbf{R}^{2.1})^{-1} {}^t\mathbf{R}^{2.1} \mathbf{R}^{Y.1}$
2. $\hat{\beta}_2 = ({}^t\mathbf{R}^{2.1} \mathbf{X}^{2.1})^{-1} {}^t\mathbf{X}^{2.1} \mathbf{Y}$

$$3. \mathbf{U} = \mathbf{R}^{Y,1} - \mathbf{R}^{2,1}\hat{\boldsymbol{\beta}}_2 = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

Preuve : (*cf.* d'Aubigny [8])

Ce théorème permet de généraliser les résultats précédents au cas des corrélations partielles entre vecteurs aléatoires. Le lecteur peut noter qu'au delà des complexités d'écriture, l'interprétation géométrique du problème reste inchangée. Il convient seulement de substituer aux notions d'indices de corrélation partielle, celle de matrice de corrélation partielle. On peut alors traiter sans grand gain de complexité la situation d'une réponse Y vectorielle, de dimension $q > 1$, c'est-à-dire le cas de la régression linéaire multivariée.

C) Test de permutation portant sur un coefficient de régression partielle

Ces décompositions géométriques demeurent pertinentes dans le cas non gaussien, c'est-à-dire lorsque le terme d'erreur ϵ ne vérifie plus les postulats $P5$. Alors, la statistique T ne suit plus nécessairement une loi de Student. C'est pourquoi, nous nous sommes intéressés à l'utilisation de tests non paramétriques et en particulier aux tests de permutation dans ce cas.

Dans le cas de la régression linéaire simple, la méthode des tests de permutation portant sur le coefficient de régression que nous avons présentée en section (1.6.2) fait consensus. On montre en effet (*cf.* Lehman [37]) que les conditions d'échangeabilité sont assurées dans ce cas et donc le test est exact et sans biais. Dans le cadre de régression multiple avec $p \geq 2$ la situation de test d'un seul coefficient de régression partielle du modèle est plus complexe parce que les propriétés d'échangeabilité des observations de la variable réponse sont perdues lors des transformations du problème. Il n'existe pas de façon unique de les restaurer par transformation du problème. Plusieurs méthodes de tests de permutation ont été proposées pour mettre à l'épreuve l'hypothèse de nullité d'un coefficient de régression partielle qui intervient dans un modèle de régression multiple conditionnellement à l'observation d'un échantillon. Les approches présentées dans la littérature reposent en fait sur des caractéristiques distinctes du modèle classique et relèvent de logiques distinctes. Le chapitre suivant présente et compare ces diverses méthodes.

Chapitre 2

Résultat des simulations

2.1 Introduction

Ce chapitre consiste en l'étude comparative des méthodes de test d'hypothèses en régression multiple, à l'aide de tests de permutation. Plus spécifiquement notre étude concerne le test de nullité du coefficient de régression partielle entre la variable réponse et l'une des variables prédictives, lorsque les autres prédicteurs du modèle sont fixés.

Dans le chapitre précédent, nous avons rappelé les propriétés principales des procédures de test statistique classique utilisées en régression linéaire multiple. Leur optimalité est assurée lorsque les données respectent les postulats $P1 - P5$.

Cependant, les tests d'hypothèse usuels, en particulier ceux qui portent sur les coefficients de régression partielle, ne sont plus applicables lorsque le processus de génération des données est trop différent du modèle classique de régression linéaire. Considérons par exemple un modèle de régression linéaire multiple d'ordre 3 :

$$Y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \epsilon_i \quad (i = 0, 1, \dots, n), \quad (2.1)$$

où X^1 et X^2 sont deux variables explicatives fixées et ϵ mesure l'écart du modèle aux données. Le modèle de régression linéaire classique est caractérisé dans le cas ci-dessus par les postulats $P1 - P5$ présentés dans le chapitre 1 dans le cas multiple.

Nous nous fixons pour objectif d'effectuer un test d'hypothèse sur le coefficient de régression partielle : $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$. Une statistique de test souvent utilisée pour évaluer cette hypothèse est la statistique de Student :

$$T = \frac{\hat{\beta}_2}{S(\hat{\beta}_2)} \sim t(n - 3), \quad (2.2)$$

où $\hat{\beta}_2$ et $S^2(\hat{\beta}_2)$ sont les estimateurs MCO respectifs de β_2 et de $\text{Var}(\hat{\beta}_2)$

Lorsque le terme d'erreur ϵ ne vérifie plus les postulats $P1 - P4$, la statistique T ne suit plus nécessairement une loi de Student. Plusieurs méthodes alternatives

existent, parmi lesquelles nous retenons l'utilisation de tests de permutation. Ces derniers sont fondés sur l'unique hypothèse que les observations du terme d'erreur ϵ_i sont échangeables.

Les premières utilisations de tests de permutation adaptés au modèle linéaire furent présentées dans la première moitié du 20ème siècle par Fisher ([18]) et Pitman ([48, 49, 50]). Leur utilisation se heurtait alors à une lourdeur de calculs certaine par comparaison aux tests classiques. Aussi, ces tests ont peu retenu l'attention des chercheurs jusqu'à l'apparition d'ordinateurs puissants et largement accessibles, rendant envisageables les calculs intensifs imposés par cette approche.

En régression linéaire simple la méthode des tests de permutation portant sur le coefficient de régression présenté en section(1.6) fait consensus. Par contre, la situation est plus confuse dans le cas d'une de régression linéaire multiple. Il n'y a plus unicité de la méthode de test de l'hypothèse de nullité d'un coefficient de régression partielle. Plusieurs méthodes ont été proposées et chaque méthode repose sur des propriétés distinctes du modèle classique. La logique sous-jacente à ces tests a été validée dans des contextes d'application distincts. Une vue synthétique de la littérature consacrée à ces question peut être acquise en consultant par exemple : Freedman et Lane ([19]); Smouse et al. ([53]); Oja, ([47]); Collins [6]; Gail, Tan et Piantadosi [20]; Welch, [61]; ter Braak; Kennedy [33]; Kennedy et Cade, [34].

Les approches présentées dans la littérature statistique peuvent être partagées en deux groupes. Le premier est constitué par les méthodes qui considèrent que les observations de la variable conditionnant X^2 (ou des covariables conditionnant lorsque nous disposons de plus d'une covariable) sont des unités échangeables sous l'hypothèse $H_0 : \beta_2 = 0$. Le deuxième groupe a pour caractéristique d'effectuer des permutations des valeurs prises sur l'échantillon par le terme d'erreur ϵ . Les méthodes du premier groupe (e.g., Oja [47]) ignorent la corrélation qui existe en général entre les variables explicatives (X^1 et X^2) au cours des permutations. Certains auteurs, tels que Anderson et Legendre [1] présentent cette caractéristique comme un désavantage, cf. Welch [61]; ter Braak [59]. Remarquons qu'il existe des méthodes de randomisation contrainte qui sont applicables dans le cas où chaque paire de valeurs X^1 et X^2 est observée de façon répétée, cf. Brown et Maritz [5]

Quatre méthodes du deuxième groupe furent comparées dans l'étude d'Anderson et Legendre [1]. Elle sont dues respectivement à : (1)Freedman et Lane, [19]; (2)Manly [40, 41]; (3)ter Braak; (4)Kennedy, [33]. Chaque méthode est fondée sur un principe propre de choix de la variable permutée. Puisqu'elles diffèrent dans leurs principes même, il est difficile de mener des comparaisons formelles de leurs propriétés et l'on doit le plus souvent se contenter de résultats asymptotiques ou d'études par simulation. Anderson et Legendre [1] ont mené une longue étude par simulation afin d'analyser les effets de quatre facteurs influant sur le risque d'erreur de première espèce, sur la puissance du test et sur la sensibilité aux points extrêmes; ces facteurs sont : a) la taille de l'échantillon; b) le degré de corrélation entre les variables explicatives $\rho(X^1, X^2)$; c) la valeur du paramètre de la covariable X^1 ; d) la distribution du terme d'erreur aléatoire ϵ_i . Chacun de ces facteurs a été discuté

de façon plus ou moins complète par les défenseurs des diverses méthodes de permutation, sans avoir été examiné de façon empirique en détail. La simulation faite par Anderson et Legendre [1] est plus complète et plus riche en information que celles qui précédaient dues à Kennedy et Cade [34] et Manly . Ces études ont pour point commun d'être limitées au cas de régression double, donc de deux prédicteurs X^1 et X^2 . Leurs conclusions indiquent en particulier que les estimations de β_2 obtenues par les méthodes de Freedman-Lane [19], et de Kennedy [33] sont égales, mais les résultats des simulations montrent des différences. Ainsi le risque d'erreur de première espèce couvert par le modèle de Freedman-Lane est inférieur à celui de Kennedy. Anderson et Legendre [1] concluent que la méthode de Freedman-Lane semblait être la meilleure proposition parce que :

- a) La méthode de Freedman-Lane conserve plus respectivement que les autres méthodes le niveau du risque d'erreur de première espèce considéré, et le risque d'erreur de première espèce obtenue par cette méthode est toujours inférieur à celui qui est obtenu par la méthode de Kennedy.
- b) Les puissances des tests comparés (la méthode Kennedy n'a pas été considérée par ces auteurs) sont presque égales.
- c) En présence d'un point extrême, le risque d'erreur de première espèce pour la méthode de Freedman-Lane est plus stable que pour les méthodes de ter Braak et de Manly (la méthode de Kennedy n'a pas été considérée)

Une question s'est alors posée à nous : pourquoi les résultats des simulations obtenus pour les méthodes de Freedman-Lane et de Kennedy sont-ils différents alors que leurs estimateurs de β_2 sont identiques ? Pour répondre à cette question, nous avons décidé d'étudier les deux méthodes théoriquement et par simulation.

Ce chapitre reprend le cadre de travail d'Anderson et Legendre [1] et se propose d'approfondir les trois points suivants :

1. Complétion des simulations d'Anderson et Legendre [1]
 Nous avons mené une étude par simulation plus complète et plus riche en information (section(2.6)). Nous avons ainsi observé des cas où le risque d'erreur de première espèce du modèle de Kennedy [33] est inférieur à celui du modèle de Freedman-Lane [19]. Cela ne s'observe pas dans la simulation d'Anderson et Legendre [1] parce que l'étendue du domaine de valeurs des facteurs considérés dans leur simulation était plus limitée que dans notre étude .
2. Au contraire de la simulation d'Anderson-Legendre, des résultats de nos simulations montrent que, lorsque les valeurs du terme d'erreur sont générées selon une loi non gaussienne, par exemple étirée à gauche (comme la loi $\Gamma(10, 0.5)$), étirée à droite (comme la loi $\text{Exp}(2)$) ou symétrique mais de variance infinie (comme la loi $\text{Student}(1)$), la méthode paramétrique classique marche bien, c'est-à-dire qu'elle donne un risque d'erreur de première espèce très peu différent de la méthode de Freedman-Lane et proche au niveau de signification considéré pour les tests ($\alpha = 0.05$). Elle donne aussi une puissance très proche

de celle de la méthode de Freedman-Lane. Ceci est peut être dû au fait que les valeurs des variables X^1 et X^2 sont générées selon une loi de Gauss.

3. A partir des résultats théoriques obtenus rapportés au chapitre 3, pour vérifier la puissance de la méthode de Kennedy, non considérée dans les simulations d'Anderson et Legendre [1], et de Legendre (2000)[35], nous avons mené une simulation en considérant les méthodes de Freedman-Lane, de Kennedy et la méthode paramétrique classique. Nous trouvons dans la plupart des cas que la puissance de la méthode de Kennedy est supérieure à celle de la méthode de Freedman-Lane. Au final nous arrivons à des conclusions globalement très semblables à celles d'Anderson et Legendre [1] : la méthode de Freedman-Lane semble fournir les meilleurs résultats.

2.2 Description des méthodes de permutation

Soit un ensemble ordonné $I_n = \{1, 2, \dots, n\}$ décrivant les indices des éléments de l'échantillon observé. On considère une permutation aléatoire π qui associe au n -uplet ordonné $i_n = (1, 2, \dots, n)$ le n -uplet ordonné $i_n^\pi = (\pi 1, \pi 2, \dots, \pi n)$:

$$i_n = (1, 2, \dots, n) \xrightarrow{\pi} I_n^\pi = (\pi 1, \pi 2, \dots, \pi n)$$

On en déduit une permutation aléatoire des valeurs prises par une variable Y sur l'échantillon :

$$\mathbf{y} = (y_1, y_2, \dots, y_n) \xrightarrow{\pi} \mathbf{y}^\pi = (y^{\pi 1}, y^{\pi 2}, \dots, y^{\pi n})$$

En régression double, on considère un échantillon de taille n d'observations appariées des variables Y, X^1 et X^2 formant le vecteur aléatoire $\mathbf{Z} = (\mathbf{Y}, \mathbf{X}^1, \mathbf{X}^2)$ comme suit :

$$\{z_1, z_2, \dots, z_n\} = \{(y_1, x_1^1, x_1^2), (y_2, x_2^1, x_2^2), \dots, (y_n, x_n^1, x_n^2)\}$$

Le test d'hypothèse $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$ au vu d'un échantillon de taille n peut utiliser l'une des méthodes de test de permutation suivantes.

Par convention, lorsque π est placé en indice haut d'une variable, par exemple dans la notation Y^π , nous exprimons le fait que la variable Y est permutée. Par contre, lorsque π est placé en indice bas d'une variable par exemple T_π , cela veut dire que la variable T est obtenue comme fonction de variables permutées et de variables non permutées.

2.2.1 Permutation des résidus du modèle contraint : Freedman-Lane

On dispose d'un échantillon de taille n de Z . Les résidus observés (observations de ϵ) du modèle (2.1)

$$Y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \epsilon_i \quad (i = 0, 1, \dots, n),$$

sont supposés échangeables. On note les paramètres du modèle (2.1) :

$$\theta = \{\beta_0, \beta_1, \beta_2, \sigma^2\}$$

Pour effectuer le test d'hypothèse $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$ Freedman-Lane [19] ont proposé une méthode du test de permutation fondée sur le principe suivant, décrit sous la forme d'un algorithme de test :

1. On réalise une régression linéaire de la variable Y en x^1 et x^2 , au sens des MCO et d'après le modèle (2.1). On obtient alors :
 - a) l'estimateur $\hat{\theta} = \{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2\}$;
 - b) la prévision $\hat{\epsilon}$;
 - c) la prévision \hat{Y} .
2. La statistique de test retenue est le T de Student :

$$T = \frac{\hat{\beta}_2}{S(\hat{\beta}_2)}$$

où $S^2(\hat{\beta}_2)$ désigne l'estimateur MCO de $\text{Var}(\hat{\beta}_2)$. La valeur observée de cette statistique T sur les observations d'origine est notée :

$$t_{obs} = \frac{b_2}{s(\hat{\beta}_2)}$$

où, b_2 est la valeur observée de $\hat{\beta}_2$ sur l'échantillon c'est-à-dire, l'estimation de β_2 , et la valeur observée de $S^2(\hat{\beta}_2)$, noté $s^2(\hat{\beta}_2)$, est l'estimation associée de la variance de cet estimateur $\text{Var}(\hat{\beta}_2)$ sur cet échantillon. Nous nous référons désormais à cette valeur de T_{obs} .

3. On réalise de plus une régression linéaire simple de la variable Y en X^1 au sens des MCO l'équation

$$Y_i = \alpha_0 + \alpha_1 x_i^1 + \epsilon_i^{Y.1}. \quad (2.3)$$

On est alors conduit à calculer :

$$\hat{Y}_i^1 = \hat{\alpha}_0 + \hat{\alpha}_1 x_i^1$$

et on obtient les résidus observés $R^{Y.1}$ comme suit :

$$R_i^{Y.1} = y_i - \hat{y}_i^1 = y_i - a_0 - a_1 x_i^1 = \hat{\epsilon}_i^{Y.1}$$

où a_0 et a_1 sont les estimations MCO respectives de α_0 et α_1 .

4. A toute permutation π , on associe alors le vecteur des valeurs permutées des résidus obtenus par régression simple de Y en X^1 , $R^{Y.1, \pi^i}$ ($i = 1, 2, \dots, n$) défini comme :

$$\mathbf{R}^{Y.1} = (R_1^{Y.1}, R_2^{Y.1}, \dots, R_n^{Y.1}) \xrightarrow{\pi} \mathbf{R}^{Y.1, \pi} = (R^{Y.1, \pi^1}, R^{Y.1, \pi^2}, \dots, R^{Y.1, \pi^n})$$

5. On constitue ensuite le vecteur des pseudo-valeurs notées $Y_{\pi i}^F$ obtenues en ajoutant les résidus permutés $R^{Y.1,\pi i}$ aux valeurs prévues par l'équation (2.3) :

$$Y_{\pi i}^F = \hat{Y}_i^1 + R^{Y.1,\pi i} = a_0 + a_1 x_i^1 + R^{Y.1,\pi i}, \quad (i = 1, 2, \dots, n)$$

l'indice haut F précise que la méthode de construction de ces pseudo-valeurs proposée est due à Freedman-Lane [19].

6. Au bilan, le modèle ajusté par Freedman-Lane s'écrit

$$Y_{\pi i}^F = \beta_{0,\pi}^F + \beta_{1,\pi}^F x_i^1 + \beta_{2,\pi}^F x_i^2 + \epsilon_i^F \quad (2.4)$$

où : $\theta^F = (\beta_{0,\pi}^F, \beta_{1,\pi}^F, \beta_{2,\pi}^F, \sigma^{2F})$ est le vecteur des paramètres du modèle, estimés selon la méthode de Freedman-Lane, et σ^{2F} est la variance du résidu ϵ^F du modèle (2.4). Chaque permutation π fournit en particulier un estimateur $\hat{\beta}_{2,\pi}^F$ de $\beta_{2,\pi}^F$ obtenu par la méthode de Freedman-Lane, défini comme l'estimateur au sens des moindres carrés ordinaires pour le modèle (2.4). De ce fait, chaque permutation fournit une statistique de test de Student :

$$T_{\pi}^F = \frac{\hat{\beta}_{2,\pi}^F}{S(\hat{\beta}_{2,\pi}^F)}$$

où, si \mathbf{e}^{π} est l'échantillon \mathbf{e} permuté par la permutation π , alors $\hat{\beta}_{2,\pi}^F = \hat{\beta}_2^F(\mathbf{e}^{\pi})$ désigne l'estimateur MCO de $\beta_{2,\pi}^F$ et $S^2(\hat{\beta}_{2,\pi}^F) = S^2(\hat{\beta}_2^F)(\mathbf{e}^{\pi})$ désigne l'estimateur MCO de $Var(\hat{\beta}_{2,\pi}^F)$. L'estimation t_{π}^F obtenue sur l'échantillon permuté s'écrit :

$$t_{\pi}^F = \frac{b_{2,\pi}^F}{s(\hat{\beta}_{2,\pi}^F)}$$

Ici, $b_{2,\pi}^F$ (valeur observée de $\hat{\beta}_{2,\pi}^F$) est l'estimation MCO de $\beta_{2,\pi}^F$, et $s^2(\hat{\beta}_{2,\pi}^F)$ (valeur observée de $S^2(\hat{\beta}_{2,\pi}^F)$) est l'estimation associée de la variance de cet estimateur.

7. On obtient la distribution permutationnelle de la statistique T_{π}^F , On répétant $n!$ fois les étapes 4, 5 et 6.
8. Pour effectuer un test de permutation bilatéral à partir des valeurs obtenues t_{π}^F , on construit la distribution des valeurs $|t_{\pi}^F|$ rangées par ordre croissant. Puis, on place la valeur $|t_{obs}|$ parmi les statistiques d'ordre de cette distribution. La p-valeur permutationnelle est $(p_{2,\pi}^F)$ calculée comme la proportion de valeurs $|t_{\pi}^F|$ qui sont supérieures ou égales à la valeur $|t_{obs}|$ dans cette distribution (cf. Hope [29]), c'est-à-dire :

$$p_{2,\pi}^F = \frac{\#(|t_{\pi}^F| \geq |t_{obs}|)}{n!}$$

2.2.2 Permutation des observations de la variable Y : méthode de Manly

La méthode de test de permutation proposée par Manly [40, 41] adopte une procédure différente pour tester l'hypothèse nulle $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$. Nous en décrivons le principe sous forme d'algorithme :

- 1-2. Les deux premières étapes sont les mêmes que pour la méthode de Freedman-Lane
3. Pour obtenir les valeurs permutées $Y^{M,\pi(i)}$, ($i = 1, 2, \dots, n$), Manly permute les valeurs y_i ,
4. On fait alors une régression de $Y^{M,\pi}$ en x^1 et x^2 (les valeurs x^1 et x^2 ne permutent pas) au sens des MCO, d'après le modèle :

$$Y^{M,\pi} = \beta_{0,\pi}^M + \beta_{1,\pi}^M x^1 + \beta_{2,\pi}^M x^2 + \epsilon^M \quad (2.5)$$

le choix de ce modèle caractérise l'approche de Manly. Alors, $\theta^M = (\beta_{0,\pi}^M, \beta_{1,\pi}^M, \beta_{2,\pi}^M, \sigma^{2M})$ est le vecteur des paramètres du modèle, estimés selon la méthode de Manly et σ^{2M} est la variance du résidu ϵ^M du modèle (2.5). Chaque permutation π fournit en particulier un estimateur $\hat{\beta}_{2,\pi}^M$ de $\beta_{2,\pi}^M$ obtenu par la méthode de Manly, défini comme l'estimateur au sens des moindres carrés ordinaires pour le modèle (2.5). De ce fait, chaque permutation fournit une statistique de test de Student :

$$T_\pi^M = \frac{\hat{\beta}_{2,\pi}^M}{S(\hat{\beta}_{2,\pi}^M)}$$

où, si \mathbf{e}^π est l'échantillon \mathbf{e} permuté par la permutation π , alors $\hat{\beta}_{2,\pi}^M = \hat{\beta}_2^M(\mathbf{e}^\pi)$ désigne l'estimateur MCO de $\beta_{2,\pi}^M$ et $S^2(\hat{\beta}_{2,\pi}^M) = S^2(\hat{\beta}_2^M)(\mathbf{e}^\pi)$ désigne l'estimateur MCO de $Var(\hat{\beta}_{2,\pi}^M)$. L'estimation t_π^M obtenue sur l'échantillon permuté s'écrit :

$$t_\pi^M = \frac{b_{2,\pi}^M}{s(\hat{\beta}_{2,\pi}^M)}.$$

Ici, $b_{2,\pi}^M$ (valeur observée de $\hat{\beta}_{2,\pi}^M$) est l'estimation MCO de $\beta_{2,\pi}^M$, et $s^2(\hat{\beta}_{2,\pi}^M)$ (valeur observée de $S^2(\hat{\beta}_{2,\pi}^M)$) est l'estimation associée de la variance de cet estimateur.

5. On obtient la distribution permutationnelle de la statistique T_π^M , en répétant $n!$ fois les étapes 3 et 4.
6. Pour effectuer un test de permutation bilatéral à partir des valeurs obtenues t_π^M , on construit la distribution des valeurs $|t_\pi^M|$ rangées par ordre croissant puis on place la valeur $|t_{obs}|$ parmi les statistiques d'ordre de cette distribution. La

p-valeur permutationnelle ($p_{2,\pi}^M$) est calculée comme proportion de valeurs $|t_\pi^M|$ qui sont supérieures ou égales à la valeur $|t_{obs}|$ dans cette distribution, c'est-à-dire :

$$p_{2,\pi}^M = \frac{\#(|t_\pi^M| \geq |t_{obs}|)}{n!}$$

2.2.3 Permutation des résidus du modèle général : ter Braak

ter Braak (1990, 1992) utilise des résidus du modèle complet de régression linéaire (2.1) pour tester l'hypothèse nulle $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$.

- 1-2 Les deux premières étapes sont les mêmes que pour la méthode de Freedman-Lane.
3. On permute les résidus obtenus en étape 1, c'est-à-dire lors du modèle sans contrainte dit modèle général. On opère comme suit :

$$R_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_i^1 - b_2 x_i^2$$

où b_0 , b_1 et b_2 sont respectivement les estimations de MCO de β_0 , β_1 et β_2 . On associe alors des valeurs permutées $R^{\pi i}$ ($i = 1, 2, \dots, n$) définies comme :

$$\mathbf{R} = (R_1, R_2, \dots, R_n) \xrightarrow{\pi} \mathbf{R}^\pi = (R^{\pi 1}, R^{\pi 2}, \dots, R^{\pi n})$$

4. On constitue ensuite le vecteur des pseudo-valeurs, noté $Y_{\pi i}^B$, obtenu en ajoutant les résidus permutés $R^{\pi i}$ aux valeurs prévues :

$$Y_{\pi i}^B = \hat{y}_i + R^{\pi i} = b_0 + b_1 x_i^1 + b_2 x_i^2 + R^{\pi i} \quad (i = 1, 2, \dots, n)$$

l'indice haut B précise que la méthode de construction de ces pseudo-valeurs proposée est due à ter Braak .

5. Alors, ter Braak ajuste le modèle :

$$Y_{\pi i}^B = \beta_{0,\pi}^B + \beta_{1,\pi}^B x_i^1 + \beta_{2,\pi}^B x_i^2 + \epsilon_i^B \quad (2.6)$$

où : $\theta^B = (\beta_{0,\pi}^B, \beta_{1,\pi}^B, \beta_{2,\pi}^B, \sigma^{2B})$ est le vecteur des paramètres du modèle, estimés selon la méthode de ter Braak, et σ^{2B} est la variance du résidu ϵ_π^B du modèle (2.6). Chaque permutation π fournit en particulier un estimateur $\hat{\beta}_{2,\pi}^B$ de $\beta_{2,\pi}^B$, définie par ter Braak comme l'estimateur au sens des moindres carrés ordinaires pour le modèle (2.6). De ce fait, chaque permutation fournit une statistique de test de Student :

$$T_\pi^B = \frac{\hat{\beta}_{2,\pi}^B - \hat{\beta}_2}{S(\hat{\beta}_{2,\pi}^B)}$$

où, si \mathbf{e}^π est l'échantillon \mathbf{e} permuté par la permutation π , alors $\hat{\beta}_{2,\pi}^B = \hat{\beta}_2^B(\mathbf{e}^\pi)$ désigne l'estimateur MCO de $\beta_{2,\pi}^B$ et $S^2(\hat{\beta}_{2,\pi}^B) = S^2(\hat{\beta}_2^B)(\mathbf{e}^\pi)$ désigne l'estimateur MCO de $Var(\hat{\beta}_{2,\pi}^B)$ et l'estimation de t obtenue sur l'échantillon permuté

s'écrit :

$$t_{\pi}^B = \frac{b_{2,\pi}^B - b_2}{s(\hat{\beta}_{2,\pi}^B)}$$

Ici, $b_{2,\pi}^B$ (valeur observée de $\hat{\beta}_{2,\pi}^B$) est l'estimation MCO de $\beta_{2,\pi}^B$, et $s^2(\hat{\beta}_{2,\pi}^B)$ (valeur observée de $S(\hat{\beta}_{2,\pi}^B)$) est l'estimation MCO associée de la variance de cet estimateur.

6. On obtient la distribution permutationnelle de la statistique T_{π}^B , en répétant $n!$ fois les étapes 3, 4 et 5.

7. Pour effectuer un test de permutation bilatéral à partir des valeurs obtenues t_{π}^B , on construit la distribution des valeur $|t_{\pi}^B|$ rangées par ordre croissant puis on place la valeur $|t_{obs}|$ parmi les statistiques d'ordre de cette distribution. La p-valeur permutationnelle ($p_{2,\pi}^B$) est calculée comme proportion de valeurs $|t_{\pi}^B|$ qui sont supérieures ou égales à la valeur $|t_{obs}|$ dans cette distribution, c'est-à-dire :

$$p_{2,\pi}^B = \frac{\#(|t_{\pi}^B| \geq |t_{obs}|)}{n!}$$

2.2.4 Permutation des résidus du modèle contraint : Kennedy

Comme Freedman-Lane [19], Kennedy [33] utilise des résidus du modèle contraint par l'hypothèse H_0 . Il propose la procédure suivante de test de l'hypothèse nulle $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$.

1-3 Les trois premières étapes sont les mêmes que pour la méthode de Freedman-Lane.

4. On effectue une régression linéaire simple de la variable X^2 sur X^1 au sens MCO et l'équation

$$X^2 = \gamma_0 + \gamma_1 x^1 + \epsilon^{2.1} \quad (2.7)$$

conduit à calculer la prévision :

$$\hat{X}^2 = \hat{\gamma}_0 + \hat{\gamma}_1 x^1$$

On calcule alors des résidus observé $R^{2.1}$ comme suit :

$$R_i^{2.1} = x_i^2 - \hat{x}_i^2 = x_i^2 - g_0 - g_1 x_i^1$$

où g_0 et g_1 sont les estimations MCO respectives de γ_0 et γ_1 .

5. Cette étape est la même que l'étape 4 de la méthode de Freedman-Lane

6. Alors le modèle ajusté par Kennedy est :

$$R^{Y,1,\pi i} = \beta_{2,\pi}^K R_i^{2,1} + \epsilon_i^K \quad (2.8)$$

où $\theta^K = (\beta_{0,\pi}^K, \beta_{1,\pi}^K, \beta_{2,\pi}^K, [\sigma^K]^2)$ est le vecteur des paramètres du modèle, estimés selon la méthode de Kennedy et $[\sigma^K]^2$ est la variance du résidu ϵ_π^K du modèle (2.8). Chaque permutation π fournit en particulier un estimateur de $\beta_{2,\pi}^K$, noté $\hat{\beta}_{2,\pi}^K$ et obtenu par Kennedy comme l'estimateur aux moindres carrés ordinaires pour le modèle (2.8). De ce fait, chaque permutation fournit une valeur de la statistique de test de Student :

$$T_\pi^K = \frac{\hat{\beta}_{2,\pi}^K}{S(\hat{\beta}_{2,\pi}^K)}$$

où, si \mathbf{e}^π est l'échantillon \mathbf{e} permuté par la permutation π , alors $\hat{\beta}_{2,\pi}^K = \hat{\beta}_2^K(\mathbf{e}^\pi)$ désigne l'estimateur MCO de $\beta_{2,\pi}^K$ et $S^2(\hat{\beta}_{2,\pi}^K) = S^2(\hat{\beta}_2^K)(\mathbf{e}^\pi)$ désigne l'estimateur MCO de $Var(\hat{\beta}_{2,\pi}^K)$. L'estimation t_π^K obtenue sur l'échantillon permuté s'écrit :

$$t_\pi^K = \frac{b_{2,\pi}^K}{s(\hat{\beta}_{2,\pi}^K)}$$

Ici, la valeur observée $b_{2,\pi}^K$ de $\hat{\beta}_{2,\pi}^K$ est l'estimation MCO de $\beta_{2,\pi}^K$, et la valeur observée $s^2(\hat{\beta}_{2,\pi}^K)$ de $S^2(\hat{\beta}_{2,\pi}^K)$ est l'estimation MCO associée de la variance de cet estimateur.

7. On obtient la distribution permutationnelle de la statistique T_π^K , en répétant $n!$ fois les étapes 5 et 6.
8. Pour effectuer un test de permutation bilatéral à partir des valeurs obtenues t_π^K , on construit la distribution des valeurs $|t_\pi^K|$ rangées par ordre croissant. Puis, on place la valeur $|t_{obs}|$ parmi les statistiques d'ordre de cette distribution. La p-valeur permutationnelle ($p_{2,\pi}^K$) est calculée comme proportion de valeurs $|t_\pi^K|$ qui sont supérieures ou égales à la valeur $|t_{obs}|$ dans cette distribution . c'est-à-dire :

$$p_{2,\pi}^K = \frac{\#(|t_\pi^K| \geq |t_{obs}|)}{n!}$$

2.3 Comparaison des méthodes

2.3.1 La méthode de Freedman et Lane

Edgington [13] rappelle que les méthodes de randomisation considèrent les résidus d'un modèle linéaire (ou non linéaire) comme des unités échangeables. Opérer sur les

résidus peut être généralement considéré comme un principe de base des méthodes de test de permutation.

Seule la méthode de Manly ne respecte pas ce principe et permute les valeurs de la variable Y .

D'un point de vue formel ϵ_i est un terme stochastique et n'est pas associé à chaque valeur particulière de Y_i (c.f. Kempthorne [31]). Deux méthodes distinctes de permutation ont été décrites qui fournissent un test respectant cette logique. La première méthode fut présentée par Freedman-Lane [19] et la deuxième est proposée par Smouse et al. [53] dans le contexte des comparaisons de plusieurs matrices de distance puis par Kennedy [33] dans le contexte plus général de la régression linéaire multiple.

A-Le principe de la méthode de Freedman-Lane

Considérons le modèle complet (l'équation (2.1))

$$Y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \epsilon_i$$

On suppose que le terme d'erreur ϵ_i est échangeable. Par ailleurs sous $H_0 : \beta_2 = 0$ on a le modèle qui respecte la contrainte H_0 et est appelé le modèle réduit :

$$Y_i = \alpha_0 + \alpha_1 x_i^1 + \epsilon_i^{Y:1}$$

ce qui montrent que les $\epsilon_i^{Y:1}$ sont échangeables comme les ϵ_i , Freedman et Lane [19].

B-Avantages et inconvénients de la méthode de Freedman et Lane

1. La méthode de permutation de Freedman-Lane préserve les covariances entre Y et x^1 , entre x^1 et x^2 et plus généralement les covariances entre les covariables deux à deux, s'il y a $p \geq 2$ covariables dans le modèle, lors des permutations, mais cette méthode ne peut pas préserver la covariance entre Y et x^2 , cf. Anderson et Legendre [1].
2. Freedman-Lane [19] ont formulé trois conditions suffisantes pour avoir une bonne approximation de $p_{2,\pi}^F$ (cf. le théorème (2)) :
 - a) les données ne doivent pas contenir des points extrêmes ;
 - b) les variables x^1 et x^2 ne doivent pas être fortement corrélées ;
 - c) la taille de l'échantillon n doit être relativement grande.

3. Anderson et Legendre [1] montrent que les trois modèles ci-dessous sont équivalents au modèle de Freedman-Lane du point de vue permutational, pour réaliser l'étape (6)

- a) $\mathbb{E}(R^{Y.1,\pi i}) = \beta_{0,\pi} + \beta_{1,\pi}x^1 + \beta_{2,\pi}x^2$
- b) $\mathbb{E}(R^{Y.1,\pi i}) = \beta_{0,\pi} + \beta_{1,\pi}x^1 + \beta_{2,\pi}R^{2.1}$
- c) $\mathbb{E}(Y_{\pi i}^F) = \beta_{0,\pi} + \beta_{1,\pi}x^1 + \beta_{2,\pi}R^{2.1}$

Cette équivalence s'entend au sens qu'ils donnent la même estimation de $\beta_{2,\pi}$ ($b_{2,\pi}^F$) et la même valeur de statistique de test (t_{π}^F) que le modèle de Freedman-Lane pour chaque permutation.

2.3.2 La méthode de Manly

Manly [40] a présenté cette méthode de permutation en utilisant une statistique de test non pivotale : par exemple le numérateur de l'estimateur de β_2 , (c.f Manly [40]). Pourtant, l'importance du caractère pivotale de la statistique de test telle que le t de Student, a été montrée dans le cas des tests de permutation d'un terme figurant dans les modèles complexes (Stapel et ter Braak [55] ; Kennedy et Cade [34]) et pour la technique de ré-échantillonnage concurrente de test par la méthode du Bootstrapp (Hall et Titterington [24] ; Fisher et Hall [17])

A- Le principe de la méthode de Manly

Le principe de cette méthode repose sur le postulat que les valeurs originales de Y sont échangeables et le fait que chaque valeur de Y peut être associée à n'importe quelle combinaison des valeurs (X^1, X^2).

B- Avantages et inconvénients de la méthode de Manly

1. Kennedy [33] puis Kennedy et Cade [34] considèrent la méthode de permutation de Y comme justifiée seulement lorsque le paramètre de la covariable conditionnant X^1 est nul, c'est-à-dire $\beta_1 = 0$. Leur argument se fonde sur l'affirmation que les permutations de Y ignorent le paramètre de la covariable, ce qui est rarement vrai. En particulier, lorsque $\beta_1 \neq 0$, la distribution de Y est radicalement différente de celle de ϵ . Alors, le test d'hypothèse $\beta_2 = 0$ donne un résultat biaisé. D'ailleurs ces auteurs ont montré par une simulation limitée que lorsque $\beta_1 \neq 0$, l'erreur de première espèce augmente quand on introduit un point extrême parmi les valeurs observées de la covariable X^1 .
2. L'erreur associée à chaque valeur de Y permute de la même manière que Y car on la considère comme partie intégrante de cette observation (Anderson et Legendre [1]). Ce point de vue est exprimé par Edgington [13] pour les tests de randomisation utilisés sur un plan de expérience.

3. La méthode de permutation de Y préserve la covariance entre X^1 et X^2 , et plus généralement entre les covariables intervenant dans un modèle à $p \geq 2$ covariables.

2.3.3 La méthode de ter Braak

A- La principe de la méthode de ter Braak

Le principe de cette méthode repose sur le postulat d'échangeabilité des résidus ϵ_i . Cette méthode utilise dans le calcul de la statistique de T , à l'étape (5) l'estimation b_2 du coefficient β_2 obtenue par régression MCO du modèle général. L'estimation de β_2 est ici fondée sur l'hypothèse H_a . C'est pour cela qu'on parle de méthode d'estimation sous l'alternative b_2 invariant par permutation.

B- Avantages et inconvénients de la méthode de ter Braak

1. Dans cette méthode, la statistique de test de Student est calculée pour chaque permutation sous l'hypothèse alternative $H_1 : \beta_2 = b_2$. On obtient donc une distribution permutationnelle de $\beta_{2,\pi}^T$ symétrique autour de b_2 , et pour cette raison, l'approche de ter Braak [58] est appelée "la permutation sous hypothèse alternative" (Anderson et Legendre [1]).
2. Comme pour la méthode de Freedman-Lane, la variable X^1 est considérée dans le modèle à l'étape 5.
3. La méthode de permutation de ter Braak préserve les covariances entre Y et x^1 , entre Y et X^2 , entre x^1 et x^2 et plus généralement les covariances entre les covariables prise deux à deux, s'il y a plus d'une covariable dans le modèle, dans toutes les permutations (Anderson et Legendre [1]).
4. Dans le contexte de la régression multiple, cette méthode permet de tester la signification de plusieurs coefficients de régression partielle, voire de tous les coefficients de régression partielle, en utilisant une seule série de permutations.
5. Le modèle ci-dessous est équivalent au modèle de ter Braak.

$$R^{\pi i} = \beta_0'' + \beta_1'' x^1 + \beta_2'' x^2 + \epsilon''$$

C'est-à-dire qu'il donne l'estimation MCO de β_2' notée $b_{2,\pi}'$ qui est différente de $b_{2,\pi}^B$, et une valeur de $s(b_{2,\pi}')$ (où $s^2(b_{2,\pi}')$ est l'estimation MCO de la variance de l'estimateur MCO de β_2') qui est différente de $s^2(b_{2,\pi}^B)$ et $t_\pi' = b_{2,\pi}'/s(b_{2,\pi}')$ qui est égal à t_π^B pour chaque permutation (cf. Anderson et Legendre [1]).

2.3.4 Méthode de Kennedy

A- La principe de la méthode de Kennedy

Comme déjà dit, le principe de la méthode de Kennedy [33] est identique à celui de la méthode de Freedman et Lane [19].

B- Avantages et inconvénients de la méthode de Kennedy

Ceux-ci seront présentés au chapitre (3).

2.4 Simulation d'Anderson et Legendre

Dans cette section nous rappelons le cadre de la simulation faite par Anderson et Legendre [1]. Ensuite nous exprimons les résultats obtenus par cette simulation. Reprenons le modèle(2.1) :

$$Y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \epsilon_i \quad (i = 0, 1, \dots, n),$$

L'objectif est d'effectuer le test de l'hypothèse $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$.

Anderson et Legendre [1] ont mené une étude par simulation afin d'analyser les effets de quatre facteurs :

- a) la taille de l'échantillon ;
- b) le degré de corrélation entre les variables explicatives $\rho(X^1, X^2)$;
- c) la valeur du paramètre de la covariable X^1 ;
- d) la distribution du terme d'erreur aléatoire ϵ_i .

sur le risque d'erreur de première espèce, sur la puissance du test et du point de vue de l'effet de l'introduction de points extrêmes sur le risque d'erreur de première espèce. Nous avons retenu les mêmes facteurs pour notre simulation. Remarquons que chacun de ces facteurs a été discuté dans une certaine mesure par les défenseurs des diverses méthodes de permutation sans avoir pour autant été examiné de façon empirique et avec assez de détail.

2.4.1 Analyse du risque d'erreur de première espèce du test

A- Le cadre de la simulation d'Anderson et Legendre

Anderson et Legendre ont étudié le risque d'erreur de première espèce pour l'ensemble de modalités des facteurs ci-dessous :

- a) la taille de l'échantillon : ils retiennent les tailles d'échantillon suivantes :
 $n = \{9, 18, 36, 54, 72, 90\}$;

- b) la valeur du coefficient de la covariable X^1 est : $\beta_1 = \{0, 1, 2.41\}$
- c) le degré de corrélation entre les variables explicatives X^1 et X^2 est :
 $\rho(X^1, X^2) = \{0, 0.5, 0.9\}$;
- d) la distribution du terme d'erreur aléatoire ϵ est : $\{LG(0, 1), Exp(1), Exp(1)^3\}$.
où $Exp(\lambda)$ désigne la loi exponentielle de paramètre λ et $LG(0, 1)$ désigne la loi de Gauss centrée réduite.

Ils construisirent deux variables Z_1 et Z_2 à valeurs dans l'ensemble $\{1, 2, 3\}$. Les 9 combinaisons ainsi obtenues par croisement de ces 2 variables apparaissent dans la simulation de façon équilibrées. Par exemple pour $n = 9$

$$\mathbf{W} = [\mathbf{Z}^1 | \mathbf{Z}^2] = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 2 & 1 \\ 2 & 2 \\ 2 & 3 \\ 3 & 1 \\ 3 & 2 \\ 3 & 3 \end{bmatrix} \implies \mathbf{W}_c = [\mathbf{Z}_c^1 | \mathbf{Z}_c^2] \quad (2.9)$$

où \mathbf{Z}_c^1 et \mathbf{Z}_c^2 sont les vecteurs centrés et réduits. Donc :

- Pour $n = 18$ chaque ligne de cette matrice est répétée deux fois ;
Pour $n = 27$ chaque ligne de cette matrice est répétée trois fois ;
 \vdots
Pour $n = 90$ chaque ligne de cette matrice est répétée dix fois.

Remarquons que $\rho(Z_c^1, Z_c^2) = 0$. A partir des \mathbf{Z}_c^1 et \mathbf{Z}_c^2 , ces auteurs construisent les vecteurs \mathbf{X}^1 et \mathbf{X}^2 de corrélation $\rho(X^1, X^2)$ de la façon suivante :

$$\mathbf{W}_R = [\mathbf{X}^1 | \mathbf{X}^2] = \mathbf{W}_c \mathbf{R}^{1/2} \quad (2.10)$$

où :

$$\mathbf{R} = \begin{bmatrix} 1 & \rho(X^1, X^2) \\ \rho(X^1, X^2) & 1 \end{bmatrix}.$$

Enfin, les valeurs de Y sont générées par l'équation :

$$\mathbf{Y} = \mathbf{W}_R \boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.11)$$

où : $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ est le vecteur des paramètres. Pour calculer le risque d'erreur de première espèce, on doit prendre $\beta_2 = 0$. $\boldsymbol{\epsilon}$ est le vecteur du terme d'erreur aléatoire

généralisé selon une distribution de l'ensemble $\{LG(0, 1), Exp(1), Exp(1)^3\}$.

Pour chaque combinaison de facteurs, 10000 ensembles de données sont générés. Puis, pour chacune des quatre méthodes, chaque ensemble de données est permuté 999 fois. Les statistiques de tests permutacionnelles ($t_{2,\pi}$) de ces méthodes sont calculées sur chaque ensemble de données permutées. Ensuite, après 999 permutations, la p-valeur permutacionnelle pour chacune des 4 méthodes est calculée par la formule :

$$p - valeur_{\pi} = \frac{\#(|t_{\pi}| \geq t_{obs})}{1000}$$

Le niveau de signification des tests a été fixé à $\alpha = 0,05$. Alors, pour chaque ensemble de données quand la $p - valeur_{\pi}$ est inférieure à 0.05 on rejette l'hypothèse $H_0 : \beta_2 = 0$. Sur les 10000 $p - valeur_{\pi}$ le pourcentage des $P - valeur_{\pi}$ (pour chaque méthode) inférieures à 0.05 donne le risque d'erreur de première espèce. En fait, ce pourcentage estime la probabilité de rejeter $H_0 : \beta_2 = 0$ lorsque H_0 est vraie (parce que les données sont générés sous $\beta_2 = 0$). De plus, l'intervalle de confiance à 95% est calculé à partir du risque d'erreur de première espèce (calculé sur 10.000 ensembles).

Afin de compléter la comparaison, on a utilisé pour chaque ensemble de données le test paramétrique classique qui consiste à calculer une statistique de Student et sa p-valeur données par la théorie Gaussienne. De la même façon, sur les 10000 ensembles de données, le risque d'erreur de première espèce et l'intervalle de confiance ont été calculés par la théorie Gaussienne.

B-Les résultats de la simulation d'Anderson et Legendre

Nous rapportons ici brièvement les résultats des simulations effectuées afin de reproduire les études présentées par Anderson et Legendre.

B1- Cas des données générées pour le terme d'erreur selon une loi de Gauss :

1. Avec la méthode de Kennedy, le risque d'erreur de première espèce augmente surtout lorsque la taille d'échantillon est petite ($n = 9, n = 18$). Lorsque la taille d'échantillon augmente, le risque se stabilise à 0.05, quelles que soient les valeurs des facteurs.
2. Lorsque la taille d'échantillon est $n = 9$, le risque d'erreur de première espèce est très supérieur à 0.05 mais cet écart diminue pour des tailles d'échantillon plus grandes ($n = 18, 13, 54, 72, 90$), quelles que soient les valeurs des facteurs.
3. Quand la taille d'échantillon est supérieure ou égale à 54, le risque d'erreur de première espèce produit par la méthode de Kennedy est très proche de $\alpha = 0.05$.

4. Pour toutes les combinaisons des facteurs, il n'y pas de différence significative entre les méthodes de Freedman-Lane, de Manly et de ter Braak.
5. Dans tous les cas, le risque d'erreur de première espèce de la méthode de Kennedy est supérieur à celui de Freedman-Lane.

B2- Cas des données générées pour le terme d'erreur selon une loi de Exp(1) :

Les résultats dans ce cas sont très semblables au cas précédent où les données sont générées selon une loi de Gauss.

B3- Cas où les données générées pour le terme d'erreur selon une loi radicalement non symétrique comme $Exp(1)^3$:

1. La méthode paramétrique classique est conservatrice et son risque d'erreur de première espèce est toujours inférieur à $\alpha = 0.05$. Cependant, lorsque la taille d'échantillon croît, le risque se stabilise à 0.05.
2. Lorsque la taille d'échantillon est inférieure à 54, selon toutes les méthodes (sauf Kennedy et lorsque $\rho = 0$ Manly) le risque d'erreur de première espèce est inférieur à 0.05, fournissant ainsi un test conservateur, mais en augmentant la taille d'échantillon ce risque converge vers 0.05 (Figure 3 p. 290).
3. Pour de petites tailles d'échantillon ($n < 54$), lorsque le coefficient de la covariable est zéro ($\beta_1 = 0$) la méthode de Manly conserve un risque d'erreur de première espèce égale à 0.05 (Figure 3-a, 3-c p. 290), ce qui confirme la prévision Kennedy et Cade [34]. Par contre, en augmentant le coefficient de la covariable (β_1) et lorsque $\rho(X^1, X^2) > 0$, les méthodes de Freedman-Lane et de Kennedy conservent un risque d'erreur de première espèce égal à 0.05 alors que celui de la méthode de Manly est inférieur à 0.05, fournissant un test conservateur.
4. Lorsque la taille d'échantillon n augmente, les risques d'erreur de première espèce de toutes les méthodes de permutation, convergent vers le niveau de signification considéré ($\alpha = 0.05$) beaucoup plus rapidement que la méthode paramétrique classique.
5. Dans tous les cas, le risque d'erreur de première espèce de la méthode de Freedman-Lane est inférieur à celui de la méthode de Kennedy.

2.4.2 Analyse de la puissance du test

A- Le cadre de la simulation d'Anderson et Legendre

Anderson-Legendre n'ont pas considéré la méthode de Kennedy pour contrôler la puissance d'un test parce que selon eux la méthode de Kennedy augmente le risque d'erreur de première espèce.

Les ensembles de données sont générés comme dans le cas précédent mais en choisissant désormais $\beta_2 \neq 0$. Anderson-Legendre ont retenu des valeurs de β_2 déterminées en fonction de la taille d'échantillon afin de permettre de mesurer et de comparer la puissance. Le paramètre β_2 a été choisi assez grand pour pouvoir rejeter l'hypothèse nulle mais pas trop grand afin de ne pas obtenir un taux de 100% de rejets.

Comme le risque d'erreur de première espèce, la puissance est calculée comme le pourcentage de rejet de $H_0 : \beta_2 = 0$, pour un test de taille $\alpha = 0.05$ en générant 10000 ensembles de données générées avec $\beta_2 \neq 0$.

Donc l'ensemble des modalités des facteurs considérées sont :

- a) La valeur du paramètre $\beta_2 = 0.75$ pour $n = \{9, 18\}$, $\beta_2 = 0.4$ pour $n = \{36, 54\}$ et $\beta_2 = 0.3$ pour $n = \{72, 90\}$;
- b) La valeur du coefficient de la covariable X^1 est : $\beta_1 = \{0, 0.5, 1, 2.5\}$;
- c) Le degré de la corrélation entre les variables explicatives X^1 et X^2 est : $\rho(X^1, X^2) = \{0.1, 0.9\}$;
- d) La distribution du terme d'erreur aléatoire ϵ est : $\{LG(0, 1), Exp(1), Exp(1)^3\}$.

Pour chaque combinaison des modalités des facteurs, 10000 ensembles de données sont générés. Puis, pour chacune des trois méthodes, chaque ensemble de données est permuté 999 fois.

B- Les résultats de la simulation d'Anderson et Legendre

B1- Cas des données générées pour le terme d'erreur selon une loi de Gauss

1. En augmentant le coefficient de corrélation $\rho(X^1, X^2)$, la puissance augmente pour chacune des trois méthodes permutationnelles, ainsi que pour la méthode paramétrique classique.
2. Lorsque la taille d'échantillon croît, la puissance augmente pour chacune des trois méthodes permutationnelles, ainsi que pour la méthode paramétrique classique.
3. Il n'y a pas de différence significative entre les trois méthodes permutationnelles et la méthode paramétrique classique.

B2- Cas des données générées pour le terme d'erreur selon une loi de Exp(1)

Les résultats obtenus sont analogues au cas des données générées selon une loi de Gauss.

B3- Cas des données générées pour le terme d'erreur selon une loi radicalement non symétrique comme $Exp(1)^3$

1. On retrouve les trois premiers résultats obtenus dans le cas des données générées selon une loi de Gauss.
2. Quand $\rho(X^1, X^2) = 0$ et $\rho(X^1, X^2) = 0.5$, la puissance de la méthode paramétrique classique est significativement plus petite que pour les méthodes permutacionnelles. Par ailleurs, pour $\rho(X^1, X^2) = 0.9$, la puissance de la méthode paramétrique classique est en moyenne plus petite que celle des méthodes permutacionnelles mais pas significativement.
3. Lorsque la taille d'échantillon est $n = 9$, pour $0 \leq \beta_1 < 0.3$, la puissance de la méthode Freedman-Lane est supérieure à celle de la méthode ter Braak, par contre pour $0.3 \leq \beta_1 \leq 1$, la méthode ter Braak est plus puissante que la méthode Freedman-Lane, mais en accroissant la taille d'échantillon la différence s'estompe.

2.4.3 Analyse d'effet d'un point extrême

Kennedy et Cade [34] et Manly [41], en comparant les méthodes de Manly et de Kennedy, ont obtenus des résultats différents au sujet de l'effet d'un point extrême pour la variable X^1 sur le risque d'erreur de première espèce. La simulation faite par Anderson et Legendre [1] est plus complète et plus riche en information que celles qui sont dues à Kennedy et Cade [34] et à Manly [41].

Andersson-Legendre ont généré des données de la même façon que Manly ou Kennedy et Cade mais, du fait que l'étendue de valeurs attribuées aux facteurs dans la simulation faite par Anderson-Legendre [1] est plus grande que dans les simulations faites par Manly [41] et par Kennedy et Cade [34], la simulation faite par Anderson et Legendre [1] est plus complète et plus riche en information que celles faites par Kennedy et Cade [34] et à Manly [41]. Par ailleurs comme la section (2.4.2) où est étudié la puissance, ici, ils n'ont pas considéré la méthode de Kennedy.

A- Le cadre de la simulation d'Anderson et Legendre

Anderson-Legendre ont étudié l'effet d'un point extrême sur le risque d'erreur de première espèce pour l'ensemble des modalités des facteurs suivant :

- a) La taille de l'échantillon, $n = \{10, 20, 100\}$.
- b) La valeur du coefficient de la covariable X^1 est : $\beta_1 = \{0, 2, 4, 5, 6, 8, 10, 20\}$.
- c) La distribution du terme d'erreur aléatoire ϵ est : $\{LG(0, 1), Exp(1), Exp(1)^3\}$.

Afin de considérer l'effet de la corrélation entre des variables X^1 et X^2 , ils ont généré des données de deux façons différentes.

A1- Cas où les variables X^1 et X^2 sont non corrélées

Considérons le cadre de la simulation pour le risque d'erreur de première espèce. Au lieu de construire deux variables X^1 et X^2 à valeurs dans l'ensemble $\{1, 2, 3\}$, ils ont générées $n - 1$ valeurs de la variable X^1 d'une loi uniforme sur $[0, 3]$ et la n -ème valeur considérée est une valeur extrême comme 55. C'est-à-dire, $X_1^1, X_2^1, \dots, X_{n-1}^1 \sim u[0, 3]$, et la valeur $X_n^1 = 55$. Puis n valeurs de la variable X^2 sont générées d'une loi uniforme sur $[0, 3]$. C'est-à-dire, $X_1^2, X_2^2, \dots, X_n^2 \sim u[0, 3]$. Ensuite, ils construisirent deux vecteurs $\mathbf{X}^1, \mathbf{X}^2$ et la matrice \mathbf{W} comme suit :

$$\mathbf{W} = [\mathbf{X}^1 | \mathbf{X}^2] = \begin{bmatrix} X_1^1 & X_1^2 \\ X_2^1 & X_2^2 \\ \vdots & \vdots \\ X_n^1 & X_n^2 \end{bmatrix} \quad (2.12)$$

Enfin, les valeurs de Y sont générées par l'équation :

$$\mathbf{Y} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.13)$$

où : $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ est le vecteur des paramètres. Pour calculer le risque d'erreur de première espèce on doit prendre $\beta_2 = 0$. $\boldsymbol{\epsilon}$ est le vecteur du terme d'erreur aléatoire généré selon la distribution considérée.

Pour chaque combinaison des modalités des facteurs, 10000 ensembles de données ont été générés. Ensuite, pour chacune des trois méthodes, chaque ensemble de données est permuté 999 fois.

Remarque 2.4.1. *Compte tenu du fait que les variables X^1 et X^2 suivent une loi uniforme $u[0, 3]$, on obtient :*

1. $\mathbb{E}(X^1) = \mathbb{E}(X^2) = 1.5$
2. $\text{Var}(X^1) = \text{Var}(X^2) = 0.75$
3. $\rho(X^1, X^2) = 0$ car X^1 et X^2 sont indépendantes

A2- Cas où les variables X^1 et X^2 sont corrélées

Ici, l'ensemble des données de la variable X^1 est généré comme dans le cas B1, c'est-à-dire, $X_1^1, X_2^1, \dots, X_{n-1}^1 \sim u[0, 3]$, et la valeur $X_n^1 = 55$. Puis, n valeurs de la variable Z^2 sont générées selon une loi de uniforme sur $[0, 3]$. C'est-à-dire, $Z_1, Z_2, \dots, Z_n \sim u[0, 3]$. Puis, on définit une variable X^2 comme suit :

$$X^2 = X^1 + Z \quad (2.14)$$

Ensuite, on construit deux vecteurs $\mathbf{X}^1, \mathbf{X}^2$ et la matrice \mathbf{W}' comme suit :

$$\mathbf{W}' = [\mathbf{X}^1 | \mathbf{X}^2] = \begin{bmatrix} X_1^1 & X_1^2 \\ X_2^1 & X_2^2 \\ \vdots & \vdots \\ X_n^1 & X_n^2 \end{bmatrix} \quad (2.15)$$

Enfin, les valeurs de Y sont générées par l'équation :

$$\mathbf{Y} = \mathbf{W}'\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.16)$$

où : $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ est le vecteur des paramètres et pour calculer le risque d'erreur de première espèce il faut prendre $\beta_2 = 0$. $\boldsymbol{\epsilon}$ est le vecteur du terme d'erreur aléatoire généré selon la distribution considérée.

Pour chaque combinaison des modalités des facteurs, 10000 ensembles de données ont été générés. Puis, pour chacune des trois méthodes, chaque ensemble de données a été permuté 999 fois.

Remarque 2.4.2. *Compte tenu du fait que les variables X^1 et Z suivent une loi de $u[0, 3]$, on obtient :*

1. $\mathbb{E}(X^1) = \mathbb{E}(Z) = 1.5$
2. $\mathbb{E}(X^2) = \mathbb{E}(X^1) + \mathbb{E}(Z) = 3$
3. $\mathbb{V}ar(X^1) = \mathbb{V}ar(Z) = \frac{9}{12}$
- 4.

$$\begin{aligned} Cov(X^1, X^2) &= Cov(X^1 + X^1 + Z) \\ &= Cov(X^1, X^1) + Cov(X^1, Z) \\ &= \mathbb{V}ar(X^1) \text{ car } X^1 \text{ et } Z \text{ sont indépendantes} \end{aligned} \quad (2.17)$$

¹Remarquons que Manly a défini la variable X^2 telle que $X^2 = 0.5X^1 + Z$, mais Anderson-Legendre ont considéré une corrélation entre les variables X^1 et X^2 plus forte que celle de Manly. Anderson-Legendre ont défini la variable X^2 par cette équation

5. Par définition de la variance on obtient :

$$\begin{aligned}
 \mathbb{V}ar(X^2) &= \text{Cov}(X^1 + Z, X^1 + Z) \\
 &= \mathbb{V}ar(X^1) + \mathbb{V}ar(Z) \\
 &= 2\mathbb{V}ar(X^1) \\
 &= \frac{9}{6}
 \end{aligned} \tag{2.18}$$

6. Par définition de la corrélation on obtient :

$$\begin{aligned}
 \rho(X^1, X^2) &= \frac{\text{Cov}(X^1, X^2)}{\sqrt{\mathbb{V}ar(X^1)}\sqrt{\mathbb{V}ar(X^2)}} \\
 &= \frac{\mathbb{V}ar(X^1)}{\sqrt{\mathbb{V}ar(X^1)}\sqrt{2\mathbb{V}ar(X^1)}} \\
 &= \frac{1}{\sqrt{2}} \approx 0.7
 \end{aligned} \tag{2.19}$$

B- Les résultats de la simulation d'Anderson et Legendre

B1- Cas des données générées pour le terme d'erreur selon une loi de Gauss

1. Lorsque $\beta_1 \neq 0$, le risque d'erreur de première espèce de la méthode de Manly est dissymétrique instable, c'est-à-dire qu'il est tantôt supérieur à 0.05, et tantôt inférieur à 0.05
2. En augmentant la taille d'échantillon, le risque d'erreur de première espèce de la méthode du Manly se stabilise vers 0.05.
3. Les méthodes de Freedman-Lane, de ter Braak et la méthode paramétrique classique donnent des risques d'erreur de première espèce proches de 0.05.
4. Il n'y a pas de différence significative entre les deux méthodes permutacionnelles de Freedman-Lane, ter Braak et la méthode paramétrique classique.

B2- Cas des données générées pour le terme d'erreur selon une loi de $Exp(1)$

1. Le risque d'erreur de première espèce de la méthode de ter Braak est instable, mais en augmentant la taille d'échantillon il se stabilise vers 0.05.
2. Pour toutes les combinaisons des modalités des facteurs, la méthode de Freedman-Lane donne un risque d'erreur de première espèce proche de 0.05.
3. Le risque d'erreur de première espèce de la méthode de Manly est instable et en augmentant la taille d'échantillon il ne se stabilise pas.

B3- Cas des données générées pour le terme d'erreur selon une loi radicalement non symétrique comme $Exp(1)^3$

1. Les risques d'erreur de première espèce des méthodes de ter Braak et paramétrique classique sont instables avec la même évolution, mais pas avec la même étendue.
2. Le risque d'erreur de première espèce de la méthode de Manly est instable et en augmentant la taille d'échantillon, il ne se stabilise pas.
3. Pour toutes combinaisons des modalités des facteurs, la méthode Freedman-Lane donne un risque d'erreur de première espèce proche de 0.05.
4. Le risque d'erreur de première espèce de la méthode de Manly est opposé à ceux des méthodes ter Braak et paramétrique classique, c'est-à-dire lorsque les méthodes de ter Braak et paramétrique classique donnent un risque inférieur à 0.05 le risque de la méthode de Manly est supérieur à 0.05 et vice versa.
5. En augmentant la taille d'échantillon, la méthode de ter Braak donne un risque d'erreur de première espèce proche de 0.05 mais la méthode paramétrique classique reste conservatrice.

2.5 Notre simulation

2.5.1 Analyse du risque d'erreur de première espèce du test

Nous avons complété les résultats obtenus par Anderson-Legendre, en considérant en plus des distributions choisies par ces auteurs $\{LG(0, 1), Exp(1), Exp(1)^3\}$, des distributions symétriques avec variance finie $Student(5)$, symétrique avec variance infinie $Student(1)$ à dissymétrie droite comme $\{\Gamma(1, 0.5), \Gamma(1.5, 0.5), \Gamma(1.5, 2)\}$, radicalement à dissymétrie droite comme $\{\Gamma(0.5, 0.5), \Gamma(0.5, 2), \Gamma(1, 2), Exp(2)\}$ à dissymétrie à gauche $\{\Gamma(10, 2)\}$ de radicalement à dissymétrie gauche comme $\{\Gamma(10, 0.5)\}$ pour le terme d'erreur.

Par ailleurs, afin de faire une analyse critique des simulations d'Anderson et Legendre qui génèrent les valeurs de X^1 et X^2 selon une loi uniforme, nous avons généré les valeurs de X^1 et X^2 selon la loi de Gauss $LG(0, 1)$.

A- Le cadre de notre simulation

Nous avons étudié le risque d'erreur de première espèce pour l'ensemble des modalités des facteurs ci-dessous :

- a) Pour vérifier les effets de taille d'échantillon et comparer les résultats obtenus par Anderson-Legendre aux nôtres, nous avons tiré les mêmes tailles d'échantillon sauf pour $n = 54$. Donc, $n = \{9, 18, 36, 72, 90\}$

- b) Afin de vérifier les effets du degré de corrélation entre les variable X^1 et X^2 , nous avons considéré une corrélation faible comme 0.1 et une corrélation forte comme 0.9. Alors, $\rho(X^1, X^2) = \{0.1, 0.9\}$
- c) Afin de vérifier les effets du degré de liaison avec la covariable et comparer les résultats obtenus par Anderson-Legendre aux notres, nous avons considéré les mêmes coefficients sauf pour $\beta_1 = 1$. Donc, $\beta_1 = \{0, 2.41\}$.
- d) La distribution du terme d'erreur aléatoire ϵ : $\{LG(0, 1), Exp(0.5), Exp(2), Student(1), Student(5), \Gamma(a, b)$ avec $a \in \{0.5, 1, 1.5, 10\}$ et $b \in \{0.5, 2\}\}$

Nous générons deux vecteurs de dimension n \mathbf{Z}^1 et \mathbf{Z}^2 selon la loi de Gauss $LG(0, 1)$, comme suit :

$$\mathbf{W} = [\mathbf{Z}^1 | \mathbf{Z}^2] = \begin{bmatrix} Z_1^1 & Z_1^2 \\ Z_2^1 & Z_2^2 \\ \vdots & \vdots \\ Z_n^1 & Z_n^2 \end{bmatrix} \implies \mathbf{W}_c = [\mathbf{Z}_c^1 | \mathbf{Z}_c^2] \quad (2.20)$$

où \mathbf{Z}_c^1 et \mathbf{Z}_c^2 sont des vecteurs centrés réduits de valeurs des vecteurs \mathbf{Z}^1 , \mathbf{Z}^2 . Remarquons que $\rho(Z_c^1, Z_c^2) = 0$. A partir des \mathbf{Z}_c^1 et \mathbf{Z}_c^2 , nous construisons les vecteurs \mathbf{X}^1 et \mathbf{X}^2 de corrélation $\rho(X^1, X^2)$ de la façon suivante :

$$\mathbf{W}_R = [\mathbf{X}^1 | \mathbf{X}^2] = \mathbf{W}_c \mathbf{R}^{1/2} \quad (2.21)$$

où : $\mathbf{R} = \begin{bmatrix} 1 & \rho(X^1, X^2) \\ \rho(X^1, X^2) & 1 \end{bmatrix}$. Enfin, les valeurs de Y furent générées par l'équation :

$$\mathbf{Y} = \mathbf{W}_R \boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.22)$$

où : $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ est le vecteur des paramètres. Pour calculer le risque d'erreur de première espèce, il faut fixer $\beta_2 = 0$ et $\boldsymbol{\epsilon}$ est le vecteur du terme d'erreur aléatoire généré selon la distribution considérée.

Pour chaque combinaison de facteurs, 10000 ensembles de données ont été générés. Puis, pour chacune des deux méthodes, chaque ensemble de données a été permuté 999 fois. Les statistiques de tests permutationnelles ($t_{2,\pi}$) de ces méthodes sont calculées sur chaque ensemble de données permutées. Ensuite, après 999 permutations, la p-valeur permutationnelle pour chacune des 2 méthodes a été calculée par la formule :

$$p - valeur_\pi = \frac{\#(|t_\pi| \geq |t_{obs}|)}{1000}$$

Les autre étapes sont identiques à celles de la simulation d'Anderson-Legendre.

Le niveau de signification pour les tests a été fixé à $\alpha = 0, 05$. Alors, pour chaque ensemble de données quand la $p - valeur_\pi$ est inférieure à 0.05 on rejette l'hypothèse

$H_0 : \beta_2 = 0$. Sur les 10000 p -valeur $_{\pi}$, le pourcentage des P -valeur $_{\pi}$ (pour chaque méthode) inférieures à 0.05 donne le risque d'erreur de première espèce. En fait, ce pourcentage estime la probabilité de rejeter $H_0 : \beta_2 = 0$ lorsque H_0 est vrai (parce que les données sont générées sous $\beta_2 = 0$). De plus, l'intervalle de confiance à 95% est calculé à partir du risque d'erreur de première espèce (calculé sur 10.000 ensembles).

Afin de compléter la comparaison pour chaque ensemble des données, on a utilisé les tests paramétriques classiques qui consistent à calculer une statistique de Student et sa p -valeur donnée par la théorie Gaussienne a également été calculée. De la même façon, sur les 10000 ensembles de données le risque d'erreur de première espèce et son intervalle de confiance sont calculés par la théorie Gaussienne.

B- Les résultats de nos simulations

Nous présentons ici les résultats des simulations effectuées par nos soins en les comparant à ceux d'Anderson-Legendre.

B1- Cas des données générées pour le terme d'erreur selon une loi de Gauss :

1. Lorsque la taille d'échantillon est $n = 9$, le risque d'erreur de première espèce obtenu par la méthode de Kennedy est nettement supérieur à 0.05 en comparaison aux résultats obtenus par la même méthode pour des tailles d'échantillon plus grandes (Figure 2.1-a,b et c)
2. Lorsque $\rho = 0.9$ pour $n \geq 18$ les figures 2.1-b et d montrent des cas où le risque d'erreur de première espèce de la méthode paramétrique classique donne un risque d'erreur de première espèce supérieur à ceux des méthodes de Kennedy et de Freedman-Lane.
3. Pour la taille d'échantillon $n = 9$, le risque d'erreur de première espèce obtenu par la méthode de Kennedy est toujours supérieur à celui de la méthode de Freedman-Lane et cela quelles que soient les valeurs attribuées aux facteurs (Figure 2.1).
4. Les méthodes de Freedman-Lane et de Kennedy donnent des risques d'erreur de première espèce très proches, pour les tailles d'échantillon supérieures à 18 (Figure 2.1).
5. Les figures 2.1-a et b pour $n = 72$; 2.1-c pour $n = 36$ et 2.1-d pour $n = 90$ montrent qu'il existe des cas où le risque d'erreur de première espèce de la méthode Kennedy est inférieur à celui qu'on obtient avec la méthode de Freedman-Lane, ce qui n'est pas observé dans les simulations d'Anderson-Legendre.

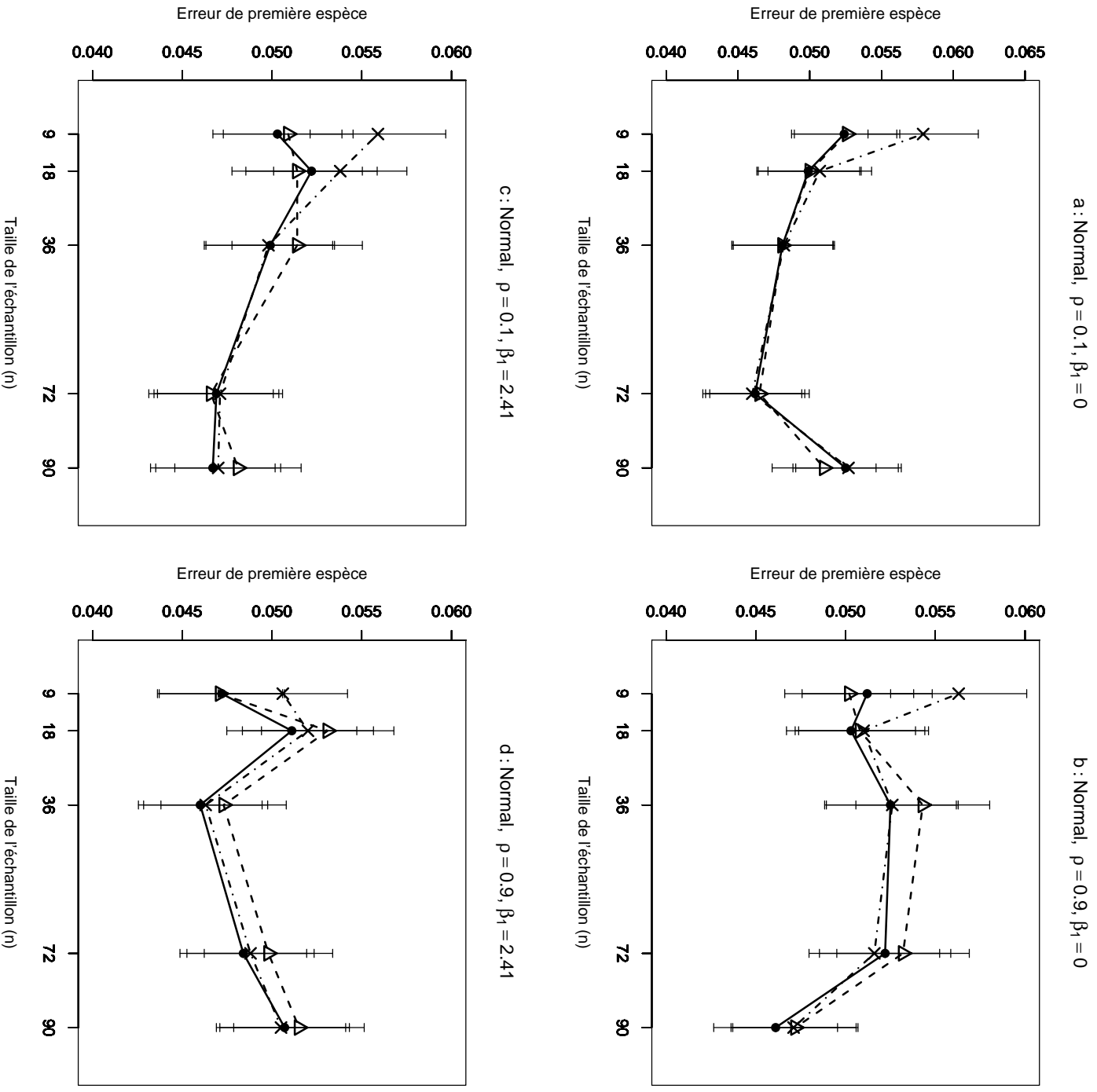


FIG. 2.1 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, Freedman et Lane— \bullet — et Kennedy— \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu e ont été simulés suivant une loi de Gauss $(0,1)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$. On a retenu $n=9,18,36,72,90$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1, \beta_1 = 0$; figure b : $\rho = 0.9, \beta_1 = 0$; figure c : $\rho = 0.1, \beta_1 = 2.41$, figure d : $\rho = 0.9, \beta_1 = 2.41$.

Pour comparer les effet de tailles d'échantillon nous avons considéré l'ensemble des cinq tailles d'échantillon $G = \{9, 18, 36, 72, 90\}$. Ensuite, nous avons partagé l'ensemble G en deux sous-ensembles, l'un des petites tailles $G_1 = \{9, 18\}$ et l'autre des grandes tailles $G_2 = \{36, 72, 90\}$. Ensuite, pour chaque méthode nous avons calculé la moyenne et l'étendue du risque d'erreur de première espèce sur G , sur G_1 et sur G_2 . Pour simplifier nous avons défini :

M : La moyenne des risques d'erreur de première espèce sur G .

R : L'étendue des risques d'erreur de première espèce sur G .

M1 : La moyenne des risques d'erreur de première espèce sur G_1 .

R1 : L'étendue des risques d'erreur de première espèce sur G_1 .

M2 : La moyenne des risques d'erreur de première espèce sur G_2 .

R2 : L'étendue des risques d'erreur de première espèce sur G_2 .

a : Normal, $\rho = 0.1, \beta_1 = 0$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0496	0.0061	0.0513	0.0026	0.0485	0.0045
Méthode de Freedman et Lane	0.0498	0.0063	0.0511	0.0025	0.0489	0.0063
Méthode de Kennedy	0.0511	0.0119	0.0543	0.0072	0.0489	0.0067
b : Normal, $\rho = 0.9, \beta_1 = 0$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0511	0.0071	0.0505	0.0006	0.0515	0.0071
Méthode de Freedman et Lane	0.0505	0.0064	0.0507	0.0009	0.0503	0.0064
Méthode de Kennedy	0.0517	0.0092	0.0536	0.0053	0.0504	0.0055
c : Normal, $\rho = 0.1, \beta_1 = 2.41$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0496	0.0048	0.0511	0.0005	0.0487	0.0048
Méthode de Freedman et Lane	0.0492	0.0055	0.0512	0.0019	0.0478	0.0032
Méthode de Kennedy	0.0507	0.0089	0.0548	0.0021	0.0479	0.0028
d : Normal, $\rho = 0.9, \beta_1 = 2.41$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0497	0.0061	0.0501	0.0060	0.0495	0.0042
Méthode de Freedman et Lane	0.0486	0.0051	0.0491	0.0039	0.0483	0.0047
Méthode de Kennedy	0.0496	0.0057	0.0513	0.0014	0.0485	0.0042

TAB. 2.1 – La moyenne et l'étendue des risques d'erreur de première espèce de chaque méthode sur les cinq tailles d'échantillon , sur les petites tailles $G_1 = \{9, 18\}$ et sur les grandes tailles $G_2 = \{36, 72, 90\}$. Figure a : $\rho = 0.1, \beta_1 = 0$; figure b : $\rho = 0.9, \beta_1 = 0$; figure c : $\rho = 0.1, \beta_1 = 2.41$; figure d : $\rho = 0.9, \beta_1 = 2.41$.

1. Dans les cas a, b et c, pour le groupe G , la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche à 0.05 que celle de la méthode de Kennedy.
2. Dans tous les cas, pour le groupe G , l'étendue du risque d'erreur de première espèce de la méthode de Freedman-Lane est inférieure à celle de la méthode de Kennedy.

3. Dans tous les cas, pour le groupe G_1 , la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche à 0.05 que celle de la méthode de Kennedy.
4. Dans les cas a, b et c, pour le groupe G_1 , l'étendue du risque d'erreur de première espèce de la méthode de Freedman-Lane est inférieure à celle de la méthode de Kennedy.
5. Dans tous les cas, pour le groupe G_2 , les méthodes de Freedman-Lane et Kennedy donnent presque la même moyenne du risque d'erreur de première espèce.
6. Dans la plupart des cas, pour le groupe G_2 , la méthode de Kennedy donne une étendue du risque d'erreur de première espèce inférieure à celle de la méthode de Freedman-Lane.
7. Dans tous les cas, la moyenne du risque d'erreur de première espèce de la méthode de Kennedy est plus proche de 0.05 pour le groupe G que pour le groupe G_1 .
8. Dans tous les cas, l'étendue du risque d'erreur de première espèce de la méthode de Kennedy pour le groupe G est supérieure à celle de la même méthode pour le groupe G_1 .
9. Dans tous les cas, la moyenne du risque d'erreur de première espèce de la méthode de Kennedy est plus proche de 0.05 pour le groupe G_2 que pour le groupe G_1 .
10. Dans tous les cas, l'étendue du risque d'erreur de première espèce de la méthode de Kennedy pour le groupe G_2 est inférieure à celle de la même méthode pour le groupe G_1 .
11. Dans tous les cas, les moyennes du risque d'erreur de première espèce de la méthode de Freedman-Lane pour les groupes G_1 et G_2 sont proches de 0.05 avec un même écart.
12. Dans tous les cas, l'étendue du risque d'erreur de première espèce de la méthode de Freedman-Lane pour le groupe G_1 est inférieure à celle de la même méthode pour le groupe G_2 .
13. Dans la plupart des cas pour les groupes G , G_1 et G_2 , la méthode Gaussienne (paramétrique classique) et celle de Freedman-Lane donnent presque les mêmes moyennes et étendues.
14. Dans la plupart des cas pour les ensembles G , G_1 et G_2 , la moyenne du risque d'erreur de première espèce de la méthode Gaussienne est plus proche de 0.05 que pour la méthode Kennedy. On constate aussi que l'étendue du risque d'erreur de première espèce de la méthode Gaussienne est inférieure à celle de la méthode de Kennedy.

B2- Cas des données générées pour le terme d'erreur selon une loi symétrique de variance infinie comme la loi de Student à 1 degré de liberté :

1. Lorsque la taille d'échantillon est $n = 9$, pour $\rho = 0.9$, le risque d'erreur de première espèce obtenu par la méthode de Kennedy est nettement supérieur à 0.05 en comparaison aux résultats obtenus par la même méthode pour des tailles d'échantillon plus grandes (Figure 2.2-a et c)
2. Même lorsque la variance des données générées pour le terme d'erreur est infinie, la méthode de Freedman-Lane conserve un risque d'erreur de première espèce très proche de 0.05, et cela quelles que soient les valeurs attribuées aux facteurs (Figure 2.2).
3. Même lorsque la variance des données générées pour le terme d'erreur est infinie, la méthode paramétrique classique donne le même risque d'erreur de première espèce que la méthode de Freedman-Lane (Figure 2.2).
4. Pour la taille d'échantillon $n = 9$, le risque d'erreur de première espèce obtenu par la méthode de Kennedy est toujours supérieur à celui de la méthode de Freedman-Lane et cela quelles que soient les valeurs attribuées aux facteurs (Figure 2.2).
5. Les méthodes de Freedman-Lane et de Kennedy donnent des risques d'erreur de première espèce très proches, pour les tailles d'échantillon supérieure à 18 (Figure 2.2).
6. Les figures 2.2-b pour $n = \{36, 72, 90\}$, 2.2-c pour $n = \{18, 72, 90\}$ et 2.1-d pour $n = \{36, 72\}$ montrent qu'il existe des cas où le risque d'erreur de première espèce de la méthode de Kennedy est inférieur à celui qu'on obtient avec la méthode de Freedman-Lane, ce qui n'est pas observé dans les simulations d'Anderson-Legendre. On remarque que ceci se produit souvent lorsque la taille d'échantillon est grande ($n > 18$). Nous en expliquerons les raisons dans le chapitre (3).

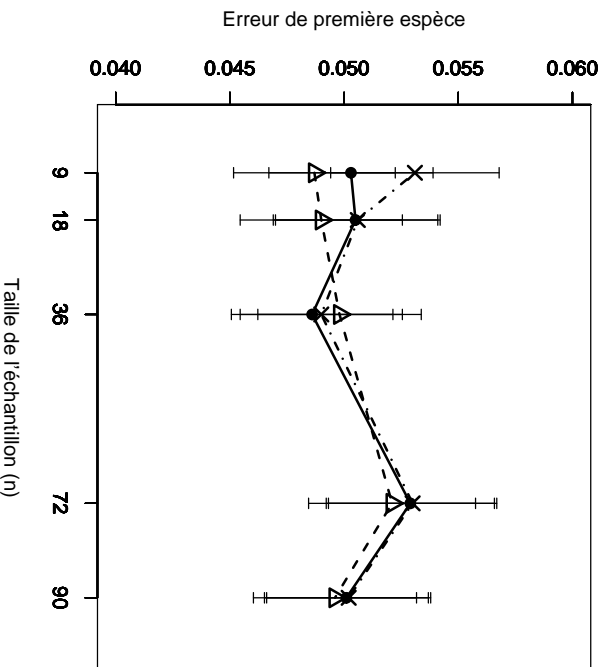
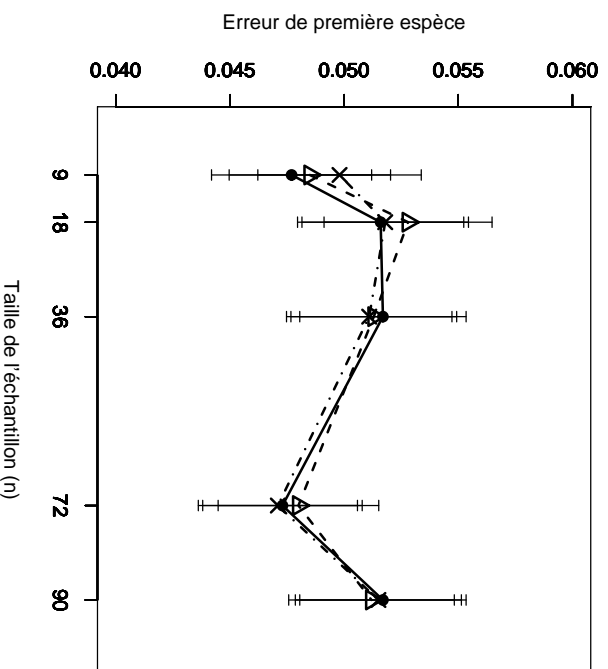
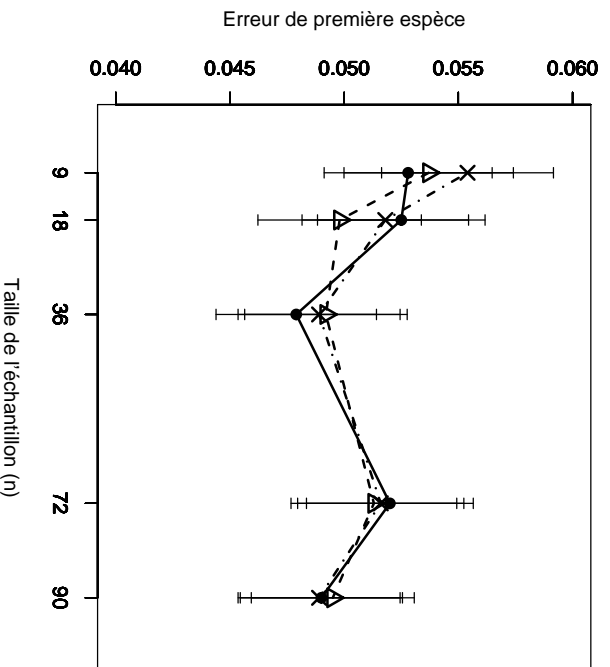
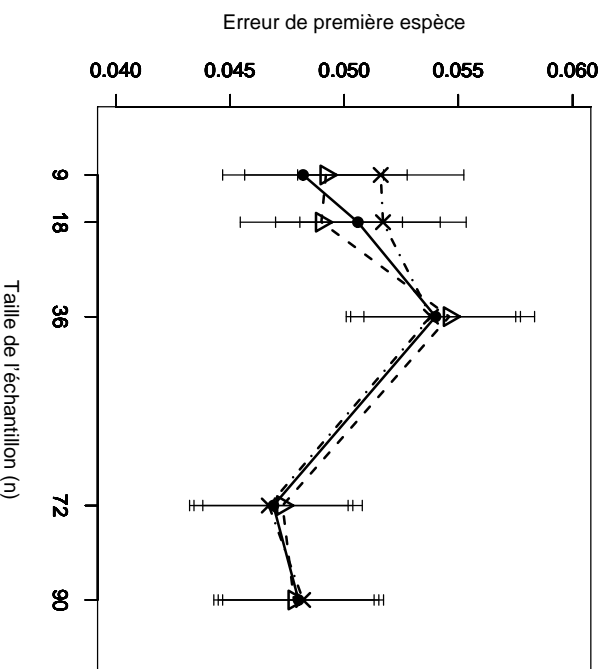
a : Student(1), $\rho = 0.1, \beta_1 = 0$ b : Student(1), $\rho = 0.9, \beta_1 = 0$ c : Student(1), $\rho = 0.1, \beta_1 = 2.41$ d : Student(1), $\rho = 0.9, \beta_1 = 2.41$ 

FIG. 2.2 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— Δ —, Freedman et Lane — \bullet — et Kennedy — \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu e ont été simulés suivant une loi de Student(1), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $n=9,18,36,72,90$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1, \beta_1 = 0$; figure b : $\rho = 0.9, \beta_1 = 0$; figure c : $\rho = 0.1, \beta_1 = 2.41$, figure d : $\rho = 0.9, \beta_1 = 2.41$.

Comme dans le cas B1, pour comparer les effets des tailles d'échantillon, nous avons partagé de la même façon l'ensemble G des cinq tailles considérées en deux sous-ensemble, l'un des petites tailles $G_1 = \{9, 18\}$ et l'autre des grandes tailles $G_2 = \{36, 72, 90\}$. Ensuite, pour chaque méthode nous avons calculé la moyenne et l'étendue sur G , sur G_1 et sur G_2 .

a : Student(1), $\rho = 0.1, \beta_1 = 0$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0498	0.0034	0.0488	0.0003	0.0505	0.0025
Méthode de Freedman et Lane	0.0504	0.0043	0.0504	0.0002	0.0505	0.0043
Méthode de Kennedy	0.0511	0.0041	0.0518	0.0025	0.0507	0.0040
b : Student(1), $\rho = 0.9, \beta_1 = 0$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0503	0.0043	0.0506	0.0043	0.0501	0.0033
Méthode de Freedman et Lane	0.0500	0.0044	0.0496	0.0039	0.0502	0.0044
Méthode de Kennedy	0.0502	0.0047	0.0508	0.0020	0.0499	0.0044
c : Student(1), $\rho = 0.1, \beta_1 = 2.41$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0507	0.0042	0.0517	0.0039	0.0500	0.0021
Méthode de Freedman et Lane	0.0508	0.0049	0.0526	0.0003	0.0496	0.0041
Méthode de Kennedy	0.0513	0.0065	0.0536	0.0036	0.0498	0.0027
d : Student(1), $\rho = 0.9, \beta_1 = 2.41$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0495	0.0073	0.0491	0.0002	0.0499	0.0073
Méthode de Freedman et Lane	0.0495	0.0071	0.0494	0.0034	0.0496	0.0071
Méthode de Kennedy	0.0504	0.0071	0.0516	0.0001	0.0495	0.0071

TAB. 2.2 – La moyenne et l'étendue des risques d'erreur de première espèce de chaque méthode sur toutes les tailles d'échantillon, sur les petites tailles $G_1 = \{9, 18\}$ et sur les grandes tailles $G_2 = \{36, 72, 90\}$. Figure a : $\rho = 0.1, \beta_1 = 0$; figure b : $\rho = 0.9, \beta_1 = 0$; figure c : $\rho = 0.1, \beta_1 = 2.41$; figure d : $\rho = 0.9, \beta_1 = 2.41$.

1. Dans les cas a, b et c, pour le groupe G , la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche de 0.05 que celle de la méthode de Kennedy.
2. Dans tous les cas, pour le groupe G , l'étendue du risque d'erreur de première espèce de la méthode de Freedman-Lane est inférieure ou égale à celle de la méthode de Kennedy.
3. Dans tous les cas, pour le groupe G_1 , la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche à 0.05 que celle de la méthode de Kennedy.
4. Dans les cas b et d (où $\rho = 0.9$), pour le groupe G_1 , l'étendue du risque d'erreur de première espèce de la méthode de Freedman-Lane est supérieure à celle de la méthode de Kennedy.

5. Dans tous les cas, pour le groupe G_2 , les méthode de Freedman-Lane et de Kennedy donnent presque la même moyenne du risque d'erreur de première espèce.
6. Dans la plupart des cas, pour le groupe G_2 , la méthode de Kennedy donne une étendue du risque d'erreur de première espèce inférieure ou égale à celle de la méthode de Freedman-Lane
7. Dans tous les cas, la moyenne du risque d'erreur de première espèce de la méthode de Kennedy est plus proche de 0.05 pour le groupe G que pour le groupe G_1 .
8. Dans tous les cas, l'étendue du risque d'erreur de première espèce de la méthode de Kennedy pour le groupe G est supérieure à celle de la même méthode pour le groupe G_1 .
9. Dans tous les cas, la moyenne du risque d'erreur de première espèce de la méthode de Kennedy est plus proche de 0.05 pour le groupe G_2 que pour le groupe G_1 .
10. Dans tous les cas, l'étendue du risque d'erreur de première espèce de la méthode de Kennedy pour le groupe G_2 est inférieure à celle de la même méthode pour le groupe G_1 .
11. Dans les cas a, b et d, les moyennes du risque d'erreur de première espèce de la méthode de Freedman-Lane pour les groupes G_1 et G_2 sont proches de 0.05 avec un même écart.
12. Dans tous les cas, l'étendue du risque d'erreur de première espèce de la méthode de Freedman-Lane pour le groupe G_1 est inférieure à celle de la même méthode pour le groupe G_2 .
13. Dans la plupart des cas, pour les groupes G , G_1 et G_2 , les méthodes Gaussienne (paramétrique classique) et Freedman-Lane donnent presque les mêmes moyennes et étendues.
14. Dans tous les cas, pour le groupe G , les méthodes de Kennedy et Gaussienne donnent des moyennes du risque d'erreur de première espèce proches de 0.05 avec presque un même écart et dans la plupart des cas, l'étendue du risque d'erreur de première espèce de la méthode de Gaussienne est inférieure à celle de la méthode de Kennedy. Par ailleurs, pour les groupes G_1 et G , la moyenne du risque d'erreur de première espèce de la méthode de Gaussienne est plus proche de 0.05 que la méthode de Kennedy, mais, on ne peut pas distinguer ses étendues.

B3- Cas des données générées pour le terme d'erreur selon une loi radicalement à dissymétrie droite comme la loi Exp(2) :

1. Comme dans le cas Gaussien, lorsque la taille d'échantillon est $n = 9$, le risque d'erreur de première espèce obtenu par la méthode de Kennedy est nettement supérieur à 0.05 en comparaison aux résultats obtenus par la même méthode pour des tailles d'échantillon plus grandes, et cela quelles que soient les valeurs attribuées aux facteurs.(Figure 2.3)
2. Pour la taille d'échantillon $n = 9$, le risque d'erreur de première espèce obtenu par la méthode de Kennedy est toujours supérieur à celui de la méthode de Freedman-Lane et cela quelles que soient les valeurs attribuées aux facteurs (Figure 2.3).
3. Contrairement à la simulation d'Anderson-Legendre pour laquelle la méthode de Freedman-Lane et la méthode paramétrique classique étaient conservatrices, nos simulations montrent ici que ces méthodes conservent des risques d'erreur de première espèce autour de 0.05.
4. Contrairement à la simulation d'Anderson-Legendre, la méthode paramétrique classique donne un risque d'erreur de première espèce très peu différent de la méthode de Freedman-Lane. Ceci est peut être dû au fait que les valeurs des variables X^1 et X^2 sont générées par nous selon une loi de Gauss et par Anderson-Legendre selon une loi uniforme discrète.
5. Les figures 2.3-a pour $n = 72$, 2.3-d pour $n = 36$ où le risque d'erreur de première espèce de la méthode Kennedy est inférieur à celui de la méthode de Freedman-Lane.

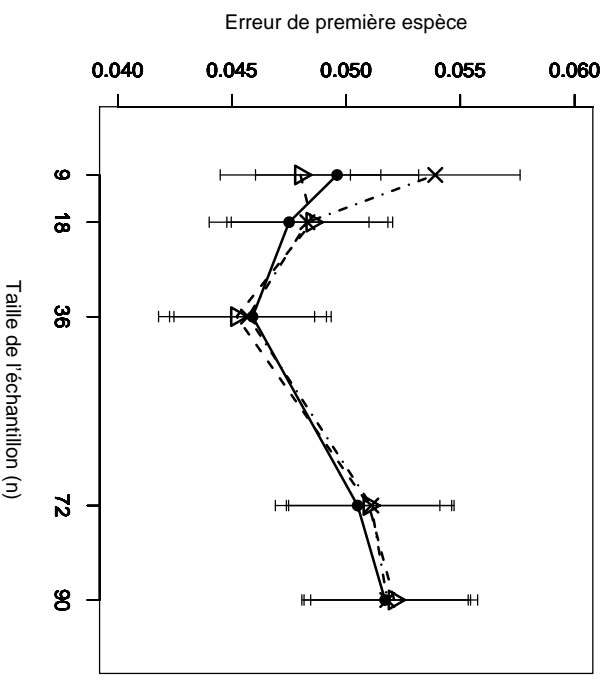
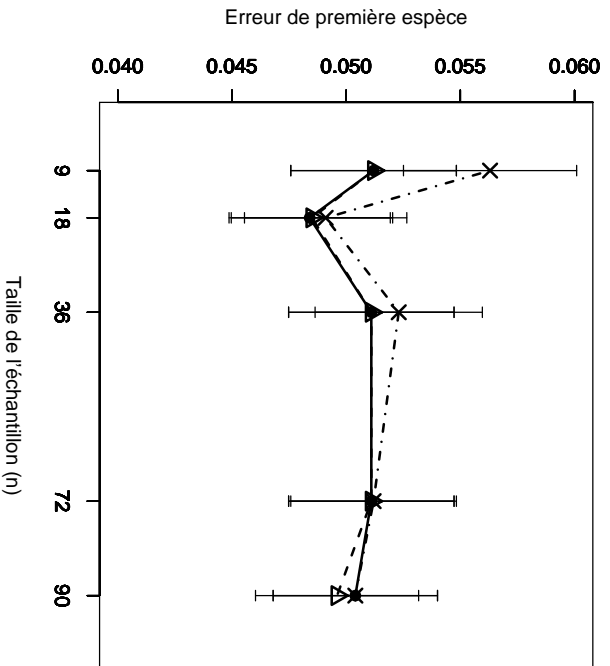
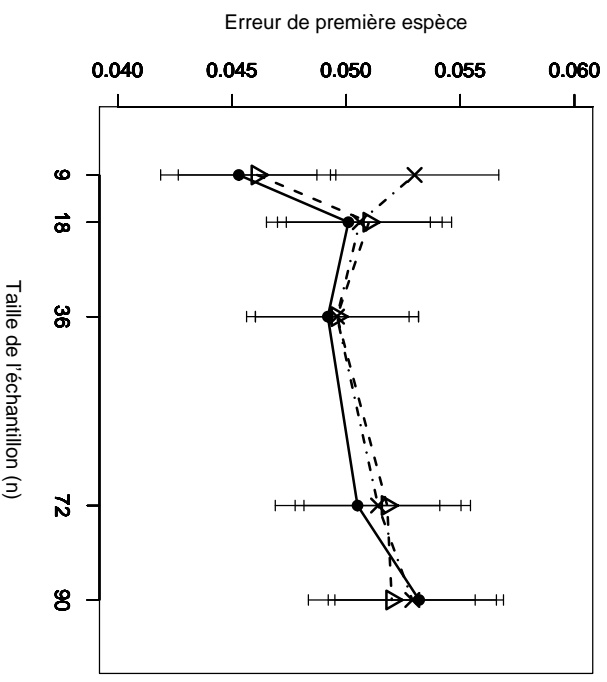
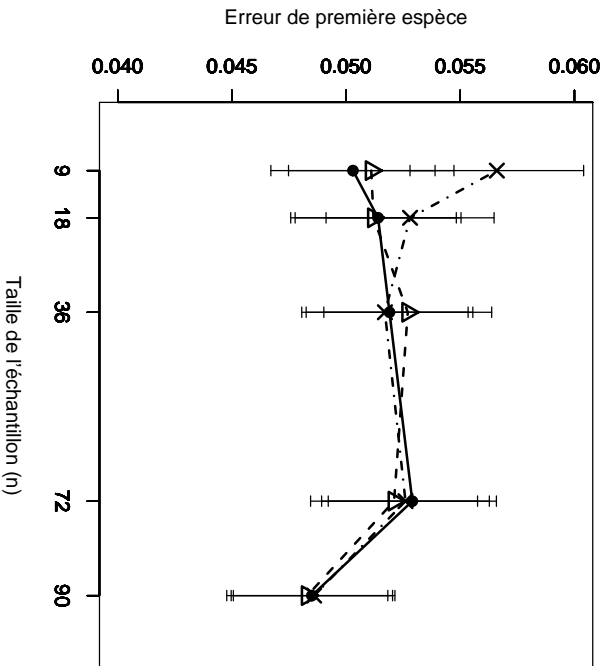


FIG. 2.3 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, Freedman et Lane— \bullet — et Kennedy— \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu e ont été simulés suivant une loi de Exp(2), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $n=9,18,36,72,90$. Les segments verticaux figurent l'intervalle de confiance à 95%.
Figure a : $\rho = 0.1, \beta_1 = 0$; figure b : $\rho = 0.9, \beta_1 = 0$; figure c : $\rho = 0.1, \beta_1 = 2.41$, figure d : $\rho = 0.9, \beta_1 = 2.41$.

Afin de comparer les effets des tailles d'échantillon, pour chaque méthode, nous avons calculé la moyenne et l'étendue sur les 3 groupes G , G_1 et G_2 .

a : Exp(2), $\rho = 0.1, \beta_1 = 0$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0501	0.0044	0.0511	0.0001	0.0510	0.0044
Méthode de Freedman et Lane	0.0510	0.0044	0.0508	0.0011	0.0511	0.0044
Méthode de Kennedy	0.0524	0.0080	0.0547	0.0038	0.0509	0.0044
b : Exp(2), $\rho = 0.9, \beta_1 = 0$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0501	0.0059	0.0485	0.0049	0.0511	0.0024
Méthode de Freedman et Lane	0.0496	0.0079	0.0477	0.0048	0.0509	0.004
Méthode de Kennedy	0.0515	0.0034	0.0518	0.0024	0.0513	0.0033
c : Exp(2), $\rho = 0.1, \beta_1 = 2.41$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0503	0.0027	0.0498	0.0027	0.0506	0.0015
Méthode de Freedman et Lane	0.0504	0.0027	0.0498	0.0028	0.0508	0.0007
Méthode de Kennedy	0.0518	0.0072	0.0527	0.0072	0.0513	0.0019
d : Exp(2), $\rho = 0.9, \beta_1 = 2.41$	M	R	M1	R1	M2	R2
Méthode Gaussienne	0.0498	0.0041	0.0482	0.0005	0.0494	0.0069
Méthode de Freedman et Lane	0.0490	0.0058	0.0485	0.0021	0.0493	0.0060
Méthode de Kennedy	0.0501	0.0082	0.0511	0.0056	0.0495	0.0061

TAB. 2.3 – La moyenne et l'étendue des risques d'erreur de première espèce de chaque méthode sur toutes les tailles d'échantillon, sur les 5 petites tailles $G_1 = \{9, 18\}$ et sur les grandes tailles $G_2 = \{36, 72, 90\}$. Figure a : $\rho = 0.1, \beta_1 = 0$; figure b : $\rho = 0.9, \beta_1 = 0$; figure c : $\rho = 0.1, \beta_1 = 2.41$; figure d : $\rho = 0.9, \beta_1 = 2.41$.

1. Dans tous les cas, pour le groupe G , la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche de 0.05 que celle de la méthode de Kennedy.
2. Dans les cas a, c et d, pour le groupe G , l'étendue du risque d'erreur de première espèce de la méthode de Freedman-Lane est inférieure à celle de la méthode de Kennedy.
3. Dans les cas b et d (où $\rho = 0.1$), pour le groupe G_1 , la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche de 0.05 que celle de la méthode de Kennedy.
4. Dans les cas a, c et d, pour le groupe G_1 , l'étendue du risque d'erreur de première espèce de la méthode de Freedman-Lane est supérieure à celle de la méthode de Kennedy.
5. Dans tous les cas, pour le groupe G_2 , les méthode de Freedman-Lane et de Kennedy donnent presque la même moyenne du risque d'erreur de première espèce.

6. Dans tous les cas, pour le groupe G_2 , la méthode de Freedman-Lane donne une étendue du risque de erreur de première espèce inférieure ou égale à celle de la méthode de Kennedy
7. Dans tous les cas, la moyenne du risque d'erreur de première espèce de la méthode de Kennedy est plus proche de 0.05 pour le groupe G que pour le groupe G_1 .
8. Dans tous les cas, l'étendue du risque d'erreur de première espèce de la méthode de Kennedy pour le groupe G est supérieure ou égale à celle de la même méthode pour le groupe G_1 .
9. Dans tous les cas la moyenne du risque d'erreur de première espèce de la méthode de Kennedy est plus proche de 0.05 pour le groupe G_2 que pour le groupe G_1 .
10. Dans les cas a, b et d, l'étendue du risque d'erreur de première espèce de la méthode de Kennedy pour le groupe G_2 est supérieure à celle de la même méthode pour le groupe G_1 .
11. Dans les cas b et d (où $\rho = 0.09$), les moyennes du risque d'erreur de première espèce de la méthode de Freedman-Lane sont plus proches de 0.05 pour le groupes G_2 que pour le groupe G_1 .
12. Dans les cas b et c, l'étendue du risque d'erreur de première espèce de la méthode de Freedman-Lane pour le groupe G_2 est inférieure à celle de la même méthode pour le groupe G_1 .
13. Dans la plupart des cas, pour le groupe G , la moyenne du risque d'erreur de première espèce de la méthode de Gaussienne est plus proche de 0.05 que celle de la méthode de Freedman-Lane, mais pour les groupes G_1 et G_2 les deux méthodes donnent presque les mêmes moyennes. Par ailleurs, dans la plupart des cas, l'étendue du risque d'erreur de première espèce de la méthode de Gaussienne est inférieure à celle de la méthode de Freedman-Lane.
14. Dans la plupart des cas pour les groupes G , G_1 et G_2 , la moyenne du risque d'erreur de première espèce de la méthode de Gaussienne est plus proche de 0.05 que celle de la méthode de Kennedy. De plus, l'étendue du risque d'erreur de première espèce de la méthode de Gaussienne est inférieure à celle de la méthode de Kennedy.

Des résultats plus complets, utilisant d'autres combinaisons des facteurs sont présentés dans l'annexe 1.

B-4 Cas des données générées pour le terme d'erreur selon une loi radicalement à dissymétrie gauche comme la loi $\Gamma(10, 0.5)$:

Les résultats dans ce cas sont les mêmes que dans le cas de données générées pour que le terme d'erreur suive une loi de $Exp(2)$.

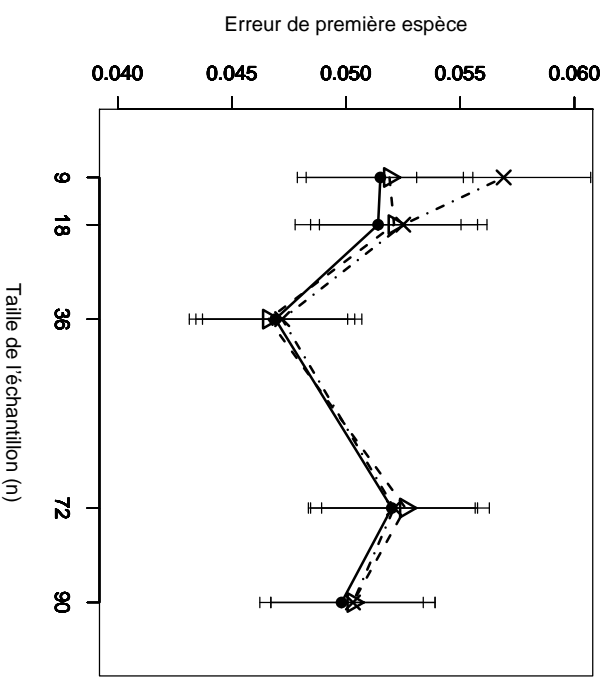
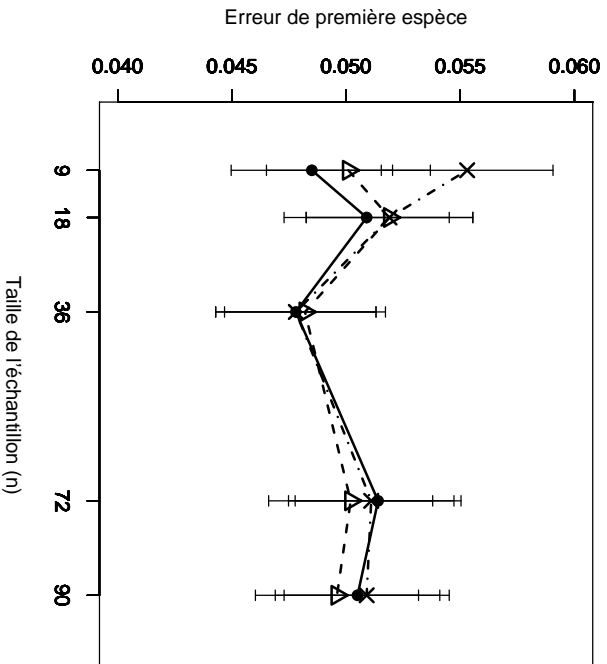
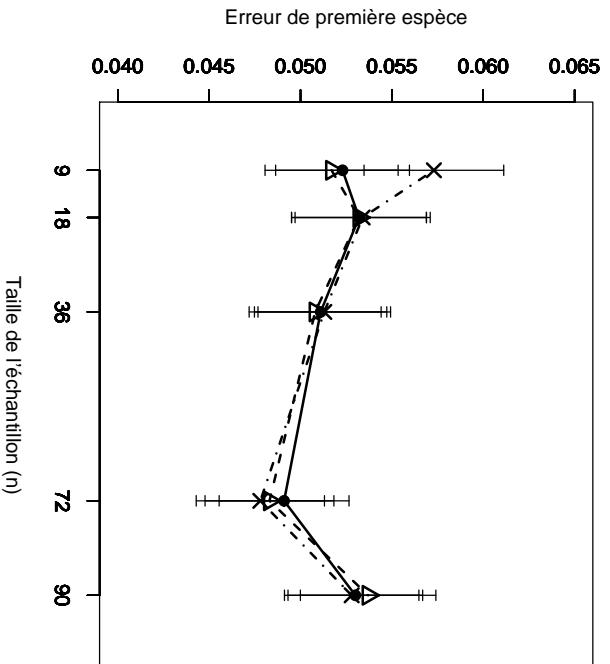
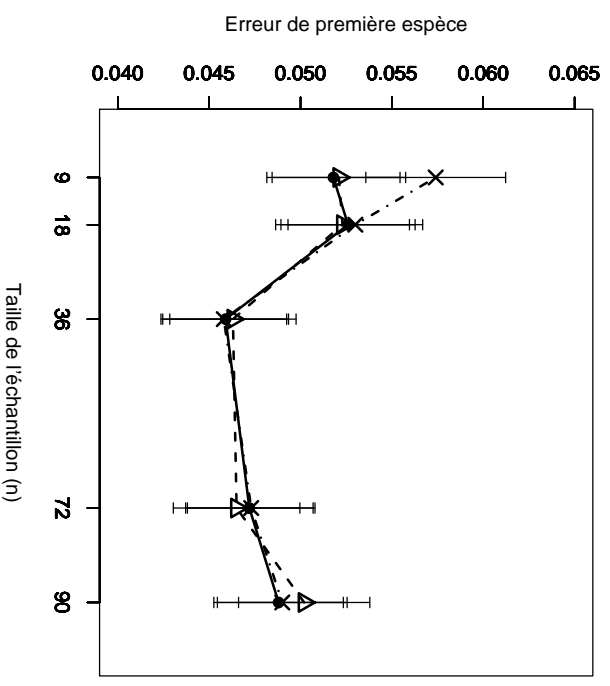
c : $\Gamma(10, 0.5)$, $\rho = 0.1, \beta_1 = 2.41$ d : $\Gamma(10, 0.5)$, $\rho = 0.9, \beta_1 = 2.41$ 

FIG. 2.4 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $-\triangle-$, Freedman et Lane $-\bullet-$ et Kennedy $-\times\cdot-$, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $\Gamma(10, 0.5)$, X^1 et X^2 sont toujours Gaussiennes $(0, 1)$. On a retenu $n=9, 18, 36, 72, 90$. Les segments verticaux figurent l'intervalle de confiance à 95%.
Figure a : $\rho = 0.1, \beta_1 = 0$; figure b : $\rho = 0.9, \beta_1 = 0$; figure c : $\rho = 0.1, \beta_1 = 2.41$, figure d : $\rho = 0.9, \beta_1 = 2.41$.

C- Résumé des principaux résultats :

- a) Dans tous les cas, pour $n = 9$ la méthode de Kennedy donne le risque d'erreur de première espèce nettement supérieure à 0.05 en comparaison aux résultats obtenus par la même méthode pour des tailles d'échantillon plus grandes, et cela quelles que soient les valeurs attribuées aux facteurs.
- b) La méthode de Freedman-Lane donne un risque d'erreur de première espèce plus stable que la méthode de Kennedy en restant conservant proche du niveau de signification considéré $\alpha = 0.05$.
- c) Contrairement au résultat de la simulation d'Anderson-Legendre (et peut être à cause des valeurs des variables X^1 et X^2 générées dans nos simulations selon une loi de Gauss), quand les données générées pour le terme d'erreur suivent une loi radicalement dissymétrique comme la loi $Exp(2)$ ou $\Gamma(10, 0.5)$, dans la plupart des cas la méthode paramétrique classique donne un risque d'erreur de première espèce proche du niveau de signification considéré $\alpha = 0.05$.
- d) Contrairement à la simulation d'Anderson-Legendre pour laquelle la méthode de Freedman-Lane était conservatrice, lorsque les données générées pour le terme d'erreur suivaient une loi radicalement dissymétrique, nos simulations montrent ici que cette méthode conserve des risques d'erreur de première espèce autour 0.05.
- e) Contrairement à la méthode de Freedman-Lane, pour les petites tailles d'échantillon $G_1 = 9, 18$ (surtout pour $n = 9$) la méthode de Kennedy donne un risque d'erreur de première espèce supérieur du niveau de signification considéré $\alpha = 0.05$.
- f) Pour les petites tailles d'échantillon $G_1 = \{9, 18\}$ la méthode paramétrique classique donne un risque d'erreur de première espèce plus proche de 0.05 que la méthode de Kennedy.
- g) Pour les grandes tailles d'échantillon $G_2 = \{36, 72, 90\}$ les méthodes de Freedman-Lane et de Kennedy donnent des risques d'erreur de première espèce proches du niveau de signification considéré $\alpha = 0.05$.
- h) Contrairement aux résultats d'Anderson-Legendre, nous avons observé des cas où le risque d'erreur de première espèce de la méthode Kennedy est inférieur à celui de Freedman-Lane lorsque la taille d'échantillon était grande ($n > 18$). Nous en expliquerons les raisons Dans le chapitre (3).
- i) Au bilan, la méthode Freedman-Lane donne un risque d'erreur de première espèce plus stable que les autres méthodes et restant proche du niveau de signification considéré $\alpha = 0.05$.

2.5.2 Analyse de la puissance du test

En s'inspirant de la simulation d'Anderson-Legendre nous n'avons pas considéré la méthode Kennedy pour étudier la puissance. Comme pour le risque d'erreur de première espèce, ici nous avons complété les résultats obtenus par Anderson-

Legendre, en considérant en plus des distributions choisies par ces auteurs $\{LG(0, 1), Exp(1), Exp(1)^3\}$ des distributions de terme d'erreur symétriques avec variance finie $Student(5)$, symétrique avec variance infinie $Student(1)$ à dissymétrie droite comme $\{\Gamma(1, 0.5), \Gamma(1.5, 0.5), \Gamma(1.5, 2)\}$, radicalement à dissymétrie droites comme $\{\Gamma(0.5, 0.5), \Gamma(0.5, 2), \Gamma(1, 2)\}$ à dissymétrie gauche $\{\Gamma(10, 2)\}$ et radicalement à dissymétrie gauches comme $\{\Gamma(10, 0.5)\}$.

Par ailleurs, afin de comparer les effets des valeurs des données X^1 et X^2 qui sont générées par Anderson-Legendre suivant une loi uniforme discrète et celles générées selon une loi continue, nous avons généré des données X^1 et X^2 provenant d'une loi de Gauss $(0, 1)$.

A- Le cadre de notre simulation

Nous avons étudié la puissance pour l'ensemble de modalités des facteurs ci-dessous :

- a) Pour vérifier les effets de taille d'échantillon et aussi pour permettre la comparaison des résultats obtenus par Anderson-Legendre aux nôtres, nous avons tiré les mêmes tailles d'échantillon. Comme dans le cadre d'Anderson-Legendre nous avons retenu des valeurs de β_2 déterminées en fonction de la taille d'échantillon afin de permettre de mesurer et de comparer la puissance : β_2 a été choisi assez grand pour pouvoir rejeter l'hypothèse nulle mais pas trop grand afin de ne pas obtenir un taux de 100% de rejets. Donc, on a retenu comme valeur du paramètre $\beta_2 = 0.75$ pour $n = \{9, 18\}$, $\beta_2 = 0.4$ pour $n = \{36, 54\}$ et $\beta_2 = 0.3$ pour $n = \{72, 90\}$.
- b) Pour étudier les effets du degré de corrélation entre les variables X^1 et X^2 , nous avons considéré une corrélation faible comme 0.1 et une corrélation forte comme 0.9. Alors, $\rho(X^1, X^2) = \{0.1, 0.9\}$
- c) Afin de vérifier les effets du degré d'influence du coefficient de la covariable conditionnant et comparer les résultats obtenus par Anderson-Legendre avec nos résultats, nous avons considéré les mêmes coefficients. Donc, $\beta_1 = \{0, 0.5, 1, 2.5\}$.
- d) La distribution du terme d'erreur aléatoire ϵ est choisie parmi les suivantes : $\{LG(0, 1), Exp(0.5), Exp(2), Student(1), Student(5), \Gamma(a, b)$ avec $a \in \{0.5, 1, 1.5, 10\}$ et $b \in \{0.5, 2\}\}$

Puis des valeurs de Y furent calculées selon l'équation (2.11)

$$\mathbf{Y} = \mathbf{W}_R \boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

où : $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ est le vecteur des paramètres. Pour calculer la puissance il faut

fixer $\beta_2 \neq 0$ et ϵ est le vecteur du terme d'erreur aléatoire généré selon l'une des distributions considérées.

Pour chaque combinaison de facteurs, 10000 ensembles de données ont été générés. Chaque ensemble de données est permuté 999 fois. La statistique de test permutationnelle ($t_{2,\pi}^F$) de la méthode de Freedman-Lane a été calculée sur chaque ensemble de données permutoées. Ensuite, après 999 permutations, la p-valeur permutationnelle pour la méthode de Freedman-Lane a été calculée par la formule :

$$p - valeur_{\pi} = \frac{\#(|t_{\pi}| \geq t_{obs})}{1000}$$

Le niveau de signification pour les tests a été fixé à $\alpha = 0,05$. Alors, pour chaque ensemble de données quand la $p - valeur_{\pi}$ est inférieure à 0.05 on rejette l'hypothèse $H_0 : \beta_2 = 0$. Sur les 10000 $p - valeur_{\pi}$ le pourcentage des $P - valeur_{\pi}$ de la méthode de Freedman-Lane inférieures à 0.05 donne le risque d'erreur de première espèce. En fait, ce pourcentage estime la probabilité de rejeter $H_0 : \beta_2 = 0$ lorsque H_0 est vraie (parce que les données sont générées sous $\beta_2 = 0$). De plus, l'intervalle de confiance à 95% est calculé à partir du risque d'erreur de première espèce calculé sur 10000 ensembles.

Afin de compléter la comparaison, pour chaque ensemble de données on a utilisé les tests paramétriques classiques qui consistent à calculer une statistique de Student. Sa p-valeur donnée par la théorie Gaussienne a également été calculée. De la même façon, sur les 10000 ensembles de données le risque d'erreur de première espèce et son intervalle de confiance sont calculés par la théorie Gaussienne.

B- Les résultats de nos simulations

B1- Cas des données générées pour le terme d'erreur selon une loi de Gauss :

1. En comparant les figures 2.5-a où $n = 9$ et 2.5-c où $n = 18$, on observe qu'en augmentant la taille d'échantillon (les valeurs des autres facteurs restant fixées), les puissances des méthodes permutationnelle de Freedman-Lane et paramétrique classique augmentent nettement. On peut trouver la même relation pour les figures 2.5-b où $n = 9$ et 2.5-d où $n = 18$. Pour les autres combinaisons des facteurs, dans les figures 2.6 et 2.7 on trouve encore la même relation. C'est-à-dire qu'en augmentant la taille d'échantillon (les valeurs des autres facteurs restant fixées), les puissances de la méthode permutationnelle de Freedman-Lane et de la méthode paramétrique classique augmentent nettement.
2. En comparant les figures 2.5-a où $\rho = 0.1$ et 2.5-b où $\rho = 0.9$, on observe qu'en augmentant le coefficient de corrélation $\rho(X^1, X^2)$ (les valeurs des

autres facteurs restant fixées), les puissances de la méthode permutatonnelle de Freedman-Lane et de la méthode paramétrique classique diminuent nettement. On peut trouver la même relation pour les figures 2.5-c où $\rho = 0.1$ et 2.5-d où $\rho = 0.9$. Pour les autres combinaison des facteurs, dans les figures 2.6 et 2.7 on trouve encore, la même relation. C'est-à-dire qu'en augmentant le degré du corrélation (les valeurs des autres facteurs restant fixées), les puissances de la méthode permutatonnelle de Freedman-Lane et de la méthode paramétrique classique diminuent nettement.

3. Ici, la méthode de Freedman-Legendre a presque la même puissance que celle obtenue dans la simulation d'Anderson-Legendre quelles que soient les valeurs des facteurs.
4. La puissance de la méthode paramétrique classique est presque la même que celle de la méthode Freedman-Lane, et cela quelles que soient les valeurs attribuées aux facteurs. On remarque que ceci est acceptable puisque les données générées pour le terme d'erreur suivent une loi de Gauss $(0, 1)$.
5. En considérant les figures 2.5-a, 2.6-a et 2.7-a où pour les trois figures ρ est fixé à 0.1, on obtient que la puissance des méthodes Freedman-Lane et paramétrique classique augmentent. On peut trouver la même relation pour les figures de la partie b, mais pour les figures des parties c et d, les puissances sont conservées respectivement entre les intervalles $(0.78-0.81)$ et $(0.22-0.24)$.

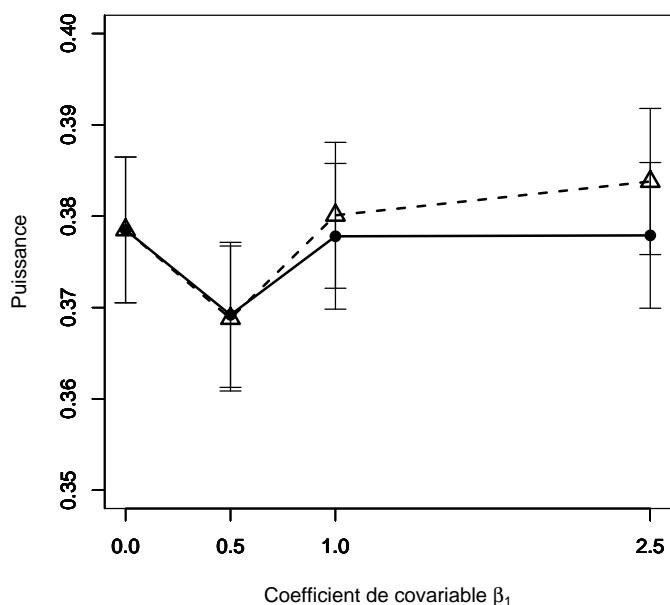
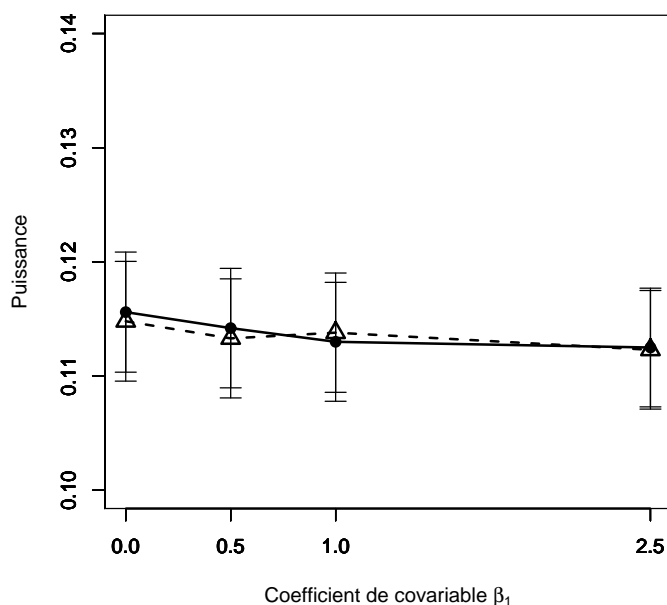
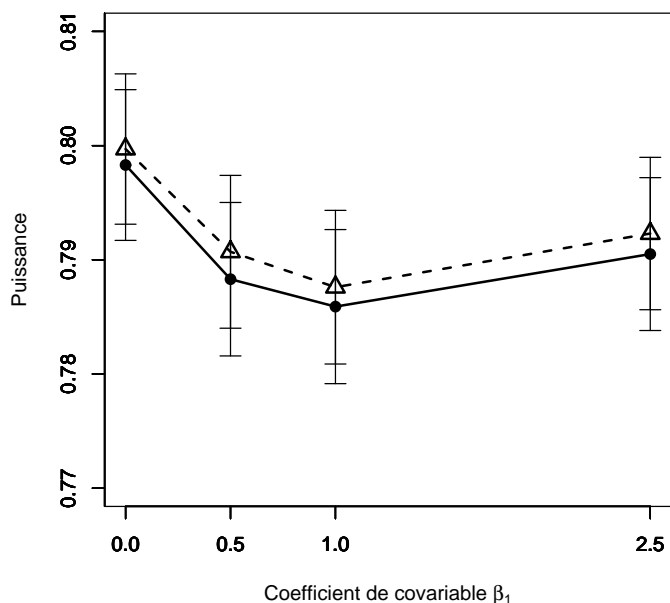
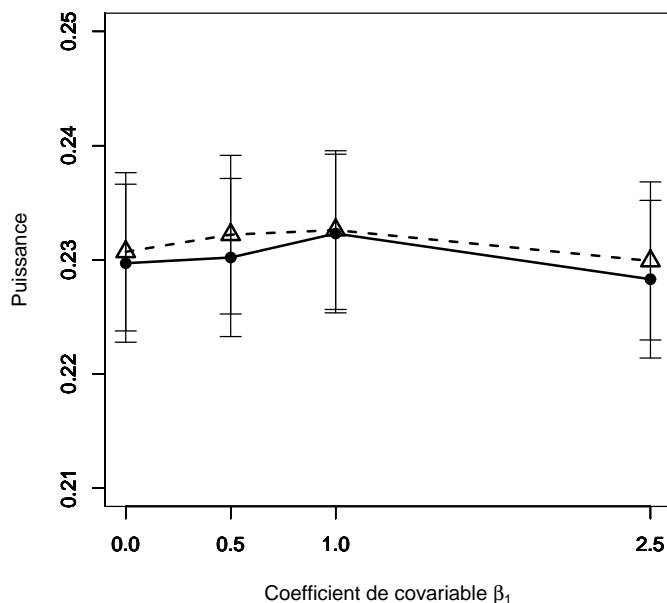
a: Gauss N(0,1), $\rho = 0.1$, $\beta_2 = 0.75$, $n = 9$ b: Gauss N(0,1), $\rho = 0.9$, $\beta_2 = 0.75$, $n = 9$ c: Gauss N(0,1), $\rho = 0.1$, $\beta_2 = 0.75$, $n = 18$ d: Gauss N(0,1), $\rho = 0.9$, $\beta_2 = 0.75$, $n = 18$ 

FIG. 2.5 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $---\bullet---$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Gauss (0,1), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2 = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2 = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2 = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2 = 0.75$, $n = 18$.

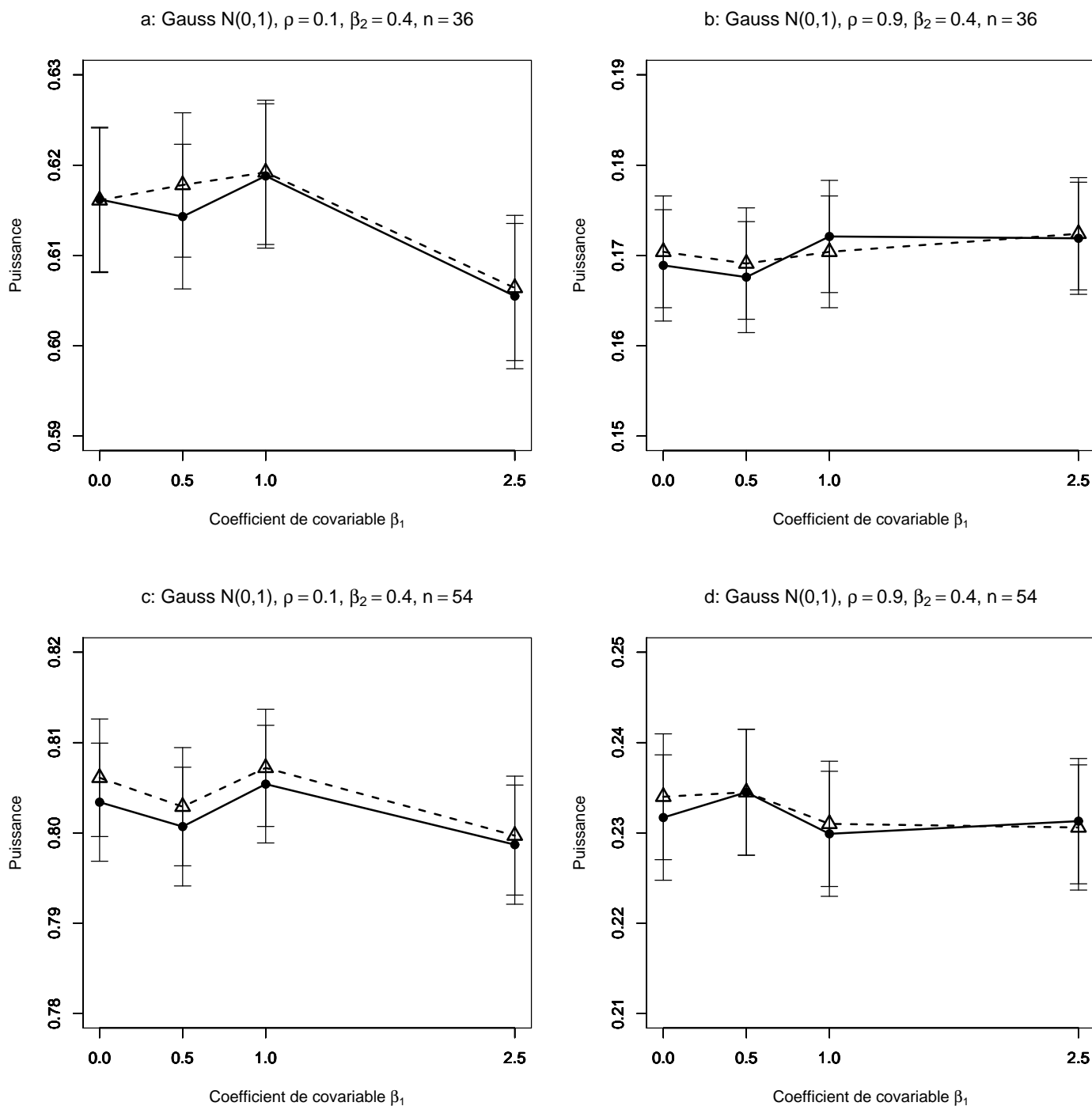


FIG. 2.6 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Gauss $(0,1)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2 = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2 = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2 = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2 = 0.4$, $n = 54$.

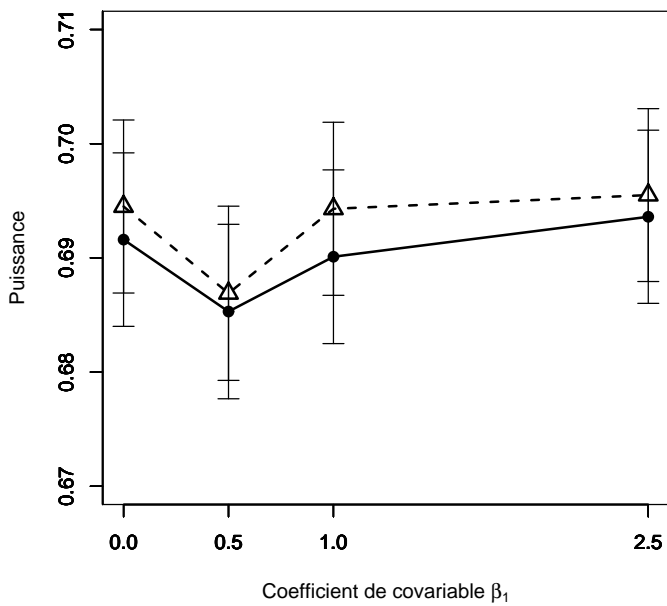
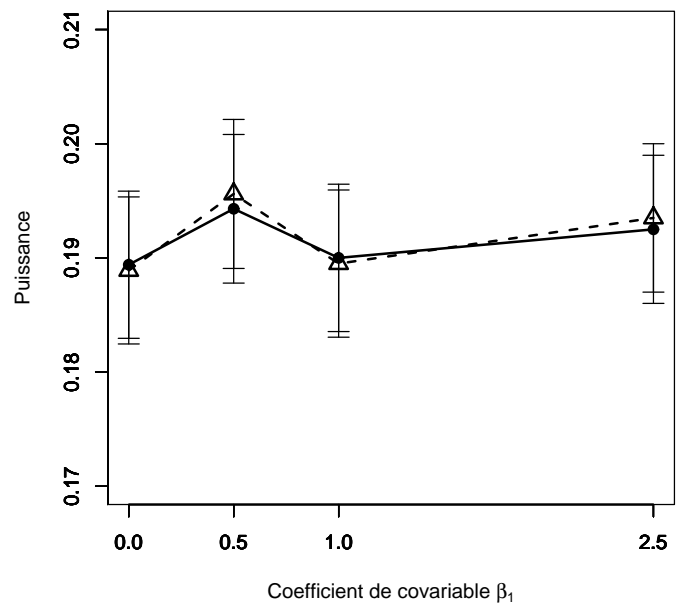
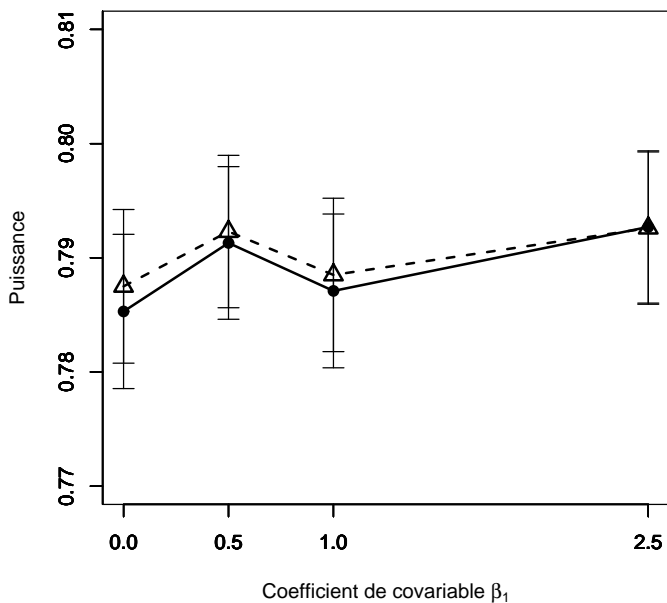
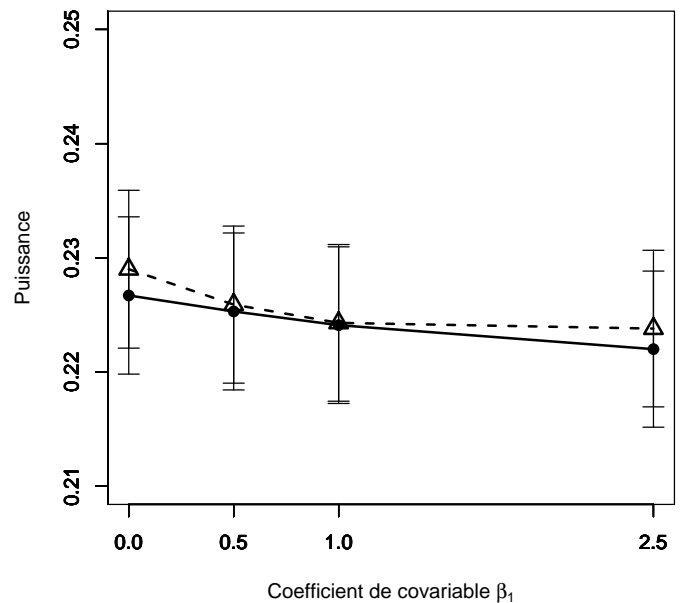
a: Gauss $N(0,1)$, $\rho = 0.1$, $\beta_2 = 0.3$, $n = 72$ b: Gauss $N(0,1)$, $\rho = 0.9$, $\beta_2 = 0.3$, $n = 72$ c: Gauss $N(0,1)$, $\rho = 0.1$, $\beta_2 = 0.3$, $n = 90$ d: Gauss $N(0,1)$, $\rho = 0.9$, $\beta_2 = 0.3$, $n = 90$ 

FIG. 2.7 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Gauss $(0,1)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2 = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2 = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2 = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2 = 0.3$, $n = 90$.

B2- Cas où les données générées pour le terme d'erreur suivent une loi symétrique de variance infinie comme la loi de Student à 1 degré de liberté :

1. En comparant les figures 2.8-a où $n = 9$ et 2.8-c où $n = 18$ on trouve qu'en augmentant la taille de l'échantillon (les valeurs des autres facteurs restant fixées), les puissances de la méthode permutationnelle de Freedman-Lane et de la méthode paramétrique classique augmentent peu. On peut trouver la même relation pour les figures 2.8-b où $n = 9$ et 2.8-d où $n = 18$. Pour les autres combinaisons des facteurs, dans les figures 2.9 et 2.10 on trouve encore la même relation. C'est-à-dire qu'en augmentant la taille d'échantillon (les valeurs des autres facteurs restant fixées), les puissances de la méthode permutationnelle de Freedman-Lane et de la méthode paramétrique classique augmentent peu.
2. En comparant les figures 2.8-a où $\rho = 0.1$ et 2.8-b où $\rho = 0.9$ on obtient qu'en augmentant le degré du coefficient de corrélation $\rho(X^1, X^2)$ (les valeurs des autres facteurs restant fixées), les puissances de la méthode permutationnelle de Freedman-Lane, et de la méthode paramétrique classique diminuent peu. On peut trouver la même relation pour les figures 2.9-c où $\rho = 0.1$ et 2.9-d où $\rho = 0.9$. Pour les autres combinaisons des facteurs, dans les figures 2.9 et 2.10 on trouve encore la même relation. C'est-à-dire qu'en augmentant le degré de corrélation (les valeurs des autres facteurs restant fixées), les puissances de la méthode permutationnelle de Freedman-Lane et de la méthode paramétrique classique diminuent peu.
3. Quelles que soient les valeurs attribuées aux facteurs, les puissances de la méthode de Freedman-Lane et de la méthode paramétrique classique par rapport aux puissances obtenues par les mêmes méthodes dans le cas précédent (B1) diminuent nettement.
4. Même lorsque la variance des données générées pour le terme d'erreur est infinie, quelles que soient les valeurs attribuées aux facteurs, la méthode paramétrique classique donne presque la même puissance que la méthode de Freedman-Lane. Ceci peut être dû au fait que les valeurs des variables X^1 et X^2 sont générées selon une loi de Gauss $(0, 1)$.
5. Contrairement au cas précédent (B1), en considérant les figures 2.8-a, 2.9-a et 2.10-a où pour les trois figures ρ est fixé à 0.1, on constate ici que les puissances de la méthode permutationnelle de Freedman-Lane, et de la méthode paramétrique classique diminuent peu. On peut trouver presque la même relation pour la partie c, mais pour les parties b et d, les puissances restent respectivement dans les intervalles $(0.05-0.06)$ et $(0.04-0.07)$.

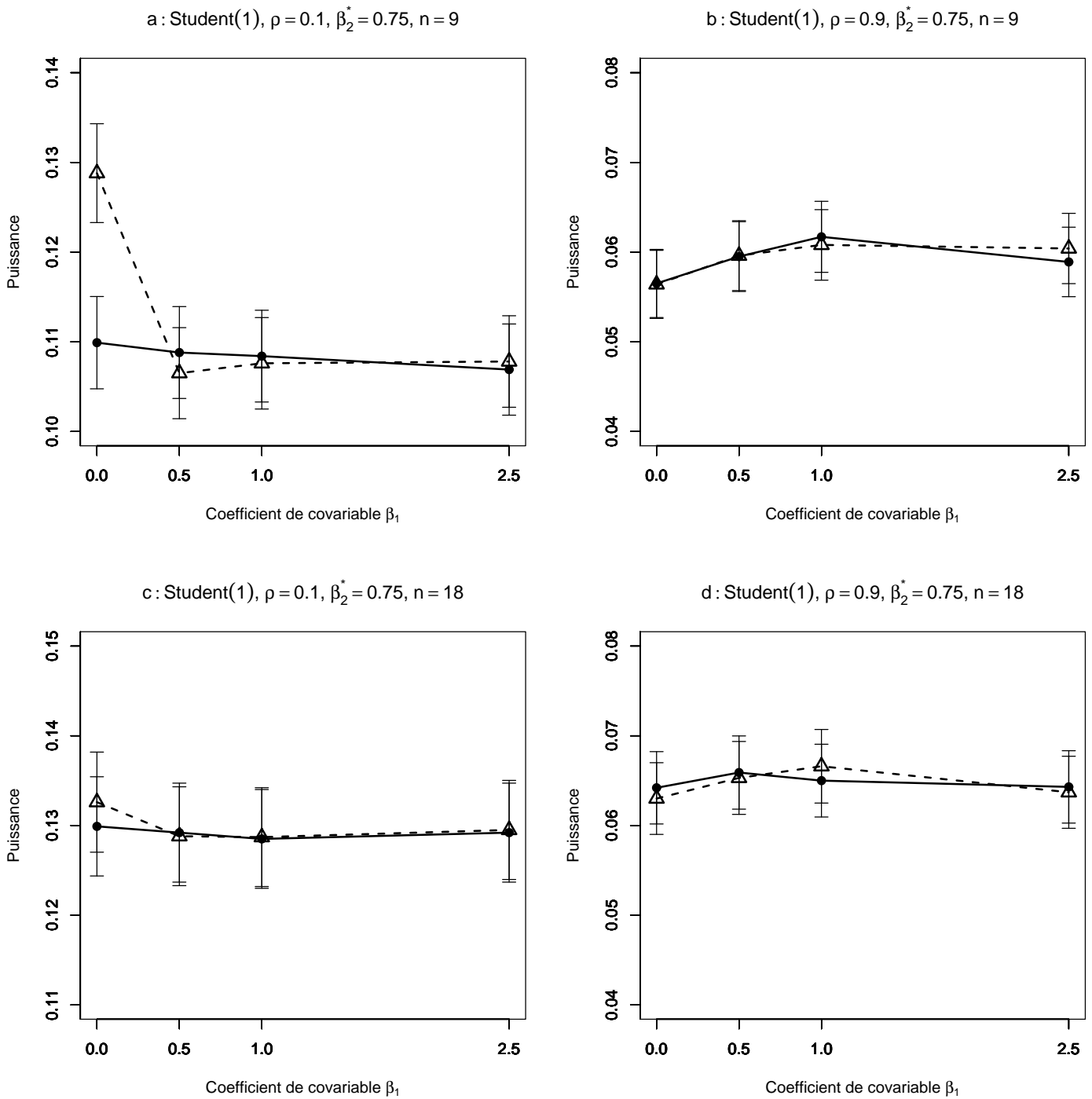


FIG. 2.8 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Student(1), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2 = 0.4$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2 = 0.4$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2 = 0.4$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2 = 0.4$, $n = 18$.

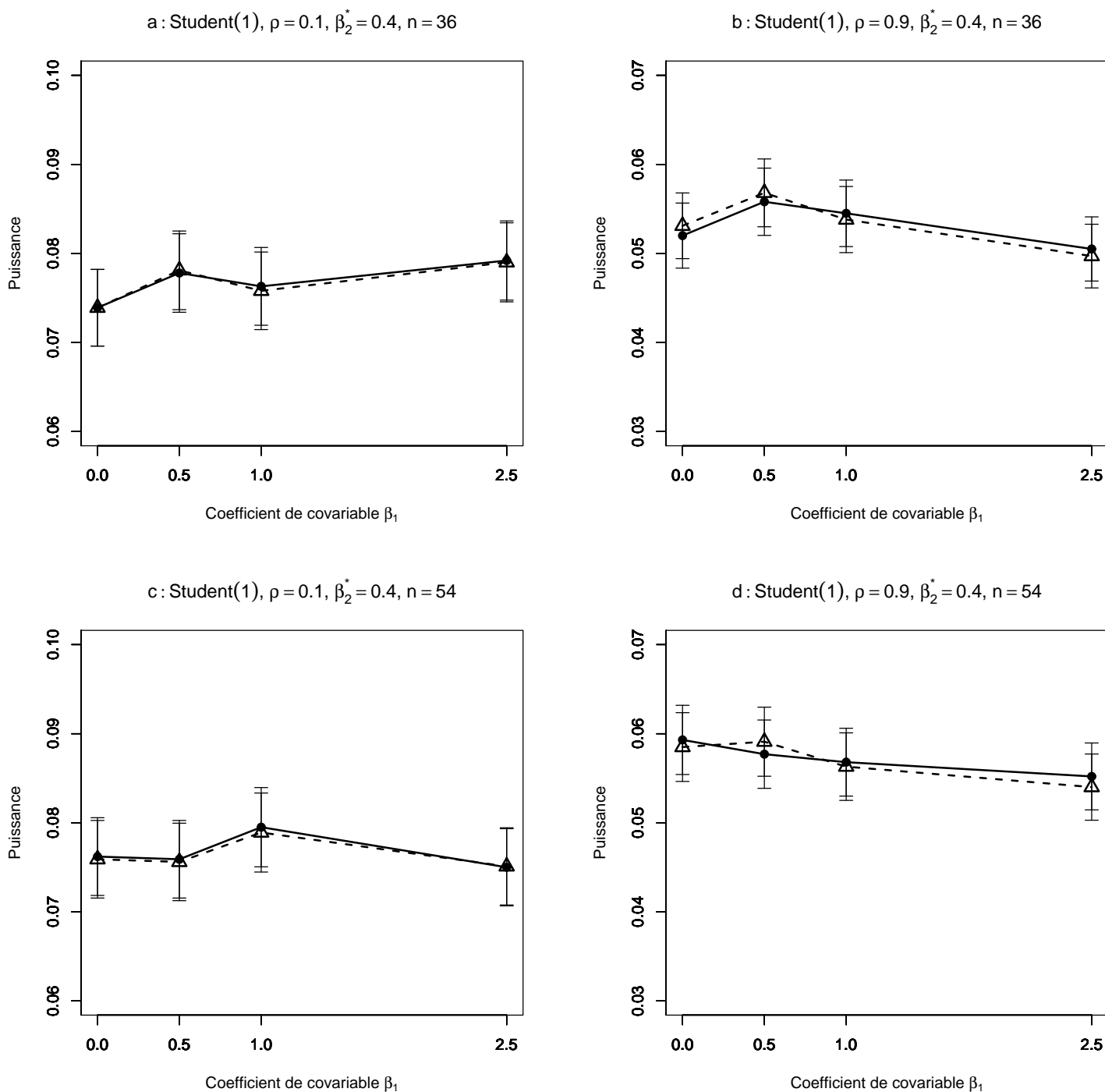


FIG. 2.9 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Student(1), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2 = 0.3$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2 = 0.3$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2 = 0.3$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2 = 0.3$, $n = 54$.

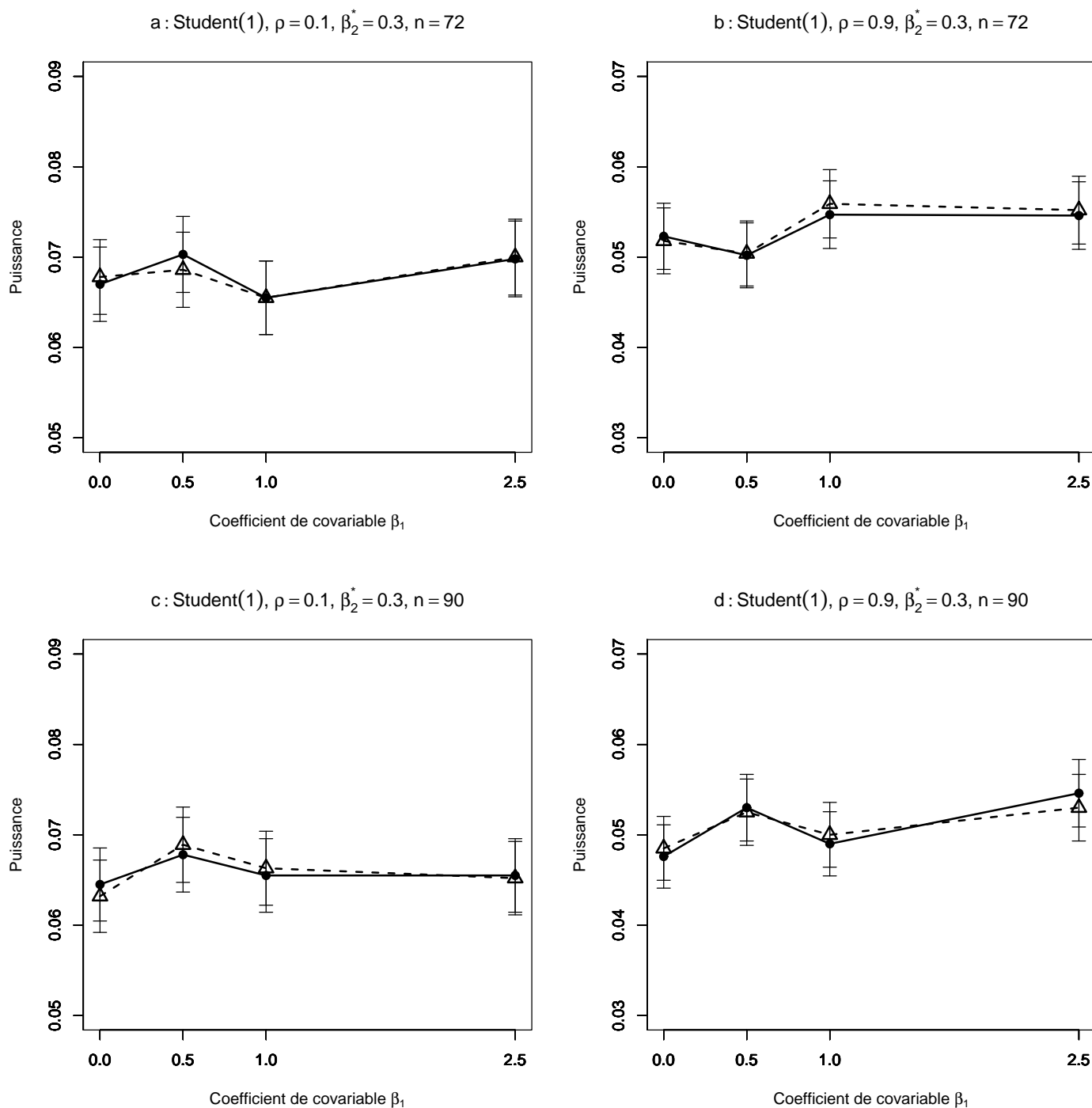


FIG. 2.10 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Student(1), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

B3- Cas des données générées pour le terme d'erreur selon une loi radicalement à dissymétrie droite comme la loi $Exp(2)$:

1. Comme dans le cas B1, en comparant les figures 2.11-a où $n = 9$ et 2.11-c où $n = 18$ on trouve qu'en augmentant la taille d'échantillon (les valeurs des autres facteurs restant fixées), les puissances de la méthode permutacionnelle de Freedman-Lane, et de la méthode paramétrique classique augmentent nettement. On peut trouver la même relation pour les figures 2.11-b où $n = 9$ et 2.11-d où $n = 18$. Pour les autres combinaisons des facteurs, dans les figures 2.12 et 2.13 on trouve encore la même relation. C'est-à-dire qu'en augmentant la taille d'échantillon (les valeurs des autres facteurs restant fixées), les puissances de la méthode permutacionnelle de Freedman-Lane et de la méthode paramétrique classique augmentent nettement.
2. Comme dans le cas B1, en comparant les figures 2.11-a où $\rho = 0.1$ et 2.11-b où $\rho = 0.9$ on constate qu'en augmentant le degré du coefficient de corrélation $\rho(X^1, X^2)$ (les valeurs des autres facteurs restant fixées), les puissances de la méthode permutacionnelle de Freedman-Lane, et de la méthode paramétrique classique diminuent nettement. On peut trouver la même relation pour les figures 2.11-c où $\rho = 0.1$ et 2.11-d où $\rho = 0.9$. Pour les autres combinaisons des facteurs, dans les figures 2.12 et 2.13 on trouve la même relation. C'est-à-dire qu'en augmentant le degré de corrélation (les valeurs des autres facteurs restant fixées), les puissances des méthodes permutacionnelle de Freedman-Lane et paramétrique classique diminuent nettement.
3. Même lorsque les données générées pour le terme d'erreur suivent une loi radicalement à dissymétrie droite comme la loi $Exp(2)$, quelles que soient les valeurs attribuées aux facteurs, la méthode paramétrique classique marche bien. C'est-à-dire que elle donne presque la même puissance que la méthode Freedman-Lane. Ceci peut être dû au fait que les valeurs des variables X^1 et X^2 sont générées selon une loi de Gauss $(0, 1)$.
4. En comparant les cas B1 et B2 avec ce cas, on constate que, quelles que soient les valeurs attribuées aux facteurs, les puissances de la méthode de Freedman-Lane et de la méthode paramétrique classique sont nettement supérieures à celles obtenues par les mêmes méthodes dans les cas précédents (B1 et B2). Ceci peut être dû à la variance finie (assez petite) du terme d'erreur.
5. En considérant les figures 2.11-a, 2.12-a et 2.13-a où, pour les trois figures ρ est fixé à 0.1, on constate que les puissances de la méthode de Freedman-Lane et de la méthode paramétrique classique augmentent. On peut trouver la même relation pour les figures de la partie b, mais pour les figures des parties c et d, les puissances restent respectivement dans les intervalles $(0.99-1)$ et $(0.68-0.73)$.

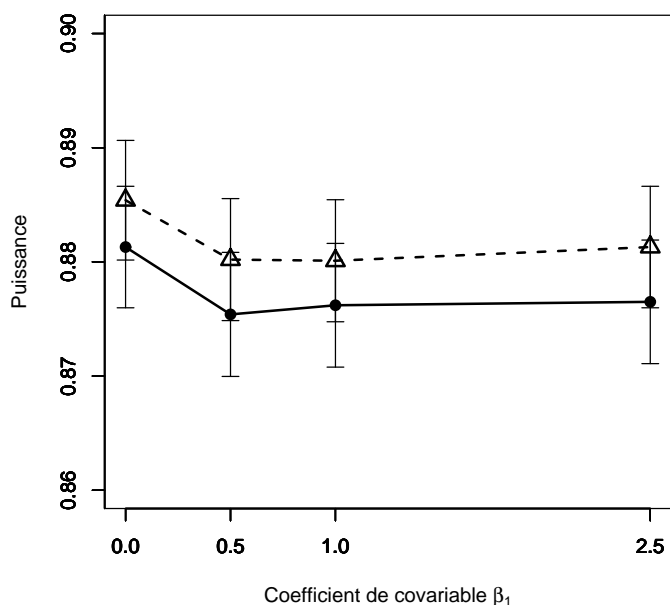
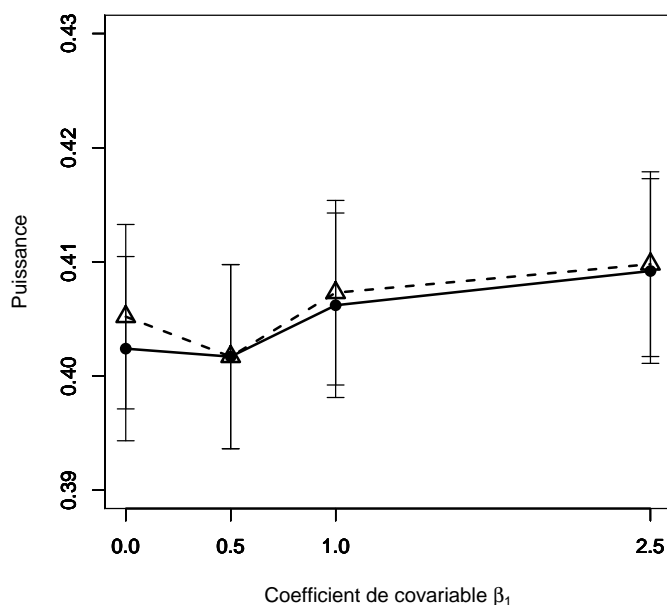
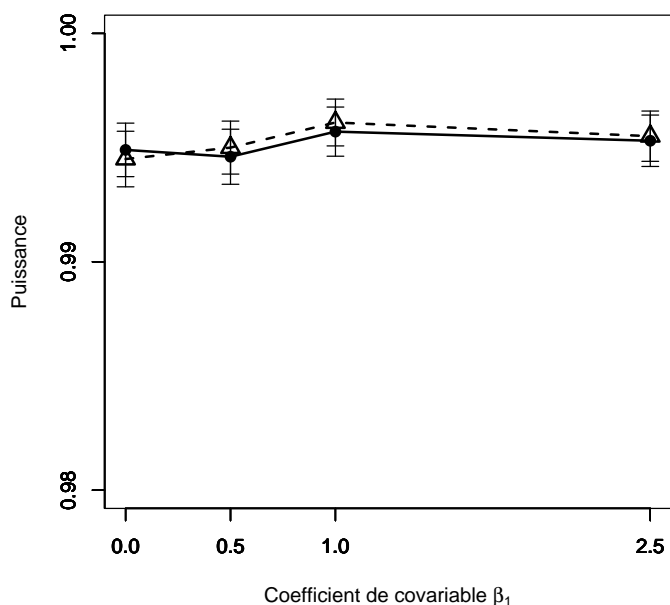
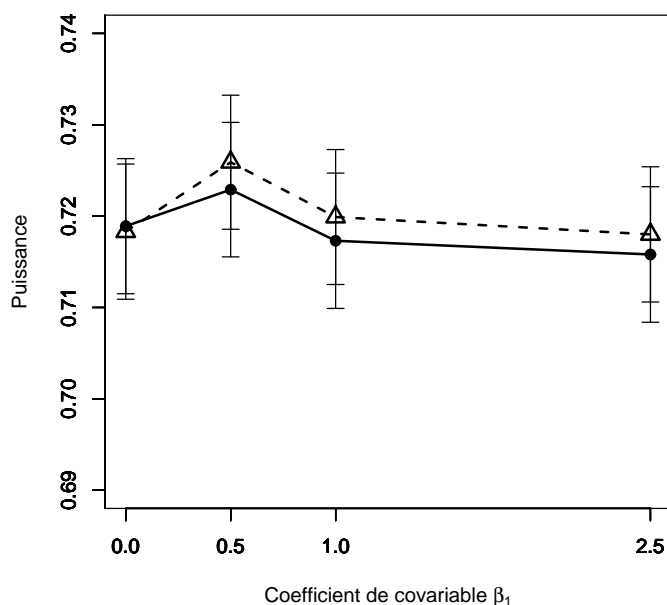
a : Exp(2), $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$ b : Exp(2), $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$ c : Exp(2), $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$ d : Exp(2), $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$ 

FIG. 2.11 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $---\bullet---$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi Exponentielle(2), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

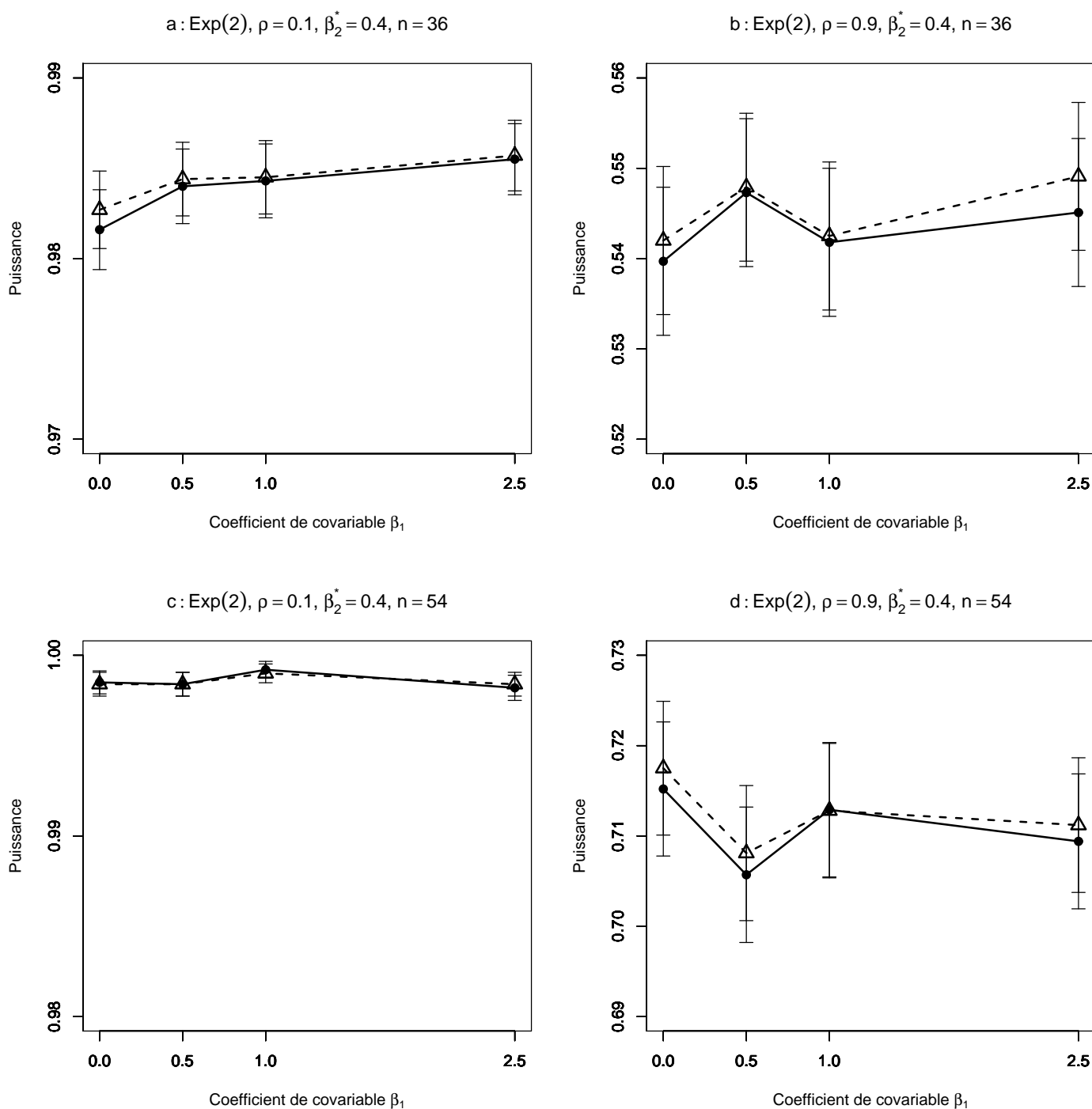


FIG. 2.12 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne—△— et Freedman et Lane—●—, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi Exponentielle(2), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

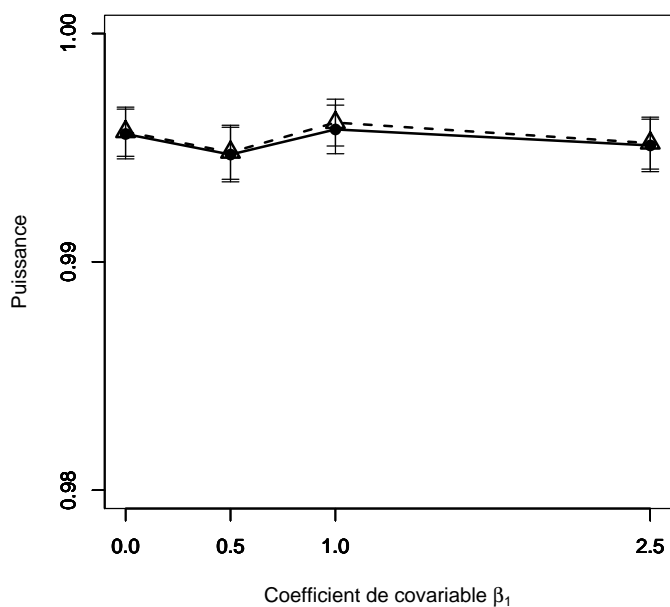
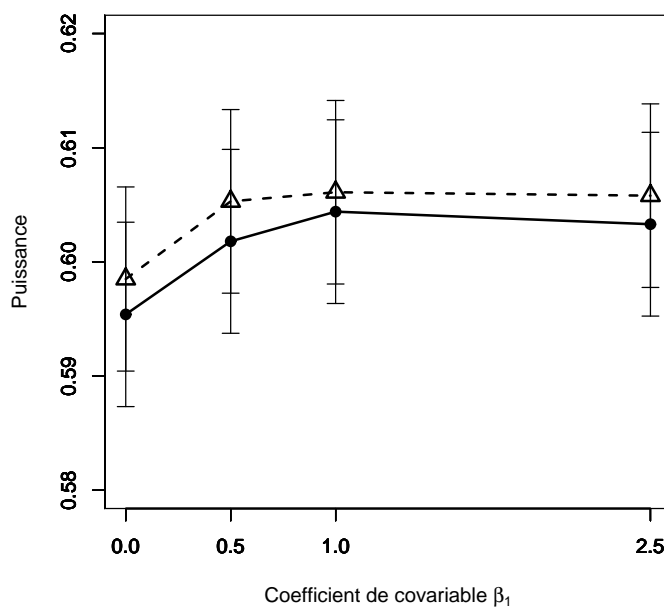
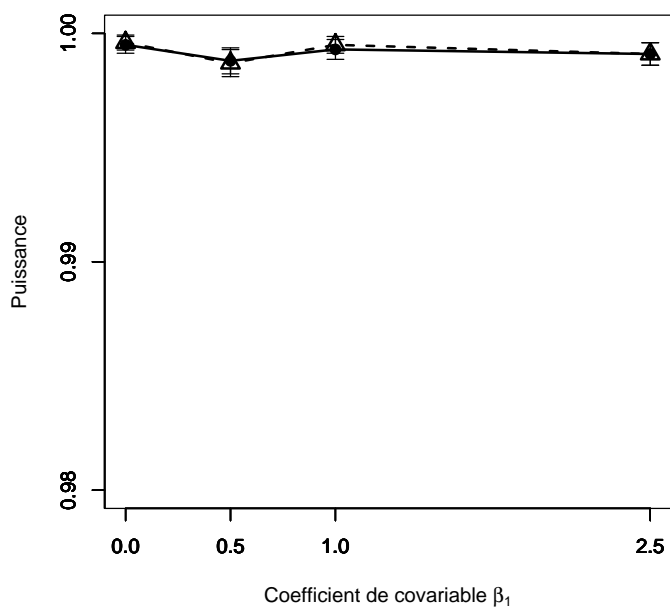
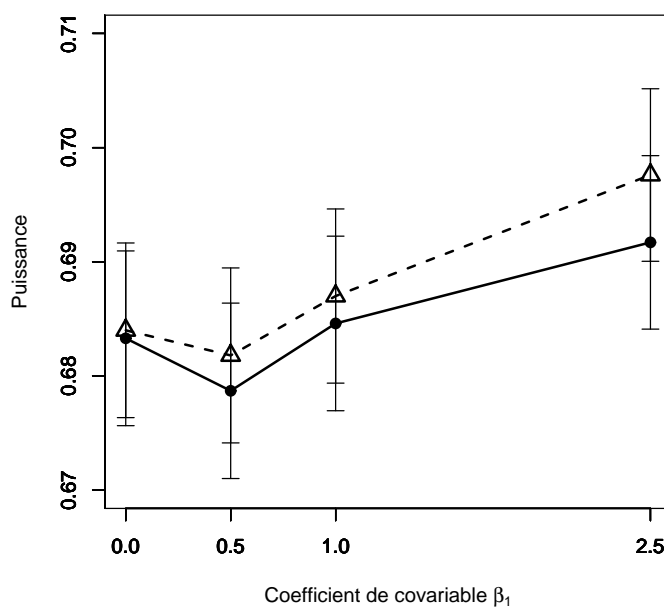
a : Exp(2), $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$ b : Exp(2), $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$ c : Exp(2), $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$ d : Exp(2), $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$ 

FIG. 2.13 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi Exponentielle(2), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

B-4 Cas des données générées pour le terme d'erreur selon une loi radicalement à dissymétrie gauche comme la loi $\Gamma(10, 0.5)$:

1. On ne peut pas trouver les deux premiers résultats obtenus dans le cas B2.
2. Contrairement aux cas B3 (où les données générées pour le terme d'erreur suivent une loi radicalement à dissymétrie droite comme la loi $Exp(2)$), lorsque (les données générées pour le terme d'erreur suivent une loi radicalement à dissymétrie gauche comme la loi $\Gamma(10, 0.5)$), les puissances de la méthode permutationnelle de Freedman-Lane et de la méthode paramétrique classique diminuent nettement. Remarquons que ceci peut être dû au fait que la variance de la loi $Exp(2)$ est 0.25 et la variance de la loi $\Gamma(10, 0.5)$ est 40.
3. Comme le cas B2, ici, les puissances de la méthode de Freedman-Lane et de la méthode paramétrique classique diminuent nettement par rapport aux cas B1 et B3.
4. Contrairement aux cas B1, B2 et B3, ici, en considérant les figures 2.8-a, 2.9-a et 2.10-a où pour les trois figures ρ est fixé à 0.1, on constate que les puissances de la méthode permutationnelle de Freedman-Lane et de la méthode paramétrique classique restent dans l'intervalle (0.05-0.07).

Des résultats plus complets, utilisant d'autres combinaisons des facteurs, sont présentés dans l'annexe 2.

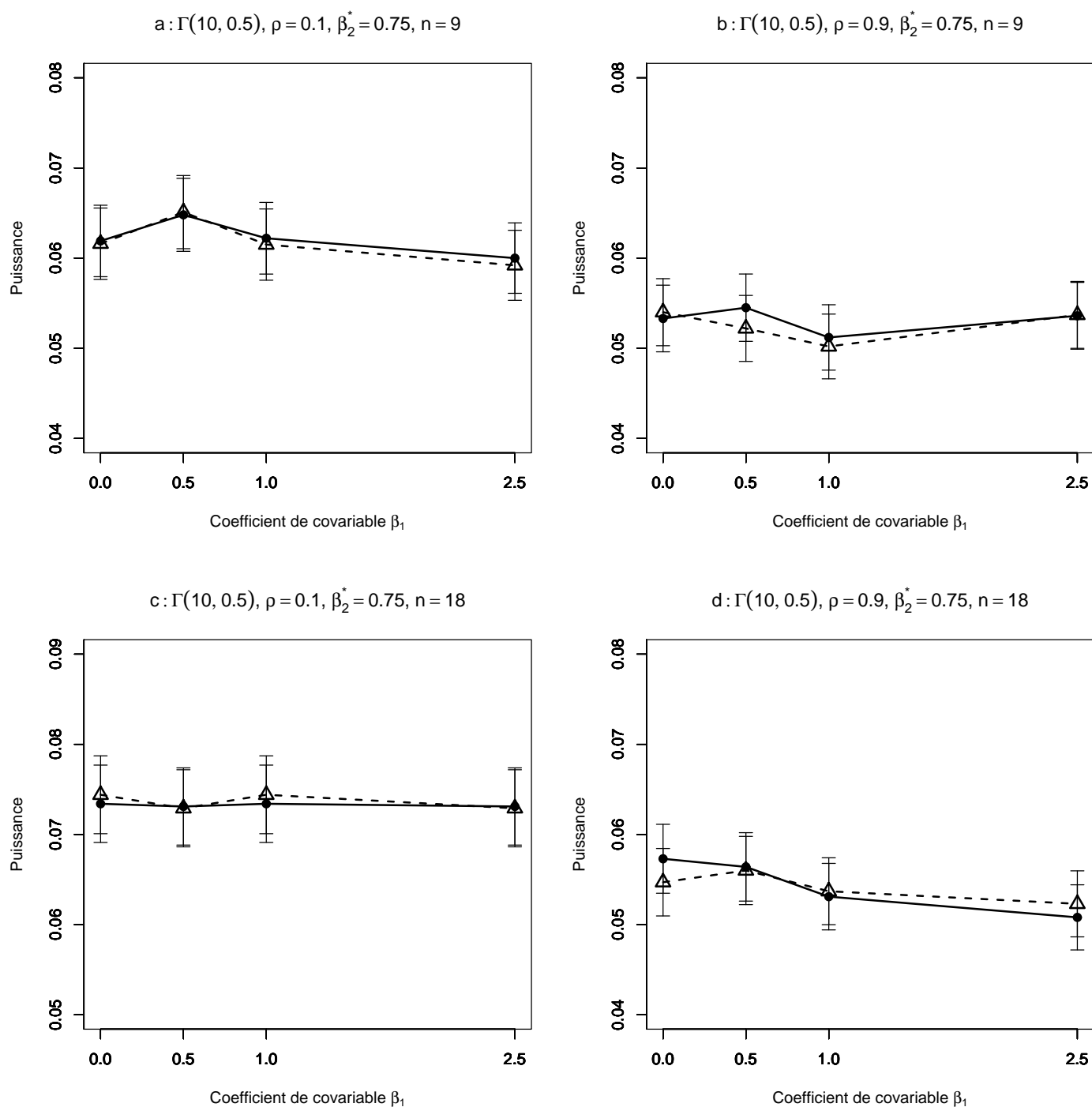


FIG. 2.14 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi $\Gamma(10, 0.5)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

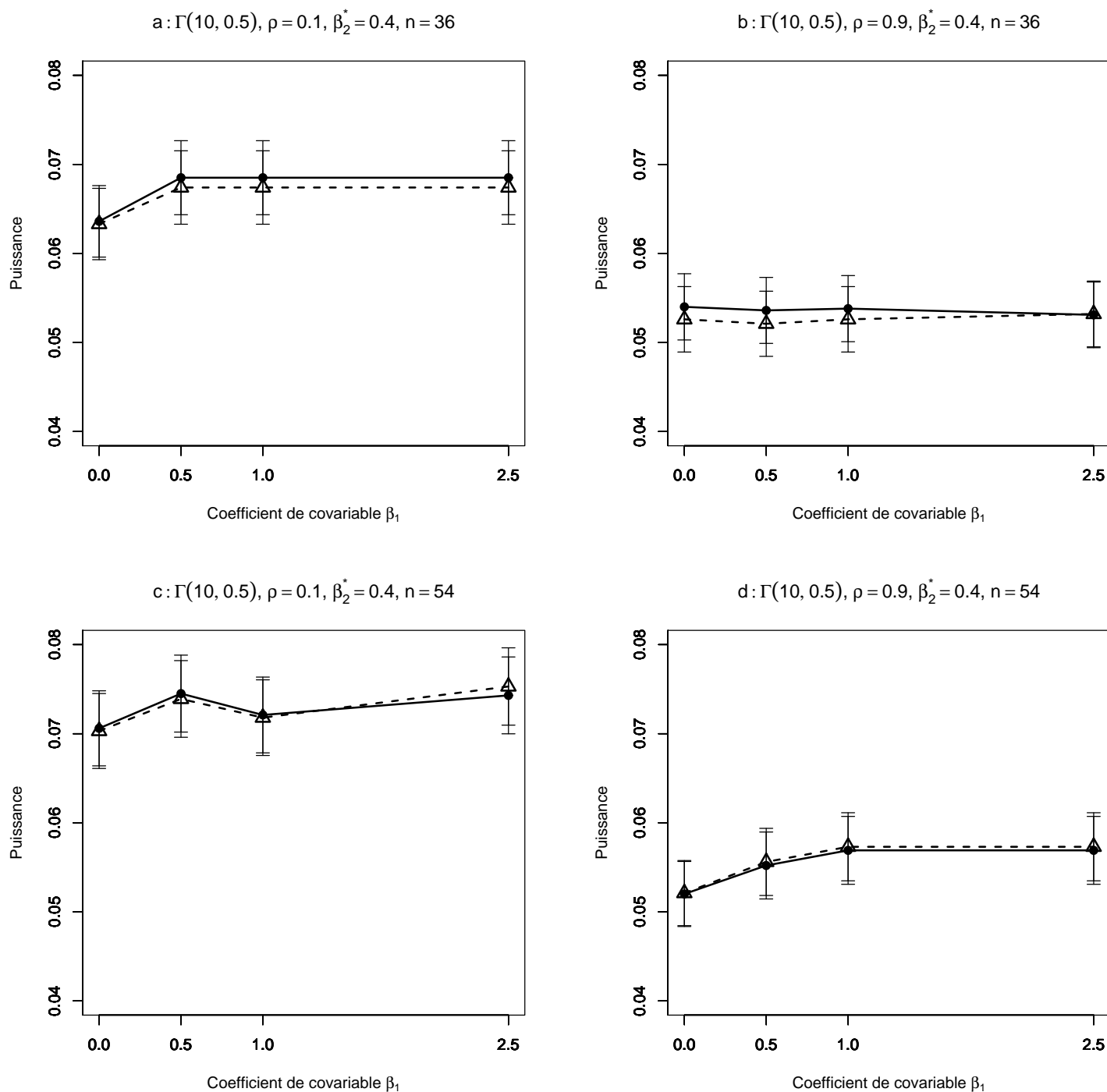


FIG. 2.15 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi $\Gamma(10, 0.5)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$; figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

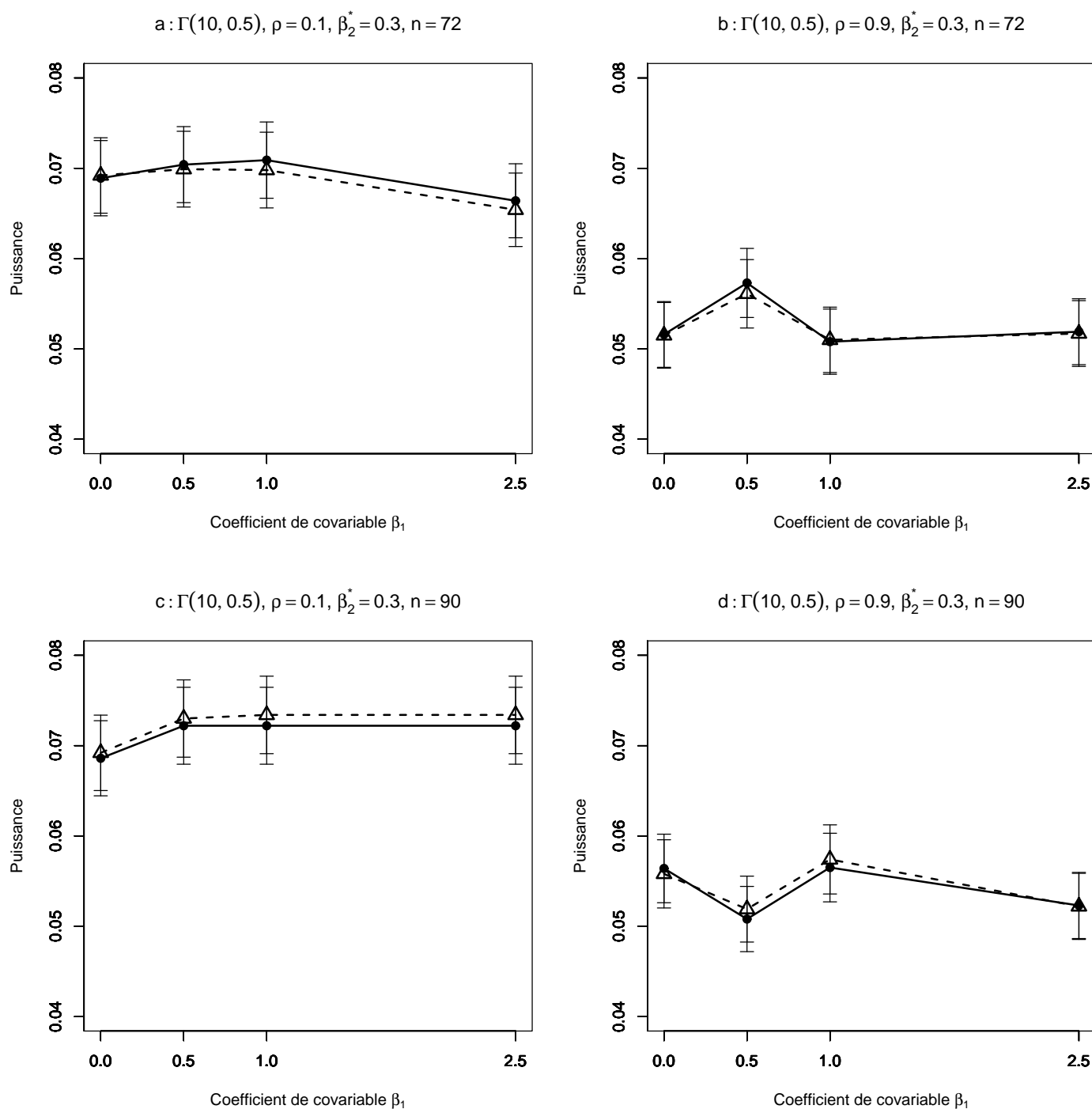


FIG. 2.16 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi $\Gamma(10, 0.5)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

2.5.3 Étude la puissance de la méthode de Kennedy

D'après les résultats théoriques que nous avons montrés au chapitre 3, nous prévoyons que contrairement au risque d'erreur de première espèce, la puissance de la méthode de Freedman-Lane est dans la plupart des cas inférieure à celle de la méthode de Kennedy. Mais il existe des cas où la puissance de la méthode de Kennedy est inférieure à celle de la méthode de Freedman-Lane quand la taille d'échantillon est grande ($n > 9$). Afin de vérifier cette prévision, dans le cadre du travail que nous avons défini pour la simulation sur la puissance de la méthode Freedman-Lane (section(2.4.2)), nous avons mené une simulation sur les puissances des méthodes permutacionnelles de Freedman-Lane, et de Kennedy ainsi que sur la méthode paramétrique classique.

A- Le cadre de la simulation

Nous avons considéré presque le même cadre de simulation que dans la section (2.5.2), sauf que les valeurs du coefficient de covariable β_1 sont $\{0, 2.5, 5\}$. Par ailleurs, pour économiser le temps nous avons considéré les quatre distributions : symétrique avec variance finie comme la loi de Gauss $(0, 1)$, symétrique avec variance infinie comme la loi de *Student*(1), radicalement à dissymétrie droite comme la loi *Exp*(2) et radicalement à dissymétrie gauche comme la loi $\Gamma(10, 0.5)$. Donc, l'ensemble des modalités des facteurs sont :

- a) La valeur du paramètre $\beta_2 = 0.75$ pour $n = \{9, 18\}$, $\beta_2 = 0.4$ pour $n = \{36, 54\}$ et $\beta_2 = 0.3$ pour $n = \{72, 90\}$.
- b) La valeur du coefficient de la covariable X^1 est : $\beta_1 = \{0, 2.5, 5\}$.
- c) Le degré de corrélation entre les variables explicatives X^1 et X^2 , $\rho(X^1, X^2) = \{0.1, 0.9\}$
- d) La distribution du terme d'erreur aléatoire ϵ : $LG(0, 1), Exp(2), Student(1)\}$

Pour chaque combinaison de facteurs, 10000 ensembles de données ont été générés. Chaque ensemble de données est permuté 999 fois. La statistique de test permutacionnelle ($t_{2,\pi}^F$) de la méthode de Freedman-Lane a été calculée sur chaque ensemble de données permutées. Ensuite, après 999 permutations, la p-valeur permutacionnelle pour la méthode de Freedman-Lane a été calculée par la formule :

$$p - valeur_{\pi} = \frac{\#(|t_{\pi}| \geq t_{obs})}{1000}$$

Le niveau de signification pour les tests a été fixé à $\alpha = 0, 05$. Alors, pour chaque ensemble de données quand la $p - valeur_{\pi}$ est inférieure à 0.05 on rejette l'hypothèse

$H_0 : \beta_2 = 0$. Sur les 10000 p -valeur $_{\pi}$ le pourcentage des P -valeur $_{\pi}$ de la méthode de Freedman-Lane inférieures à 0.05 donne le risque d'erreur de première espèce. En fait, ce pourcentage estime la probabilité de rejeter $H_0 : \beta_2 = 0$ lorsque H_0 est vraie (parce que les données sont générées sous $\beta_2 = 0$). De plus, l'intervalle de confiance à 95% est calculé à partir du risque d'erreur de première espèce, calculé sur les 10.000 ensembles.

Afin de compléter la comparaison, pour chaque ensemble de données, on a utilisé les tests paramétriques classiques qui consistent à calculer une statistique de Student. Sa p -valeur donnée par la théorie Gaussienne a également été calculée. De la même façon, sur les 10000 ensembles de données le risque d'erreur de première espèce et son intervalle de confiance ont été calculés par la théorie Gaussienne.

Les autres étapes sont identiques à celles de la section (2.5.2).

B- Les résultats de nos simulations

B1- Cas des données générées pour le terme d'erreur selon une loi de Gauss :

1. En comparant les figures 2.17-a où $n = 9$ et 2.17-c où $n = 18$, on trouve, comme avec la méthode de Freedman-Lane et la méthode paramétrique classique qu'en augmentant la taille d'échantillon (les valeurs des autres facteurs restant fixées), la puissance de la méthode de Kennedy augmente. On peut trouver la même relation pour les figures 2.17-b où $n = 9$ et 2.17-d où $n = 18$. Pour les autres combinaisons de facteurs, dans les figures 2.18 et 2.19, on trouve encore la même relation. C'est-à-dire qu'en faisant croître la taille de l'échantillon (les valeurs des autres facteurs restant fixées), la puissance de la méthode de Kennedy comme les méthodes de Freedman-Lane et la méthode paramétrique classique, augmente nettement.
2. En comparant les figures 2.17-a où $\rho = 0.1$ et 2.17-b où $\rho = 0.9$, on constate comme dans les méthodes de Freedman-Lane et la méthode paramétrique classique qu'en augmentant le degré de corrélation $\rho(X^1, X^2)$ (les valeurs des autres facteurs restant fixées), la puissance de la méthode de Kennedy diminue nettement. On peut trouver la même relation pour les figures 2.17-c où $\rho = 0.1$ et 2.17-d où $\rho = 0.9$. Pour les autres combinaisons de facteurs, dans les figures 2.18 et 2.19 on trouve la même relation. C'est-à-dire qu'en augmentant le degré de corrélation (les valeurs des autres facteurs restant fixées), la puissance de la méthode de Kennedy, comme celle de la méthode de Freedman-Lane et celle de la méthode paramétrique classique diminue nettement.
3. Comme nous l'avons prévu, la puissance de la méthode de Kennedy est supérieure à celle de la méthode de Freedman-Lane (figure 2.17) dans la plupart des cas (surtout lorsque la taille d'échantillon est petite ($n = 9, 18$)).

4. En augmentant la taille d'échantillon, les deux méthodes permutationnelles ainsi que la méthode paramétrique classique donnent la même puissance.
5. Comme nous l'avons prévu, les figures 2.18-c où $\beta_1 = 2.5$, 2.18-d où $\beta_1 = 2.5, 5$, 2.19-b où $\beta_1 = 5$ et 2.19-d où $\beta_1 = 2.5$ montrent des cas où la puissance de la méthode de Kennedy est inférieure à celle de la méthode de Freedman-Lane.

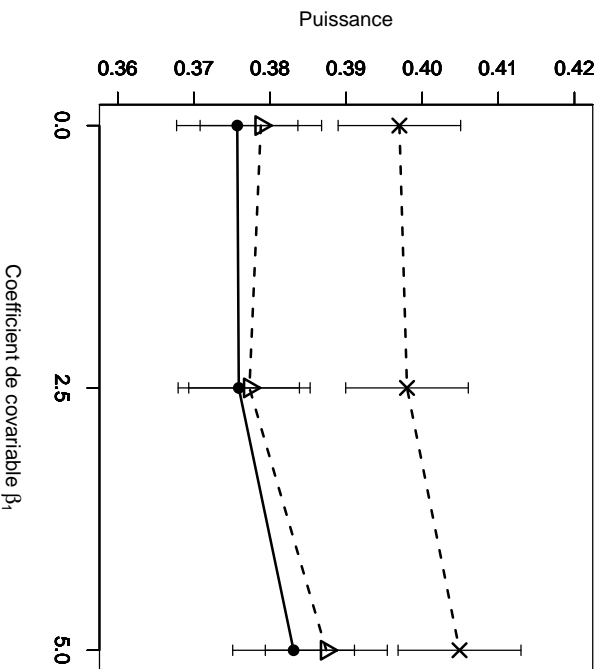
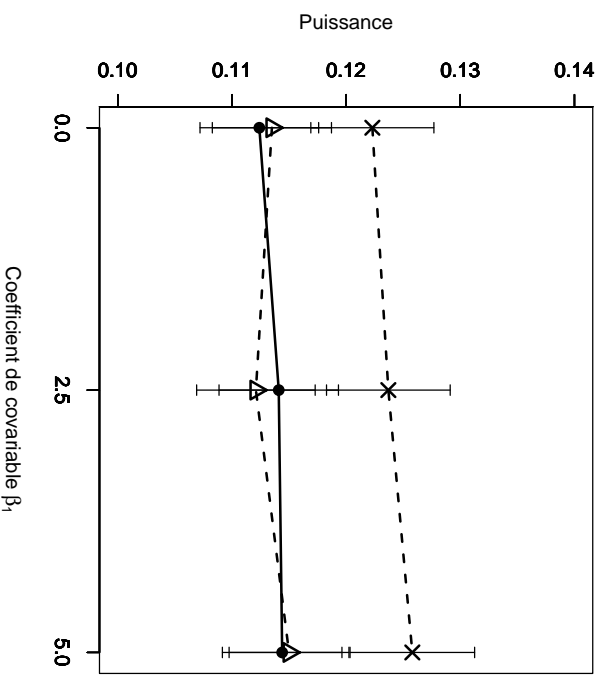
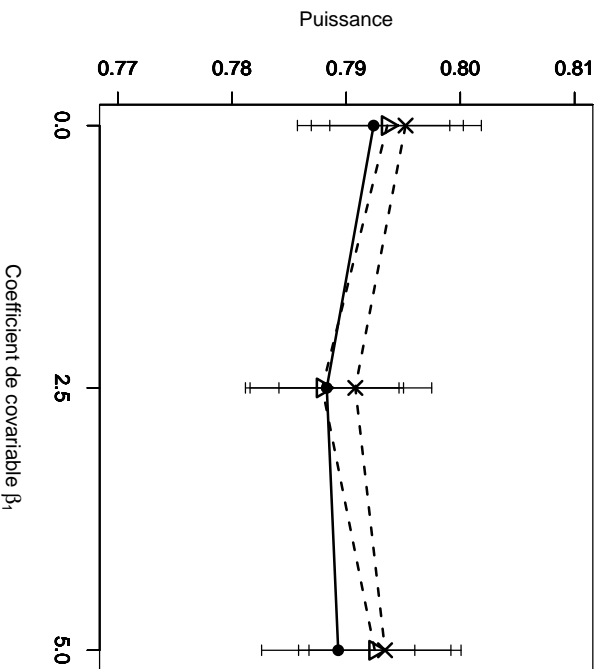
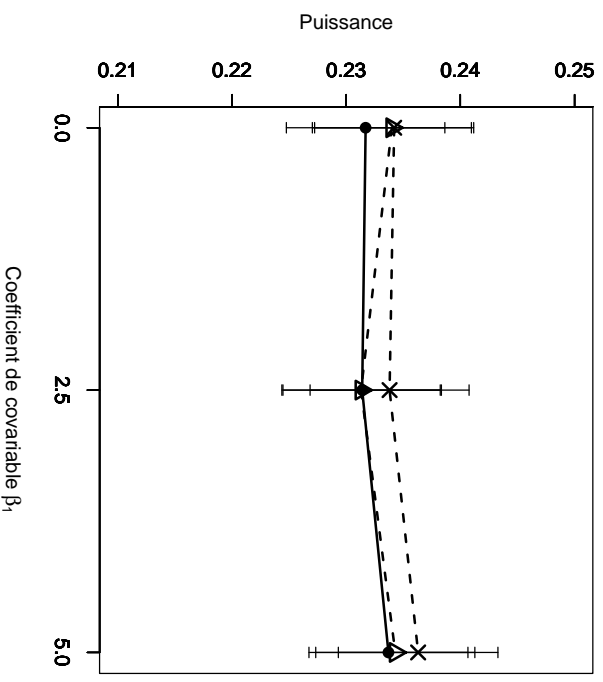
a: Gauss $N(0,1)$, $\rho = 0.1$, $\beta_2 = 0.75$, $n = 9$ b: Gauss $N(0,1)$, $\rho = 0.9$, $\beta_2 = 0.75$, $n = 9$ c: Gauss $N(0,1)$, $\rho = 0.1$, $\beta_2 = 0.75$, $n = 18$ d: Gauss $N(0,1)$, $\rho = 0.9$, $\beta_2 = 0.75$, $n = 18$ 

FIG. 2.17 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne—△— et Freedman et Lane—●— et Kennedy—×— en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi Gaussienne $(0,1)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$. On a retenu $\beta_1=0$, 2.5, 5. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

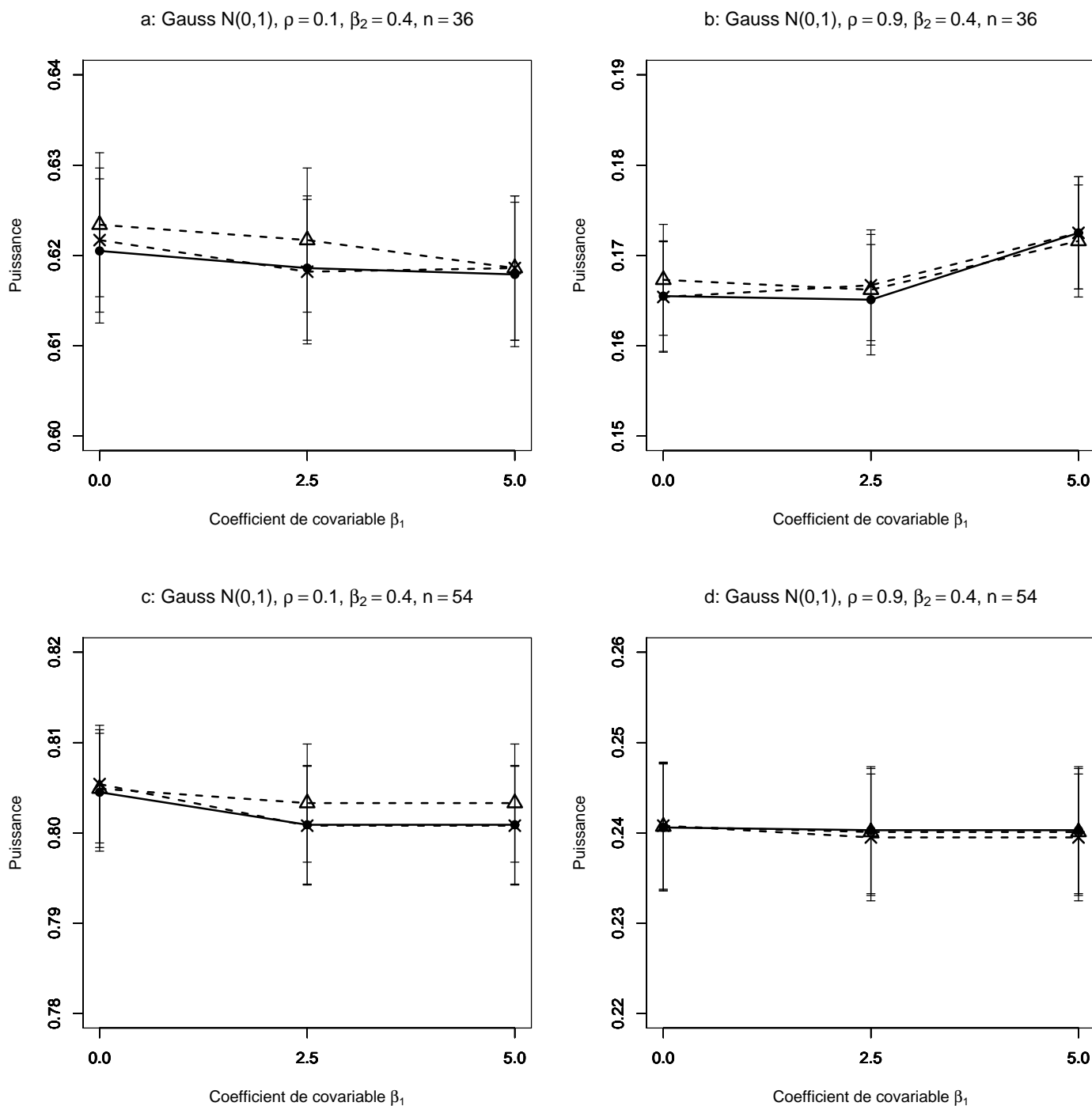


FIG. 2.18 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$ et Kennedy $-\cdot\times-$ en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi Gaussiennes $(0,1)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$. On a retenu $\beta_1=0, 2.5, 5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

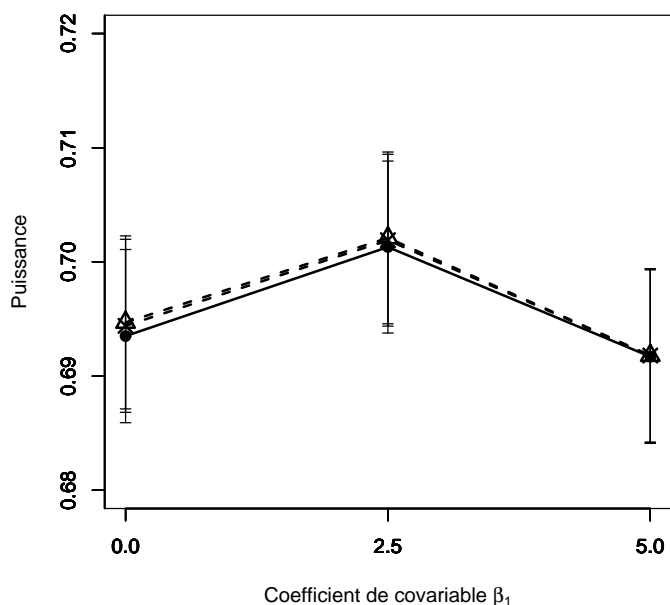
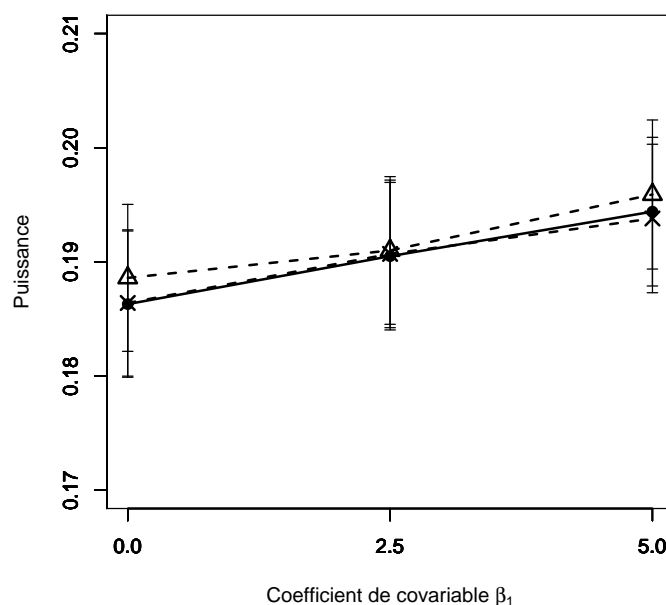
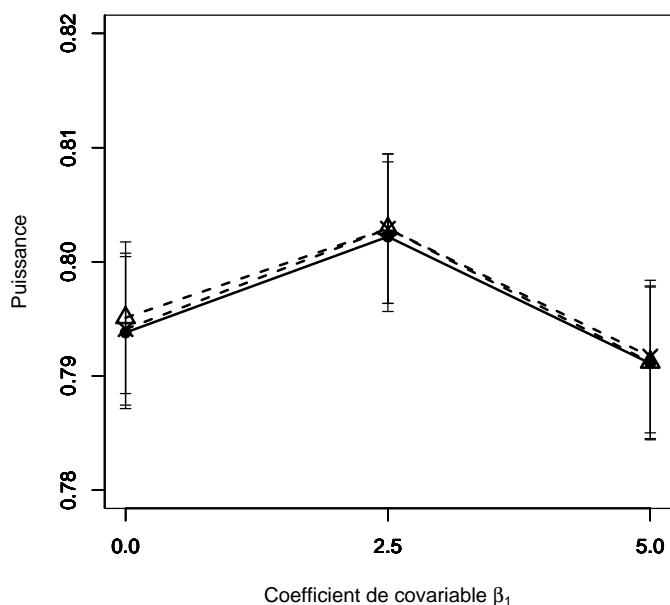
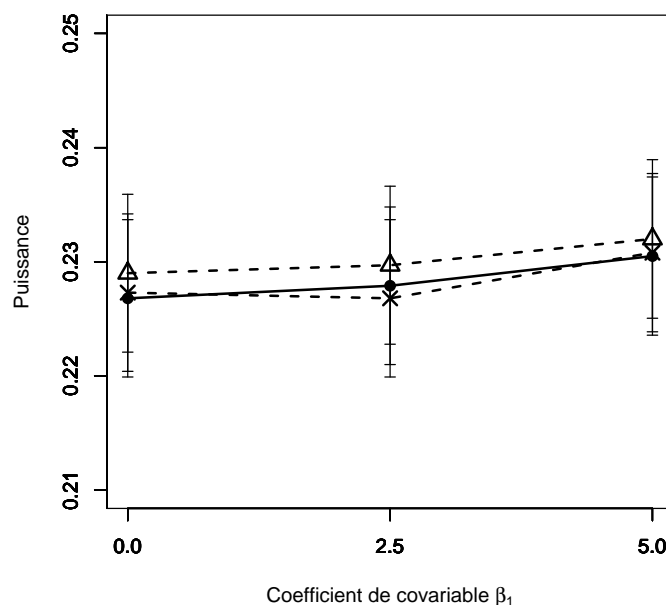
a: Gauss $N(0,1)$, $\rho = 0.1$, $\beta_2 = 0.3$, $n = 72$ b: Gauss $N(0,1)$, $\rho = 0.9$, $\beta_2 = 0.3$, $n = 72$ c: Gauss $N(0,1)$, $\rho = 0.1$, $\beta_2 = 0.3$, $n = 90$ d: Gauss $N(0,1)$, $\rho = 0.9$, $\beta_2 = 0.3$, $n = 90$ 

FIG. 2.19 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne --- Δ --- et Freedman-Lane — \bullet — et Kennedy - \times - en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi Gaussiennes $(0,1)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$. On a retenu $\beta_1=0, 2.5, 5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

B2- Cas des données générées pour le terme d'erreur selon une loi symétrique de variance infinie comme la loi de Student à 1 degré de liberté :

1. En comparant les figures 2.20-a où $n = 9$ et 2.20-c où $n = 18$, on trouve comme pour les méthodes de Freedman-Lane et pour la méthode paramétrique classique qu'en augmentant la taille d'échantillon (les valeurs des autres facteurs restant fixées), la puissance de la méthode de Kennedy augmente peu. On peut trouver la même relation pour les figures 2.17-b où $n = 9$ et 2.17-d où $n = 18$. Pour les autres combinaisons de facteurs, dans les figures 2.18 et 2.19, on trouve la même relation. C'est-à-dire qu'en augmentant la taille d'échantillon (les valeurs des autres facteurs restant fixées), la puissance de la méthode de Kennedy, comme celle des méthodes de Freedman-Lane et celle de la méthode paramétrique classique, augmente peu.
2. En comparant les figures 2.17-a où $\rho = 0.1$ et 2.17-b où $\rho = 0.9$, on constate comme dans les méthodes de Freedman-Lane et la méthode paramétrique classique qu'en augmentant le degré de corrélation $\rho(X^1, X^2)$ (les valeurs des autres facteurs restant fixées), la puissance de la méthode de Kennedy diminue peu. On peut trouver la même relation pour les figures 2.17-c où $\rho = 0.1$ et 2.17-d où $\rho = 0.9$. Pour les autres combinaisons de facteurs, dans les figures 2.18 et 2.19 on trouve la même relation. C'est-à-dire qu'en augmentant le degré de corrélation (les valeurs des autres facteurs restant fixées), la puissance de la méthode de Kennedy, comme celle des méthodes de Freedman-Lane et celle de la méthode paramétrique classique, diminue peu.
3. Comme la méthode de Freedman-Lane (on l'a déjà vu dans la section (2.5.2)), quelles que soient les valeurs attribuées aux facteurs, la puissance de la méthode de Kennedy par rapport à la puissance obtenue par la même méthode mais pour des données générées selon une loi de Gauss diminue nettement.
4. Comme nous l'avons prévu dans la plupart des cas (surtout lorsque la taille d'échantillon est petit ($n = 9$)), la puissance de la méthode de Kennedy est supérieure à celle de la méthode de Freedman-Lane (figure 2.20).
5. En augmentant la taille d'échantillon, les deux méthodes permutationnelles ainsi que la méthode paramétrique classique donnent la même puissance.
6. Les figures 2.20-d où $\beta_1 = 5$, 2.21-b où $\beta_1 = 0$, 2.21-c où $\beta_1 = 0$, 2.21-d, où $\beta_1 = \{0, 1, 2.5\}$, 2.22-b, où $\beta_1 = \{2.5, 5\}$, 2.22-c où $\beta_1 = 5$ et 2.22-d où $\beta_1 = \{0, 2.5\}$ montrent des cas où la puissance de la méthode de Kennedy est inférieure à celle de la méthode de Freedman-Lane.

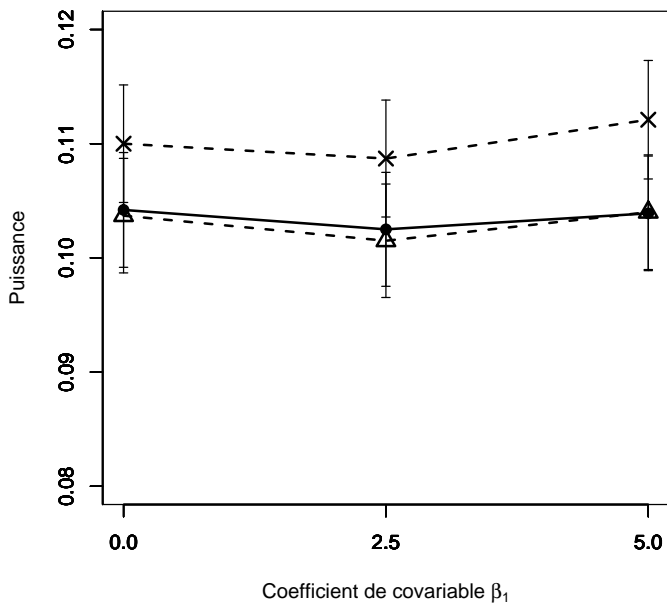
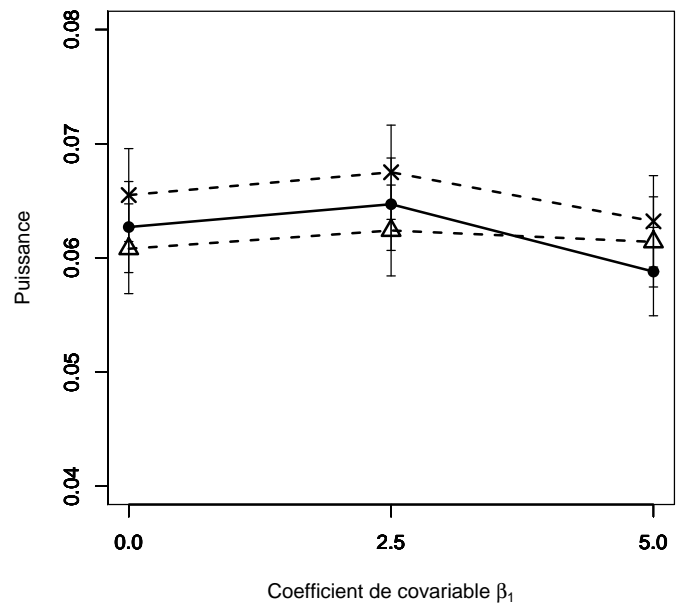
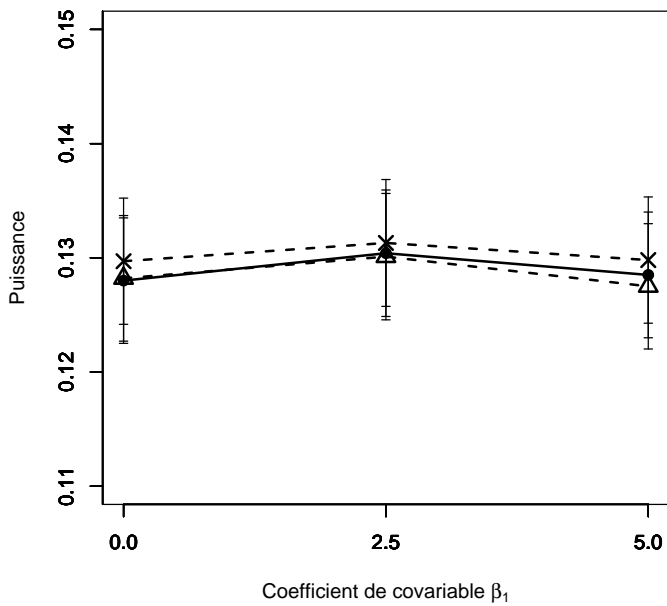
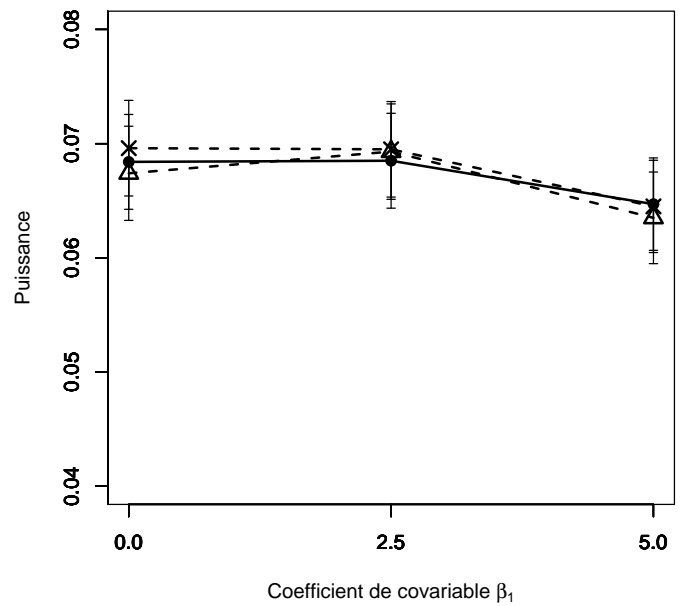
a : Student(1), $\rho = 0.1$, $\beta_2 = 0.75$, $n = 9$ b : Student(1), $\rho = 0.9$, $\beta_2 = 0.75$, $n = 9$ c : Student(1), $\rho = 0.1$, $\beta_2 = 0.75$, $n = 18$ d : Student(1), $\rho = 0.9$, $\beta_2 = 0.75$, $n = 18$ 

FIG. 2.20 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$ et Kennedy $-\times-$ en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Student(1), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1=0, 2.5, 5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

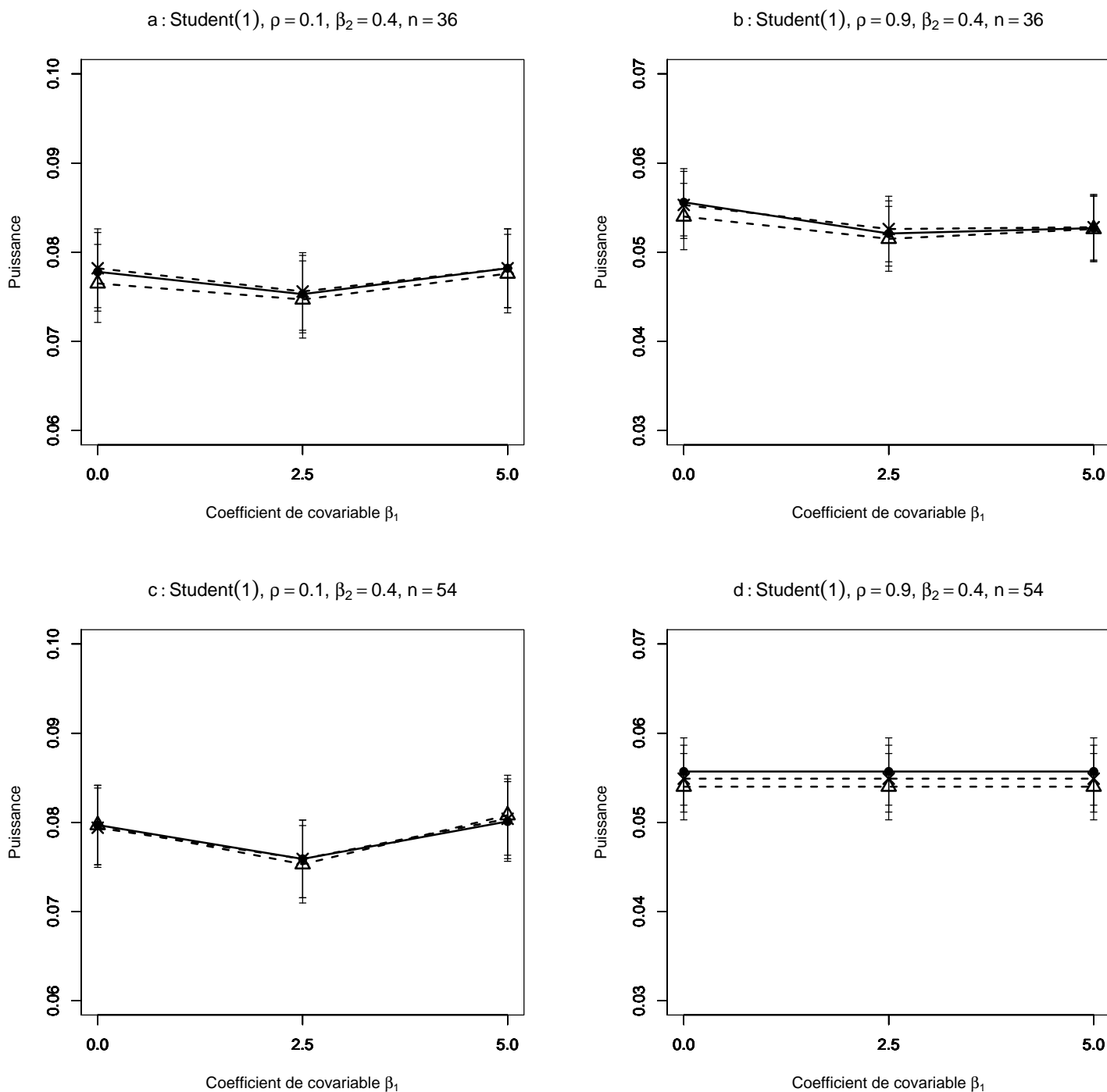


FIG. 2.21 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$ et Kennedy $- \cdot \times \cdot -$ en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Student(1), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1 = 0, 2.5, 5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

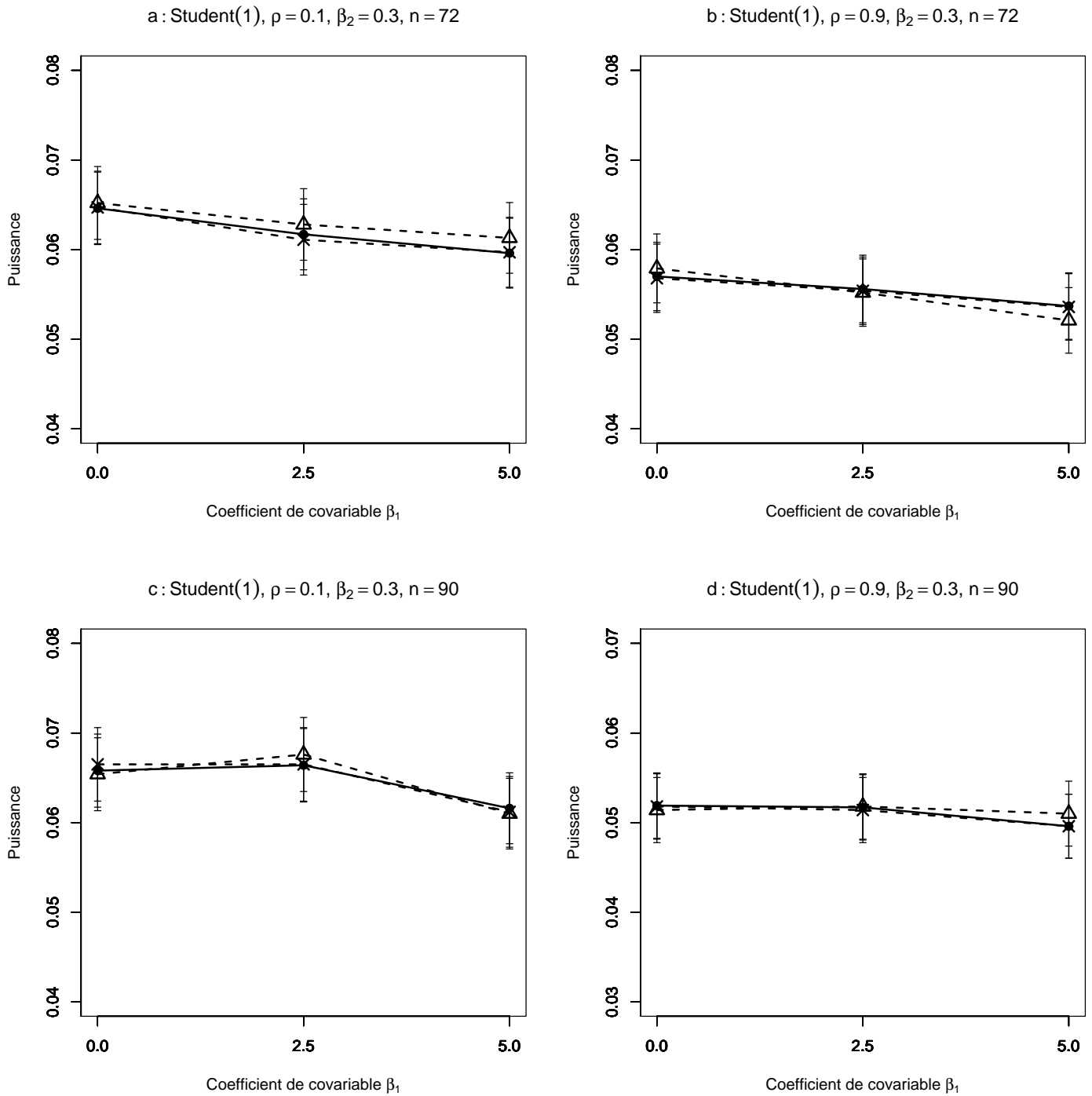


FIG. 2.22 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$ et Kennedy $-\cdot-\cdot-$ en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Student(1), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1=0, 2.5, 5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

B3- Cas des données générées pour le terme d'erreur selon une loi radicalement à dissymétrie droite comme la loi $Exp(2)$:

1. On peut trouver les deux premiers résultats obtenus dans le cas B1.
2. En comparant les cas B1 et B2 avec ce cas, on observe que quelles que soient les valeurs attribuées aux facteurs, la puissance de la méthode de Kennedy (comme celle de la méthode de Freedman-Lane, et celle de la méthode paramétrique classique) est nettement supérieure à celles obtenues par la même méthode dans les simulations précédentes (les cas B1 et B2). Remarquons que ceci peut être dû à la variance du terme d'erreur (loi $Exp(2)$) est 0.25 qui est plus petite que celles des cas B1 et B2.

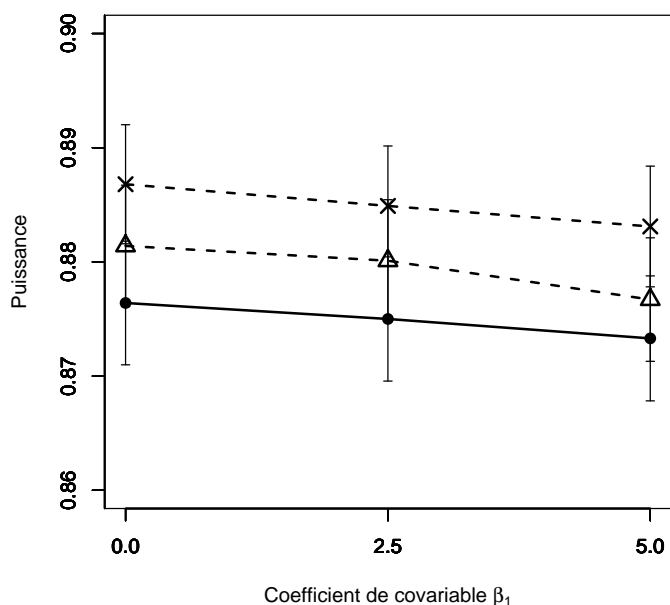
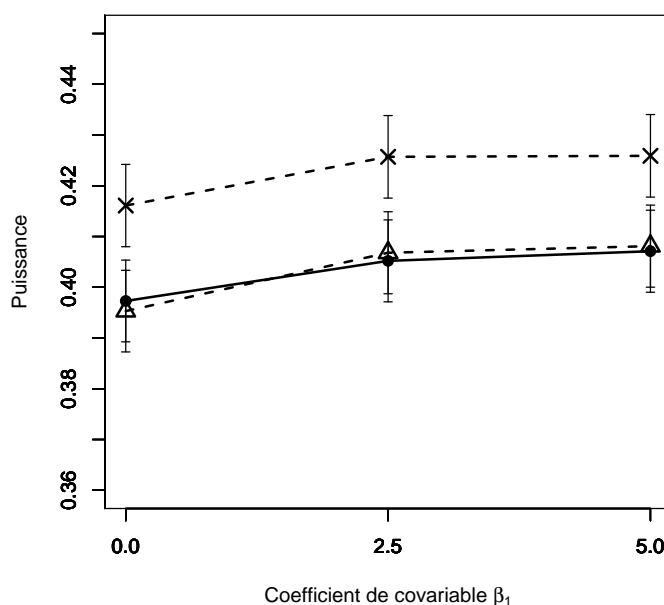
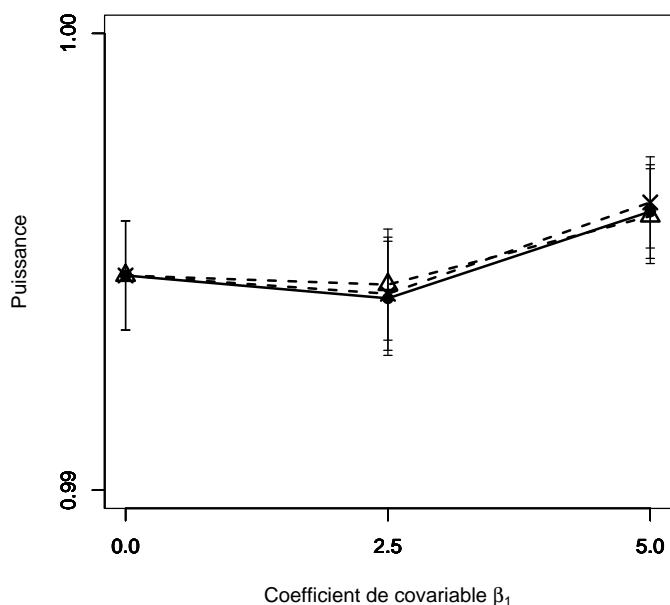
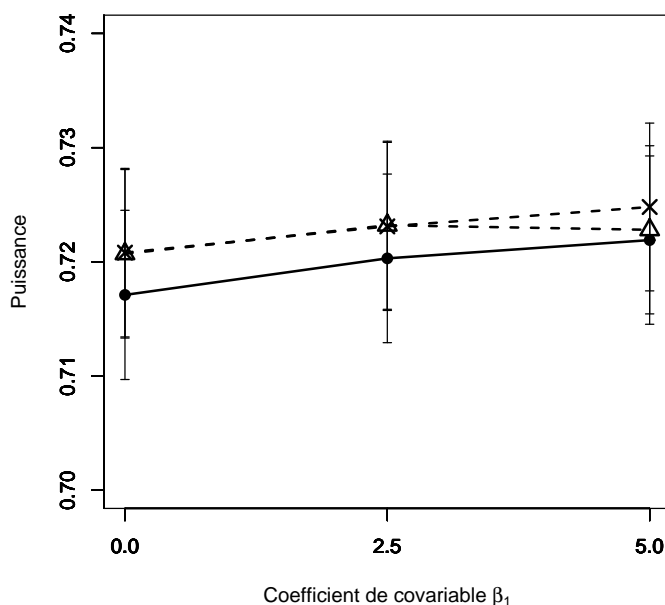
a : Exp(2), $\rho = 0.1$, $\beta_2 = 0.75$, $n = 9$ b : Exp(2), $\rho = 0.9$, $\beta_2 = 0.75$, $n = 9$ c : Exp(2), $\rho = 0.1$, $\beta_2 = 0.75$, $n = 18$ d : Exp(2), $\rho = 0.9$, $\beta_2 = 0.75$, $n = 18$ 

FIG. 2.23 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$ et Kennedy $-\times-$ en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi Exp(2), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1 = 0, 2.5, 5$. Les segments verticaux figurent l'intervalle de confiance à 95%.
 Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

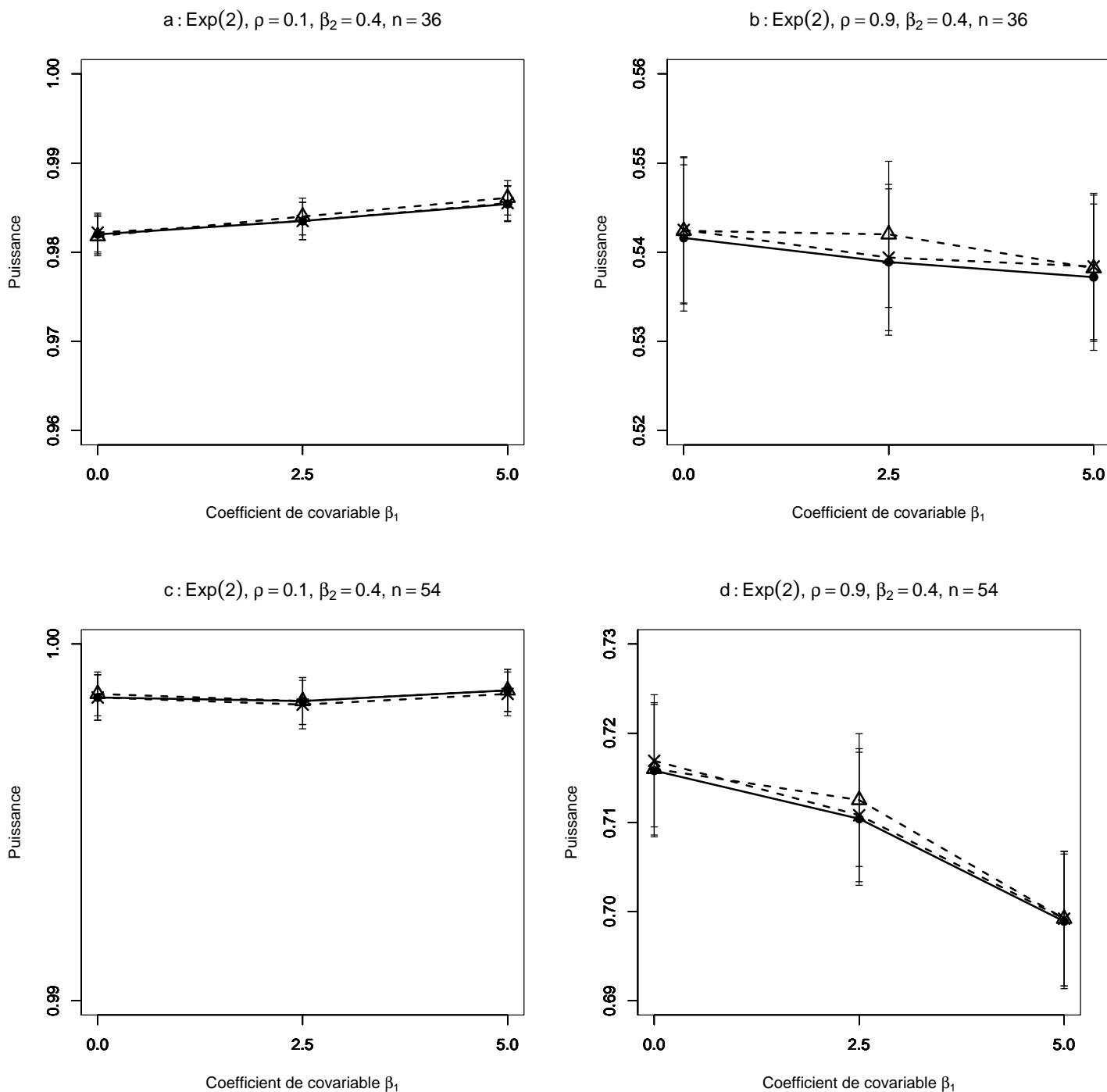


FIG. 2.24 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$ et Kennedy $\cdot\times\cdot$ en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi Exp (2), X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1=0, 2.5, 5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

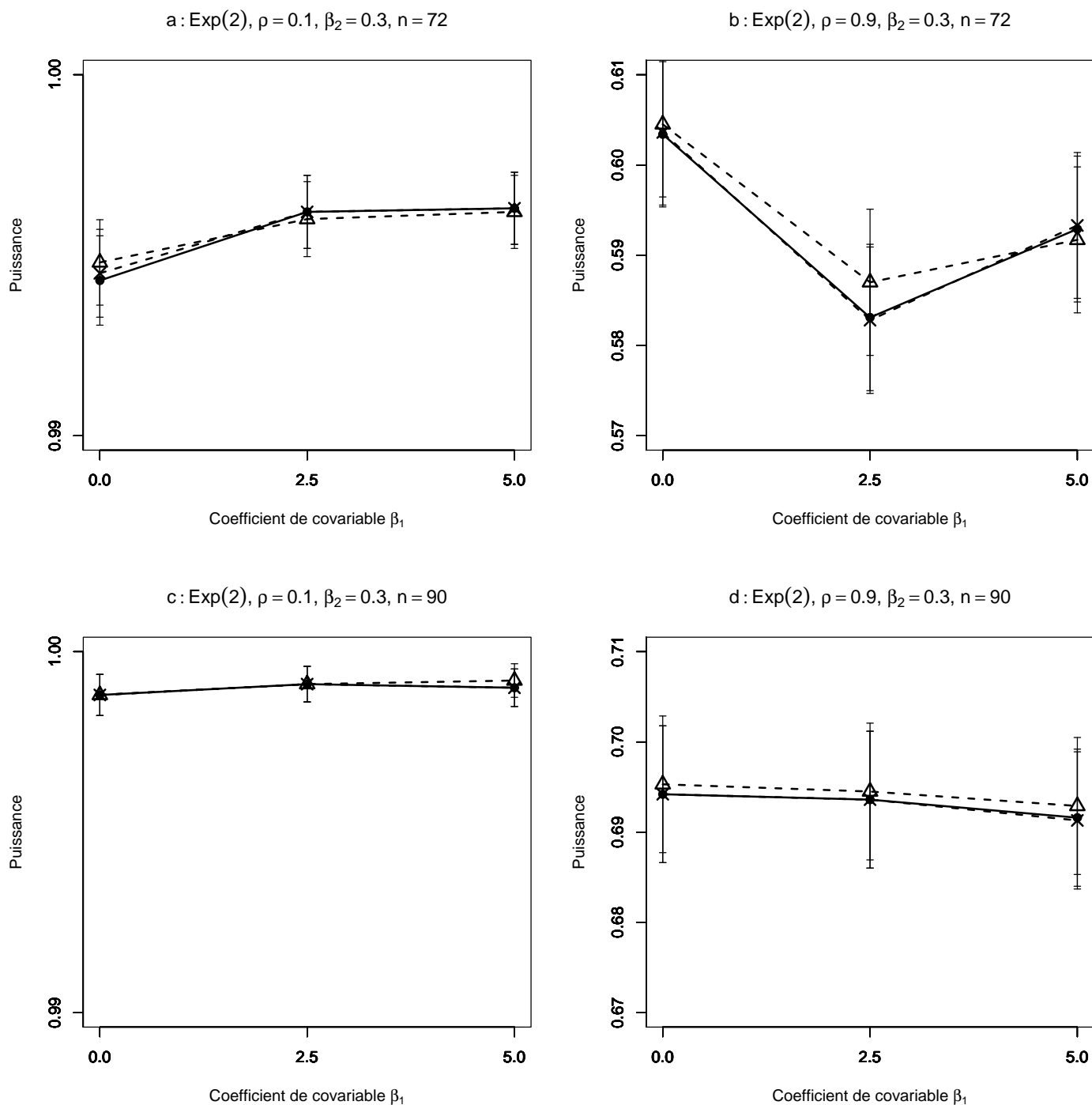


FIG. 2.25 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$ et Kennedy $-\times-$ en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi $\text{Exp}(2)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$. On a retenu $\beta_1 = 0, 2.5, 5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

B-4 Cas des données générées pour le terme d'erreur selon une loi radicalement à dissymétrie gauche comme la loi $\Gamma(10, 0.5)$:

1. On ne peut pas trouver les deux premiers résultats obtenus dans le cas B2.
2. Comme dans les cas précédents (B1, B2 et B3), dans la plupart des cas (surtout, lorsque la taille d'échantillon est petite ($n = 9$)), la puissance de la méthode de Kennedy est supérieure à celle de la méthode de Freedman-Lane (figure 2.26).
3. Contrairement aux cas B3 (où les données générées pour le terme d'erreur suivent une loi radicalement à dissymétrie droite comme la loi $\text{Exp}(2)$), ici, la puissance de la méthode de Kennedy comme celle de la méthode permutationnelle de Freedman-Lane et celle de la méthode paramétrique classique (comme on l'a vu déjà dans la section (2.5.2)) diminue nettement. Remarquons que ceci peut être dû au fait que la variance de la loi $\text{Exp}(2)$ est 0.25 et la variance de la loi $\Gamma(10, 0.5)$ est 40.
4. En augmentant la taille d'échantillon, les deux méthodes permutationnelles ainsi que la méthode paramétrique classique donnent la même puissance.
5. Les figures 2.27-c où $\beta_1 = 5$, 2.27-d où $\beta_1 = 5$, 2.28-a,b où $\beta_1 = 2.5$, 2.28-d, où $\beta_1 = 5$, montrent des cas où la puissance de la méthode de Kennedy est inférieure à celle de la méthode de Freedman-Lane.

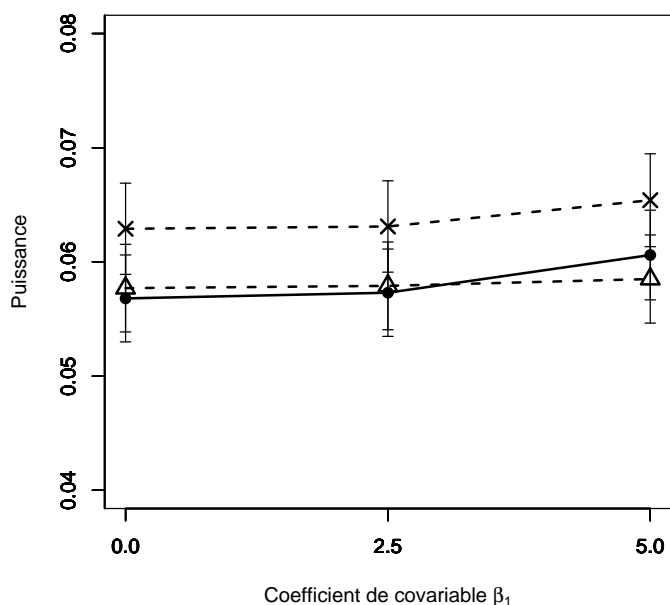
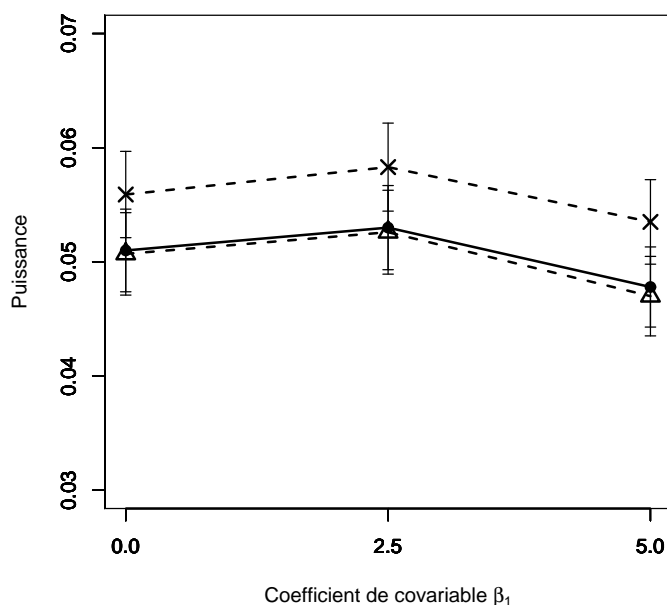
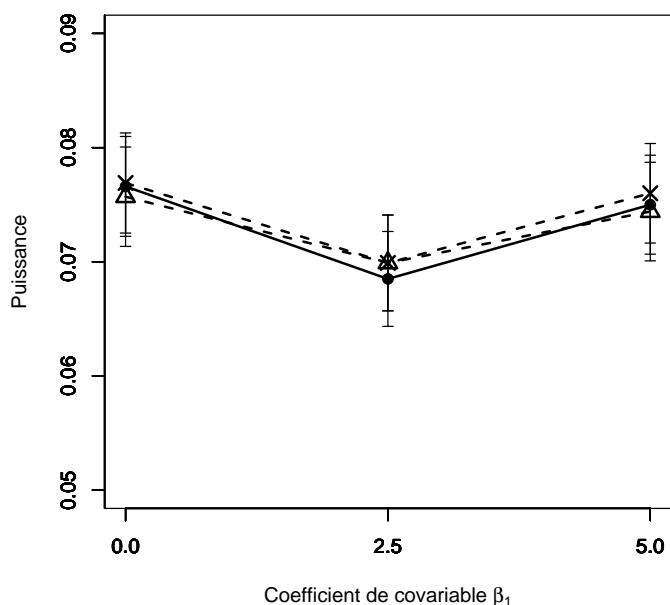
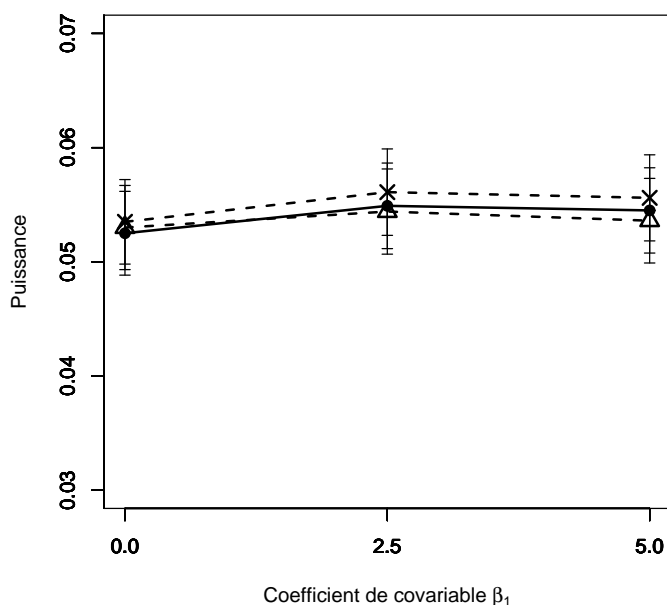
a : $\Gamma(10, 0.5)$, $\rho = 0.1$, $\beta_2 = 0.75$, $n = 9$ b : $\Gamma(10, 0.5)$, $\rho = 0.9$, $\beta_2 = 0.75$, $n = 9$ c : $\Gamma(10, 0.5)$, $\rho = 0.1$, $\beta_2 = 0.75$, $n = 18$ d : $\Gamma(10, 0.5)$, $\rho = 0.9$, $\beta_2 = 0.75$, $n = 18$ 

FIG. 2.26 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$ et Kennedy $-\times-$ en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi $\Gamma(10, 0.5)$, X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1 = 0, 2.5, 5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

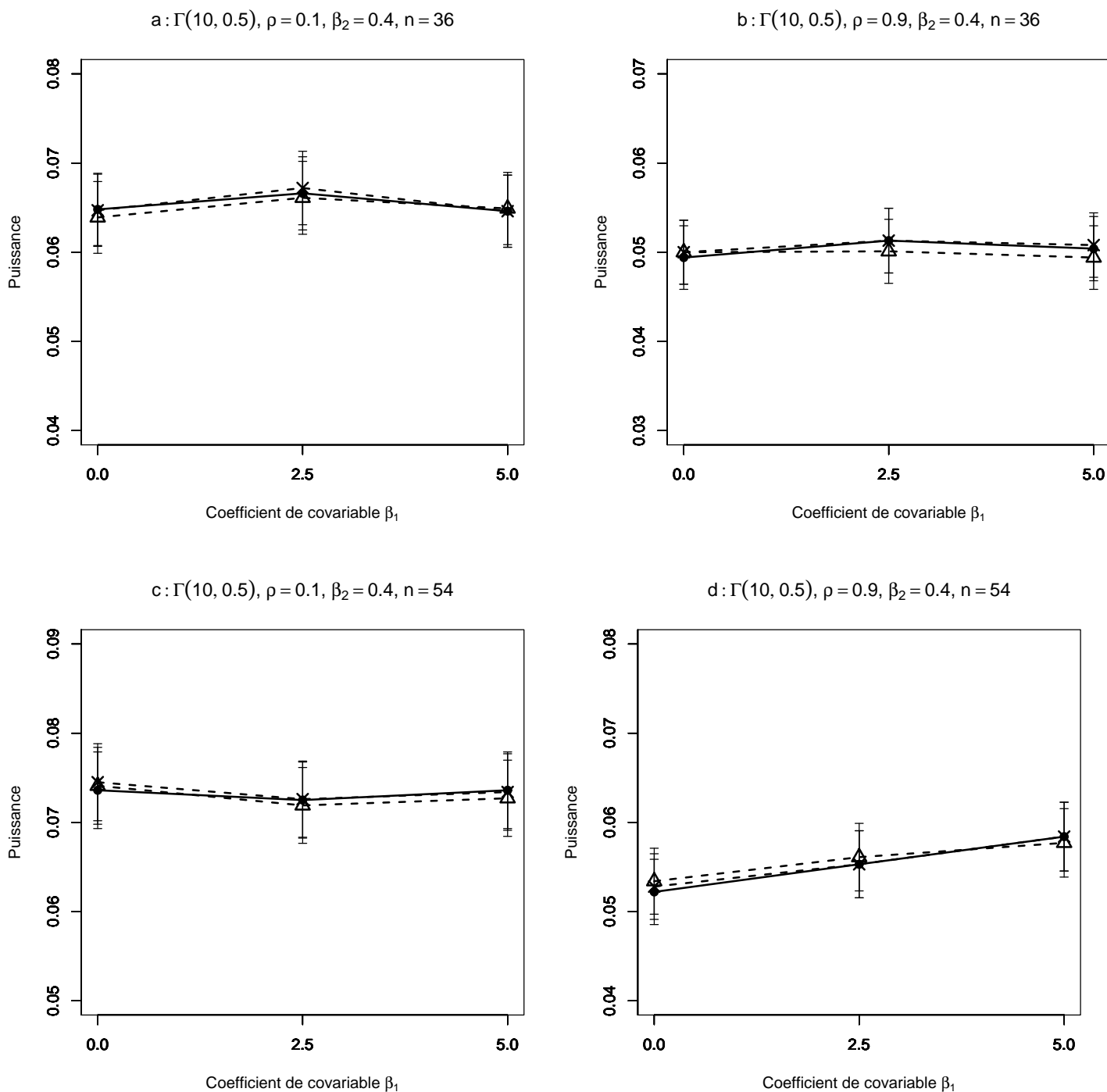


FIG. 2.27 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$ et Kennedy $\cdot\times\cdot$ en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi $\Gamma(10, 0.5)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$. On a retenu $\beta_1 = 0, 2.5, 5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

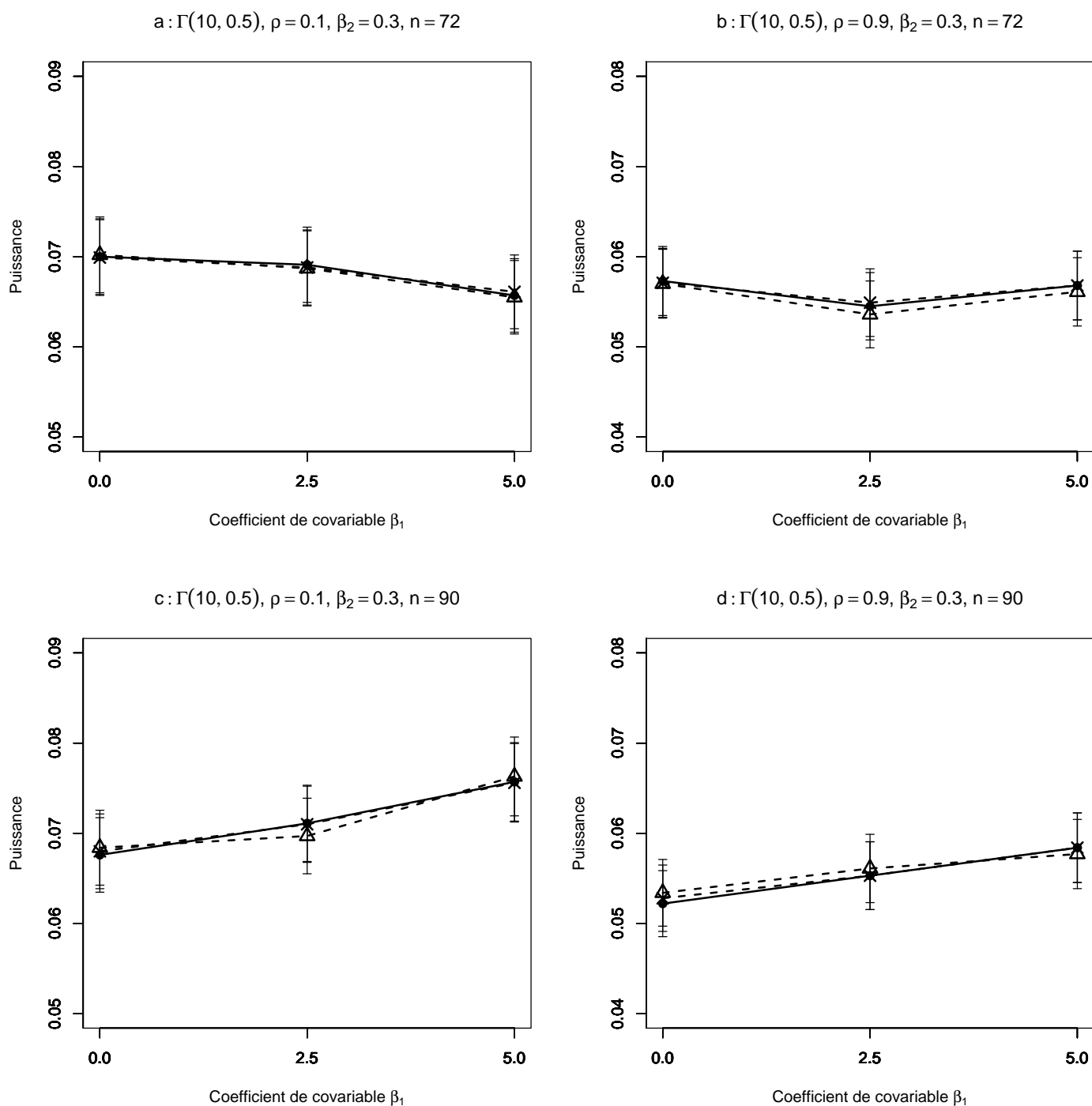


FIG. 2.28 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$ et Kennedy $-\cdot-\cdot-$ en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi $\Gamma(10, 0.5)$, X^1 et X^2 sont toujours Gaussiennes (0,1). On a retenu $\beta_1 = 0, 2.5, 5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

C- Résumé des principaux résultats

- a) Quelles que soient les valeurs attribuées aux facteurs, dans les cas B1 et B3 où les données générées pour le terme d'erreur suivent une loi symétrique de variance finie comme la loi de Gauss $LG(0, 1)$ ou une loi radicalement à dissymétrie droite comme la loi $Exp(2)$, en augmentant la taille de l'échantillon (les valeurs des autres facteurs restant fixées), les puissances des méthodes permutationnelles de Freedman-Lane, de Kennedy et la méthode paramétrique classique (en comparant les cas B2 et B4) augmentent nettement. Ceci peut être dû à la variance assez petite du terme d'erreur. Remarquons que la variance de la loi de Gauss $LG(0, 1)$ est 1, la variance de la loi de $Student(1)$ est infinie, la variance de la loi $Exp(2)$ est 0.25 et la variance de loi $\Gamma(10, 0.5)$ est 40.
- b) Lorsque le terme d'erreur suit d'une loi symétrique comme la loi de Student(1) ou une loi radicalement à dissymétrie gauche comme la loi $\Gamma(10, 0.5)$, en augmentant la taille d'échantillon (les valeurs des autres facteurs restant fixées), les puissances des méthodes permutationnelles de Freedman-Lane, de Kennedy et la méthode paramétrique classique (en comparant les cas B1 et B3) augmentent peu. Ceci peut être dû à la variance infini (ou assez grande) du terme d'erreur.
- c) Lorsque les données générées pour le terme d'erreur suivent une loi symétrique de variance finie comme la loi de Gauss $(0, 1)$ ou une loi radicalement à dissymétrie droite comme la loi $Exp(2)$, en augmentant le degré de coefficient de corrélation $\rho(X^1, X^2)$ (les valeurs des autres facteurs restant fixées), les puissances des méthodes permutationnelles de Freedman-Lane, de Kennedy et la méthode paramétrique classique (en comparant les cas B2 et B4) diminuent nettement. Ceci peut être dû à la variance assez petite du terme d'erreur.
- d) Quelles que soient les valeurs attribuées aux facteurs, dans les cas B2 et B4 où les données générées pour le terme d'erreur suivent une loi symétrique de variance infinie comme la loi de $Student(1)$ ou une loi radicalement à dissymétrie gauche comme la loi de $\Gamma(10, 0.5)$, en augmentant le degré de coefficient de corrélation $\rho(X^1, X^2)$ (les valeurs des autres facteurs restant fixées), les puissances des méthodes permutationnelles de Freedman-Lane, de Kennedy et la méthode paramétrique classique (en comparant les cas B1 et B3) diminuent peu. Ceci peut être dû à la variance infinie ou trop grande du terme d'erreur.
- e) Même lorsque les données générées pour le terme d'erreur suivent une loi symétrique de variance infinie comme la loi de Student(1) ou radicalement à dissymétrie droite comme la loi $Exp(2)$ ou radicalement à dissymétrie gauche comme la loi $\Gamma(10, 0.5)$, la méthode paramétrique classique marche bien. C'est-à-dire qu'elle donne une puissance presque identique à celle des méthodes permutationnelles de Freedman-Lane et de Kennedy. Ceci peut être dû à des valeurs des variables X^1 et X^2 générées selon une loi de Gauss $(0, 1)$.
- f) Lorsque la variance des données générées pour le terme d'erreur est infinie ou trop grande (soit d'une loi symétrique comme Student(1) ou d'une loi radicalement à dissymétrie gauche comme la loi $\Gamma(10, 0.5)$), les puissances des

méthodes permutacionnelles de Freedman-Lane, de Kennedy et de la méthode paramétrique classique diminuent nettement en comparaison aux puissances obtenues par les mêmes méthodes mais où la variance des données générés pour le terme d'erreur est assez petite comme la loi de Gauss $(0, 1)$ de variance 1 ou comme la loi radicalement à dissymétrie droite comme la loi $Exp(2)$ de variance 0.25.

- g) Comme nous l'avons prévu, dans la plupart des cas, la puissance de la méthode de Kennedy est supérieure à celle de Freedman-Lane, surtout pour $n = 9, 18$
- h) Comme nous l'avons prévu, nous observons des cas où la puissance de la méthode Kennedy est inférieure à celle du test de Freedman-Lane où la taille de l'échantillon était grande ($n > 18$). Nous en expliquerons les raisons dans le chapitre 3.

2.5.4 Effet d'un point extrême

En s'inspirant de la simulation d'Anderson-Legendre, nous n'avons pas considéré la méthode de Kennedy pour vérifier l'effet d'un point extrême sur le risque d'erreur de première espèce. Comme dans le risque d'erreur de première espèce et dans la puissance, ici, nous avons complété les résultats obtenus par Anderson-Legendre, en considérant en plus des distributions choisies par Anderson-Legendre $\{LG(0, 1), Exp(1), Exp(1)^3\}$ des distributions symétriques avec variance finie $Student(5)$, symétrique avec variance infinie $Student(1)$ à dissymétrie droite comme $\{\Gamma(1, 0.5), \Gamma(1.5, 0.5), \Gamma(1.5, 2)\}$, radicalement à dissymétrie droite comme $\{\Gamma(0.5, 0.5), \Gamma(0.5, 2), \Gamma(1, 2)\}$ à dissymétrie gauche $\{\Gamma(10, 2)\}$ et radicalement à dissymétrie gauche comme $\{\Gamma(10, 0.5)\}$ de terme d'erreur. Par ailleurs, afin de comparer les effets des valeurs des données X^1 et X^2 nous avons générées les valeurs de X^1 et X^2 selon une loi de Gauss. Les autres paramètres sont presque les mêmes que dans Anderson-Legendre.

A- Le cadre de notre simulation

- a) Afin d'étudier les effets de taille d'échantillon et comparer les résultats obtenus par Anderson-Legendre aux nôtres, nous avons tiré les mêmes tailles d'échantillons que celles qui furent retenues dans la simulation d'Anderson-Legendre. Donc, $n = \{10, 20, 100\}$
- b) Pour vérifier les effets du degré d'influence du coefficient de covariable et aussi pour permettre la comparaison des résultats obtenus par Anderson-Legendre avec nos résultats, nous avons considéré les mêmes valeurs du coefficient β_1 que celles de ces auteurs. Donc, $\beta_1 = \{0, 2, 4, 5, 6, 8, 10, 20\}$.

- c) La distribution du terme d'erreur aléatoire ϵ appartient à l'ensemble : $\{LG(0, 1), Exp(0.5), Exp(2), Student(1), Student(5), \Gamma(a, b)$ avec $a \in \{0.5, 1, 1.5, 10\}$ et $b \in \{0.5, 2\}\}$

Afin de comparer les effets des valeurs de X^1 et X^2 générées par Anderson-Legendre selon une loi uniforme (0,3) () et les nôtres générées selon une loi de Gauss (0, 1) mais aussi afin d'étudier les effets du degré de corrélation entre les variables X^1 et X^2 , nous avons considéré les mêmes corrélations que celles qui furent retenues par la simulation d'Anderson-Legendre. Pour cela, nous avons généré les données des variables X^1 et X^2 de deux façons différentes.

A1- Cas où les variables X^1 et X^2 sont non corrélées

D'après remarque (2.4.1) on a $\rho(X^1, X^2) = 0$. Donc, pour la variable X^1 nous générons $n - 1$ valeurs d'une loi de Gauss (0, 1) et nous posons $X_n^1 = 55$. Puis nous générons n valeurs de la variable X^2 d'une loi de Gauss (0, 1). De ce fait, la corrélation entre les variables X^1 et X^2 est zéro. Ensuite, nous construisons deux vecteurs $\mathbf{X}^1, \mathbf{X}^2$ et la matrice \mathbf{W} par l'équation :

$$\mathbf{W} = [\mathbf{X}^1 | \mathbf{X}^2] = \begin{bmatrix} X_1^1 & X_1^2 \\ X_2^1 & X_2^2 \\ \vdots & \vdots \\ X_n^1 & X_n^2 \end{bmatrix} \quad (2.23)$$

Enfin, les valeurs de Y sont générées par l'équation :

$$\mathbf{Y} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.24)$$

où : $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ est le vecteur des paramètres. Pour calculer le risque d'erreur de première espèce il faut poser $\beta_2 = 0$ et $\boldsymbol{\epsilon}$ est le vecteur du terme d'erreur aléatoire généré selon une distribution considérée.

Pour chaque combinaison des modalités des facteurs, 10000 ensembles de données ont été générés. Puis, pour la méthode de Freedman-Lane, chaque ensemble de données a été permuté 999 fois. La statistique de test permutacionnelle ($t_{2,\pi}$) de la méthodes de Freedman-Lane a été calculée sur chaque ensemble de données permutes. Ensuite, après 999 permutations, la p -valeur permutacionnelle pour la méthode de Freedman-Lane a été calculée par la formule :

$$p - valeur_{\pi} = \frac{\#(|t_{\pi}| \geq t_{obs})}{1000}$$

Le niveau de signification pour les tests a été fixé à $\alpha = 0,05$. Alors, pour chaque ensemble de données quand la $p - valeur_{\pi}$ est inférieure à 0.05 on rejette l'hypothèse $H_0 : \beta_2 = 0$. Sur les 10000 $p - valeur_{\pi}$ le pourcentage des $P - valeur_{\pi}$ de la méthode de Freedman-Lane inférieures à 0.05 donne le risque d'erreur de première espèce. En

fait, ce pourcentage estime la probabilité de rejeter $H_0 : \beta_2 = 0$ lorsque H_0 est vraie (parce que les données sont générés sous $\beta_2 = 0$). De plus, l'intervalle de confiance à 95% est calculé à partir du risque d'erreur de première espèce (calculé sur 10000 ensembles).

Afin de compléter la comparaison, pour chaque ensemble de données on a utilisé les tests paramétriques classiques qui consistent à calculer une statistique de Student. Sa p-valeur donnée par théorie Gaussienne a également été calculée. De la même façon, sur les 10000 ensembles de données le risque d'erreur de première espèce et son intervalle de confiance sont calculés par la théorie Gaussienne.

Les autres étapes sont identiques à celles où nous avons étudié le risque d'erreur de première espèce.

A2- Cas des variables X^1 et X^2 sont corrélées

D'après remarque (2.4.2) on a $\rho(X^1, X^2) \approx 0.7$. De ce fait, nous générons $n - 1$ valeurs d'une loi de Gauss $(0, 1)$ pour la variable Z^1 , et nous posons $Z_n^1 = 55$. Puis nous générons n valeurs d'une loi de Gauss $(0, 1)$ pour la variable Z^2 . Ensuite, nous construisons deux vecteurs $\mathbf{Z}^1, \mathbf{Z}^2$ et la matrice \mathbf{W}' par l'équation :

$$\mathbf{W}' = [\mathbf{Z}^1 | \mathbf{Z}^2] = \begin{bmatrix} Z_1^1 & Z_1^2 \\ Z_2^1 & Z_2^2 \\ \vdots & \vdots \\ Z_n^1 & Z_n^2 \end{bmatrix} \quad (2.25)$$

Remarquons que $\rho(Z^1, Z^2) = 0$. A partir des \mathbf{Z}^1 et \mathbf{Z}^2 , ces auteurs construisent les vecteurs \mathbf{X}^1 et \mathbf{X}^2 de corrélation $\rho(X^1, X^2) = 0.7$ de la façon suivante :

$$\mathbf{W}'_{\mathbf{R}} = [\mathbf{X}^1 | \mathbf{X}^2] = \mathbf{W}' \mathbf{R}^{1/2} \quad (2.26)$$

où $\mathbf{R} = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}$. Enfin, les valeurs de Y sont générées par l'équation :

$$\mathbf{Y} = \mathbf{W}'_{\mathbf{R}} \boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.27)$$

où $\boldsymbol{\epsilon} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ est le vecteur des paramètres. Pour calculer le risque d'erreur de première espèce il faut prendre $\beta_2 = 0$ et $\boldsymbol{\epsilon}$ est le vecteur du terme d'erreur aléatoire généré selon la distribution considérée.

Pour chaque combinaison des modalités des facteurs, 10000 ensembles de données ont été générées. Chaque ensemble de données a été permuté 999 fois. La statistique de test permutacionnelle ($t_{2,\pi}^F$) de la méthode de Freedman-Lane est calculée sur chaque ensemble de données permutes. Ensuite, après 999 permutations, la p-valeur permutacionnelle pour la méthode de Freedman-Lane a été calculée par la formule :

$$p - \text{valeur}_{\pi} = \frac{\#(|t_{\pi}| \geq t_{obs})}{1000}$$

Les autres étapes sont identiques à celle du cas A1.

B- Les résultats de nos simulations**B1- Cas des données générées pour le terme d'erreur selon une loi $LG(0, 1)$**

1. Comme dans la simulation d'Anderson-Legendre, pour toutes les valeurs de β_1 , la méthode de Freedman-Lane donne un risque d'erreur de première espèce proche de 0.05. C'est-à-dire que le point extrême (avec $\beta_1 \neq 0$) n'a pas d'effet sur le risque d'erreur de première espèce.
2. En comparant les figures 2.26-a,c,e où $\rho = 0$, et 2.26-b,d,f où $\rho = 0.7$ on constate que dans les deux cas la méthode de Freedman-Lane donne presque le même risque d'erreur de première espèce. C'est-à-dire que même s'il existe un point extrême, la corrélation entre les variables X^1 et X^2 , n'a pas d'effet sur le risque d'erreur de première espèce de la méthode de Freedman-Lane.
3. En considérant la figure 2.26-a,b où la taille d'échantillon est petite $n = 10$, le risque d'erreur de première espèce de la méthode de Freedman-Lane est proche de 0.05. C'est-à-dire que lorsque la taille d'échantillon est petite même s'il existe un point extrême, la méthode de Freedman-Lane donne un risque proche de 0.05.
4. Comme dans la simulation d'Anderson-Legendre, il n'y a pas de différence significative entre la méthode permutationnelle de Freedman-Lane et la méthode paramétrique classique. Ceci est acceptable puisque les données générées pour le terme d'erreur suivent une loi de Gauss $(0, 1)$.

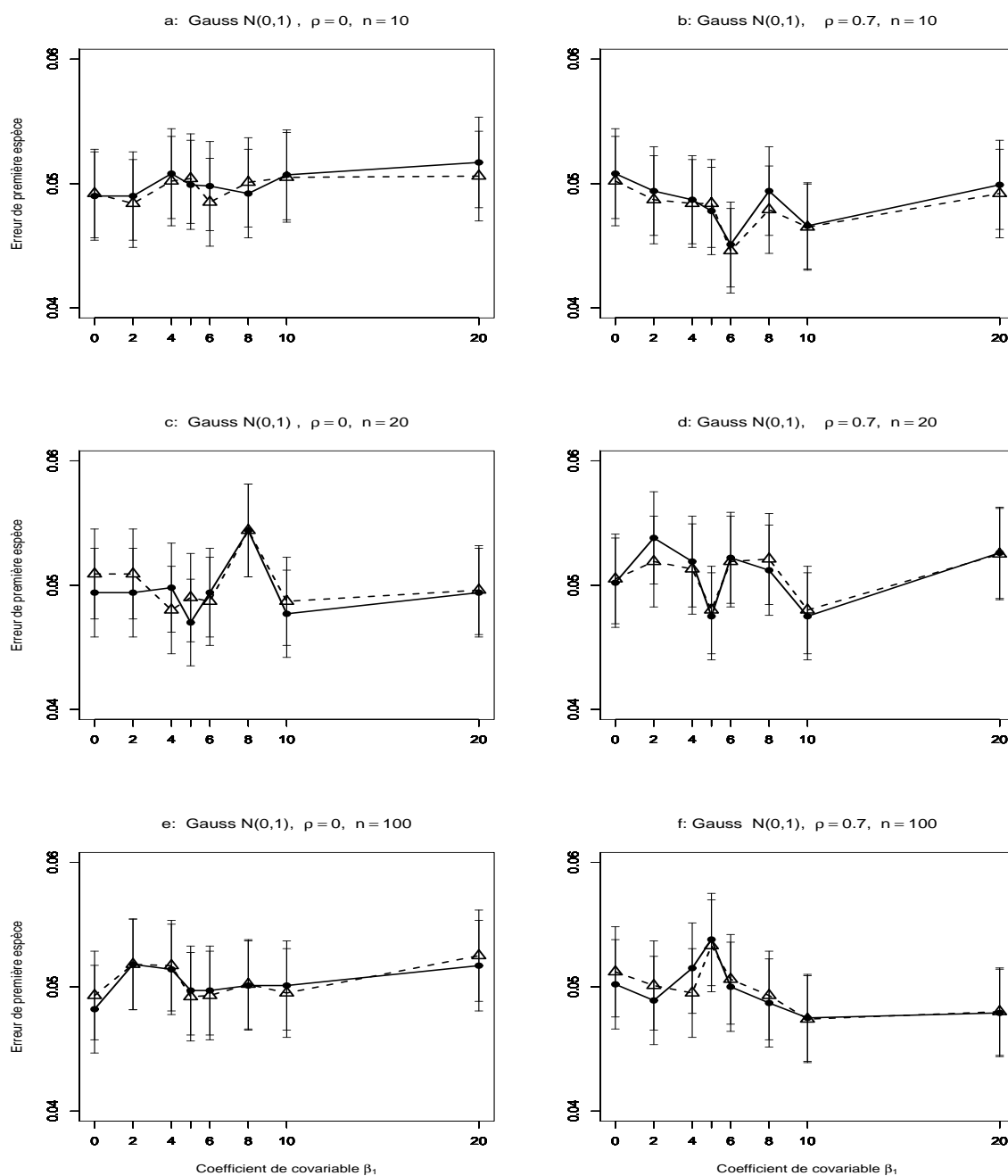


FIG. 2.29 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $-\triangle-$, et Freedman et Lane $-\bullet-$, en fonction du coefficient de covariable. 10000 n -échantillons du résidu ϵ ont été simulés suivant une loi de Gauss $(0,1)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$ où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

B2- Cas des données générées pour le terme d'erreur selon une loi symétrique de variance infinie comme la loi de Student à 1 degré de liberté :

1. Pour toutes les combinaisons de valeurs de facteurs, la méthode du Freedman-Lane donne un risque d'erreur de première espèce proche de 0.05. C'est-à-dire qu'elle se stabilise vers 0.05.
2. Comme dans le cas précédent, en comparant les figures 2.27-a,c,d où $\rho = 0$, et 2.27-b,d,f où $\rho = 0.7$ on constate que dans les deux cas la méthode de Freedman-Lane donne presque le même risque d'erreur de première espèce, c'est-à-dire que la corrélation entre les variables X^1 et X^2 , n'a pas d'effet sur le risque d'erreur de première espèce de la méthode de Freedman-Lane.
3. En considérant la figure 2.27-a,b où la taille d'échantillon est petite $n = 10$, le risque d'erreur de première espèce de la méthode de Freedman-Lane est proche de 0.05. C'est-à-dire que même lorsque les données générées pour le terme d'erreur suivent une loi de variance infinie comme la loi de Student(1), lorsque la taille de l'échantillon est petite et même s'il existe un point extrême, la méthode de Freedman-Lane donne un risque proche de 0.05.
4. Contrairement à la simulation menée pour étudier les puissances des méthodes permutationnelles, ici la grandeur de la variance du terme d'erreur, n'a pas d'effet sur le risque d'erreur de première espèce.
5. Comme dans le cas précédent il n'y a pas de différence significative entre la méthode permutationnelle de Freedman-Lane et la méthode paramétrique classique.
6. Même lorsque les données générées pour le terme d'erreur suivent une loi symétrique de variance infinie comme la loi de Student(1), la méthode paramétrique classique donne un risque d'erreur de première espèce proche de 0.05. Ceci peut être dû au fait que les valeurs des variables X^1 et X^2 sont générées selon une loi de Gauss $(0, 1)$.

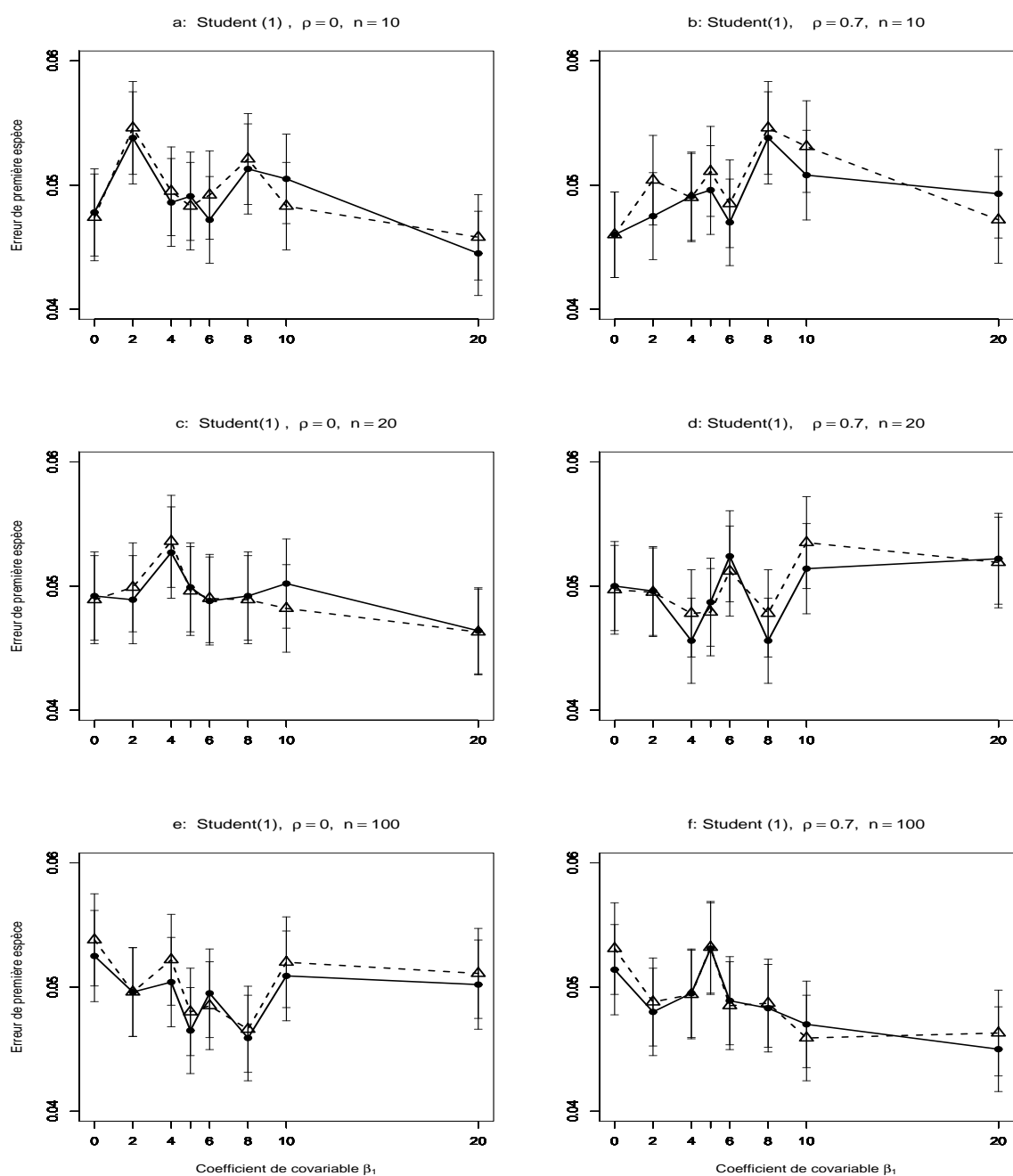


FIG. 2.30 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $-\triangle-$, et Freedman et Lane $-\bullet-$, en fonction du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Student(1)$, X^1 et X^2 sont toujours Gaussiennes (0,1) où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

B3- Cas des données générées pour le terme d'erreur selon une loi radicalement à dissymétrie droite comme la loi $Exp(2)$:

1. Comme dans la simulation d'Anderson-Legendre, pour toutes les combinaisons des valeurs de facteurs, la méthode de Freedman-Lane donne un risque d'erreur de première espèce proche de 0.05. C'est-à-dire qu'elle se stabilise vers 0.05.
2. Comme dans les cas précédents (B1 et B2), en comparant les figures 2.28-a,c,e où $\rho = 0$, et 2.28-b,d,f où $\rho = 0.7$ on constate que dans les deux cas la méthode de Freedman-Lane donne presque le même risque d'erreur de première espèce, c'est-à-dire que la corrélation entre les variables X^1 et X^2 , n'a pas d'effet sur le risque d'erreur de première espèce de la méthode de Freedman-Lane.
3. En considérant la figure 2.28-a,b où la taille d'échantillon est petite $n = 10$, comme dans les cas précédents (B1 et B2) le risque d'erreur de première espèce de la méthode de Freedman-Lane est proche de 0.05. C'est-à-dire que lorsque la taille d'échantillon est petite et lorsqu'il existe un point extrême, la méthode de Freedman-Lane donne un risque proche de 0.05.
4. Contrairement à la simulation d'Anderson-Legendre où la méthode paramétrique classique était instable, ici cette méthode se stabilise vers 0.05. Ceci peut être dû au fait que les valeurs des variables X^1 et X^2 sont générées selon une loi de Gauss $(0, 1)$.
5. Contrairement à la simulation menée pour étudier les puissances des méthodes permutationnelles, ici la magnitude de la variance du terme d'erreur n'a pas d'effet sur le risque d'erreur de première espèce.
6. Comme dans le cas précédent il n'y a pas de différence significative entre la méthode permutationnelle de Freedman-Lane, et la méthode paramétrique classique. Ceci peut-être dû au fait que les valeurs des variables X^1 et X^2 sont générées selon une loi de Gauss $(0, 1)$.

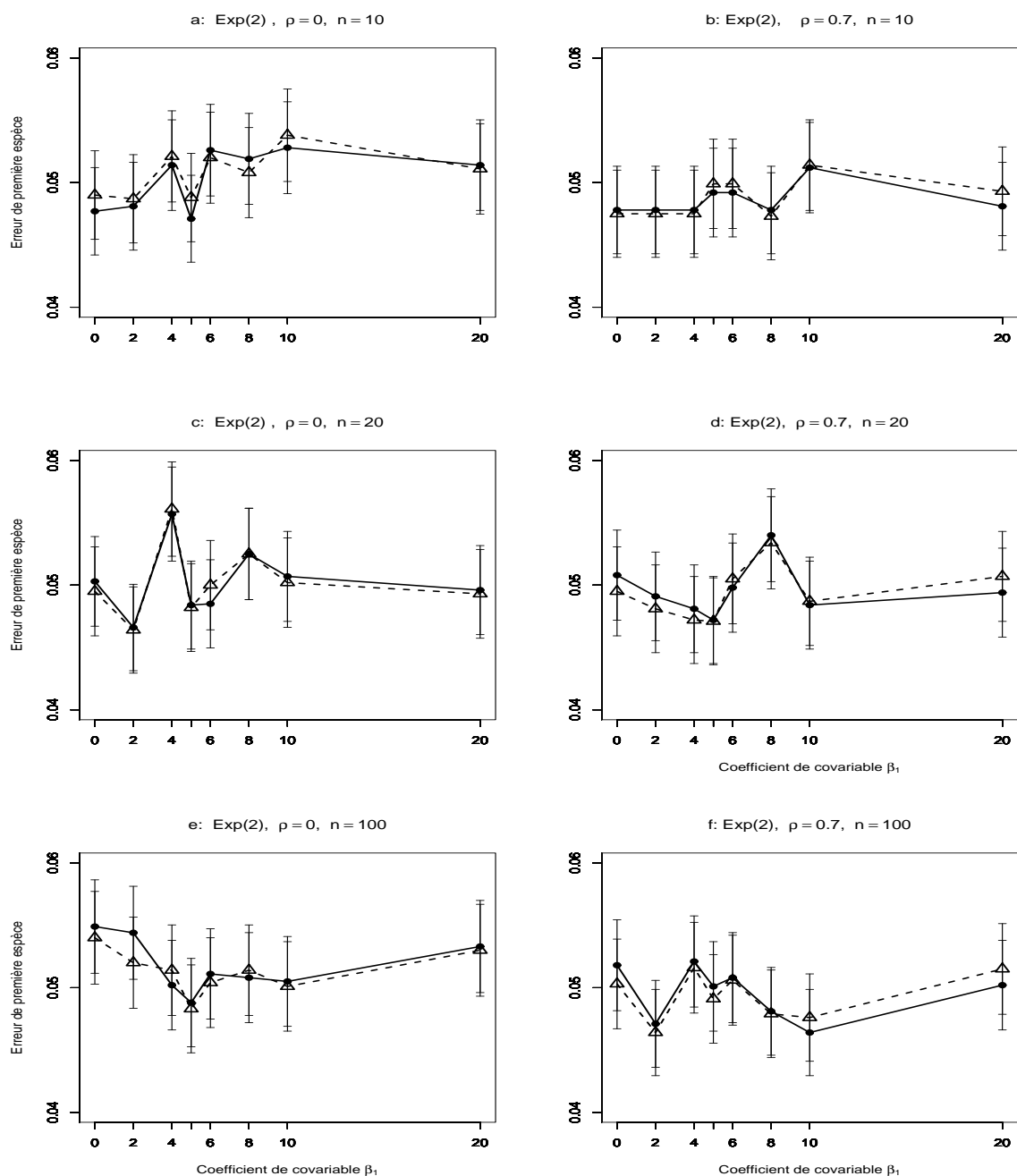


FIG. 2.31 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $-\triangle-$, et Freedman et Lane $-\bullet-$, en fonction du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi $Exp(2)$, X^1 et X^2 sont toujours Gaussiennes (0,1) où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

B-4 Cas des données générées pour le terme d'erreur selon une loi radicalement à dissymétrie gauche comme la loi $\Gamma(10, 0.5)$:

1. Pour toutes les valeurs de β_1 , la méthode de Freedman-Lane donne un risque d'erreur de première espèce proche de 0.05. C'est-à-dire qu'elle se stabilise vers 0.05.
2. Comme dans le cas précédent, en comparant les figures 2.29-a,c,e où $\rho = 0$, et 2.29-b,d,f où $\rho = 0.7$ on constate que pour les deux cas la méthode de Freedman-Lane donne presque le même risque d'erreur de première espèce, c'est-à-dire que la corrélation entre les variables X^1 et X^2 n'a pas d'effet sur le risque d'erreur de première espèce de la méthode de Freedman-Lane.
3. En considérant la figure 2.32-a,b où la taille d'échantillon est petite $n = 10$, le risque d'erreur de première espèce de la méthode de Freedman-Lane est proche de 0.05. C'est-à-dire que la méthode de Freedman-Lane se stabilise aussi autour de 0.05 quand les données générées pour le terme d'erreur suivent une loi de variance assez grande (40) par exemple la loi $\Gamma(10, 0.5)$ avec une taille échantillon petite ($n = 10$), et la présence d'un point extrême.
4. Comme dans le cas précédent il n'y a pas de différence significative entre la méthode permutacionnelle de Freedman-Lane et la méthode paramétrique classique.
5. Même lorsque les données générées pour le terme d'erreur suivent une loi dissymétrique de variance assez grande (40) comme la loi $\Gamma(10, 0.5)$, la méthode paramétrique classique donne un risque d'erreur de première espèce proche de 0.05. Ceci peut être dû au fait que les valeurs des variables X^1 et X^2 sont générées selon une loi de Gauss $(0, 1)$.

Des résultats plus complets, utilisant d'autres combinaisons des facteurs, sont présentés dans l'annexe 3.

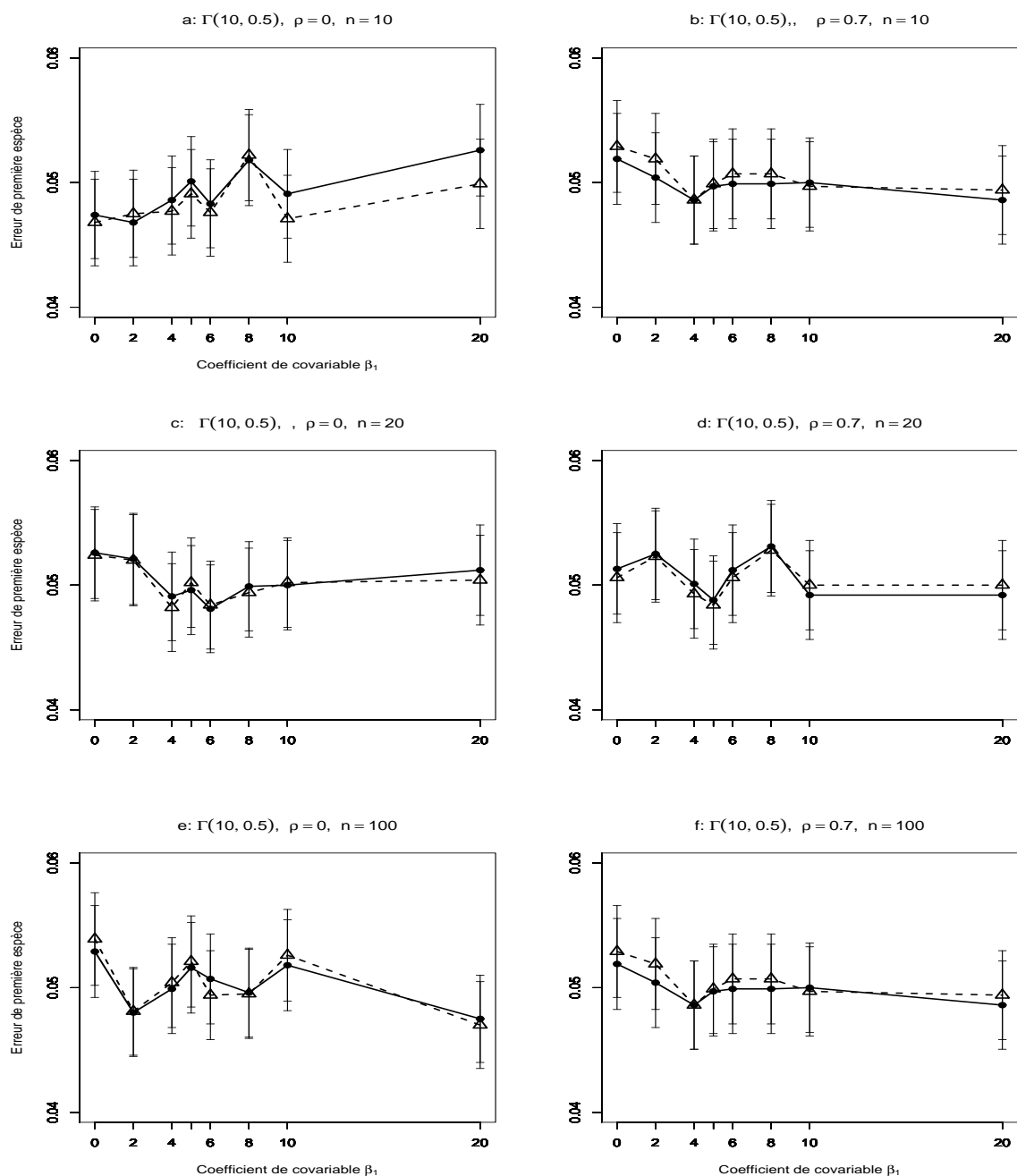


FIG. 2.32 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, et Freedman et Lane— \bullet —, en fonction du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi $\Gamma(10, 0.5)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$ où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; figure d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

2.5.5 Étude d'effet d'un point extrême pour la méthode de Kennedy

Compte tenu des résultats théoriques présentés au chapitre 3, nous avons décidé d'étudier (contrairement à la section (2.5.4) et à la simulation d'Anderson-Legendre) l'effet d'un point extrême sur le risque d'erreur de première espèce de la méthode de Kennedy. Même dans la section (2.4.4) nous avons considéré la même corrélation (entre les variables X^1 et X^2) que celle d'Anderson-Legendre, mais nous avons oublié de considérer les mêmes moyennes et variances pour les variables X^1 et X^2 que celles d'Anderson-Legendre. Alors, dans cette section, en considérant les mêmes moyennes et variances (pour les variables X^1 et X^2) et corrélation $\rho(X^1, X^2)$ que celles dans la simulation d'Anderson-Legendre, nous étudions l'effet d'un point extrême sur le risque d'erreur de première espèce pour les méthodes permutationnelles de Freedman-Lane, de Kennedy et aussi la méthode paramétrique classique.

A- Le cadre de la simulation

Nous avons considéré presque le même cadre de simulation que dans la section (2.5.4), sauf que pour économiser du temps nous avons choisi les quatre distributions suivantes pour le terme d'erreur : symétrique avec variance finie comme la loi de Gauss $LG(0, 1)$, symétrique avec variance infinie comme la loi de *Student*(1), radicalement à dissymétrie droite comme la loi *Exp*(2) et radicalement à dissymétrie gauche comme la loi $\Gamma(10, 0.5)$. Donc, les ensembles des facteurs sont :

- a) Afin d'étudier les effets de tailles d'échantillons et de comparer les résultats obtenus par Anderson-Legendre aux nôtres, nous avons tiré les mêmes tailles d'échantillon que celles dans la simulation d'Anderson-Legendre. Donc, $n = \{10, 20, 100\}$
- b) Pour vérifier les effets du degré de relation et aussi permettre la comparaison des résultats obtenus par Anderson-Legendre avec nos résultats, nous avons considéré les mêmes coefficients de β_1 que dans la simulation d'Anderson-Legendre. Donc, $\beta_1 = \{0, 2, 4, 5, 6, 8, 10, 20\}$.
- c) La distribution du terme d'erreur aléatoire ϵ appartient à l'une des lois suivant : $\{LG(0, 1), Exp(2), Student(1), \Gamma(10, 0.5)\}$

Afin de comparer les effets des valeurs des données X^1 et X^2 générées par Anderson-Legendre selon une loi uniforme (0,3) et les nôtres générées suivant une loi de Gauss, afin de plus, d'étudier les effets du degré de corrélation entre les variables X^1 et X^2 , nous avons considéré les mêmes corrélations, moyenne et variance que celles d'Anderson-Legendre. Pour cela, nous avons généré les données des variables X^1 et X^2 de deux façons différentes.

A1- Cas des variables X^1 et X^2 non corrélées

D'après la remarque (2.4.1) on a :

$$\mathbb{E}(X^1) = \mathbb{E}(X^2) = 1.5, \quad \text{Var}(X^1) = \text{Var}(X^2) = 0.75, \quad \rho(X^1, X^2) = 0.$$

Donc, pour la variable X^1 nous générons $n - 1$ valeurs d'une loi de Gauss (1.5, 0.75) et nous posons $X_n^1 = 55$. Puis nous générons n valeurs de la variable X^2 d'une loi de Gauss (1.5, 0.75). Ainsi, la corrélation entre X^1 et X^2 est zéro. De ce fait, la moyenne, la variance et la corrélation des variables X^1 et X^2 sont les mêmes que celles de la simulation d'Anderson-Legendre. Ensuite, nous construisons deux vecteurs $\mathbf{X}^1, \mathbf{X}^2$ et la matrice \mathbf{W} par l'équation :

$$\mathbf{W} = [\mathbf{X}^1 | \mathbf{X}^2] = \begin{bmatrix} X_1^1 & X_1^2 \\ X_2^1 & X_2^2 \\ \vdots & \vdots \\ X_n^1 & X_n^2 \end{bmatrix} \quad (2.28)$$

Enfin, les valeurs de Y sont générées par l'équation :

$$\mathbf{Y} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.29)$$

où : $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ est le vecteur des paramètres. Pour calculer le risque d'erreur de première espèce il faut choisir $\beta_2 = 0$ et $\boldsymbol{\epsilon}$ est le vecteur du terme d'erreur aléatoire généré selon la distribution considérée.

Pour chaque combinaison de facteurs, 10000 ensembles de données ont été générés. Pour chacune des deux méthodes, chaque ensemble de données a été permuté 999 fois. Les statistiques de test permutationnelle ($t_{2,\pi}$) de ces méthodes ont été calculées sur chaque ensemble de données permutoées. Ensuite, après 999 permutations, la p-valeur permutationnelle pour chacune des 2 méthodes a été calculée par formule :

$$p - \text{valeur}_\pi = \frac{\#(|t_\pi| \geq t_{obs})}{1000}$$

Les autre étapes sont identiques à celles de la section (2.5.4).

A2- Cas des variables X^1 et X^2 corrélées

D'après la remarque (2.4.2) on a :

$$\mathbb{E}(X^1) = \mathbb{E}(Z) = 1.5, \quad \mathbb{E}(X^2) = 3, \quad \text{Var}(X^1) = \text{Var}(Z) = 0.75$$

$$\text{Var}(X^2) = 1.5, \quad \rho(X^2, X^1) \approx 0.7$$

De ce fait, pour la variable Z^1 nous générons $n - 1$ valeurs d'une loi de Gauss (1.5, 0.75) et nous posons $Z_n^1 = 55$. Puis nous générons n valeurs de la variable Z^2

d'une loi de Gauss (3, 1.5). Ensuite, nous construisons deux vecteurs $\mathbf{Z}^1, \mathbf{Z}^2$ et la matrice \mathbf{W}' par l'équation :

$$\mathbf{W}' = [\mathbf{Z}^1 | \mathbf{Z}^2] = \begin{bmatrix} Z_1^1 & Z_1^2 \\ Z_2^1 & Z_2^2 \\ \vdots & \vdots \\ Z_n^1 & Z_n^2 \end{bmatrix} \quad (2.30)$$

Remarquons que $\rho(Z^1, Z^2) = 0$. A partir des Z^1 et Z^2 . On peut construire les variables X^1 et X^2 de corrélation $\rho(X^1, X^2) = 0.7$ de la façon suivante :

$$\mathbf{W}'_{\mathbf{R}} = [\mathbf{X}^1 | \mathbf{X}^2] = \mathbf{W}' \mathbf{R}^{1/2} \quad (2.31)$$

où : $\mathbf{R} = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}$. Enfin, les valeurs de Y sont générées par l'équation :

$$\mathbf{Y} = \mathbf{W}'_{\mathbf{R}} \boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.32)$$

où : $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ est le vecteur des paramètres (pour calculer le risque d'erreur de première espèce il faut mettre $\beta_2 = 0$) et $\boldsymbol{\epsilon}$ est le vecteur du terme d'erreur aléatoire généré selon la distribution considérée.

Les autre étapes sont identiques à celles du cas A1.

B- Les résultats de la simulation

B1- Cas des données générées pour le terme d'erreur selon une loi de Gauss :

1. Dans les figures 2.33-a,b où $n = 10$ (comme dans la section 2.5.1 où nous avons étudié le risque d'erreur de première espèce) pour toutes les valeurs de β_1 , le risque d'erreur de première espèce de la méthode de Kennedy est supérieur à celui de la méthode de Freedman-Lane.
2. En augmentant la taille de l'échantillon (comme dans la section 2.5.1 où nous avons étudié le risque d'erreur de première espèce) le risque d'erreur de première espèce de la méthode de Kennedy est proche de celui de la méthode de Freedman-Lane. Surtout, lorsque $n = 100$, les deux méthodes donnent presque le même risque qui est proche de 0.05.
3. En comparant les figures 2.33-c,e où $\rho = 0$, et 2.33-d,f où $\rho = 0.7$ on constate que la méthode de Kennedy comme la méthode de Freedman-Lane donne presque le même risque d'erreur de première espèce dans les deux cas. C'est-à-dire que même s'il existe un point extrême, la corrélation entre les variables X^1 et X^2 , n'a pas d'effet sur le risque d'erreur de première espèce.

4. Les risques d'erreur de première espèce des méthodes de Kennedy et de Freedman-Lane suivent la même évolution.
5. Dans tous les cas, lorsque $n > 10$, en présence d'un point extrême, la méthode de Kennedy comme la méthode de Freedman-Lane donne un risque proche de 0.05. C'est-à-dire que le point extrême n'a pas d'effet sur le risque d'erreur de première espèce de la méthode de Kennedy.
6. Les figures 2.33-c où $\beta_1 = 4$, 2.33-e où $\beta_1 = \{0, 6, 20\}$ et 2.33-f où $\beta_1 = 4$, montrent des cas où le risque d'erreur de première espèce de la méthode de Kennedy est inférieur à celui de la méthode de Freedman-Lane.
7. Comme dans la section 2.5.1 où nous avons étudié le risque d'erreur de première espèce, il n'y pas de différence significative entre les deux méthodes permutationnelles de Freedman-Lane, et de Kennedy et la méthode paramétrique classique. Ceci est acceptable puisque les données générées pour le terme d'erreur suivent une loi de Gauss $(0, 1)$.
8. Dans tous les cas, lorsque $n > 10$, même s'il existe un point extrême, la méthode de Kennedy comme la méthode de Freedman-Lane donne un risque proche de 0.05. C'est-à-dire que le point extrême n'a pas d'effet sur le risque d'erreur de première espèce de la méthode de Kennedy.

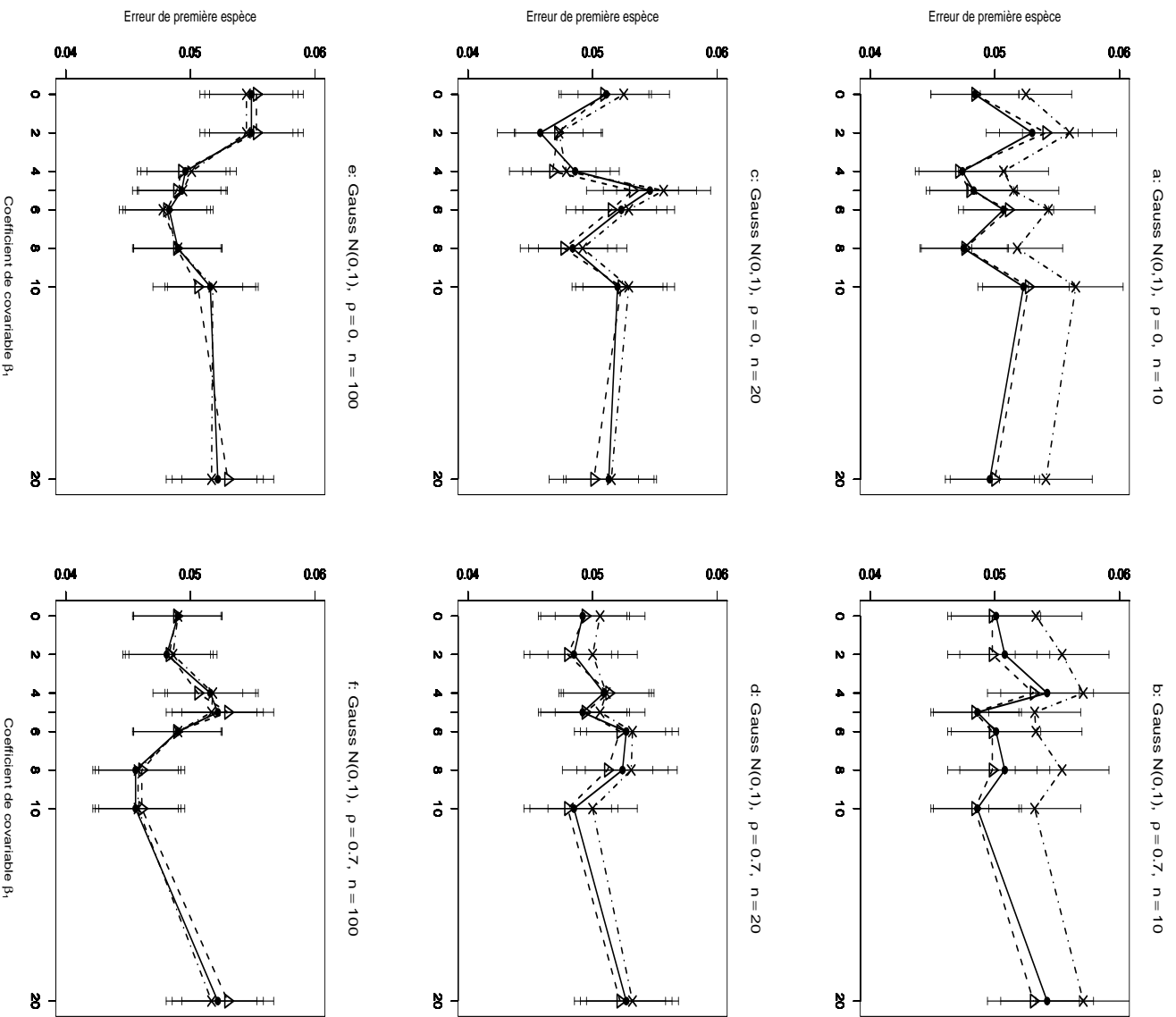


FIG. 2.33 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, Freedman et Lane— \bullet — et Kennedy— \times —, en fonction du coefficient de covariable. 10000 n-échantillons du résidu e ont été simulés suivant une loi de Gauss $(0,1)$, X^1 et X^2 sont défini dans le cadre A où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

Afin de comparer les effets de la valeur du coefficients β_1 , pour chaque méthode, nous avons considéré 48 risques d'erreur de première espèce (8 risques pour chacun des 6 cas a,b,...,f) des valeurs de β_1), dans le première groupe. Puis, pour chaque méthode, nous avons considéré 6 risque d'erreur de première espèce (pour les 6 cas a,b,...,f) où $\beta_1 = 0$, dans le deuxième groupe. Ensuite, pour chaque méthode nous avons considéré 42 risques d'erreur de première espèce (7 risques pour chacun des 6 cas a,b,...,f) où $\beta_1 \neq 0$, dans le troisième groupe. Enfin, pour chacune des méthodes nous avons calculé la moyenne et l'écart type, dans chacun des trois groupes :

$M_{\beta_1=0}$: La moyenne des risques d'erreur de première espèce pour $\beta_1 = 0$, les valeurs des autres facteurs restant fixés.

$SD_{\beta_1=0}$: La variance des risques d'erreur de première espèce pour $\beta_1 = 0$, les valeurs des autres facteurs restant fixés.

$M_{\beta_1 \neq 0}$: La moyenne des risques d'erreur de première espèce pour $\beta_1 \neq 0$, les valeurs des autres facteurs restant fixés.

$SD_{\beta_1 \neq 0}$: La variance des risques d'erreur de première espèce pour $\beta_1 \neq 0$, les valeurs des autres facteurs restant fixés.

M : La moyenne des risques d'erreur de première espèce sur toute les combinaisons des facteurs.

SD : La variance des risques d'erreur de première espèce sur toute les combinaisons des facteurs.

Gauss (0,1)	$M_{\beta_1=0}$	$SD_{\beta_1=0}$	$M_{\beta_1 \neq 0}$	$SD_{\beta_1 \neq 0}$	M	SD
Méthode Gaussienne	0.05045	0.00252	0.05010	0.00235	0.05014	0.00234
Méthode de Freedman et Lane	0.05045	0.00237	0.05031	0.00241	0.05033	0.00238
Méthode de Kennedy	0.05206	0.00196	0.05183	0.00289	0.05186	0.00278

TAB. 2.4 –

1. Pour $\beta_1 = 0$ la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche de 0.05 que celle de la méthode de Kennedy. En effet, on obtient une moyenne de 0.05206 pour la méthode de Kennedy et de 0.05045 pour la méthode de Freedman-Lane. Par contre, l'écart type du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.00237) est plus grand que celui de la méthode de Kennedy (0.00196).

2. Lorsque $\beta_1 \neq 0$, la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche de 0.05 que celle de la méthode de Kennedy. En effet, pour la méthode de Freedman-Lane elle est de 0.05031 et pour la méthode de Kennedy elle est de 0.05186. De plus, l'écart type du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus petit que celui de la méthode de Kennedy.

3. Pour toutes les valeurs de β_1 , la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche de 0.05 que celle de la méthode de Kennedy, c'est-à-dire que pour la méthode de Freedman-Lane est 0.05033 et pour la méthode de Kennedy elle est 0.05186. D'autre part, l'écart type du risque d'erreur

de première espèce de la méthode de Freedman-Lane est plus petit que celui de la méthode de Kennedy.

4. Pour $\beta_1 \neq 0$, la méthode paramétrique classique est plus efficace que les méthodes permutationnelles de Freedman-Lane, et de Kennedy, c'est-à-dire que la moyenne du risque d'erreur de première espèce de la méthode paramétrique classique est plus proche de 0.05 que celles des méthodes de Freedman-Lane et de Kennedy. De plus, son écart type est plus petit que celui des méthodes de Freedman-Lane et de Kennedy. Ceci est acceptable puisque les données générées pour le terme d'erreur suivent une loi de Gauss $(0, 1)$.

B2- Cas des données générées pour le terme d'erreur selon une loi symétrique de variance infinie comme la loi de Student à 1 degré de liberté :

1. Contrairement au cas précédent, dans les figures 2.34-a,b où $n = 10$, la méthode de Kennedy est une méthode conservatrice. En effet, les risques d'erreur de première espèce des méthodes de Kennedy et de Freedman-Lane ont la même évolution. Mais pour toutes les valeurs de β_1 , le risque d'erreur de première espèce de la méthode de Kennedy est inférieur à celui de la méthode de Freedman-Lane. Cela ne s'observe pas dans la section 2.5.1 où nous avons étudié le risque d'erreur de première espèce.
2. Dans la figure 2.34-a le risque d'erreur de première espèce de la méthode paramétrique classique est opposé à celui de la méthode de Kennedy. C'est-à-dire lorsque la méthode paramétrique classique donne un risque supérieur à 0.05, le risque de la méthode de Kennedy est inférieur à 0.05.
3. En augmentant la taille de l'échantillon, le risque d'erreur de première espèce de la méthode de Kennedy est proche de celui de la méthode de Freedman-Lane. Surtout, lorsque $n = 100$, les deux méthodes donnent presque le même risque qui est proche de 0.05.
4. Comme dans le cas précédent, en comparant les figures 2.34-c,e où $\rho = 0$, et 2.34-d,f où $\rho = 0.7$ on constate que dans les deux cas, la méthode de Kennedy donne presque le même risque d'erreur de première espèce. C'est-à-dire que même s'il existe un point extrême, la corrélation entre les variables X^1 et X^2 , n'a pas d'effet sur le risque d'erreur de première espèce de la méthode de Freedman-Lane.
5. Dans tous les cas, les risques d'erreur de première espèce des méthodes de Kennedy et de Freedman-Lane ont la même évolution.
6. Même lorsque la variance des données générées pour le terme d'erreur est infinie, il n'y a pas de différence significative entre les deux méthodes permutationnelles de Freedman-Lane, et de Kennedy, et la méthode paramétrique classique. Ceci peut-être dû au fait que les valeurs des variables X^1 et X^2 sont générées selon une loi de Gauss.
7. Dans tous les cas, lorsque $n > 10$, même s'il existe un point extrême, la méthode de Kennedy comme la méthode de Freedman-Lane donne un risque proche de 0.05. C'est-à-dire que le point extrême n'a pas d'effet sur le risque d'erreur de première espèce de la méthode de Kennedy.

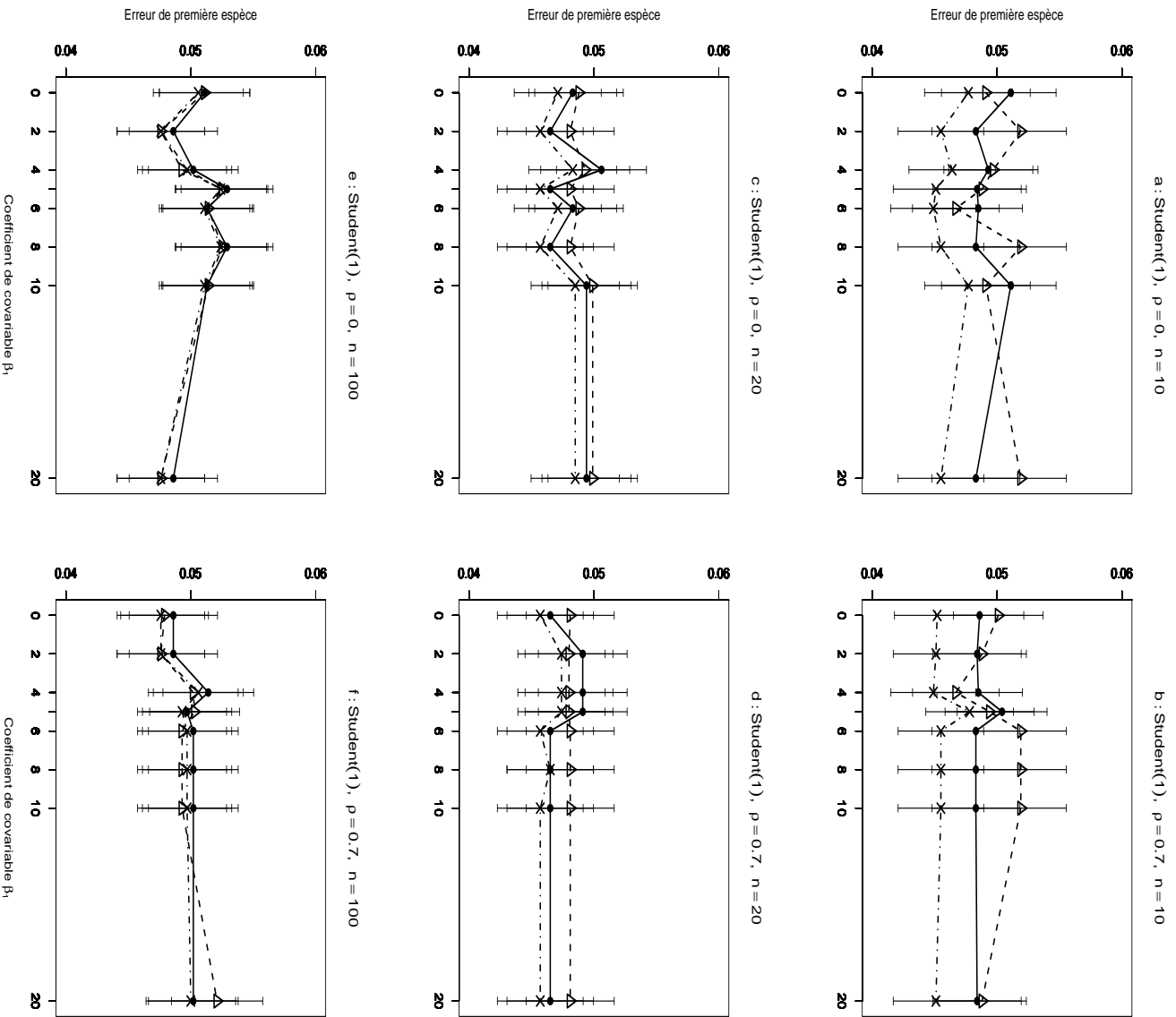


FIG. 2.34 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, Freedman et Lane — \bullet — et Kennedy — \times —, en fonction du coefficient de covariable. 10000 n-échantillons du résidu e ont été simulés suivant une loi de *Student*(1), X^1 et X^2 sont défini dans le cadre A où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

Comme dans le cas B1, afin de comparer les effets du coefficient β_1 , pour chaque méthode, nous avons calculé la moyenne et l'écart type dans chacun des trois groupes définis :

<i>Student</i> (1)	$M_{\beta_1=0}$	$SD_{\beta_1=0}$	$M_{\beta_1 \neq 0}$	$SD_{\beta_1 \neq 0}$	M	SD
Méthode Gaussienne	0.04918	0.00122	0.04948	0.00169	0.04944	0.00163
Méthode de Freedman et Lane	0.04903	0.00178	0.04903	0.00166	0.04903	0.00165
Méthode de Kennedy	0.04731	0.00190	0.04747	0.00216	0.04745	0.00214

TAB. 2.5 –

1. Comme dans le cas précédent, pour $\beta_1 = 0$, la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche de 0.05 que celle de la méthode de Kennedy. Contrairement au cas précédent, ici, la méthode de Kennedy avec une moyenne de 0.04731 est plus conservatrice que la méthode de Freedman-Lane avec une moyenne de 0.04903. D'autre part, l'écart type du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.00178) est plus petit que celui de la méthode de Kennedy (0.00190).
2. Lorsque $\beta_1 \neq 0$, la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche de 0.05 que celle de la méthode de Kennedy. En effet pour la méthode de Freedman-Lane est 0.04903 et pour la méthode de Kennedy elle est 0.04747, c'est-à-dire que là encore, la méthode de Kennedy est plus conservatrice. De plus l'écart type du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.00166) est plus petite que celui de la méthode de Kennedy.
3. Pour toutes les valeurs de β_1 , la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane est plus proche de 0.05 que celle de la méthode de Kennedy, c'est-à-dire que pour la méthode de Freedman-Lane elle est 0.04903 et pour la méthode de Kennedy est 0.04745. Par ailleurs, l'écart type du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.00165) est plus petit que celui de la méthode de Kennedy (0.00214).
4. Pour $\beta_1 \neq 0$, la méthode paramétrique classique, est plus efficace que les méthodes permutationnelles de Freedman-Lane et de Kennedy. C'est-à-dire que la moyenne du risque d'erreur de première espèce de la méthode paramétrique classique (0.04948) est plus proche de 0.05 que celles des méthodes de Freedman-Lane (0.04903) et de Kennedy (0.04747). Par ailleurs, l'écart type du risque d'erreur de première espèce de la méthode paramétrique classique (0.00169) est plus petit que celui de la méthode de Kennedy (0.00216), et il est un peu plus grand que celui de la méthode de Freedman-Lane (0.00166).

B3- Cas des données générées pour le terme d'erreur selon une loi radicalement à dissymétrie droite comme la loi $Exp(2)$:

1. Comme dans le cas B1 et contrairement au cas précédent, dans les figure 2.35-a,b pour toutes les valeurs de β_1 , le risque d'erreur de première espèce de la méthode de Kennedy est supérieur à celui de la méthode de Freedman-Lane.

2. Comme dans les cas B1 et B2, en augmentant la taille d'échantillon le risque d'erreur de première espèce de la méthode de Kennedy est proche de celui de la méthode de Freedman-Lane. Surtout, lorsque $n = 100$, les deux méthodes donnent presque le même risque qui est proche de 0.05.

3. Comme dans le cas précédent, en comparant les figures 2.35-c,e où $\rho = 0$, et 2.35-d,f où $\rho = 0.7$ on constate que dans les deux cas la méthode de Kennedy donne presque le même risque d'erreur de première espèce. C'est-à-dire que même s'il existe un point extrême, la corrélation entre les variables X^1 et X^2 , n'a pas d'effet sur le risque d'erreur de première espèce de la méthode de Freedman-Lane.

4. Comme dans les cas précédents, les risques d'erreur de première espèce des méthodes de Kennedy et de Freedman-Lane ont la même évolution.

5. Contrairement à la simulation d'Anderson-Lane, ici, la méthode paramétrique classique est une méthode stable, et donne un risque d'erreur de première espèce proche de 0.05. C'est-à-dire qu'il n'y pas de différence significative entre les deux méthodes permutationnelles de Freedman-Lane et de Kennedy, et la méthode paramétrique classique. Ceci peut-être à cause des valeurs des variables X^1 et X^2 générées selon une loi de Gauss.

6. Dans tous les cas, lorsque $n > 10$, même s'il existe un point extrême, la méthode de Kennedy comme la méthode de Freedman-Lane donne un risque proche de 0.05. C'est-à-dire un point extrême n'a pas d'effet sur le risque d'erreur de première espèce de la méthode de Kennedy.

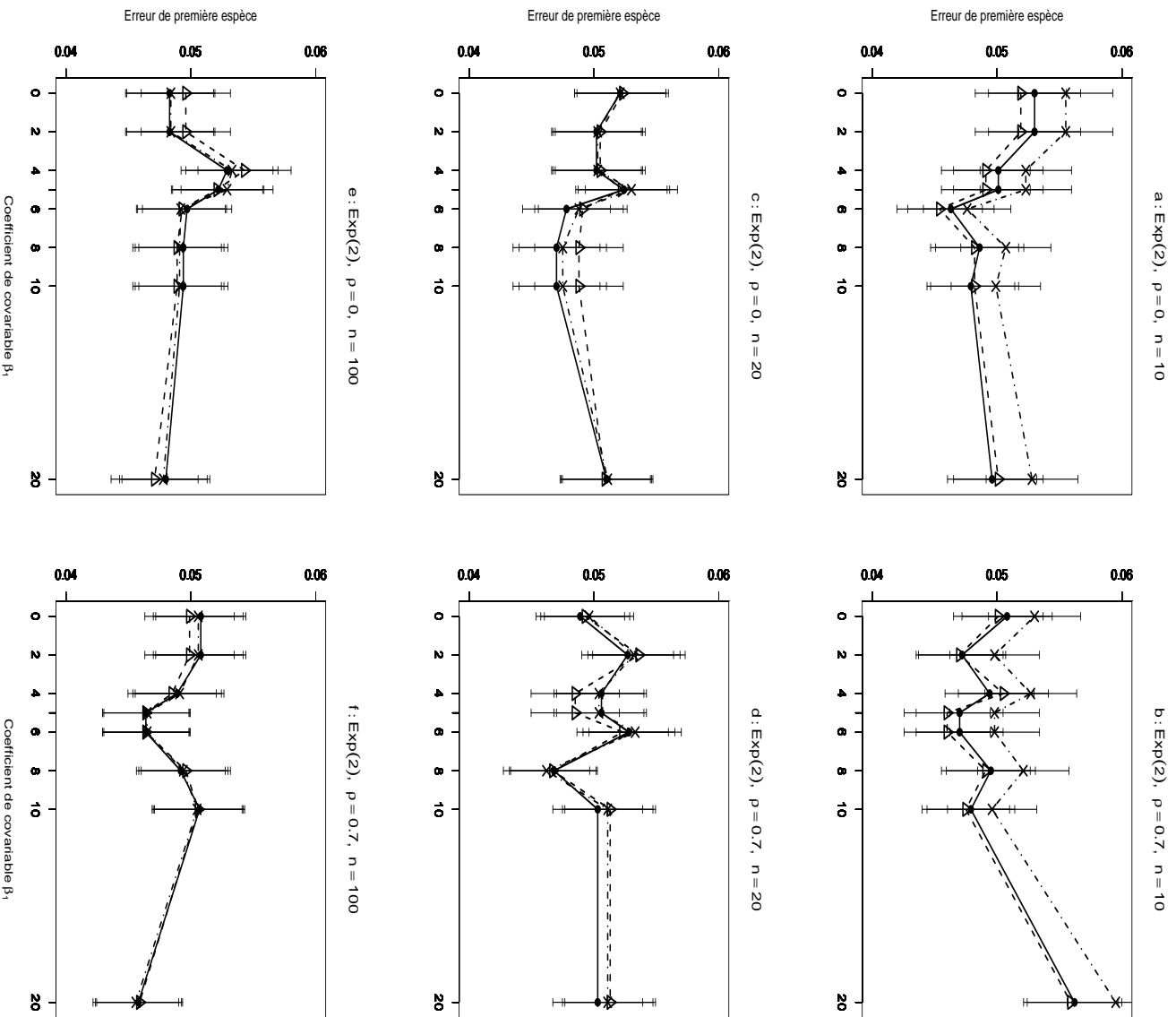


FIG. 2.35 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, Freedman et Lane — \bullet — et Kennedy — \ast —, en fonction du coefficient de covariable. 10000 n-échantillons du résidu e ont été simulés suivant une loi $Exp(2)$, X^1 et X^2 sont défini dans le cadre A où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

Afin de comparer les effets des coefficients β_1 , pour chaque méthode, nous avons calculé la moyenne et l'écart type sur les 3 groupes :

Exp(2)	$M_{\beta_1=0}$	$SD_{\beta_1=0}$	$M_{\beta_1 \neq 0}$	$SD_{\beta_1 \neq 0}$	M	SD
Méthode Gaussienne	0.05051	0.00126	0.04939	0.00235	0.04953	0.00226
Méthode de Freedman et Lane	0.05065	0.00180	0.04946	0.00226	0.04961	0.00222
Méthode de Kennedy	0.05153	0.00255	0.05039	0.00264	0.05053	0.00263

TAB. 2.6 –

1. Pour $\beta_1 = 0$ la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.05065) est plus proche de 0.05 que celle de la méthode de Kennedy (0.05153). En effet pour la méthode de Kennedy elle est plus grande que 0.05 et que celle de la méthode de Freedman-Lane. D'autre part, l'écart type du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.00180) est plus petite que celui de la méthode de Kennedy (0.00255).

2. Lorsque $\beta_1 \neq 0$, la moyenne celui du risque d'erreur de première espèce de la méthode de Kennedy (0.05039) est plus proche de 0.05 que celle de la méthode de Freedman-Lane (0.04946). De plus l'écart type du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.00226) est plus petit que celui de la méthode de Kennedy (0.00264).

3. Pour toutes les valeurs de β_1 , la méthode de Freedman-Lane est plus efficace que la méthode de Kennedy. C'est-à-dire que la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.04961) est plus proche de 0.05 que celle de la méthode de Kennedy (0.05053), par ailleurs, l'écart type du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.00222) est plus petit que celui de la méthode de Kennedy (0.0263).

4. Pour $\beta_1 \neq 0$, la méthode Freedman-Lane est plus efficace que la méthode paramétrique classique. C'est-à-dire, la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.04946) est plus proche de 0.05 que celle de la méthodes paramétrique classique (0.04939). D'autre part, l'écart type du risque d'erreur de première espèce de la méthode Freedman-Lane (0.00226) est plus petit que celui de la méthode de paramétrique classique (0.00235).

De plus, la moyenne du risque d'erreur de première espèce de la méthode de Kennedy (0.05039) est plus proche de 0.05 que celle de la méthode paramétrique classique (0.04939), mais l'écart type du risque d'erreur de première espèce de la méthode paramétrique classique (0.00235) est plus petit que celui de la méthode de Kennedy (0.00264).

B-4 Cas des données générées pour le terme d'erreur selon une loi radicalement à dissymétrie gauche comme la loi $\Gamma(10, 0.5)$:

On peut trouver les résultats obtenus dans le cas B1 et B3.

Ici, comme les cas précédents, pour comparer les effets des coefficients β_1 , pour chaque méthode, nous avons calculé la moyenne et l'écart type sur les trois groupes :

$\Gamma(10, 0.5)$	$M_{\beta_1=0}$	$SD_{\beta_1=0}$	$M_{\beta_1 \neq 0}$	$SD_{\beta_1 \neq 0}$	M	SD
Méthode de Gaussienne	0.04955	0.00186	0.04991	0.00218	0.04986	0.00213
Méthode de Freedman et Lane	0.04935	0.00197	0.04963	0.00214	0.04960	0.00210
Méthode de Kennedy	0.05091	0.00257	0.05108	0.00214	0.05106	0.00217

TAB. 2.7 –

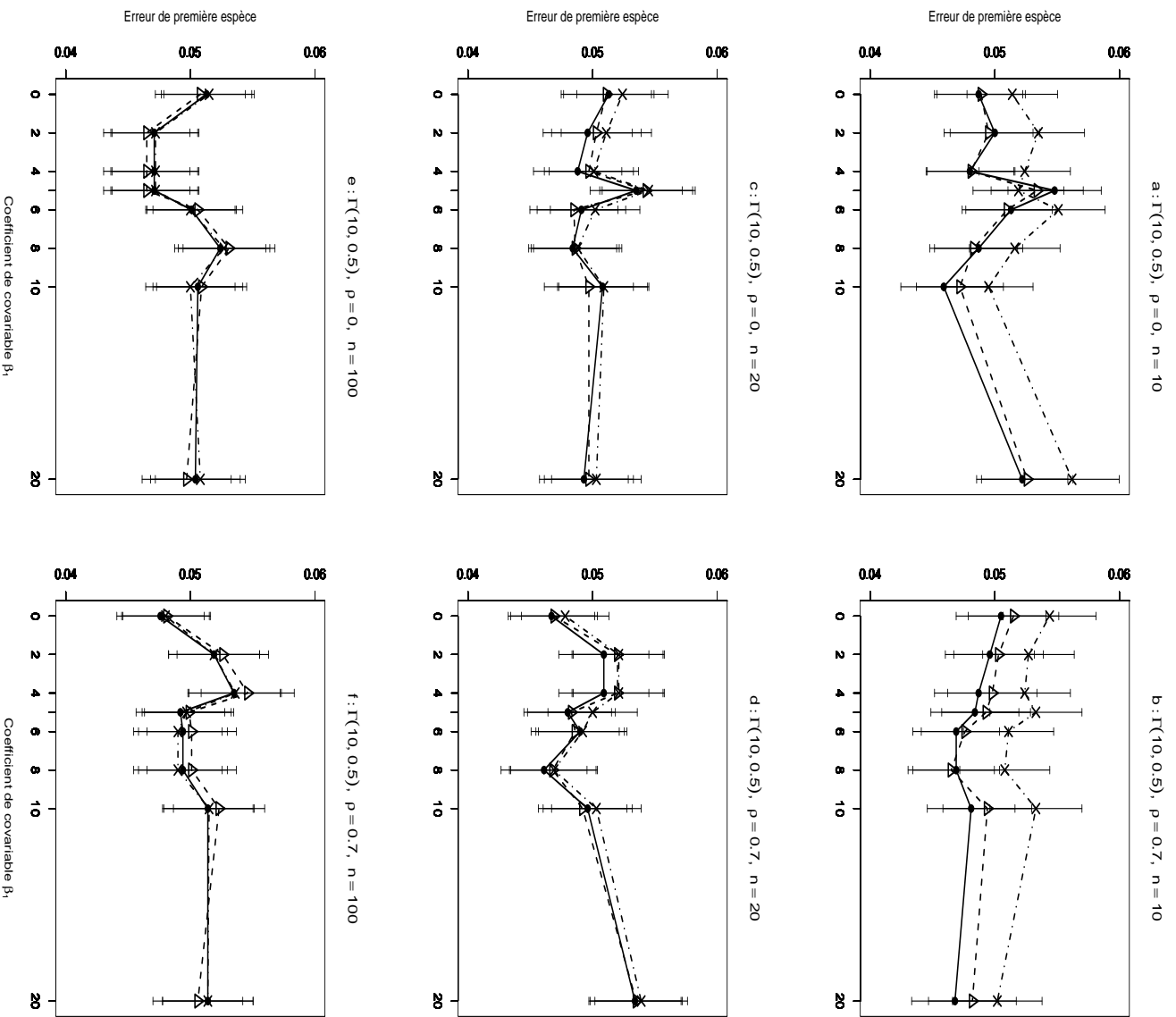


FIG. 2.36 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, Freedman et Lane — \bullet — et Kennedy — \times —, en fonction du coefficient de covariable. 10000 n-échantillons du résidu e ont été simulés suivant une loi $\Gamma(10, 0.5)$, X^1 et X^2 sont défini dans le cadre A où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

C- Résumé des principaux résultats

1. Comme dans la section 2.5.1 où nous avons étudié le risque d'erreur de première espèce, lorsque la taille de l'échantillon est petite $n = 10$ (sauf le cas B2), pour toutes les valeurs de β_1 , le risque d'erreur de première espèce de la méthode de Kennedy est supérieur à celui de la méthode de Freedman-Lane.
2. Comme dans la simulation d'Anderson-Legendre, la méthode de Freedman-Lane est une méthode stable, c'est-à-dire pour toutes les valeurs de β_1 , avec $\beta_1 = 0$ ou $\beta_1 \neq 0$ donnent un risque d'erreur de première espèce proche de 0.05. En effet, un point extrême n'a pas d'effet sur le risque d'erreur de première espèce.
3. En augmentant la taille de l'échantillon (comme dans la section 2.5.1 où nous avons étudié le risque d'erreur de première espèce) les deux méthodes permutationnelle de Freedman-Lane et de Kennedy donnent presque le même risque d'erreur de première espèce qui est proche de 0.05, c'est-à-dire qu'un point extrême n'a pas d'effet sur le risque d'erreur de première espèce de la méthode de Kennedy. Donc pour $n > 10$, la méthode de Kennedy est une méthode stable.
4. Même s'il existe un point extrême, la corrélation entre les variables X^1 et X^2 , n'a pas d'effet sur le risque d'erreur de première espèce des méthodes de Freedman-Lane et de Kennedy.
5. Même s'il existe un point extrême, le degré de corrélation β_1 n'a pas d'effet sur le risque d'erreur de première espèce des méthodes de Freedman-Lane et de Kennedy.
6. Dans tous les cas, les risques d'erreur de première espèce des méthodes de Kennedy et de Freedman-Lane ont la même évolution.
7. Lorsque les données générées pour le terme d'erreur suivent une loi de *Student*(1) avec $n = 10$, contrairement aux autres cas, la méthode de Kennedy est une méthode conservatrice. En effet, les risques d'erreur de première espèce des méthodes de Kennedy et de Freedman-Lane ont la même évolution. Mais pour toutes les valeurs de β_1 , le risque d'erreur de première espèce de la méthode de Kennedy est inférieur à celui de la méthode de Freedman-Lane, cela ne s'observe pas dans la section 2.5.1 où nous avons étudié le risque d'erreur de première espèce.
8. Comme dans la section 2.5.1 où nous avons étudié le risque d'erreur de première espèce et nous avons observé des cas où le risque d'erreur de première espèce de la méthode de Kennedy est inférieur à celui de la méthode de Freedman-Lane, ici aussi, on observe des cas où le risque d'erreur de première espèce de la méthode de Kennedy est inférieur à celui de la méthode de Freedman-Lane.
9. Il n'y a pas de différence significative entre les deux méthodes permutationnelles de Freedman-Lane et de Kennedy et la méthode paramétrique classique. Lorsque les données générées pour le terme d'erreur suivent une loi de Gauss

(0, 1), ceci est acceptable. Mais dans d'autre cas, ceci peut-être à cause des valeurs des variables X^1 et X^2 générées selon une loi de Gauss.

10. Dans tous les cas, même s'il existe un point extrême, la méthode de Freedman-Lane est plus stable que la méthode de Kennedy puisque pour $n = 10$, la méthode de Kennedy est instable. C'est-à-dire tantôt un risque est supérieur à 0.05, et tantôt inférieur à 0.05.

Afin d'avoir un bilan, nous avons calculé pour chaque méthode la moyenne et l'écart type dans les trois groupes mais en confondant les résultats obtenus pour les quatre lois :

sur toute les quatre loi	$M_{\beta_1=0}$	$SD_{\beta_1=0}$	$M_{\beta_1 \neq 0}$	$SD_{\beta_1 \neq 0}$	M	SD
Méthode Gaussienne	0.04992	0.00177	0.04972	0.00216	0.04974	0.00211
Méthode de Freedman et Lane	0.04987	0.00199	0.04961	0.00216	0.04964	0.00214
Méthode de Kennedy	0.05045	0.00284	0.05019	0.00296	0.05022	0.00294

TAB. 2.8 –

1. Lorsque $\beta_1 = 0$ la moyenne du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.04987) et plus proche de 0.05 que celle de la méthode de Kennedy (0.05045). D'autre part, l'écart type du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.00199) est plus petit que celui de la méthode de Kennedy (0.00284).

2. Pour $\beta_1 \neq 0$, la moyenne du risque d'erreur de première espèce de la méthode de Kennedy (0.05019) et plus proche de 0.05 que celle de la méthode de Freedman-Lane (0.04961), mais l'écart type du risque d'erreur de première espèce de la méthode de Freedman-Lane (0.00216) est plus petit que celui de la méthode de Kennedy (0.00296).

3. Lorsque $\beta_1 \neq 0$, la méthode paramétrique classique est plus efficace que la méthode de Freedman-Lane. C'est-à-dire que, la moyenne du risque d'erreur de première espèce de la méthode paramétrique classique (0.04972) est plus proche de 0.05 que celle de la méthode de Kennedy (0.05019), par ailleurs, les écart type sont égaux (0.00216).

D'autre part, la moyenne du risque d'erreur de première espèce de la méthode de Kennedy (0.05019) est plus proche de 0.05 que celle de la méthode paramétrique classique (0.04972), mais l'écart type du risque d'erreur de première espèce de la méthode paramétrique classique (0.00216) est plus petit que celui de la méthode de Kennedy (0.00296).

Chapitre 3

Résultats théoriques

3.1 Introduction

Les simulations, présentées dans le chapitre précédent montrent pour chaque combinaison des facteurs retenus, que le risque d'erreur de première espèce encouru par la méthode de Freedman-Lane est différent de celui de la méthode de Kennedy. Dans la plupart des cas le risque d'erreur de première espèce de la méthode de Freedman-Lane est inférieur à celui de Kennedy. Mais, dans les simulations d'Anderson et Legendre, le risque d'erreur de première espèce de la méthode de Freedman et Lane est toujours inférieur à celui de la méthode de Kennedy. Toutes deux utilisent un principe de test de Student. D'un point de vue formel, les statistiques de tests permutacionnelles utilisées par ces deux méthodes sont différentes :

$$T_{2,\pi}^F = \frac{\hat{\beta}_{2,\pi}^F}{S(\hat{\beta}_{2,\pi}^F)} \neq T_{2,\pi}^K = \frac{\hat{\beta}_{2,\pi}^K}{S(\hat{\beta}_{2,\pi}^K)} \quad (3.1)$$

où $S^2(\hat{\beta}_{2,\pi}^F)$ et $S^2(\hat{\beta}_{2,\pi}^K)$ sont respectivement les estimateurs MCO de $\text{Var}(\hat{\beta}_{2,\pi}^F)$ et de $\text{Var}(\hat{\beta}_{2,\pi}^K)$. Anderson et Legendre (1999) précisent : *La raison de la différence entre les deux méthodes est que la méthode de Kennedy en régressant Y sur X^1 selon le modèle (3.3), enlève l'effet de la variable X^1 qui n'a pas d'intérêt pour effectuer le test et donc au cours des permutations le coefficient de régression associé à X^1 reste fixe :*

$$R^{Y,1,\pi i} = \beta_{2,\pi} R_i^{2,1} + \epsilon_i^K \quad (3.2)$$

Par contre, bien qu'elle enlève l'effet de la variable X^1 comme la méthode de Kennedy, la méthode de Freedman-Lane, en maintenant la variable X^1 dans le modèle de régression :

$$Y^{\pi,F} = \beta_{0\pi}^F + \beta_{1\pi}^F x_i^1 + \beta_{2\pi}^F x_i^2 + \epsilon_i^F. \quad (3.3)$$

prend en considération l'effet de la variable X^1 au cours des permutations.

Cette argumentation nous semble ignorer un autre aspect très important de la question : Les simulations, aussi rigoureuses soient-elles, n'ont pas force de preuve. Aussi, ce chapitre est-il consacré à l'étude mathématique des causes des différences observées par simulation. En particulier, nous présentons dans ce chapitre, une étude analytique des variances estimées des statistiques de test retenues pour ces simulations, de façon à démontrer formellement le bien fondé des observations empiriques présentées aux chapitres précédents. En particulier :

- a) Nous démontrons que les estimateurs de β_2 issus des deux méthodes analysées sont égaux. C'est-à-dire : $\hat{\beta}_{2,\pi}^F = \hat{\beta}_{2,\pi}^K$ et cet estimateur commun est désormais noté $\hat{\beta}_{2,\pi}$.
- b) Nous montrons que $S^2(\hat{\beta}_{2,\pi}^F) < S^2(\hat{\beta}_{2,\pi}^K)$ où $S^2(\hat{\beta}_{2,\pi}^F)$ et $S^2(\hat{\beta}_{2,\pi}^K)$ sont les estimateurs MCO respectifs des variances de $\hat{\beta}_{2,\pi}$ dans les modèles de Freedman-Lane et de Kennedy. On en déduit que la fonction de puissance de la méthode de Freedman-Lane est toujours plus grande que celle de la méthode de Kennedy.
- c) Nous calculons la variance permutationnelle de $\hat{\beta}_{2,\pi}$, qui est notée $\mathcal{V}ar(\hat{\beta}_{2,\pi})$. De plus, lorsque \mathcal{E} est la notation de l'espérance permutationnelle, et $S^2(\hat{\beta}_{2,\pi})_{K*}$ est l'estimateur de la variance de $\hat{\beta}_{2,\pi}$ où la somme des carrés des résidus est divisée par $(n-2)$ et n est le nombre des observations, nous montrerons que :
 - d) $\mathcal{E}(\widehat{\mathcal{V}ar}(\hat{\beta}_{2,\pi})_F) = \mathcal{V}ar(\hat{\beta}_{2,\pi})$
 - e) $\mathcal{E}(\widehat{S^2}(\hat{\beta}_{2,\pi})_K) \neq \mathcal{V}ar(\hat{\beta}_{2,\pi})$
 - f) $\lim_{n \rightarrow +\infty} \mathcal{E}(\widehat{\mathcal{V}ar}(\hat{\beta}_{2,\pi})_K) = \lim_{n \rightarrow +\infty} \mathcal{V}ar(\hat{\beta}_{2,\pi})$.
 - g) $S^2(\widehat{\hat{\beta}}_{2,\pi})_K \geq S^2(\hat{\beta}_{2,\pi})_{K*}$
 - h) $\mathcal{E}(\widehat{\mathcal{V}ar}(\hat{\beta}_{2,\pi})_{K*}) = \mathcal{V}ar(\hat{\beta}_{2,\pi})$

Ensuite, nous avons étendu au cas du modèle général de régression linéaire multiple (où on a $p > 2$ prédicteurs) les résultats que nous avons établis dans le cas du modèle de régression double.

3.2 Estimateurs MCO de $\beta_{2,\pi}^F$ et $\beta_{2,\pi}^K$

3.2.1 Cas du modèle de Freedman et Lane

Afin de calculer les estimateurs MCO des coefficients d'un modèle nous adoptons les notations suivantes :

définition 3.2.1.

1. $S^2(X^j) = S(X^j, X^j) = \frac{1}{n} \sum_i^n (x_i^j - \bar{x}^j)^2, \quad j = 1, 2$
2. $S(X^1, X^2) = \frac{1}{n} \sum_i^n (x_i^1 - \bar{x}^1)(x_i^2 - \bar{x}^2)$

$$3. S(X^j, Y_\pi^F) = \frac{1}{n} \sum_i^n (x_i^j - \bar{x}^j)(Y_{\pi i}^F - \bar{Y}_\pi^F), \quad j = 1, 2$$

$$4. a_1 = \frac{S(X^1, Y)}{S^2(X^1)}$$

Lemme 3.2.1. *Pour le modèle de Freeman-Lane.*

$$Y_{\pi i}^F = \beta_{0,\pi}^F + \beta_{1,\pi}^F x_i^1 + \beta_{2,\pi}^F x_i^2 + \epsilon_i^F. \quad (3.4)$$

sous les postulats P1-P4, on vérifie les identités suivantes :

1. $\bar{Y}_\pi^F = \bar{Y}$
2. $Y_{\pi i}^F - \bar{Y}_\pi^F = a_1(x_i^1 - \bar{x}^1) + R^{Y,1,\pi}$
3. $S(X^j, Y_\pi^F) = a_1 S(X^1, X^j) + S(X^j, R^{Y,1,\pi}) \quad j = 1, 2$

1. Par définition,

$$Y_{\pi i}^F = \hat{Y}_{1i} + R^{Y,1,\pi i} = a_0 + a_1 x_i^1 + R^{Y,1,\pi i} \quad (3.5)$$

et par ailleurs $a_0 = \bar{Y} - a_1 \bar{x}^1$, donc :

$$Y_{\pi i}^F = \hat{Y} + a_1(x_i^1 - \bar{x}^1) + R^{Y,1,\pi i}. \quad (3.6)$$

Par suite :

$$\begin{aligned} \bar{Y}_\pi^F &= \frac{1}{n} \sum_{i=1}^n Y_{\pi i}^F \\ &= \bar{Y} + \frac{1}{n} \sum_{i=1}^n a_1(x_i^1 - \bar{x}^1) + \frac{1}{n} \sum_{i=1}^n R^{Y,1,\pi i} \\ &= \bar{Y} + a_1 \frac{1}{n} \sum_{i=1}^n (x_i^1 - \bar{x}^1) + \frac{1}{n} \sum_{i=1}^n R^{Y,1,\pi i} \\ &= \bar{Y} + a_1 0 + 0 = \bar{Y} \end{aligned}$$

2. On sait que $Y_{\pi i}^F - \bar{Y}_\pi^F = \bar{Y} - a_1(x_i^1 - \bar{x}^1) + R^{Y,1,\pi i} - \bar{Y}_\pi^F$. D'après la propriété (1) du lemme on a donc :

$$\begin{aligned} Y_{\pi i}^F - \bar{Y}_\pi^F &= \bar{Y} + a_1(x_i^1 - \bar{x}^1) + R^{Y,1,\pi i} - \bar{Y} \\ &= a_1(x_i^1 - \bar{x}^1) + R^{Y,1,\pi i} \end{aligned} \quad (3.7)$$

3. D'après les définitions (3.2.1) on a :

$$\begin{aligned}
S(X^j, Y_\pi^F) &= \frac{1}{n} \sum_i^n (x_i^j - \bar{x}^j)(Y_{\pi i}^F - \bar{Y}_\pi^F) \\
&= \frac{1}{n} \sum_i^n (x_i^j - \bar{x}^j)(a_1(x_i^1 - \bar{x}^1) + R^{Y.1, \pi i}) \quad \text{du fait de la propriété (2)} \\
&= a_1 \frac{1}{n} \sum_i^n (x_i^j - \bar{x}^j)(x_i^1 - \bar{x}^1) + \frac{1}{n} \sum_i^n (x_i^j - \bar{x}^j) R^{Y.1, \pi i} \\
&= a_1 S(X^j, X^1) + S(X^j, R^{Y.1, \pi}) \quad \square
\end{aligned} \tag{3.8}$$

La formulation de la proposition suivante nécessite l'introduction des matrices $\mathbf{C}^{\mathbf{Y}_\pi^F}$ et $\mathbf{C}^{\mathbf{R}^{Y.1, \pi}}$ définies comme suit :

$$\mathbf{C}^{\mathbf{Y}_\pi^F} = \begin{bmatrix} S^2(X^1) & S(X^1, X^2) & S(X^1, Y_\pi^F) \\ S(X^1, X^2) & S^2(X^2) & S(X^2, Y_\pi^F) \\ S(X^1, Y_\pi^F) & S(X^2, Y_\pi^F) & S^2(Y_\pi^F) \end{bmatrix}$$

$\mathbf{C}^{\mathbf{Y}_\pi^F}$ est la matrice de variance covariance empirique du vecteur (X^1, X^2, Y_π^F) . De façon analogue pour $(X^1, X^2, R^{Y.1, \pi})$ on a :

$$\mathbf{C}^{\mathbf{R}^{Y.1, \pi}} = \begin{bmatrix} S^2(X^1) & S(X^1, X^2) & S(X^1, R^{Y.1, \pi}) \\ S(X^1, X^2) & S^2(X^2) & S(X^2, R^{Y.1, \pi}) \\ S(X^1, R^{Y.1, \pi}) & S(X^2, R^{Y.1, \pi}) & S^2(R^{Y.1, \pi}) \end{bmatrix}$$

Les sous matrices $\mathbf{C}_{i,j}^{\mathbf{Y}_\pi^F}$ et $\mathbf{C}_{i,j}^{\mathbf{R}^{Y.1, \pi}}$ pour $i, j = 1, 2, 3$, sont obtenues respectivement à partir des matrices $\mathbf{C}^{\mathbf{Y}_\pi^F}$ et $\mathbf{C}^{\mathbf{R}^{Y.1, \pi}}$ en éliminant la ligne i et la colonne j . Par exemple les sous matrices $\mathbf{C}_{1,3}^{\mathbf{Y}_\pi^F}$, $\mathbf{C}_{2,3}^{\mathbf{Y}_\pi^F}$, ainsi que $\mathbf{C}_{1,3}^{\mathbf{R}^{Y.1, \pi}}$, $\mathbf{C}_{2,3}^{\mathbf{R}^{Y.1, \pi}}$ sont :

$$\begin{aligned}
\mathbf{C}_{1,3}^{\mathbf{Y}_\pi^F} &= \begin{bmatrix} S(X^1, X^2) & S^2(X^2) \\ S(X^1, Y_\pi^F) & S(X^2, Y_\pi^F) \end{bmatrix}; \quad \mathbf{C}_{1,3}^{\mathbf{R}^{Y.1, \pi}} = \begin{bmatrix} S(X^1, X^2) & S^2(X^2) \\ S(X^1, R^{Y.1, \pi}) & S(X^2, R^{Y.1, \pi}) \end{bmatrix} \\
\mathbf{C}_{2,3}^{\mathbf{Y}_\pi^F} &= \begin{bmatrix} S^2(X^1) & S(X^1, Y_\pi^F) \\ S(X^1, X^2) & S(X^2, Y_\pi^F) \end{bmatrix}; \quad \mathbf{C}_{2,3}^{\mathbf{R}^{Y.1, \pi}} = \begin{bmatrix} S^2(X^1) & S(X^1, R^{Y.1, \pi}) \\ S(X^1, X^2) & S(X^2, R^{Y.1, \pi}) \end{bmatrix}
\end{aligned}$$

On note de plus :

$$\Delta = \det(\mathbf{C}_{3,3}^{\mathbf{Y}_\pi^F}) = \det(\mathbf{C}_{3,3}^{\mathbf{R}^{Y.1, \pi}}) = \begin{vmatrix} S^2(X^1) & S(X^1, X^2) \\ S(X^1, X^2) & S(X^2) \end{vmatrix}$$

Proposition 3.2.1. *Considérons le modèle de Freedman et Lane (3.3) :*

$$Y_{\pi i}^F = \beta_{0, \pi}^F + \beta_{1, \pi}^F x_i^1 + \beta_{2, \pi}^F x_i^2 + \epsilon^F$$

les estimateurs MCO pour les coefficients de régression sont :

$$\begin{aligned}
1. \quad \widehat{\beta}_{1,\pi}^F &= a_1 + \frac{\Delta_1}{\Delta} = a_1 - \frac{\det(\mathbf{C}_{1,3}^{\mathbf{R}^{Y \cdot 1, \pi}})}{\det(\mathbf{C}_{3,3}^{\mathbf{R}^{Y \cdot 1, \pi}})} \\
2. \quad \widehat{\beta}_{2,\pi}^F &= \frac{\Delta_2}{\Delta} = \frac{\det(\mathbf{C}_{2,3}^{\mathbf{R}^{Y \cdot 1, \pi}})}{\det(\mathbf{C}_{3,3}^{\mathbf{R}^{Y \cdot 1, \pi}})} \\
3. \quad \widehat{\beta}_{0,\pi}^F &= a_0 - \frac{\Delta_1}{\Delta} \bar{x}^1 - \frac{\Delta_2}{\Delta} \bar{x}^2
\end{aligned}$$

Preuve :

1. L'estimateur MCO pour le coefficient de régression $\beta_{1,\pi}^F$ est (cf. Johnston et Dinardo [30]) :

$$\widehat{\beta}_{1,\pi}^F = \frac{S^2(X^2)S(X^1, Y_\pi^F) - S(X^2, Y_\pi^F)S(X^1, X^2)}{S^2(X^1)S^2(X^2) - S(X^1, X^2)}$$

La définition de Δ permet d'obtenir :

$$\Delta \cdot \widehat{\beta}_{1,\pi}^F = S^2(X^2)S(X^1, Y_\pi^F) - S(X^2, Y_\pi^F)S(X^1, X^2).$$

D'après le lemme 3.2.1, on obtient par substitution de $S(X^1, Y_\pi^F)$, $S(X^2, Y_\pi^F)$:

$$\begin{aligned}
\Delta \cdot \widehat{\beta}_{1,\pi}^F &= S^2(X^2)(a_1 S^2(X^1) + S(X^1, R^{Y \cdot 1, \pi})) & (3.9) \\
&- (a_1 S(X^1, X^2) + S(X^2, R^{Y \cdot 1, \pi}))S(X^1, X^2) \\
&= a_1 S^2(X^2)S^2(X^1) + S^2(X^2)S(X^1, R^{Y \cdot 1, \pi}) \\
&- a_1 S^2(X^1, X^2) - S(X^2, R^{Y \cdot 1, \pi})S(X^1, X^2) \\
&= a_1 (S^2(X^2)S^2(X^1) - S^2(X^1, X^2)) \\
&+ S^2(X^2)S(X^1, R^{Y \cdot 1, \pi}) - S(X^2, R^{Y \cdot 1, \pi})S(X^1, X^2) \\
&= a_1 \Delta + \Delta_1
\end{aligned}$$

c'es-à-dire :

$$\widehat{\beta}_{1,\pi}^F = a_1 + \frac{\Delta_1}{\Delta} = a_1 - \frac{\det(\mathbf{C}_{1,3}^{\mathbf{R}^{Y \cdot 1, \pi}})}{\det(\mathbf{C}_{1,2}^{\mathbf{R}^{Y \cdot 1, \pi}})} \quad (3.10)$$

2. De la même façon, l'estimateur MCO pour le coefficient de régression de $\widehat{\beta}_{2,\pi}^F$ est (cf. Johnston et Dinardo [30]) :

$$\widehat{\beta}_{2,\pi}^F = \frac{S^2(X^1)S(X^2, Y_\pi^F) - S(X^1, Y_\pi^F)S(X^1, X^2)}{S^2(X^1)S^2(X^2) - S(X^1, X^2)}$$

On obtient à l'aide de la définition de Δ :

$$\Delta \cdot \widehat{\beta}_{2,\pi}^F = S^2(X^1)S(X^2, Y_\pi^F) - S(X^1, Y_\pi^F)S(X^1, X^2).$$

Par substitution de $S(X^1, Y_\pi^F)$, $S(X^2, Y_\pi^F)$, on obtient :

$$\begin{aligned}\Delta \cdot \widehat{\beta}_{2,\pi}^F &= S^2(X^1)(a_1 S(X^1, X^2) + S(X^2, R^{Y.1,\pi})) \\ &\quad - (a_1 S^2(X^1) + S(X^1, R^{Y.1,\pi}))S(X^1, X^2) \\ &= S^2(X^1)S(X^2, R^{Y.1,\pi}) - S(X^1, R^{Y.1,\pi})S(X^1, X^2) \\ &= \Delta_2\end{aligned}$$

c'est-à-dire

$$\widehat{\beta}_{2,\pi}^F = \frac{\Delta_2}{\Delta} = \frac{\det(\mathbf{C}_{2,3}^{\mathbf{Y}_\pi^F})}{\Delta} \quad (3.11)$$

3. On sait que l'estimateur MCO du terme constant s'écrit :

$$\widehat{\beta}_{0,\pi}^F = \bar{Y} - \widehat{\beta}_{1,\pi}^F \bar{x}^1 - \widehat{\beta}_{2,\pi}^F \bar{x}^2$$

A partir des estimateurs $\widehat{\beta}_{1,\pi}^F$ et $\widehat{\beta}_{2,\pi}^F$ on obtient donc par substitution :

$$\widehat{\beta}_{0,\pi}^F = \bar{Y} - a_1 \bar{x}^1 - \frac{\Delta_1}{\Delta} \bar{x}^1 - \frac{\Delta_2}{\Delta} \bar{x}^2 \quad (3.12)$$

$$\widehat{\beta}_{0,\pi}^F = a_0 - \frac{\Delta_1}{\Delta} \bar{x}^1 - \frac{\Delta_2}{\Delta} \bar{x}^2 \quad \square \quad (3.13)$$

3.2.2 Cas du modèle de Kennedy

Considérons le modèle de régression :

$$X^2 = \gamma_0 + \gamma_1 x^1 + \epsilon^{2.1} \quad (3.14)$$

où $\mathbb{E}(\epsilon^{2.1} | \mathbf{X}^1) = 0$, et $\text{Var}(\epsilon^{2.1} | \mathbf{X}^1) = \sigma^{2.1}$. Nous savons que les estimateurs MCO des coefficients de γ_0 et γ_1 sont respectivement $g_1 = S(X^1, X^2)/S^2(X^1)$, et $g_0 = \bar{x}^2 - g_1 \bar{x}^1$. Par ailleurs, les résidus MCO observés par le modèle ci-dessus sont $R_i^{2.1} = x_i^2 - g_0 - g_1 x_i^1$. Ces résidus interviennent dans le calcul des estimateurs MCO des coefficients du modèle de Kennedy. Par suite, la proposition (3.2.1), ci-dessus fait intervenir le calcul des quantités $\sum_{i=1}^n R^{Y.1,\pi i} R_i^{2.1}$ et $\sum_{i=1}^n (R_i^{2.1})^2$. dont l'expression analytique en terme de déterminant est donnée par le lemme suivant :

Lemme 3.2.2.

1. $\sum_{i=1}^n R^{Y.1,\pi i} R_i^{2.1} = n \frac{\Delta_2}{S^2(X^1)}$
2. $\sum_{i=1}^n (R_i^{2.1})^2 = n \frac{\Delta}{S^2(X^1)}$

Preuve :

1. Une conséquence directe de la définition $R_i^{2.1}$ est :

$$\begin{aligned}
\sum_{i=1}^n R^{Y.1,\pi i} R_i^{2.1} &= \sum_{i=1}^n R^{Y.1,\pi i} (x_i^2 - g_0 - g_1 x_i^1) \\
&= \sum_{i=1}^n R^{Y.1,\pi i} (x_i^2 - \bar{x}^2 - d_1(x_i^1 - \bar{x}^1)) \\
&= \sum_{i=1}^n R^{Y.1,\pi i} (x_i^2 - \bar{x}^2) - \frac{S(X^1, X^2)}{S^2(X^1)} \sum_{i=1}^n R^{Y.1,\pi i} (x_i^1 - \bar{x}^1) \\
&= nS(X^2, R^{Y.1,\pi i}) - n \frac{S(X^1, X^2)}{S^2(X^1)} S(X^1, R^{Y.1,\pi i}) \\
&= n \frac{S^2(X^1)S(X^2, R^{Y.1,\pi i}) - S(X^1, X^2)S(X^1, R^{Y.1,\pi i})}{S^2(X^1)} \\
&= n \frac{\Delta_2}{S^2(X^1)} \tag{3.15}
\end{aligned}$$

2. De même, il résulte de la définitions de $R_i^{2.1}$ que :

$$\begin{aligned}
\sum_{i=1}^n (R_i^{2.1})^2 &= \sum_{i=1}^n (x_i^2 - \bar{x}^2 - g_1(x_i^1 - \bar{x}^1))^2 \\
&= \sum_{i=1}^n (x_i^2 - \bar{x}^2 - \frac{S(X^1, X^2)}{S^2(X^1)}(x_i^1 - \bar{x}^1))^2 \\
&= nS^2(X^2) + n \frac{S^2(X^1, X^2)}{S^2(X^1)} - 2n \frac{S^2(X^1, X^2)}{S^2(X^1)} \\
&= n(S^2(X^2) - \frac{S^2(X^1, X^2)}{S^2(X^1)}) = n \frac{\Delta}{S^2(X^1)} \quad \square \tag{3.16}
\end{aligned}$$

Remarquons de plus que

$$\sum_{i=1}^n R^{Y.1,\pi i} = \sum_{i=1}^n R_i^{Y.1} = 0$$

Proposition 3.2.2. *L'estimateur MCO du coefficient de régression $\beta_{2,\pi}^K$ dans le modèle de Kennedy :*

$$R^{Y.1,\pi i} = \beta_{2,\pi}^K R_i^{2.1} + \epsilon^K$$

est égal à l'estimateur MCO du coefficient de régression $\beta_{2,\pi}^F$ dans le modèle de Freedman-Lane, c'est-à-dire :

$$\hat{\beta}_{2\pi}^K = \frac{\Delta_2}{\Delta} = \hat{\beta}_{2\pi}^F$$

Preuve : L'estimateur MCO pour le coefficient de régression $\beta_{2,\pi}^K$ est (Johnston et Dinardo [30]) :

$$\hat{\beta}_{2,\pi}^K = \frac{\sum_{i=1}^n R^{Y,1,\pi i} R_i^{2,1}}{\sum_{i=1}^n (R_i^{2,1})^2}$$

et d'après le lemme (3.2.2)

$$\begin{aligned} \hat{\beta}_{2,\pi}^K &= \frac{n\Delta_2/S^2(X^1)}{n\Delta/S^2(X^1)} \\ &= \frac{\Delta_2}{\Delta} = \hat{\beta}_{2,\pi}^F \quad \square \end{aligned} \quad (3.17)$$

Nous allons désormais noter cet estimateur commun $\hat{\beta}_{2,\pi}$.

3.3 Estimateurs MCO de la Variance de $\hat{\beta}_{2,\pi}$

Dans cette section, nous donnons la variance de $\hat{\beta}_{2,\pi}$ ainsi que les deux estimateurs MCO de celle-ci, obtenus respectivement à l'aide des modèles de Freedman-Lane et de Kennedy. Puis, nous montrons que puisque la somme des carrés des résidus du modèle de Freedman-Lane est différente de celle qui est obtenue par le modèle de Kennedy, l'estimateur de la variance de $\hat{\beta}_{2,\pi}$ du modèle de Freedman-Lane diffère de celle de Kennedy. Enfin, nous démontrons mathématiquement que le modèle $\mathbb{E}(R^{Y,1,\pi i}) = \beta'_{0,\pi} + \beta'_{1,\pi}x^1 + \beta'_{2,\pi}x^2$ proposé par Anderson et Lengendre [1] et rappelé en section 2.3.1-B est équivalent au modèle de Freedman-Lane pour effectuer le test de l'hypothèse $H_0 : \beta_2 = 0$.

3.3.1 Cas du modèle de Freedman et Lane

Considérons le modèle de Freedman-Lane,

$$Y_{\pi i}^F = \beta_{0,\pi}^F + \beta_{1,\pi}^F x^1 + \beta_{2,\pi}^F x_i^2 + \epsilon_i^F \quad (3.18)$$

Soit σ^{2F} la variance de ϵ_i . Alors, La variance d'échantillonnage de l'estimateur MCO $\hat{\beta}_{2,\pi}$ est classiquement (*cf.* Johnston et Dinardo) :

$$\text{Var}(\hat{\beta}_{2,\pi})_F = \frac{S^2(X^1)\sigma^{2F}}{n[S^2(X^1)S^2(X^2) - (S(X^1, X^2))^2]} \quad (3.19)$$

$$= \frac{S^2(X^1)\sigma^{2F}}{n\Delta}. \quad (3.20)$$

L'estimateur MCO de cette variance sous le modèle de Freedman-Lane s'obtient donc par substitution à partir de l'estimateur MCO de σ^{2F} défini de façon classique pour le modèle (3.18), noté :

$$\widehat{\text{Var}}(\hat{\beta}_{2,\pi})_F = S^2(\hat{\beta}_{2,\pi})_F = \frac{S^2(X^1)S^{2F}}{n\Delta} \quad (3.21)$$

où $S^{2F} = \frac{\sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n-3}$ est l'estimateur sans biais pour σ^{2F} . Le calcul de cet estimateur nécessite donc de connaître la somme de carrés $\sum_{i=1}^n (\hat{\epsilon}_i^F)^2$. On utilise à cette fin le lemme suivant.

Lemme 3.3.1.

$$\begin{aligned} \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 &= \sum_{i=1}^n (R_i^{Y,1})^2 + n \frac{\Delta_1^2}{\Delta^2} S^2(X^1) + n \frac{\Delta_2^2}{\Delta^2} S^2(X^2) + 2n \left(\frac{\Delta_1 \Delta_2}{\Delta^2} \right) S(X^1, X^2) \\ &\quad - 2n \left(\frac{\Delta_1}{\Delta} S(X^1, R^{Y,1,\pi i}) + \frac{\Delta_2}{\Delta} S(X^2, R^{Y,1,\pi i}) \right) \end{aligned}$$

Preuve : Puisque $\hat{Y}_i^1 = \bar{Y} + a_1(x_i^1 - \bar{x}^1)$, le modèle de Freedman-Lane (3.18) permet d'obtenir

$$\begin{aligned} \hat{\epsilon}_i^F &= Y_{\pi i}^F - \hat{Y}_{\pi i}^F = (\hat{Y}_{1i} + R^{Y,1,\pi i}) - (\hat{\beta}_{0,\pi} + \hat{\beta}_{1,\pi}^F x_i^1 + \hat{\beta}_{2,\pi} x_i^2) \\ &= R^{Y,1,\pi i} + \bar{Y} + a_1(x_i^1 - \bar{x}^1) - \hat{\beta}_{0,\pi}^F - \hat{\beta}_{1,\pi}^F x_i^1 - \hat{\beta}_{2,\pi} x_i^2 \\ &= R^{Y,1,\pi i} + a_1(x_i^1 - \bar{x}^1) - \hat{\beta}_{1,\pi}^F (x_i^1 - \bar{x}^1) - \hat{\beta}_{2,\pi} (x_i^2 - \bar{x}^2) \end{aligned} \quad (3.22)$$

puisque $\bar{Y}_\pi^F = \bar{Y}$ et par ailleurs $\hat{\beta}_{0,\pi}^F = \bar{Y} - \hat{\beta}_{1,\pi}^F \bar{x}^1 - \hat{\beta}_{2,\pi} \bar{x}^2$. Mais, il résulte de la proposition (3.2.1) que $\hat{\beta}_{1,\pi}^F = a_1 + \frac{\Delta_1}{\Delta}$ et $\hat{\beta}_{2,\pi}^F = \frac{\Delta_2}{\Delta}$. Par suite :

$$\hat{\epsilon}_i^F = R_i^{Y,1,\pi} - \frac{\Delta_1}{\Delta} (x_i^1 - \bar{x}^1) - \frac{\Delta_2}{\Delta} (x_i^2 - \bar{x}^2) \quad (3.23)$$

Par conséquent :

$$\begin{aligned}
\sum_{i=1}^n (\hat{\epsilon}^F)^2 &= \sum_{i=1}^n (R^{Y.1,\pi i})^2 + \frac{\Delta_1^2}{\Delta^2} \sum_{i=1}^n (x_i^1 - \bar{x}^1)^2 + \frac{\Delta_2^2}{\Delta^2} \sum_{i=1}^n (x_i^2 - \bar{x}^2)^2 \\
&+ 2\left(\frac{\Delta_1\Delta_2}{\Delta^2}\right) \sum_{i=1}^n (x_i^1 - \bar{x}^1)(x_i^2 - \bar{x}^2) \\
&- 2\frac{\Delta_1}{\Delta} \sum_{i=1}^n (x_i^1 - \bar{x}^1)R^{Y.1,\pi i} - 2\frac{\Delta_2}{\Delta} \sum_{i=1}^n (x_i^2 - \bar{x}^2)R^{Y.1,\pi i}.
\end{aligned}$$

et puisque : $\sum_{i=1}^n (R^{Y.1,\pi i})^2 = \sum_{i=1}^n (R_i^{Y.1})^2$

$$\begin{aligned}
\sum_{i=1}^n (\hat{\epsilon}^F)^2 &= \sum_{i=1}^n (R_i^{Y.1})^2 + n\frac{\Delta_1^2}{\Delta^2}S^2(X^1) + n\frac{\Delta_2^2}{\Delta^2}S^2(X^2) + 2n\left(\frac{\Delta_1\Delta_2}{\Delta^2}\right)S(X^1, X^2) \\
&- 2n\left(\frac{\Delta_1}{\Delta}S(X^1, R^{Y.1,\pi i}) + \frac{\Delta_2}{\Delta}S(X^2, R^{Y.1,\pi i})\right) \quad \square \quad (3.24)
\end{aligned}$$

3.3.2 Cas du modèle de Kennedy

Considérons maintenant le modèle de Kennedy

$$R^{Y.1,\pi i} = \beta_{2,\pi}R_i^{2.1} + \epsilon_i^K \quad (3.25)$$

Soit σ^{2K} est la variance de ϵ^K . Alors, la variance d'échantillonnage de $\hat{\beta}_{2\pi}$ s'écrit classiquement (cf. Johnston et Dinardo [30]) :

$$\text{Var}(\hat{\beta}_{2,\pi})_K = \frac{\sigma^{2K}}{nS^2(R^{2.1})} \quad (3.26)$$

d'après le lemme (3.2.2)

$$\text{Var}(\hat{\beta}_{2,\pi})_K = S^2(\beta_{2,\pi})_K = \frac{S^2(X^1)\sigma^{2K}}{n\Delta}. \quad (3.27)$$

L'estimateur MCO pour le modèle de Kennedy s'écrit donc :

$$\widehat{\text{Var}}(\hat{\beta}_{2,\pi})_K = S^2(\beta_{2,\pi})_K = \frac{S^2(X^1)S^{2K}}{n\Delta} \quad (3.28)$$

où $S^{2K} = \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n-3}$ est l'estimateur MCO de σ^{2K} obtenu de façon classique pour le modèle (3.25). On peut aller plus loin grâce au lemme suivant :

Lemme 3.3.2. *Les résidus du modèle de Kennedy*

$$R^{Y.1,\pi i} = \beta_{2,\pi}R_i^{2.1} + \epsilon_i^K$$

vérifient :

$$\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 = \sum_{i=1}^n (R_i^{Y.1})^2 - n \frac{\Delta_2^2}{\Delta S^2(X^1)}$$

Preuve :

Les résidus du modèle ci-dessus s'écrivent :

$$\hat{\epsilon}_i^K = R^{Y.1,\pi i} - \hat{\beta}_{2,\pi} R_i^{2.1} \quad (3.29)$$

Par suite, une conséquence du lemme (3.2.2) s'écrit :

$$\begin{aligned} \hat{\epsilon}_i^K &= R^{Y.1,\pi i} - \hat{\beta}_{2,\pi} R_i^{2.1} \\ &= R^{Y.1,\pi i} - \frac{\Delta_2}{\Delta} ((x_i^2 - \bar{x}^2) - \frac{S(X^1, X^2)}{S^2(X^1)} (x_i^1 - \bar{x}^1)) \\ &= R^{Y.1,\pi i} - \frac{\Delta_2}{\Delta} (x_i^2 - \bar{x}^2) + \frac{\Delta_2}{\Delta} \frac{S(X^1, X^2)}{S^2(X^1)} (x_i^1 - \bar{x}^1) \end{aligned} \quad (3.30)$$

et donc, la somme des carrés des résidus vérifie :

$$\begin{aligned} \sum (\hat{\epsilon}_i^K)^2 &= \sum_{i=1}^n (R^{Y.1,\pi i})^2 + \frac{\Delta_2^2}{\Delta^2} \sum_{i=1}^n (x_i^2 - \bar{x}^2)^2 + \frac{\Delta_2^2}{\Delta^2} \left[\frac{S(X^1, X^2)}{S^2(X^1)} \right]^2 \sum_{i=1}^n (x_i^1 - \bar{x}^1)^2 \\ &\quad - 2 \frac{\Delta_2}{\Delta} \sum_{i=1}^n (x_i^2 - \bar{x}^2) R^{Y.1,\pi i} + 2 \frac{\Delta_2}{\Delta} \frac{S(X^1, X^2)}{S^2(X^1)} \sum_{i=1}^n (x_i^1 - \bar{x}^1) R^{Y.1,\pi i} \\ &\quad - 2 \frac{S(X^1, X^2)}{S^2(X^1)} \frac{\Delta_2^2}{\Delta^2} \sum_{i=1}^n (x_i^1 - \bar{x}^1) (x_i^2 - \bar{x}^2) \end{aligned}$$

utilisons l'identité $\sum_{i=1}^n R^{Y.1,\pi i} = \sum_{i=1}^n R_i^{Y.1}$. On obtient alors :

$$\begin{aligned} \sum (\hat{\epsilon}_i^K)^2 &= \sum_{i=1}^n (R_i^{Y.1})^2 + n \frac{\Delta_2^2}{\Delta^2} S^2(X^2) + n \frac{\Delta_1^2}{\Delta^2} \frac{[S(X^1, X^2)]^2}{S^2(X^1)} \\ &\quad - 2n \frac{\Delta_2}{\Delta} S(X^2, R^{Y.1,\pi i}) + 2n \frac{\Delta_2}{\Delta} \frac{S(X^1, X^2)}{S^2(X^1)} S(X^1, R^{Y.1,\pi i}) \\ &\quad - 2n \frac{\Delta_2^2}{\Delta^2} \frac{[S(X^1, X^2)]^2}{S^2(X^1)} \\ &= \sum_{i=1}^n (R_i^{Y.1})^2 + n \frac{\Delta_2^2}{\Delta^2} \left(S^2(X^2) + \frac{[S(X^1, X^2)]^2}{S^2(X^1)} - 2 \frac{[S(X^1, X^2)]^2}{S^2(X^1)} \right) \\ &\quad - 2n \frac{\Delta_2}{\Delta} \left(S(X^2, R^{Y.1,\pi i}) - \frac{S(X^1, X^2)}{S^2(X^1)} S(X^1, R^{Y.1,\pi i}) \right) \\ &= \sum_{i=1}^n (R_i^{Y.1})^2 - n \frac{\Delta_2^2}{\Delta S^2(X^1)} \quad \square \end{aligned} \quad (3.31)$$

Remarque 3.3.1.

Les lemmes (3.3.1) et (3.3.2) montrent en particulier qu'en général la somme des carrés des résidus du modèle de Freedman-Lane est différente de la somme des carrés des résidus du modèle de Kennedy.

$$\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 \neq \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 \quad (3.32)$$

Proposition 3.3.1. *Le modèle de régression de la variable $R^{Y,1,\pi i}$ en x^1 et x^2 :*

$$R^{Y,1,\pi i} = \beta'_{0,\pi} + \beta'_{1,\pi}x_i^1 + \beta'_{2,\pi}x_i^2 + \epsilon'_i, \quad (i = 1, 2, \dots, n) \quad (3.33)$$

où $\text{Var}(\epsilon'_i | \mathbf{x}^1, \mathbf{x}^2) = \sigma^{2'}$ et $\mathbb{E}(\epsilon'_i | \mathbf{x}^1, \mathbf{x}^2) = 0$, vérifie :

1. Les estimateurs MCO de $\beta'_{0,\pi}$, $\beta'_{1,\pi}$ et $\beta'_{2,\pi}$ sont :

$$\begin{aligned} a) \hat{\beta}'_{2,\pi} &= \frac{\Delta_2}{\Delta} = \hat{\beta}_{2,\pi} \\ b) \hat{\beta}'_{1,\pi} &= \frac{\Delta_1}{\Delta} = \hat{\beta}_{1,\pi}^F - a_1 \\ c) \hat{\beta}'_{0,\pi} &= -\frac{\Delta_1}{\Delta}\bar{x}^1 - \frac{\Delta_2}{\Delta}\bar{x}^2 = \hat{\beta}_{0,\pi}^F - a_0 \end{aligned}$$

2. Les résidus MCO observés du modèle ci-dessus sont égaux aux résidus du modèle de Freedman-Lane, c'est-à-dire :

$$\hat{\epsilon}' = \hat{\epsilon}^F$$

3. Les estimateurs MCO des variances des estimateurs des coefficients de régression du modèle ci-dessus sont égaux aux estimateurs MCO des variances des estimateurs des coefficients de régression du modèle de Freedman-Lane, c'est-à-dire :

$$\begin{aligned} a) \widehat{\text{Var}}(\hat{\beta}'_{2,\pi}) &= \widehat{\text{Var}}(\hat{\beta}_{2,\pi}^F) \\ b) \widehat{\text{Var}}(\hat{\beta}'_{1,\pi}) &= \widehat{\text{Var}}(\hat{\beta}_{1,\pi}^F) \\ c) \widehat{\text{Var}}(\hat{\beta}'_{0,\pi}) &= \widehat{\text{Var}}(\hat{\beta}_{0,\pi}^F) \end{aligned}$$

Preuve :

1. a) L'estimateur $\beta'_{2,\pi}$ au sens MCO s'écrit :

$$\hat{\beta}'_{2,\pi} = \frac{S^2(X^1)S(X^2, R^{Y,1,\pi i}) - S(X^1, X^2)S(X^1, R^{Y,1,\pi i})}{\Delta} = \frac{\Delta_2}{\Delta}$$

qui est égal à $\hat{\beta}_{2,\pi}^F$, c'est-à-dire :

$$\hat{\beta}'_{2,\pi} = \frac{\Delta_2}{\Delta} = \hat{\beta}_{2,\pi}^F \quad (3.34)$$

1. b) De façon analogue l'estimateur MCO de $\beta'_{1,\pi}$ est égal à :

$$\hat{\beta}'_{1,\pi} = \frac{S^2(X^2)S(X^1, R^{Y.1,\pi i}) - S(X^1, X^2)S(X^2, R^{Y.1,\pi i})}{\Delta} = \frac{\Delta_1}{\Delta}$$

qui en application la proposition (3.2.1) est égal à $\hat{\beta}_{1,\pi}^F - a_1$, c'est-à-dire :

$$\hat{\beta}'_{1,\pi} = \frac{\Delta_1}{\Delta} = \hat{\beta}_{1,\pi}^F - a_1 \quad (3.35)$$

1. c) L'estimateur MCO de l'intercept $\beta'_{0,\pi}$ est classiquement :

$$\hat{\beta}'_{0,\pi} = \bar{R}^{\pi,Y.1} - \hat{\beta}'_{1,\pi}\bar{x}^1 - \hat{\beta}'_{2,\pi}\bar{x}^2,$$

on sait de plus que $\bar{R}^{\pi,Y.1} = 0$ et l'expression des estimateurs $\hat{\beta}'_{1,\pi}$ et $\hat{\beta}'_{2,\pi}$, induit :

$$\hat{\beta}'_{0,\pi} = -\frac{\Delta_1}{\Delta}\bar{x}^1 - \frac{\Delta_2}{\Delta}\bar{x}^2.$$

en utilisant la proposition (3.2.1) on obtient :

$$\hat{\beta}'_{0,\pi} = \hat{\beta}_{0,\pi}^F - a_0. \quad (3.36)$$

2. Considérons maintenant le modèle (3.33). Le résidu observé est :

$$\hat{\epsilon}'_i = R^{Y.1,\pi i} - \hat{R}^{Y.1,\pi i} = R^{Y.1,\pi i} - \hat{\beta}'_{0,\pi} - \hat{\beta}'_{1,\pi}x_i^1 - \hat{\beta}'_{2,\pi}x_i^2$$

d'après les résultats obtenus dans la partie (1) et en appliquant le lemme (3.3.1) on obtient :

$$\hat{\epsilon}'_i = R^{Y.1,\pi i} - \frac{\Delta_1}{\Delta}(x_i^1 - \bar{x}^1) - \frac{\Delta_2}{\Delta}(x_i^2 - \bar{x}^2) = \hat{\epsilon}_i^F \quad (3.37)$$

Par conséquent, on a :

$$\sum_{i=1}^n (\hat{\epsilon}'_i)^2 = \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 \quad (3.38)$$

3. a) la variance de $\hat{\beta}'_{2,\pi}$ s'écrit :

$$\text{Var}(\hat{\beta}'_{2,\pi}) = \frac{S^2(X^1)\sigma^{2'}}{\Delta} \quad (3.39)$$

et l'estimateur MCO de $\text{Var}(\hat{\beta}'_{2,\pi})$ est classiquement :

$$\widehat{\text{Var}}(\hat{\beta}'_{2,\pi}) = S^2(\hat{\beta}'_{2,\pi}) = \frac{S^2(X^1)S^{2'}}{\Delta} \quad (3.40)$$

où $S^{2'} = \frac{1}{n-3} \sum_{i=1}^n (\hat{\epsilon}'_i)^2$. On obtient en application des résultats obtenus dans la partie (2) :

$$S^{2'} = \frac{\sum_{i=1}^n (\hat{\epsilon}'_i)^2}{n-3} = \frac{\sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n-3} = S^{2F}$$

alors, en considérant l'équation (3.21), on déduit :

$$\widehat{\text{Var}}(\hat{\beta}'_{2,\pi}) = \frac{S^2(X^1)S'^2}{\Delta} = \frac{S^2(X^1)S^{2F}}{\Delta} = \widehat{\text{Var}}(\hat{\beta}_{2,\pi})_F \quad (3.41)$$

3. b) De façon analogue, on a :

$$\text{Var}(\hat{\beta}'_{1,\pi}) = \frac{S^2(X^2)\sigma^{2'}}{\Delta}, \quad \text{Var}(\hat{\beta}_{1,\pi}^F) = \frac{S^2(X^2)\sigma^{2F}}{\Delta} \quad (3.42)$$

Les estimateurs MCO de $\text{Var}(\hat{\beta}'_{1,\pi})$ et de $\text{Var}(\hat{\beta}_{1,\pi}^F)$ sont respectivement :

$$\widehat{\text{Var}}(\hat{\beta}'_{1,\pi}) = S^2(\hat{\beta}'_{1,\pi}) = \frac{S^2(X^1)S'^2}{\Delta}, \quad \widehat{\text{Var}}(\hat{\beta}_{1,\pi}^F) = S^2(\hat{\beta}_{1,\pi}^F) = \frac{S^2(X^1)S^{2F}}{\Delta}$$

et puisque $S^{2'} = S^{2F}$, on obtient :

$$\widehat{\text{Var}}(\hat{\beta}'_{1,\pi}) = \frac{S^2(X^1)S'^2}{\Delta} = \frac{S^2(X^1)S^{2F}}{\Delta} = \widehat{\text{Var}}(\hat{\beta}_{1,\pi}^F) \quad (3.43)$$

3. c) Un calcul analogue permet aussi de montrer que :

$$\widehat{\text{Var}}(\hat{\beta}'_0) = \widehat{\text{Var}}(\hat{\beta}_{0,\pi}^F) \quad \square \quad (3.44)$$

Remarque 3.3.2.

La proposition (3.3.1) montre que les statistiques utilisées pour effectuer le test de permutations de l'hypothèse $H_0 : \beta_2 = 0$ sous le modèle de Freedman-Lane et sous le modèle (3.33) sont équivalentes, c'est-à-dire :

$$T'_\pi = \frac{\hat{\beta}'_{2,\pi}}{S(\hat{\beta}'_{2,\pi})} = \frac{\hat{\beta}_{2,\pi}}{S(\hat{\beta}_{2,\pi})_K} = T^F_\pi \quad (3.45)$$

3.4 Étude comparative des deux méthodes

Dans cette section, nous montrons que l'estimateur MCO de $\text{Var}(\hat{\beta}_{2,\pi})$ du modèle de Freedman-Lane est inférieur à celui de Kennedy. Ensuite, nous montrons les conséquences de cette propriété sur la statistique de test, sur la distribution de cette statistique de test, sur le risque d'erreur de première espèce et sur la puissance du test de permutation.

3.4.1 Estimation de $\text{Var}(\hat{\beta}_{2,\pi})$

Rappelons que les estimateurs MCO de $\text{Var}(\hat{\beta}_{2,\pi})_F$ et de $\text{Var}(\hat{\beta}_{2,\pi})_K$ sont respectivement pour les modèles de Freedman-Lane et de Kennedy :

$$S^2(\hat{\beta}_{2,\pi})_F = \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n(n-3)\Delta}, \quad S^2(\hat{\beta}_{2,\pi})_K = \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-3)\Delta}$$

La remarque (3.3.1) montre que la différence entre $S^2(\hat{\beta}_{2,\pi})_F$ et $S^2(\hat{\beta}_{2,\pi})_K$ provient de la différence entre $\sum_{i=1}^n (\hat{\epsilon}_i^F)^2$ et $\sum_{i=1}^n (\hat{\epsilon}_i^K)^2$. Nous devons donc comparer ces deux sommes de carrés.

Lemme 3.4.1.

$$\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 = \frac{n[S(X^1, R^{Y.1, \pi i})]^2}{S^2(X^1)} \geq 0$$

Preuve : Formons la différence des équations (3.24) et (3.31)

$$\begin{aligned} \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 &= \sum_{i=1}^n (R_i^{Y.1})^2 + n \frac{\Delta_2^2}{\Delta^2} S^2(X^2) + n \frac{\Delta_1^2}{\Delta^2} S^2(X^1) \\ &\quad + 2n \left(\frac{\Delta_1 \Delta_2}{\Delta^2} \right) S(X^1, X^2) - 2n \left(\frac{\Delta_2}{\Delta} \right) S(X^2, R^{Y.1, \pi i}) + \frac{\Delta_1}{\Delta} S(X^1, R^{Y.1, \pi i}) \end{aligned}$$

$$\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 = \sum_{i=1}^n (R_i^{Y.1})^2 - n \frac{\Delta_2^2}{\Delta S^2(X^1)}$$

Donc :

$$\begin{aligned} \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 &= -n \frac{\Delta_2^2}{\Delta S^2(X^1)} - n \frac{\Delta_2^2}{\Delta^2} S^2(X^2) - n \frac{\Delta_1^2}{\Delta^2} S^2(X^1) \\ &\quad - 2n \frac{\Delta_1 \Delta_2}{\Delta^2} S(X^1, X^2) \\ &\quad + 2n \left(\frac{\Delta_2}{\Delta} \right) S(X^2, R^{Y.1, \pi i}) + \frac{\Delta_1}{\Delta} S(X^1, R^{Y.1, \pi i}) \end{aligned}$$

$$\begin{aligned}
&= n \frac{-\Delta_2^2(\Delta + S^2(X^1)S^2(X^2)) - n\Delta_1^2[S^2(X^1)]^2 - 2n\Delta_1\Delta_2S^2(X^1)S(X^1, X^2)}{\Delta^2S^2(X^1)} \\
&+ \frac{2n\Delta S^2(X^1)[\Delta_2S(X^2, R^{Y.1, \pi i}) + \Delta_1S(X^1, R^{Y.1, \pi i})]}{\Delta^2S^2(X^1)} \\
&\text{en remplaçant } \Delta_1 \text{ et } \Delta_2 \text{ on obtient :} \\
&= -n \left[\frac{[S^2(X^1)]^2[S(X^2, R^{Y.1, \pi})]^2 (\Delta + S^2(X^1)S^2(X^2))}{S^2(X^1)\Delta^2} \right. \\
&+ \frac{[S(X^1, X^2)]^2[S(X^1, R^{Y.1, \pi})]^2 (\Delta + S^2(X^1)S^2(X^2))}{S^2(X^1)\Delta^2} \\
&- \frac{2[S^2(X^1)S(X^1, X^2)S(X^1, R^{Y.1, \pi})S(X^2, R^{Y.1, \pi}) (\Delta + S^2(X^1)S^2(X^2))]}{S^2(X^1)\Delta^2} \\
&- n \left[\frac{[[S^2(X^2)]^2[S(X^1, R^{Y.1, \pi})]^2 + [S(X^1, X^2)]^2[S(X^2, R^{Y.1, \pi})]^2] [S^2(X^1)]^2}{\Delta^2S^2(X^1)} \right. \\
&- \frac{[S^2(X^1)S(X^1, X^2)S(X^1, R^{Y.1, \pi})S(X^2, R^{Y.1, \pi}) [S^2(X^1)]^2]}{\Delta^2S^2(X^1)} \\
&- 2n \left[\frac{[(S^2(X^1)S^2(X^2)S(X^1, R^{Y.1, \pi})S(X^2, R^{Y.1, \pi})) (S^2(X^1)S(X^1, X^2))]}{\Delta^2S^2(X^1)} \right. \\
&- \frac{[(S^2(X^1)S(X^1, X^2)[S(X^2, R^{Y.1, \pi})]^2] (S^2(X^1)S(X^1, X^2))}{\Delta^2S^2(X^1)} \\
&- \frac{[(S^2(X^2)S(X^1, X^2)[S(X^1, R^{Y.1, \pi})]^2] (S^2(X^1)S(X^1, X^2))}{\Delta^2S^2(X^1)} \\
&- \left. \frac{[[S^2(X^1, X^2)]^2S(X^1, R^{Y.1, \pi})S(X^2, R^{Y.1, \pi})] (S^2(X^1)S(X^1, X^2))}{\Delta^2S^2(X^1)} \right] \\
&+ n\Delta \left[\frac{[2[S^2(X^1)]^2[S(X^2, R^{Y.1, \pi})]^2 + 2S^2(X^1)S^2(X^2)[S(X^1, R^{Y.1, \pi})]^2]}{\Delta^2S^2(X^1)} \right. \\
&- \left. \frac{[4S^2(X^1)S(X^1, X^2)S(X^1, R^{Y.1, \pi})S(X^2, R^{Y.1, \pi})]}{\Delta^2S^2(X^1)} \right] \\
&= n\Delta \left[\frac{-2S^2(X^1)S(X^1, X^2)S(X^1, R^{Y.1, \pi})S(X^2, R^{Y.1, \pi}) + [S^2(X^1)]^2[S(X^2, R^{Y.1, \pi})]^2}{\Delta^2S^2(X^1)} \right. \\
&- \left. \frac{[S(X^1, X^2)]^2[S(X^1, R^{Y.1, \pi})]^2 + 2S^2(X^1)S^2(X^2)[S(X^1, R^{Y.1, \pi})]^2}{\Delta^2S^2(X^1)} \right] \\
&+ n \frac{[S^2(X^1)]^2[S(X^1, X^2)]^2[S(X^2, R^{Y.1, \pi})]^2 + S^2(X^1)S^2(X^2)[S(X^1, X^2)]^2[S(X^1, R^{Y.1, \pi})]^2}{\Delta^2S^2(X^1)} \\
&- n \frac{2S^2(X^1)[S(X^1, X^2)]^3S(X^1, R^{Y.1, \pi})S(X^2, R^{Y.1, \pi}) + [S^2(X^1)]^3S(X^2)[S(X^2, R^{Y.1, \pi})]^2}{\Delta^2S^2(X^1)} \\
&+ n \frac{2[S^2(X^1)]^3S(X^1, X^2)S(X^1, R^{Y.1, \pi})S(X^2, R^{Y.1, \pi}) - [S^2(X^1)]^2[S(X^2)]^2[S(X^1, R^{Y.1, \pi})]^2}{\Delta^2S^2(X^1)}
\end{aligned}$$

En substituant par Δ , on obtient :

$$\begin{aligned}
 &= n \left[\frac{[S(X^1, X^2)]^4 S(X^1, R^{Y.1, \pi}) - 2S^2(X^1)S^2(X^2)[S(X^1, X^2)]^2 [S(X^1, R^{Y.1, \pi})]^2}{\Delta^2 S^2(X^1)} \right. \\
 &+ \left. \frac{[S^2(X^1)]^2 [S^2(X^2)]^2 [S(X^1, R^{Y.1, \pi})]^2}{\Delta^2 S^2(X^1)} \right] \\
 &= n \frac{\Delta^2 [S(X^1, R^{Y.1, \pi})]^2}{\Delta^2 S^2(X^1)}
 \end{aligned}$$

Au bilan, on obtient donc

$$\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 = \frac{n[S(X^1, R^{Y.1, \pi i})]^2}{S^2(X^1)} \geq 0 \quad \square \quad (3.46)$$

Proposition 3.4.1.

$$S^2(\beta_{2, \pi})_K - S^2(\beta_{2, \pi})_F = \frac{[S(X^1, R^{Y.1, \pi i})]^2}{(n-3)\Delta} \geq 0$$

Preuve : On déduit des équations (3.21) et (3.28) :

$$\begin{aligned}
 S^2(\hat{\beta}_{2, \pi})_K - S^2(\hat{\beta}_{2, \pi})_F &= \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-3)\Delta} - \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n(n-3)\Delta} \\
 &= \frac{S^2(X^1)}{n(n-3)\Delta} \left(\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 \right)
 \end{aligned}$$

Le lemme (3.4.1) fournit alors le résultat :

$$S^2(\hat{\beta}_{2, \pi})_K - S^2(\hat{\beta}_{2, \pi})_F = \frac{[S(X^1, R^{Y.1, \pi i})]^2}{(n-3)\Delta} \geq 0 \quad \square$$

Remarque 3.4.1.

La proposition (3.3.1) permet d'affirmer que pour effectuer le test de permutation de l'hypothèse $H_0 : \beta_2 = 0$, le modèle (3.33) et le modèle de Freedman-Lane sont équivalents. Par suite, comparer les modèles de Freedman-Lane et de Kennedy, est équivalent à comparer le modèle (3.33)

$$R^{Y.1, \pi i} = \beta'_0 + \beta'_1 x_i^1 + \beta'_2 x_i^2 + \epsilon'_i$$

au modèle de Kennedy :

$$R^{Y.1, \pi i} = \beta_{2K} R^{2.1} + \epsilon^K.$$

On déduit alors du lemme (3.4.1) que le modèle de Freedman-Lane est mieux ajusté que le modèle de Kennedy. Mais, il faut noter que ce lemme ne permet pas de déduire que $(\hat{\epsilon}_i^k)^2 \geq (\hat{\epsilon}_i^F)^2$ pour chaque i , $i = 1, 2, \dots, n$, puisque, bien sûr :

$$\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 \geq \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 \not\Rightarrow (\hat{\epsilon}_i^k)^2 \geq (\hat{\epsilon}_i^F)^2 \quad (3.47)$$

De ce fait, on peut dire que la variation de la variable $R^{Y,1,\pi i}$ expliquée par les deux variables X^1 et X^2 dans le modèle de Freedman-Lane est plus grande que celle de la variable $R^{2,1}$ dans le modèle de Kennedy.

3.4.2 Comparaison des estimateurs de $\text{Var}(\hat{\beta}_{2,\pi})$

Dans cette section, nous montrons les conséquences de la différence entre estimateurs de $\text{Var}(\hat{\beta}_{2,\pi})$ sur le risque d'erreur de première espèce et sur la puissance du test de permutation.

Lemme 3.4.2. *Pour chaque permutation d'un ensemble de données, la valeur absolue de la statistique de test de la méthode de Freedman-Lane est supérieure ou égale à la valeur absolue de la statistique de test correspondante de la méthode de Kennedy :*

$$|T_{2,\pi}^K| \leq |T_{2,\pi}^F|$$

Preuve : Nous avons montré dans la proposition (3.2.2) que l'estimateur MCO $\hat{\beta}_{2,\pi}$ de $\beta_{2,\pi}$ est le même dans le modèle de Freedman-Lane et dans celui de Kennedy pour chaque permutation des données. De plus, la proposition (3.4.1) dit qu'alors, l'estimateur MCO associé de la variance de cet estimateur dans le modèle de Freedman-Lane ($S^2(\hat{\beta}_{2,\pi})_F$), est inférieur ou égal à l'estimateur correspondant dans le modèle de Kennedy ($S^2(\hat{\beta}_{2,\pi})_K$), c'est-à-dire :

$$S^2(\hat{\beta}_{2,\pi})_K \geq S^2(\hat{\beta}_{2,\pi})_F$$

Alors, pour chaque permutation, la valeur absolue de la statistique de test du modèle de Freedman-Lane est supérieure ou égale à celle qui lui correspond dans le modèle de Kennedy, c'est-à-dire :

$$S^2(\hat{\beta}_{2,\pi})_K \geq S^2(\hat{\beta}_{2,\pi})_F \implies |T_{2,\pi}^K| = \frac{|\hat{\beta}_{2,\pi}|}{S(\hat{\beta}_{2,\pi})_K} \leq \frac{|\hat{\beta}_{2,\pi}|}{S(\hat{\beta}_{2,\pi})_F} = |T_{2,\pi}^F| \quad \square \quad (3.48)$$

Remarque 3.4.2.

Une conséquence directe du lemme ci-dessus est que la distribution permutationnelle de $|T_{\pi}^F|$ a une dispersion supérieure ou égale à celle de $|T_{2,\pi}^F|$

Proposition 3.4.2. *La fonction de répartition de $|T_{2,\pi}^F|$ est inférieure ou égale à celle de $|T_{2,\pi}^K|$. C'est-à-dire :*

$$F_{|T_{2,\pi}^F|}(t) \leq F_{|T_{2,\pi}^K|}(t) \quad \forall t$$

Preuve : Selon le lemme (3.4.2), on a :

$$|T_{2,\pi}^K| \leq |T_{2,\pi}^F|$$

Par suite, quel que soit $t \in \mathbb{R}$, on a :

$$\begin{aligned} \{|T_{2,\pi}^F| > t\} &\supseteq \{|T_{2,\pi}^K| > t\} \\ Pr(|T_{2,\pi}^F| > t) &\geq Pr(|T_{2,\pi}^K| > t) \Leftrightarrow F_{|T_{2,\pi}^F|}(t) \leq F_{|T_{2,\pi}^K|}(t) \quad \square \end{aligned} \quad (3.49)$$

c'est-à-dire que $|T_{2,\pi}^F|$ et $|T_{2,\pi}^K|$ sont stochastiquement ordonnées. La situation est illustrée par la figure ci-dessous :

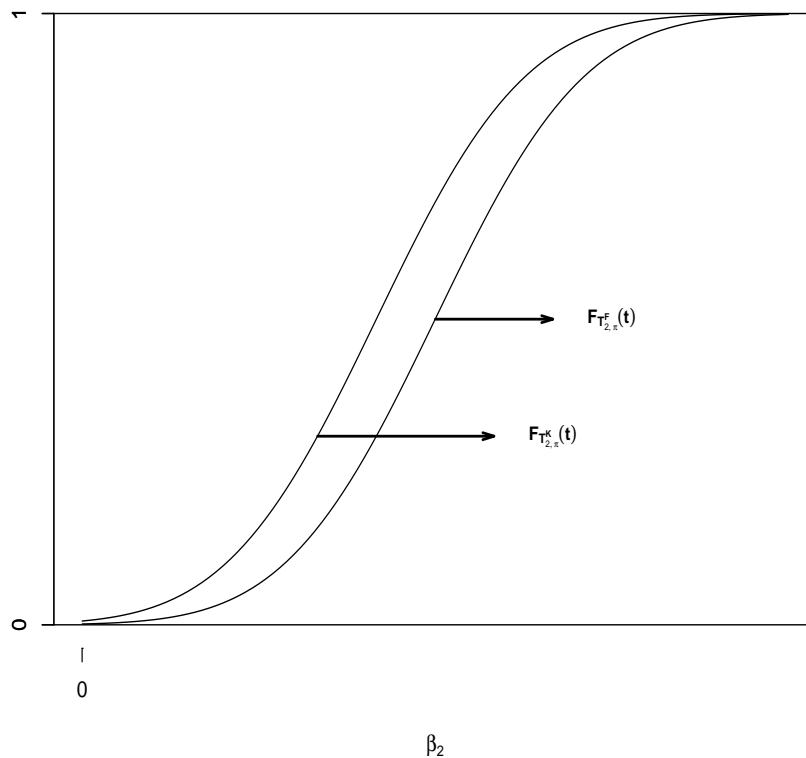


FIG. 3.1 – Les fonctions de répartition des statistiques de tests $|T_{2,\pi}^F|$ et $|T_{2,\pi}^K|$

La p-valeur permutationnelle observée associée à chaque méthode est le pourcentage de valeurs de la statistique permutationnelle (c'est-à-dire t_{π}^F et t_{π}^K) qui sont plus grandes que t_{obs} . Le corollaire suivant montre que la p-valeur permutationnelle de la méthode de Freedman-Lane ($p_{2,\pi}^F$) est supérieure ou égale à celle de la méthode de Kennedy ($p_{2,\pi}^K$). On en déduit ensuite que le risque d'erreur de première espèce de la méthode de Freedman-Lane est inférieur ou égal à celui de la méthode de Kennedy et à l'opposé, la puissance du test obtenue par la méthode de Freedman-Lane inférieure ou égale à celle de la méthode de Kennedy.

Corollaire 3.4.1. *Pour le test d'hypothèse $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$, la p-valeur permutationnelle de la méthode de Freedman-Lane est supérieure ou égale à celle de la méthode de Kennedy :*

$$p_{2,\pi}^F \geq p_{2,\pi}^K$$

De plus :

1. *Le risque d'erreur de première espèce de la méthode de Freedman-Lane est inférieur ou égal à celui de la méthode de Kennedy ;*
2. *La puissance du test obtenue par la méthode de Freedman-Lane est inférieure ou égale à celle de la méthode de Kennedy.*

Preuve : D'après la proposition (3.4.2), on obtient :

$$\begin{aligned} p_{2,\pi}^F &= Pr(|T_{2,\pi}^F| > t_{obs}) \\ &\geq Pr(|T_{2,\pi}^K| > t_{obs}) = p_{2,\pi}^K \end{aligned} \quad (3.50)$$

Par ailleurs, nous savons qu'on rejette l'hypothèse $H_0 : \beta_2 = 0$ lorsque la p-valeur observée est plus petite que le niveau α du test. La probabilité de rejeter $H_0 : \beta_2 = 0$ par la méthode de Freedman-Lane est donc inférieure ou égale à celle qu'on obtient par la méthode de Kennedy, c'est-à-dire :

$$Pr(P_{2,\pi}^F < \alpha) \leq Pr(P_{2,\pi}^K < \alpha) \quad (3.51)$$

Donc, la fonction puissance permutationnelle de la méthode de Freedman-Lane est inférieure ou égale à celle de la méthode de Kennedy. Si C est la région critique du test, on a :

$$\eta_{\pi}^F(\beta_2) = Pr_{\pi}^F(C) \leq Pr_{\pi}^K(C) = \eta_{\pi}^K(\beta_2) \quad (3.52)$$

La situation est illustrée en termes de fonction puissance permutationnelle par la figure ci-dessus.

Par conséquent, pour des données générées sous l'hypothèse $H_0 : \beta_2 = 0$, la relation ci-dessus dit que le risque d'erreur de première espèce de la méthode de Freedman-Lane est inférieur à celui de Kennedy.

$$Pr_{H_0}(P_{2,\pi}^F < \alpha) \leq Pr_{H_0}(P_{2,\pi}^K < \alpha) \quad (3.53)$$

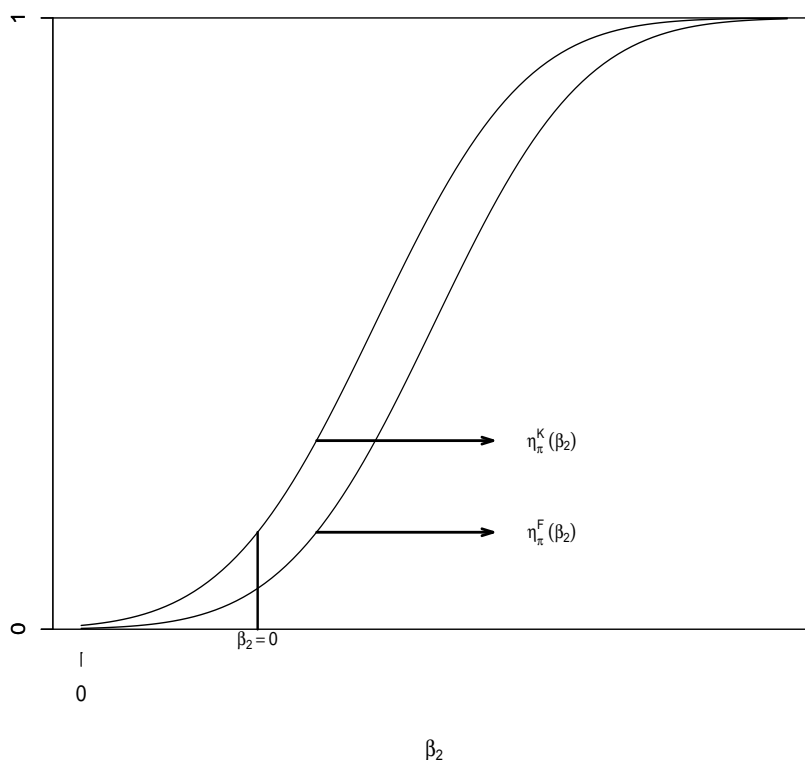


FIG. 3.2 – Les fonctions puissance des statistiques $|T_{2,\pi}^F|$ et $|T_{2,\pi}^K|$

Ces résultats donnent une justification mathématique aux résultats des simulations d'Andersson et Legendre [1]. Mais les résultats de nos propres simulations, présentés dans la section (2.5.1), montrent que cette situation simple de comparaison de puissance de test est intimement liée à la définition de la même statistique de test retenue. Nous revenons sur influence du choix de cette statistique de test dans la section suivante.

Par contre, lorsque nos données sont générées sous l'hypothèse $H_1 : \beta_2 \neq 0$ la relation ci-dessus dit que la puissance du test utilisant la méthode de Freedman-Lane est inférieure ou égale à celle de la méthode de Kennedy.

$$Pr_{H_1}(P_{2,\pi}^F < \alpha) \leq Pr_{H_1}(P_{2,\pi}^K < \alpha) \quad \square \quad (3.54)$$

Cependant, lors de la vérification des puissances de tests par des méthodes permutationnelles, la méthode de Kennedy n'est considérée ni dans les simulations d'Anderson et Legendre [1], ni dans celle de Legendre (2000). Mais, comme nous l'avons vu dans la section (2.5.3), il existe dans nos simulations, des cas où la puissance du test par la méthode de Kennedy est inférieure à celle qui est obtenue par

la méthode de Freedman-Lane. Nous montrerons plus bas pourquoi il en est ainsi.

Utilisons maintenant la proposition (3.4.1). Il vient :

$$S^2(\hat{\beta}_{2,\pi})_K - S^2(\hat{\beta}_{2,\pi})_F = \frac{S^2(X^1)S^2(R^{Y.1})[\rho(X^1, R^{Y.1,\pi})]^2}{(n-3)\Delta} \geq 0 \quad (3.55)$$

où $\rho(X^1, R^{Y.1,\pi})$ est la corrélation entre X^1 et $R^{Y.1,\pi}$. On vérifie de plus que la variance empirique est invariante par

$$S^2(R^{Y.1}) = \frac{1}{n} \sum_{i=1}^n (R_i^{Y.1})^2 = \frac{1}{n} \sum_{i=1}^n (R^{Y.1,\pi i})^2 = S^2(R^{Y.1,\pi}) \quad (3.56)$$

Les valeurs Δ , n , $S^2(X^1)$ et $S^2(R^{Y.1})$ sont déterminées, par des données. Par suite, pour chaque permutation, plus la corrélation $\rho(X^1, R^{Y.1,\pi})$ est grande et plus la différence entre les deux estimateurs est élevée. Par ailleurs, les deux estimateurs sont égaux si $\rho(X^1, R^{Y.1,\pi}) = 0$. Géométriquement, on peut dire que les deux méthodes donnent des estimateurs identiques lorsque le vecteur aléatoire $\mathbf{R}^{Y.1,\pi}$ est orthogonal à \mathbf{X}^1 .

Proposition 3.4.3. *Lorsque la taille de l'échantillon de données tend vers l'infini, la différence entre les deux estimateurs de $\mathbb{V}\text{ar}(\hat{\beta}_{2,\pi})$, tend vers zéro :*

1) Si lorsque n tend vers l'infini, la matrice $S = \begin{bmatrix} S^2(X^1) & S(X^1, X^2) \\ S(X^1, X^2) & S^2(X^2) \end{bmatrix}$ converge vers $\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$ fini,

2) si de plus, $S^2(R^{2.1})$ converge vers $S^{2*}(R^{2.1})$ fini,

3) si enfin, Δ converge vers $\Delta^* > 0$ fini

4) et $\rho(X^1, X^2)$ existe. Alors, on obtient :

$$\begin{aligned} \lim_{n \rightarrow +\infty} (S^2(\hat{\beta}_{2,\pi})_K - S^2(\hat{\beta}_{2,\pi})_F) &= \lim_{n \rightarrow +\infty} \frac{S^2(X^1)S^2(R^{Y.1})[\rho(X^1, R^{Y.1,\pi})]^2}{(n-3)\Delta} \\ &= \lim_{n \rightarrow +\infty} \frac{S^2(X^1)S^2(R^{Y.1})[\rho(X^1, R^{Y.1,\pi})]^2}{\Delta} \end{aligned} \quad (3.57)$$

$$\begin{aligned} &\times \lim_{n \rightarrow +\infty} \frac{1}{n-3} \\ &= \frac{\sigma_1^2 S^{2*}(R^{Y.1})[\rho(X^1, R^{Y.1,\pi})]^2}{\Delta^*} \times 0 \\ &= 0 \end{aligned} \quad (3.58)$$

C'est dire que les deux estimateurs sont asymptotiquement égaux. \square

Cette proposition montre en particulier que les statistiques de test associées aux méthodes de Freedman-Lane et de Kennedy deviennent indiscernable lorsque la taille de l'échantillon de données est grande.

$$T_{2,\pi}^F = \frac{\hat{\beta}_{2,\pi}}{S(\hat{\beta}_{2,\pi})_F} \approx \frac{\hat{\beta}_{2,\pi}}{S(\hat{\beta}_{2,\pi})_K} = T_{2,\pi}^K$$

Par conséquent, lorsque la taille de l'échantillon de données est grande, on doit avoir le même risque d'erreur de première espèce, ce que confirment la simulation d'Anderson et Legendre. On a aussi la même puissance.

3.4.3 Méthode de Kennedy et degrés de liberté

Considérons à nouveau le modèle de Kennedy,

$$R^{Y.1,\pi_i} = \beta_{2,\pi}^K R^{2.1} + \epsilon^K,$$

où $\text{Var}(\epsilon^K) = \sigma^{2K}$. pour chaque permutation π_i , on peut considérer $(n - 2)$ degrés de liberté pour la statistique de test permutational utilisée par la méthode de Kennedy ($T_{2,\pi}^K = \hat{\beta}_{2,\pi} / S(\hat{\beta}_{2,\pi})_K$). Ceci puisqu'on réalise une régression de la variable $R^{Y.1,\pi}$ en $R^{2.1}$ pour chaque permutation. Mais, le modèle comprend deux paramètres au lieu des trois paramètres utilisés par la méthode de Freedman-Lane. En effet, pour chaque permutation, on peut considérer $R^{Y.1,\pi_i}$ et $R^{2.1}$ comme deux nouvelles variables (Anderson et Legendre [1]). De ce fait, un estimateur MCO de σ^{2K} dans le modèle de Kennedy est $S^{2K*} = \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 / (n - 2)$, c'est-à-dire que la somme des carrés des résidus est divisée par $(n - 2)$ au lieu de $(n - 3)$. A partir de l'équation (4.21), un estimateur MCO de la variance de $\hat{\beta}_{2,\pi}$ sous le modèle de Kennedy s'écrit dans ce cas :

$$\widehat{\text{Var}}(\hat{\beta}_{2,\pi})_{K*} = S^2(\beta_{2,\pi})_{K*} = \frac{S^2(X^1)S^{2K*}}{n\Delta} = \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n - 2)\Delta} \quad (3.59)$$

Dans la proposition (3.4.1) où nous avons retenu $(n - 3)$ degrés de liberté pour la statistique de test par la méthode permutational de Kennedy et nous avons alors montré que l'estimation MCO de $\text{Var}(\hat{\beta}_{2,\pi})$ par la méthode de Freedman-Lane est inférieure à celle qui était fourni par la méthode de Kennedy. Retenons maintenant, à partir de l'équation ci-dessus, la statistique de test de Kennedy corrigée en considérant $(n - 2)$ degrés de liberté, nous montrons dans la proposition suivante qu'il existe alors des cas où l'estimateur de $\text{Var}(\hat{\beta}_{2,\pi})$ obtenu par la méthode de Freedman-Lane est supérieur à celui qui résulte de la méthode de Kennedy corrigée.

Proposition 3.4.4. *Pour la méthode de Kennedy, corrigée en retenant $(n - 2)$ degrés de liberté et non plus $(n - 3)$ dans la définition de la statistique de test, on vérifie :*

$$S^2(\hat{\beta}_{2,\pi})_{K*} - S^2(\hat{\beta}_{2,\pi})_F \begin{cases} \geq 0 & \text{si } [S(X^1, R^{Y.1,\pi_i})]^2 \geq \Delta S^2(\hat{\beta}_{2,\pi})_F \\ < 0 & \text{si } [S(X^1, R^{Y.1,\pi_i})]^2 < \Delta S^2(\hat{\beta}_{2,\pi})_F \end{cases}$$

$$\text{où } S^2(\beta_{2,\pi})_{K^*} = \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-2)\Delta} \text{ et } S^2(\beta_{2,\pi})_F = \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n(n-3)\Delta}$$

Preuve : Les équations (3.21) et (3.59) donnent :

$$\begin{aligned} S^2(\hat{\beta}_{2,\pi})_{K^*} - S^2(\hat{\beta}_{2,\pi})_F &= \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-2)\Delta} - \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n(n-3)\Delta} \\ &= \frac{S^2(X^1) \left[(n-3) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - (n-2) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 \right]}{n(n-2)(n-3)\Delta} \end{aligned}$$

les valeurs $S^2(X^1)$, Δ et n sont positives , donc :

$$S^2(\hat{\beta}_{2,\pi})_{K^*} - S^2(\hat{\beta}_{2,\pi})_F \begin{cases} \geq 0 & \text{si } \left[(n-3) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - (n-2) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 \right] \geq 0 \\ < 0 & \text{si } \left[(n-3) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - (n-2) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 \right] < 0 \end{cases}$$

Mais,

$$\begin{aligned} \left[(n-3) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - (n-2) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 \right] &\geq 0 \quad \text{ce qui confirme que} \\ \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n} &\geq \frac{n-2}{n-3} \quad \text{et donc} \\ \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{\sum_{i=1}^n (\hat{\epsilon}_i^F)^2} &\geq \frac{1}{n-3} \quad \text{d'après le lemme (3.4.1)} \\ \frac{n[S(X^1, R^{Y.1,\pi i})]^2}{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2} &\geq \frac{1}{n-3} \quad \text{selon la définition de } S^{2F} \\ [S(X^1, R^{Y.1,\pi i})]^2 &\geq \frac{S^2(X^1) S^{2F}}{n} \quad \text{d'après la définition de } S^2(\hat{\beta}_{2,\pi})_F \\ [S(X^1, R^{Y.1,\pi i})]^2 &\geq \Delta S^2(\hat{\beta}_{2,\pi})_F \end{aligned}$$

Enfin :

$$S^2(\hat{\beta}_{2,\pi})_{K^*} - S^2(\hat{\beta}_{2,\pi})_F \begin{cases} \geq 0 & \text{si } [S(X^1, R^{Y.1,\pi i})]^2 \geq \Delta S^2(\hat{\beta}_{2,\pi})_F \\ < 0 & \text{si } [S(X^1, R^{Y.1,\pi i})]^2 < \Delta S^2(\hat{\beta}_{2,\pi})_F \end{cases} \quad \square \quad (3.60)$$

Contrairement à la proposition (3.4.1) qui disait qu'on avait toujours $S^2(\hat{\beta}_{2,\pi})_K \geq S^2(\hat{\beta}_{2,\pi})_F$, on vérifie ainsi que retenir $(n - 2)$ degrés de liberté pour la statistique de test utilisée par la méthode de Kennedy corrigée, assure que si le carré de la covariance permutationnelle $[S(X^1, R^{Y.1,\pi i})]^2$ est inférieur à $\Delta S^2(\hat{\beta}_{2,\pi})_F$, on a $S^2(\hat{\beta}_{2,\pi})_{K^*} < S^2(\hat{\beta}_{2,\pi})_F$

Corollaire 3.4.2. *Pour effectuer un test d'hypothèse $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$, en considérant $(n - 2)$ degrés de liberté pour la méthode de Kennedy :*

1. *il existe des cas où :* $Pr_{H_0}(P_{2,\pi}^F < \alpha) > Pr_{H_0}(P_{2,\pi}^{K^*} < \alpha)$

2. *il existe des cas où :* $Pr_{H_1}(P_{2,\pi}^F < \alpha) > Pr_{H_1}(P_{2,\pi}^{K^*} < \alpha)$

Preuve : D'après la proposition (3.4.4) on a :

$$S^2(\hat{\beta}_{2,\pi})_{K^*} - S^2(\hat{\beta}_{2,\pi})_F \begin{cases} \geq 0 & \text{si } [S(X^1, R^{Y.1,\pi i})]^2 \geq \Delta S^2(\hat{\beta}_{2,\pi})_F \\ < 0 & \text{si } [S(X^1, R^{Y.1,\pi i})]^2 < \Delta S^2(\hat{\beta}_{2,\pi})_F \end{cases}$$

De ce fait, pour la méthode de Kennedy corrigée en considérant $(n - 2)$ degrés de liberté et non plus $(n - 3)$, il existe des cas où $S^2(\hat{\beta}_{2,\pi})_{K^*} < S^2(\hat{\beta}_{2,\pi})_F$. On en déduit qu'il existe des cas où $T_{2,\pi}^{K^*} = \hat{\beta}_{2,\pi}/S(\hat{\beta}_{2,\pi})_{K^*} > T_{2,\pi}^F = \hat{\beta}_{2,\pi}/S(\hat{\beta}_{2,\pi})_F$. Donc, contrairement au corollaire (3.4.1) on ne peut plus toujours affirmer que la p-valeur permutationnelle de la méthode de Freedman-Lane est supérieure ou égale à celle qui est déduite de la méthode de Kennedy ($p_{2,\pi}^F \geq p_{2,\pi}^{K^*}$). Autrement dit, la relation :

$$Pr(P_{2,\pi}^F < \alpha) \leq Pr(P_{2,\pi}^{K^*} < \alpha)$$

n'est pas toujours vraie si l'on retient la définition de $S^2(\hat{\beta}_{2,\pi})_{K^*}$ corrigée avec $(n - 2)$ degrés de liberté. En effet, il existe des cas où

$$Pr(P_{2,\pi}^F < \alpha) > Pr(P_{2,\pi}^{K^*} < \alpha) \quad (3.61)$$

Par conséquent,

1. pour des données générées sous l'hypothèse $H_0 : \beta_2 = 0$, la relation ci-dessus autorise l'existence de cas où :

$$Pr_{H_0}(P_{2,\pi}^F < \alpha) > Pr_{H_0}(P_{2,\pi}^{K^*} < \alpha) \quad (3.62)$$

Cette inégalité montre qu'il existe des cas où le risque d'erreur de première espèce de la méthode de Kenney avec $(n - 2)$ degrés de liberté est inférieur à celui de Freedman-Lane. C'est ce que confirment nos simulations présentées en section (2.5.1).

2. Par contre, lorsque les données sont générées sous l'hypothèse $H_1 : \beta_2 \neq 0$ la relation (3.55) autorise l'existence de :

$$Pr_{H_1}(P_{2,\pi}^F < \alpha) > Pr_{H_1}(P_{2,\pi}^{K*} < \alpha) \quad (3.63)$$

Autrement dit, il existe des cas où la puissance du test suivant la méthode de Freedman-Lane est supérieure à celle du test suivant la méthode de Kennedy corrigée. \square

Remarque 3.4.3.

1. Dans toutes les simulations que nous avons effectuées, le nombre de degrés de liberté retenu était $(n - 2)$ contrairement aux simulations d'Anderson et Legendre qui ont retenu $(n - 3)$ degrés de liberté pour la méthode de Kennedy. Comme montré par nos simulation en section(2.5.1), il existe des cas où le risque d'erreur de première espèce encouru par la méthode de Kennedy est inférieur à celui de la méthode de Freedman-Lane. La relation (3.56) donne donc une explication aux faits observés grâce à nos simulations. Notons cependant que les résultats des simulations présentées en section 2.5.1, montrent que dans la plupart des cas, le risque d'erreur de première espèce de la méthode de Freedman-Lane est plus petit que celui de la méthode de Kennedy corrigée.

2. Cependant, lors de l'étude des puissances de tests par des méthodes permutationnelles, la méthode de Kennedy n'est considérée ni dans les simulations d'Anderson et Legendre [1], ni dans celles de Legendre (2000), ni dans la notre (section (2.5.2)). Donc, en considérant la relation (3.62), nous avons repris (dans le même cadre que la section (2.5.2)), la vérification des puissances de tests par des méthodes permutationnelles de Kennedy et de Freedman-Lane, et comme nous l'avons vu dans cette section (2.5.3), il existe des cas où la puissance de test de la méthode de Kennedy est inférieure à celle oqui est btendue par la méthode de Freedman-Lane. Ce constat empirique est corroboré par le résultat mathématique décrit dans la relation (3.63).

Proposition 3.4.5. *Comme dans la proposition (3.4.3) où nous avons considéré $(n - 3)$ degrés de liberté pour la méthode de Kennedy, ici, en retenant $(n - 2)$ degrés de liberté pour la méthode de Kennedy et en augmentant la taille de l'échantillon de données, la différence entre les deux estimateurs de $\mathbb{V}ar(\hat{\beta}_{2,\pi})$, tend vers zéro. En effet :*

$$\lim_{n \rightarrow +\infty} [S^2(\hat{\beta}_{2,\pi})_{K*} - S^2(\hat{\beta}_{2,\pi})_F] = 0$$

Preuve :

$$\begin{aligned}
 [S^2(\hat{\beta}_{2,\pi})_{K^*} - S^2(\hat{\beta}_{2,\pi})_F] &= \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-2)\Delta} - \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n(n-3)\Delta} \\
 &= \frac{S^2(X^1)}{\Delta} \left[\frac{(n-3) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - (n-2) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n(n-2)(n-3)} \right]
 \end{aligned} \tag{3.64}$$

Par suite, si $S^2(X)$, $S(X^1, R^{Y.1,\pi})$ et Δ convergent vers une limite finie σ_1^2 , $\sigma_{1R^{Y.1,\pi}}^2$, $\Delta^* > 0$ et respectivement, lorsque n tend vers l'infini, on a :

$$\begin{aligned}
 \lim_{n \rightarrow +\infty} [S^2(\hat{\beta}_{2,\pi})_{K^*} - S^2(\hat{\beta}_{2,\pi})_F] &= \frac{\sigma_1^2}{\Delta^*} \lim_{n \rightarrow +\infty} \left[\frac{(n-3) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - (n-2) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n(n-2)(n-3)} \right] \\
 &\approx \frac{\sigma_1^2}{\Delta^*} \lim_{n \rightarrow +\infty} \left[\frac{(\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\epsilon}_i^F)^2)}{n^2} \right] \\
 \text{par le lemme (3.4.1)} &= \frac{\sigma_1^2}{\Delta^*} \lim_{n \rightarrow +\infty} \frac{nS(X^1, R^{Y.1,\pi})}{n^2 S^2(X^1)} \\
 &= \frac{\sigma_1^2 \sigma_{1R^{Y.1,\pi}}}{\Delta^* \sigma_1^2} \lim_{n \rightarrow +\infty} \frac{1}{n} = 0 \quad \square
 \end{aligned}$$

Les deux estimateurs sont donc asymptotiquement égaux. En pratique, lorsque la taille de l'échantillon de données est assez grande, les statistiques de test associées respectivement à la méthode de Freedman-Lane et de celle de Kennedy corrigée avec $(n-2)$ degrés de liberté, sont approximativement égales :

$$T_{2,\pi}^F = \frac{\hat{\beta}_{2,\pi}}{S(\hat{\beta}_{2,\pi})_F} \approx \frac{\hat{\beta}_{2,\pi}}{S(\hat{\beta}_{2,\pi})_{K^*}} = T_{2,\pi}^{K^*}$$

Par conséquent, dans cette situation, on doit avoir approximativement le même risque d'erreur de première espèce, ce que confirme nos simulations lorsque $n \geq 36$, cf. la section (2.5.1). De même, les deux tests ont approximativement la même puissance ce que confirme notre simulation, pour $n \geq 36$, cf. la section (2.5.3)).

3.4.4 Influence de la normalisation

Afin de vérifier l'influence de la normalisation par $(n - 2)$, ou plus généralement par $(n - k)$ si $k = p + 1$ et p est le nombre de variables explicatives dans le modèle, sur le risque d'erreur de première espèce, Anderson et Legendre [1] ont effectué une simulation sur une petite étendue du domaine de valeurs, retenant les 4 facteurs de variation des caractéristiques du test étudiées plus haut. Ils notent [1] en particulier (page 289) que *pour tous les cas, la méthode de Kennedy corrigée avec $(n - 2)$ degrés de liberté augmente moins le risque d'erreur de première espèce que celle avec $(n - k)$ degrés de liberté*. C'est dire que dans tous les cas, le risque d'erreur de première espèce de la méthode de Kennedy corrigée avec $(n - 2)$ degrés de liberté est inférieure ou égale à celui de la même méthode avec $(n - k)$ degrés de liberté. Nous en montrons la raison plus bas pour $k = 3$. Plus généralement, nous montrons dans la proposition suivante que l'estimateur de $\text{Var}(\hat{\beta}_{2,\pi})$ par la méthode de Kennedy corrigée avec $(n - 2)$ degrés de liberté, est toujours inférieur ou égal à celui de la même méthode avec $(n - 3)$ degrés de liberté. Par suite, le corollaire suivant montre que le risque d'erreur de première espèce de la méthode de Kennedy corrigée avec $(n - 2)$ degrés de liberté est toujours inférieur ou égal à celui de la même statistique divisé par $(n - 3)$ degrés de liberté. Dans la section (3.8), nous étendons ces résultats au cas général $k > 3$.

Proposition 3.4.6.

$$S^2(\hat{\beta}_{2,\pi})_K - S^2(\hat{\beta}_{2,\pi})_{K^*} = \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-2)(n-3)\Delta} \geq 0$$

Preuve : Rappelons que l'estimateur MCO de $\text{Var}(\hat{\beta}_{2,\pi})$ dans la méthode de Kennedy, s'écrit dans le cas corrigé et non corrigé respectivement,

$$\begin{aligned} S^2(\hat{\beta}_{2,\pi})_{K^*} &= \frac{S^2(X^1)S^{2K^*}}{n\Delta} = \frac{S^2(X^1)}{n\Delta} \cdot \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n-2} \\ S^2(\hat{\beta}_{2,\pi})_K &= \frac{S^2(X^1)S^{2K}}{n\Delta} = \frac{S^2(X^1)}{n\Delta} \cdot \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n-3} \end{aligned}$$

donc :

$$\begin{aligned} S^2(\hat{\beta}_{2,\pi})_K - S^2(\hat{\beta}_{2,\pi})_{K^*} &= \frac{S^2(X^1)}{n\Delta} (S^{2K} - S^{2K^*}) \\ S^2(\hat{\beta}_{2,\pi})_K - S^2(\hat{\beta}_{2,\pi})_{K^*} &= \frac{S^2(X^1)}{n\Delta} \left(\frac{\sum_{i=1}^n \epsilon_i^K}{n-3} - \frac{\sum_{i=1}^n \epsilon_i^K}{n-2} \right) \\ &= \frac{S^2(X^1)}{n\Delta} \frac{\sum_{i=1}^n \epsilon_i^K}{(n-2)(n-3)} \end{aligned} \tag{3.65}$$

Par suite, puisque $S^2(X^1)$, $\sum_{i=1}^n (\hat{\epsilon}_i^K)^2$, n et Δ sont des valeurs positive

$$S^2(\hat{\beta}_{2,\pi})_K \geq S^2(\hat{\beta}_{2,\pi})_{K^*} \quad \square \tag{3.66}$$

L'équation ci-dessus a en particulier trois conséquences utiles en pratique :

1. La valeur absolue de la statistique de test utilisée par la méthode de Kennedy corrigée par $(n - 2)$ degrés de liberté est toujours supérieure à celle de la méthode initiale de $(n - 3)$ degrés de liberté :

$$|T_{2,\pi}^{K*}| = \frac{|\hat{\beta}_{2,\pi}|}{S^2(\hat{\beta}_{2,\pi})_{K*}} \geq \frac{|\hat{\beta}_{2,\pi}|}{S^2(\hat{\beta}_{2,\pi})_K} = |T_{2,\pi}^K| \quad (3.67)$$

2. La distribution permutationnelle de $|T_{2,\pi}^{K*}|$ a une dispersion supérieure ou égale à celle de $|T_{2,\pi}^K|$.

3. La p-valeur permutationnelle de la forme corrigée de la méthode de Kennedy est inférieure ou égale à celle de la forme initiale de cette méthode, c'est-à-dire :

$$p_{2,\pi}^{K*} \geq p_{2,\pi}^K \quad (3.68)$$

Corollaire 3.4.3. *Le test de permutation de l'hypothèse $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$, pour la méthode de Kennedy vérifie les propriétés suivantes*

1. *Le risque d'erreur de première espèce de la forme corrigée est inférieur ou égal à celui de la forme initiale.*
2. *Le puissance du test de la méthode de la forme corrigée est inférieure ou égale à celle de la forme initiale du test par cette méthode avec $(n-3)$ degrés de liberté.*

Preuve : D'après la remarque ci-dessus, on a :

$$p_{2,\pi}^{K*} \geq p_{2,\pi}^K$$

De façon analogue au corollaire (3.4.1) on montre donc que la probabilité de rejeter $H_0 : \beta_2 = 0$ par la forme corrigée est inférieure ou égale à celle qui est obtenue par la forme initiale de cette méthode.

$$Pr(P_{2,\pi}^{K*} < \alpha) \leq Pr(P_{2,\pi}^K < \alpha) \quad (3.69)$$

Donc, la fonction puissance permutationnelle de la méthode de la forme corrigée est inférieure ou égale à celle qui est obtenue avec la forme initiale.

Autrement dit, si C est la région critique du test, pour les formes corrigée et initiale, on rejette $H_0 : \beta_2 = 0$ quand la statistique de test appartient à la région critique. Donc :

$$\eta_{\pi}^{K*}(\beta_2) = Pr_{\pi}^{K*}(C) \leq Pr_{\pi}^K(C) = \eta_{\pi}^K(\beta_2) \quad (3.70)$$

Par conséquent, pour des données générées sous l'hypothèse H_0 , la relation ci-dessus dit que le risque d'erreur de première espèce de la forme corrigée est inférieur à celui de la forme initiale de cette méthode.

$$Pr_{H_0}(P_{2,\pi}^{K*} < \alpha) \leq Pr_{H_0}(P_{2,\pi}^K < \alpha) \quad (3.71)$$

Ces résultats donnent un support mathématique aux résultats des simulations d'Anderson et Legendre [1](page 289). Par contre, lorsque les données sont générées sous l'hypothèse $H_1 : \beta_2 \neq 0$ la relation ci-dessus (3.70) dit que la puissance de test de la forma corrigée de la méthode de Kennedy est inférieure ou égale à celle de la forme initiale de cette méthode.

$$Pr_{H_1}(P_{2,\pi}^{K*} < \alpha) \leq Pr_{H_1}(P_{2,\pi}^K < \alpha) \quad (3.72)$$

Ceci n'est pas vérifié dans les simulations d'Anderson et Legendre.

D'autre part, ces auteurs remarquent que *dans les simulations, le risque d'erreur de première espèce de la méthode de Freedman-Lane est toujours inférieur à celui de la méthode de Kennedy, avec $(n - k)$ ou $(n - 2)$ degrés de liberté*. Nous sommes d'accord avec la partie de cette remarque qui concerne la forme du test de Kennedy en $(n - k)$ avec $k \geq 3$ et la comparaison des risques d'erreur de première espèce de la méthode de Freedman-Lane et de la méthode de Kennedy avec $(n - k)$ degrés de liberté, puisque pour $k = 3$ nous avons démontré ce résultats dans la proposition (3.2.2). Nous étendons même ce résultat au cas général $k \geq 3$ dans la section (3.8). Mais, nous contestons la partie de la remarque de ces auteurs qui étend leur constat à la forme du test de Kennedy corrigée en $(n - 2)$ degrés de liberté. Il nous semble que cette simulation d'Anderson et Legendre ne permet pas d'observer des cas où le risque d'erreur de première espèce de la méthode de Freedman-Lane est supérieur à celui de la méthode de Kennedy sous forme normalisée en $(n - 2)$, du fait de la faible étendue du domaine des valeurs des facteurs retenues par ces auteurs (*cf.* [1] page 289). C'est contraire à ce que nous avons observé dans nos simulations pour $k = 3$ (Voir section (2.5.1)) et à ce que nous avons démontré ensuite dans la proposition (3.4.6) et nous démontrerons ce résultat pour $k > 3$, plus bas.

On doit de plus remarquer que lorsque la taille de l'échantillon tend vers l'infini, les estimateurs de $\mathbb{V}ar(\hat{\beta}_{2,\pi})$ associée aux formes en $(n - 2)$ et $(n - 3)$ degrés de liberté de la méthode de Kenndy sont asymptotiquement égaux,

Proposition 3.4.7. *Lorsque la taille de l'échantillon de données tend vers l'infini, la différence entre les deux estimateurs $S^2(\beta_{2,\pi}^k)_K$ et $S^2(\beta_{2,\pi})_{K*}$ de $\mathbb{V}ar(\hat{\beta}_{2,\pi})$, tend vers zéro :*

Preuve : La proposition (3.4.6) dit que :

$$S^2(\beta_{2,\pi})_K - S^2(\beta_{2,\pi})_{K*} = \frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-2)(n-3)\Delta} \geq 0$$

$$\lim_{n \rightarrow +\infty} [S^2(\hat{\beta}_{2,\pi})_K - S^2(\hat{\beta}_{2,\pi})_{K*}] = \lim_{n \rightarrow +\infty} \left[\frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-2)(n-3)\Delta} \right]$$

Si lorsque n vers infini, la matrice $S^2(X^1)$, S^{2K} et Δ converge respectivement vers σ_1^2 , σ^{2K} et $\Delta^* > 0$. Alors, :

$$\begin{aligned} \lim_{n \rightarrow +\infty} [S^2(\hat{\beta}_{2,\pi})_K - S^2(\hat{\beta}_{2,\pi})_{K^*}] &\approx \frac{\sigma_1^2 \sigma^{2K}}{\Delta^*} \lim_{n \rightarrow +\infty} \frac{1}{n^2} \\ &\approx 0 \quad \square \end{aligned}$$

Les statistiques $|T_{2,\pi}^K|$ et $|T_{2,\pi}^{K^*}|$ sont donc asymptotiquement égales et pour effectuer un test de l'hypothèse $H_0 : \beta_2 = 0$ contre $H_0 : \beta_2 \neq 0$, ces deux statistiques donnent asymptotiquement les mêmes risques d'erreur de première espèce et les mêmes puissances. C'est que les simulations d'Anderson et Legendre (*cf.* [1] page 289) permettent de constater empiriquement.

3.5 Espérance permutationnelle des variables

Cette section est consacrée au calcul de la variance et de l'espérance permutationnelle de $\hat{\beta}_{2,\pi}$ ainsi qu'aux espérances permutationnelles des estimateurs de $\text{Var}(\hat{\beta}_{2,\pi})$ construits selon les méthodes de Freedman-Lane et de Kennedy.

3.5.1 Cas univarié

Dans cette section, nous utilisons les résultats de Czaplewski *et al.* ([7]) pour calculer les deux premiers moments permutationnels d'une variable aléatoire d'intérêt.

Lemme 3.5.1. *Soit Πi un indice permuté de l'ensemble I_n et $x^{\Pi i}$ la variable aléatoire permutée associée. Alors, les espérances permutationnelles de $x^{\Pi i}$ et de $(x^{\Pi i})^2$ calculées sur toutes les $n!$ permutations possibles sont :*

1. $\mathcal{E}(x^{\Pi i}) = \frac{1}{n} \sum_{p=1}^n x_p = \bar{x}$
2. $\mathcal{E}(x^{\Pi i})^2 = \frac{1}{n} \sum_{i=1}^n x_p^2 = \overline{x^2}$

Preuve :

1. Par définition de l'espérance mathématique, on a :

$$\mathcal{E}(x^{\Pi i}) = x_1 p(x^{\Pi i} = x_1) + x_2 p(x^{\Pi i} = x_2) + \dots + x_n p(x^{\Pi i} = x_n)$$

où $p(x^{\Pi i} = x_p)$ est la probabilité avec laquelle de la variable permutée $x^{\Pi i}$ prend la valeur observée x_p ($1 \leq p \leq n$) lors des $n!$ permutations possibles. Parmi ces $n!$ permutations possibles, il existe $(n-1)!$ permutations pour lesquelles $x^{\Pi i} = x_p$ et donc

$$p(x^{\Pi i} = x_p) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

Par suite :

$$\mathcal{E}(x^{\Pi i}) = \frac{1}{n} \sum_{p=1}^n x_p = \bar{x} \quad (3.73)$$

2. On montre de façon analogue, que :

$$\mathcal{E}(x^{\Pi i})^2 = \frac{1}{n} \sum_{i=1}^n x_p^2 = \overline{x^2} \quad \square \quad (3.74)$$

3.5.2 Cas bivarié

Dans cette section, nous calculons l'espérance permutationnelle d'un couple de variables ainsi que son moment d'ordre 2. Nous considérons un échantillon de taille n observations appariées du couple (X, Y) et nous notons $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ l'observation du vecteur aléatoire $\mathbf{Z} = (\mathbf{X}, \mathbf{Y}) \in \mathbb{R} \times \mathbb{R}$ défini par :

$$\{Z_1, Z_2, \dots, Z_n\} = \{(X_1, Y_1), (X_2, Y_2) \dots (X_n, Y_n)\} \quad (3.75)$$

Toute permutation π définie sur \mathbf{I}_n induit une permutation des coordonnées du vecteur \mathbf{z} indicé par \mathbf{I}_n que l'on note encore π :

$$\mathbf{z} = (z_1, z_2, \dots, z_n) \xrightarrow{\pi} \mathbf{z}^\pi = (z^{\pi 1}, z^{\pi 2}, \dots, z^{\pi n}) \quad (3.76)$$

Ici, $z = (x^i, y^i)$ et donc $z^{\pi i} = (x^{\pi i}, y^{\pi i})$. Autrement dit :

$$\mathbf{z} = ((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)) \xrightarrow{\pi} \mathbf{z}^\pi = ((x^{\pi 1}, y^{\pi 1}), (x^{\pi 2}, y^{\pi 2}), \dots, (x^{\pi n}, y^{\pi n})) \quad (3.77)$$

Lemme 3.5.2. $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ un vecteur aléatoire bivarié observé, défini sur (Ω, \mathcal{F}, P) . Soit \mathbf{z} un échantillon de taille n tiré suivant la loi de \mathbf{Z} indicé par l'ensemble \mathbf{I}_n . Soit $(x^{\Pi i}, y^{\Pi i})$ la variable aléatoire permutée associée. Alors, les espérances permutationnelles de $(x^{\Pi i} y^{\Pi i})$ et de $((x^{\Pi i})^2 (y^{\Pi i})^2)$, calculées sur toutes les $n!$ permutations possibles de \mathbf{z} , sont :

1. $\mathcal{E}(x^{\Pi i} y^{\Pi i}) = \frac{1}{n} \sum_{p=1}^n x_p y_p = \overline{xy}$
2. $\mathcal{E}((x^{\Pi i})^2 (y^{\Pi i})^2) = \frac{1}{n} \sum_{p=1}^n x_p^2 y_p^2 = \overline{x^2 y^2}$

Preuve :

1. Par définition de l'espérance :

$$\mathcal{E}(x^{\Pi i} y^{\Pi i}) = \sum_{k=1}^n x_k y_k p(x^{\Pi i} = x_k, y^{\Pi i} = y_k)$$

où $p(x^{\Pi i} = x_k, y^{\Pi i} = y_k)$ est la probabilité avec laquelle la variable permutée $(x^{\Pi i}, y^{\Pi i})$, $(1 \leq i \leq n)$ prend les valeurs observées (x_p, y_p) , $(1 \leq p \leq n)$ lors des $n!$ permutations possibles. On sait que $(n-1)!$ permutations parmi les $n!$ permutations possibles sont telles que $x^{\Pi i} = x_p, y^{\Pi i} = y_p$. Par suite :

$$p(x^{\Pi i} = x_p, y^{\Pi i} = y_p) = \frac{(n-1)!}{n!} = \frac{1}{n}, \quad (1 \leq p \leq n)$$

et donc :

$$\mathcal{E}(x^{\Pi i} \cdot y^{\Pi i}) = \frac{1}{n} \sum_{p=1}^n x_p y_p = \overline{xy} \quad (3.78)$$

2. On montre de façon analogue, que :

$$\mathcal{E}((x^{\Pi i})^2 \cdot (y^{\Pi i})^2) = \frac{1}{n} \sum_{p=1}^n x_p^2 y_p^2 = \overline{x^2 y^2} \quad \square \quad (3.79)$$

Lorsqu'elle est appliquée aux n composantes du vecteur \mathbf{x}

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \xrightarrow{\Pi} \mathbf{x}^{\Pi} = (x^{\Pi 1}, x^{\Pi 2}, \dots, x^{\Pi n})$$

la permutation aléatoire Π , définit n variables aléatoires $x^{\Pi i}$. Donc, on peut étudier la structure de répartition permutationnelle. Nous monterons dans le lemme suivant que ces variables $x^{\Pi 1}, x^{\Pi 2}, \dots, x^{\Pi n}$ sont dépendantes.

Lemme 3.5.3. *Considérons la permutation aléatoire Π , définie sur l'ensemble $\mathbf{x} = (x_1, x_2, \dots, x_n)$,*

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \xrightarrow{\Pi} \mathbf{x}^{\Pi} = (x^{\Pi 1}, \dots, x^{\Pi i}, \dots, x^{\Pi j}, \dots, x^{\Pi n})$$

Alors, pour tout couple $i, j = 1, \dots, n$, $i \neq j$, l'espérance permutationnelle $(x^{\Pi i} x^{\Pi j})$ calculée sur les $n!$ permutations possibles, vérifie :

$$\mathcal{E}(x^{\Pi i} \cdot x^{\Pi j}) = \frac{n}{n-1} \bar{x}^2 - \frac{1}{n-1} \overline{x^2} \neq \mathcal{E}(x^{\Pi i}) \mathcal{E}(x^{\Pi j}) = \bar{x}^2$$

Preuve :

$$\mathcal{E}(x^{\Pi i} \cdot x^{\Pi j}) = \sum_{p=1}^n \sum_{q \neq p=1}^n x_p x_q p(x^{\Pi i} = x_p, x^{\Pi j} = x_q)$$

où

$$p(x^{\Pi i} = x_p, x^{\Pi j} = x_q) = p(x^{\Pi j} = x_q | x^{\Pi i} = x_p) p(x^{\Pi i} = x_p).$$

Par ailleurs, nous savons que lorsque la marge $X^{\pi i}$ de la variable bivariée $(x^{\Pi i}, x^{\Pi j})$ prend la valeur x_p , la variable $x^{\Pi j}$ ($j \neq i$) ne peut pas prendre cette valeur x_p , puisque

x_p est prise par la variable $x^{\Pi i}$. Parmi les $n!$ permutations possibles, il en existe donc $(n-1)!$ telles que $x^{\Pi i} = x_p$. Alors,

$$p(x^{\Pi i} = x_p) = p(x^{\Pi i} = x_p, x^{\Pi i} = x_p) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

De ce fait, seules $(n-1)!$ permutations parmi $n!$ possibles sont admissibles pour $(n-1)$ variables aléatoires $(x^{\Pi k})$, ($k \neq i$) et parmi ces $(n-1)!$ permutations admissible il existent $(n-2)!$ permutation telles que $x^{\Pi j} = x_q$ ($q \neq p$). Par suite :

$$p(x^{\Pi j} = x_q | x^{\Pi i} = x_p) = \frac{(n-2)!}{(n-1)!} = \frac{1}{n-1}$$

et donc :

$$p(x^{\Pi i} = x_p, x^{\Pi j} = x_q) = p(x^{\Pi j} = x_q | x^{\Pi i} = x_p) p(x^{\Pi i} = x_p) = \frac{1}{n(n-1)}$$

Par conséquent :

$$\begin{aligned} \mathcal{E}(x^{\Pi i} . x^{\Pi j}) &= \frac{1}{n(n-1)} \sum_{p=1}^n \sum_{q \neq p=1}^n x_p x_q = \frac{1}{n(n-1)} \sum_{p=1}^n x_p \left(\sum_{q=1}^n x_q - x_p \right) \\ \mathcal{E}(x^{\Pi i}, x^{\Pi j}) &= \frac{n}{n-1} (\bar{x})^2 - \frac{1}{n-1} \overline{x^2} \quad \square \end{aligned} \quad (3.80)$$

L'expression suivante concerne l'espérance permutationnelle des produits associés de puissances de $x^{\Pi i}$ et $x^{\Pi j}$. On peut les démontrer dans le cas général, mais nous avons seulement besoin ici du résultat particulier suivant :

Lemme 3.5.4. *Soit Π une permutation aléatoire défini sur l'ensemble $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, $i, j = 1, 2, \dots, n$, et s'écrit $i \neq j$*

1. $\mathcal{E}((x^{\Pi i})^2 . (x^{\Pi j})^2) = \frac{n}{n-1} (\overline{x^2})^2 - \frac{1}{n} \overline{x^4}$
2. $\mathcal{E}((x^{\Pi i})^2 . x^{\Pi j}) = \frac{n}{n-1} \overline{x^2} \bar{x} - \frac{1}{n} \overline{x^3}$

3.6 Moments permutationnels de $\hat{\beta}_{2,\pi}$

Dans cette section, nous utilisons les résultats obtenus sous forme de lemmes dans la section précédente pour calculer l'espérance et la variance permutationnelle de $\hat{\beta}_{2,\pi}$ lorsque cette statistique est l'estimateur MCO de $\beta_{2,\pi}$ par la méthode de Freedman-Lane et par la méthode de Kennedy. Nous commençons par l'espérance de $\beta_{2,\pi}$

3.6.1 Espérance permutationnelle de $\hat{\beta}_{2,\pi}$

D'après la proposition (3.2.1) on a :

$$\Delta \cdot \hat{\beta}_{2,\pi} = \Delta_2 = S^2(X^1)S(X^2, R^{Y.1,\pi}) - S(X^1, R^{Y.1,\pi})S(X^1, X^2)$$

Donc, afin de calculer l'espérance permutationnelle de $\hat{\beta}_{2,\pi}$, nous devons connaître les espérances permutationnelles des statistiques $S(X^1, R^{Y.1,\pi})$ et $S(X^2, R^{Y.1,\pi})$.

Lemme 3.6.1.

Pour toute permutation Π définie sur \mathbf{I}_n , le vecteur des résidus $\mathbf{R}^{Y.1,\pi}$ utilisé dans les méthodes de Freedman-Lane et de Kennedy vérifie :

1. $\mathcal{E}(S(X^1, R^{Y.1,\pi})) = 0$
2. $\mathcal{E}(S(X^2, R^{Y.1,\pi})) = 0$

Preuve :

1. Par définition :

$$\begin{aligned} \mathcal{E}(S(X^1, R^{Y.1,\pi})) &= \mathcal{E}\left(\sum_{i=1}^n (x_i^1 - \bar{x}^1) R^{Y.1,\pi i}\right) \\ &= \sum_{i=1}^n (x_i^1 - \bar{x}^1) \mathcal{E}(R^{Y.1,\pi i}) \\ &= \sum_{i=1}^n (x_i^1 - \bar{x}^1) \frac{1}{n} \sum_{j=1}^n (R_j^{Y.1}) \end{aligned}$$

Puisque $\sum_i R^{Y.1,\pi i} = \sum_{i=1}^n R_i^{Y.1} = 0$

$$\begin{aligned} \mathcal{E}(S(X^1, R^{Y.1,\pi i})) &= \sum_{i=1}^n (x_i^1 - \bar{x}^1) \times 0 \\ &= 0 \times 0 \\ &= 0 \end{aligned}$$

2. On montre bien sûr de la même façon, que :

$$\mathcal{E}(S(X^2, R^{Y.1,\pi})) = 0 \quad \square \tag{3.81}$$

Proposition 3.6.1. L'espérance permutationnelle de $\hat{\beta}_{2,\pi}$ est nulle : $\mathcal{E}(\hat{\beta}_{2,\pi}) = 0$

Preuve : Rappelons que Δ est invariante par permutation et donc :

$$\begin{aligned}\mathcal{E}(\Delta \cdot \hat{\beta}_{2\pi}) &= \Delta \mathcal{E}(\hat{\beta}_{2\pi}) \\ &= \mathcal{E}(\Delta_2) \\ &= \mathcal{E}(S^2(X^1)S(X^2, R^{Y.1, \pi}) - S(X^1, R^{Y.1, \pi})S(X^1, X^2))\end{aligned}$$

et donc ces expressions, $S^2(X^1)$ et $S(X^1, X^2)$ sont invariantes par permutation

$$\mathcal{E}(\Delta \cdot \hat{\beta}_{2\pi}) = S^2(X^1)\mathcal{E}(S(X^2, R^{Y.1, \pi})) - S(X^1, X^2)\mathcal{E}(S(X^1, R^{Y.1, \pi}))$$

on obtient alors la nullité de $\mathcal{E}(\hat{\beta}_{2, \pi})$ comme conséquence du lemme (3.6.1). On en déduit le résultat souhaité :

$$\mathcal{E}(\hat{\beta}_{2\pi}) = 0 \quad \square \quad (3.82)$$

3.6.2 La Variance permutationnelles de $\hat{\beta}_{2, \pi}$

Dans cette section, nous évaluons la variance permutationnelle de $\hat{\beta}_{2, \pi}$. Pour cela, nous devons d'abord démontrer quelques lemmes techniques.

Lemme 3.6.2. Soit $R^{Y.1, \pi}$ le vecteur résidu permuté et X^1 est de variance empirique non nulle on a :

$$\begin{aligned}1. \sum_{i=1}^n (R_i^{Y.1})^2 &= n \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \\ 2. \sum_{i=1}^n \sum_{j \neq i=1}^n R_i^{Y.1} R_j^{Y.1} &= -n \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right)\end{aligned}$$

Preuve :

$$\begin{aligned}1. \sum_{i=1}^n (R_i^{Y.1})^2 &= \sum_{i=1}^n (Y_i - a_0 - a_1 x_i^1)^2 \\ &= \sum_{i=1}^n (Y_i - \bar{Y} - a_1(x_i^1 - \bar{x}^1))^2 \\ &= \sum_{i=1}^n (Y_i - \bar{Y})^2 + a_1^2 \sum_{i=1}^n (x_i^1 - \bar{x}^1)^2 - 2a_1 \sum_{i=1}^n (x_i^1 - \bar{x}^1)(Y_i - \bar{Y}) \\ &= nS^2(Y) + n \frac{[S(X^1, Y)]^2}{[S^2(X^1)]^2} S^2(X^1) - 2n \frac{S(X^1, Y)}{S^2(X^1)} S(X^1, Y) \\ &= n \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \quad (3.83)\end{aligned}$$

Notons que si l'on suppose que $S^2(X^1) \neq 0$.

$$\begin{aligned}
2. \sum_{i=1}^n \sum_{j \neq i=1}^n R_i^{Y,1} R_j^{Y,1} &= \sum_{i=1}^n \sum_{j \neq i=1}^n (Y_i - a_0 - a_1 x_i^1)(Y_j - a_0 - a_1 x_j^1) \\
&= \sum_{i=1}^n \sum_{j \neq i=1}^n ((Y_i - \bar{Y}) - a_1(x_i^1 - \bar{x}^1))((Y_j - \bar{Y}) - a_1(x_j^1 - \bar{x}^1)) \\
&= \sum_{i=1}^n \sum_{j \neq i=1}^n ((Y_i - \bar{Y})(Y_j - \bar{Y}) + a_1^2 \sum_{i=1}^n \sum_{j \neq i=1}^n (x_i^1 - \bar{x}^1)(x_j^1 - \bar{x}^1) \\
&\quad - 2a_1 \sum_{i=1}^n \sum_{j \neq i=1}^n (Y_i - \bar{Y})(x_j^1 - \bar{x}^1)) \\
&= -nS^2(Y) - n \frac{[S(X^1, Y)]^2}{[S^2(X^1)]^2} S^2(X^1) + 2n \frac{S(X^1, Y)}{S^2(X^1)} S(X^1, Y) \\
&= -n \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \quad \square \tag{3.84}
\end{aligned}$$

Lemme 3.6.3. Si $R^{Y,1,\pi}$, le vecteur résidu permuté et X^1 , est de variance empirique non nulle, on a :

$$\begin{aligned}
1. \mathcal{E}(R^{Y,1,\pi i})^2 &= S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \\
2. \mathcal{E}(R^{Y,1,\pi i} R^{Y,1,\pi j}) &= -\frac{1}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right)
\end{aligned}$$

Preuve :

1. D'après du lemme (3.5.1), on a :

$$\mathcal{E}((R^{Y,1,\pi i})^2) = \frac{1}{n} \sum_{i=1}^n (R_i^{Y,1})^2$$

le lemme (3.6.2) permet d'écrire :

$$\mathcal{E}(R^{Y,1,\pi i})^2 = S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \tag{3.85}$$

2. D'après le lemme (3.5.3), on a de plus :

$$\mathcal{E}(R^{Y,1,\pi i} R^{Y,1,\pi j}) = \frac{n(\bar{R}^{Y,1})^2}{n-1} - \frac{\sum_{p=1}^n (R_p^{Y,1})^2}{n(n-1)}$$

et l'on sait que $\bar{R}^{Y,1} = 0$. Par suite, le résultat suivant est une conséquence du lemme (3.6.2) :

$$\mathcal{E}(R^{Y,1,\pi i} R^{Y,1,\pi j}) = -\frac{1}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \quad \square \tag{3.86}$$

Lemme 3.6.4. *On suppose que $S^2(X^1) \neq 0$. La variable $R^{Y,1,\pi}$ vérifie alors :*

1. $\mathcal{E}(S(X^1, R^{Y,1,\pi}))^2 = \frac{S^2(X^1)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right)$
2. $\mathcal{E}(S(X^2, R^{Y,1,\pi}))^2 = \frac{S^2(X^2)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right)$
3. $\mathcal{E}(S(X^1, R^{Y,1,\pi})S(X^2, R^{Y,1,\pi})) = \frac{S(X^1, X^2)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right)$

Preuve :

$$\begin{aligned}
1. \quad & \mathcal{E}(S(X^1, R^{Y,1,\pi}))^2 = \mathcal{E}\left(\frac{1}{n} \sum_{i=1}^n (X_i^1 - \bar{X}^1) R^{Y,1,\pi i}\right)^2 \\
&= \mathcal{E}\left(\frac{1}{n^2} \sum_{i=1}^n (X_i^1 - \bar{X}^1)^2 (R^{Y,1,\pi i})^2\right) + \mathcal{E}\left(\frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i=1}^n (X_i^1 - \bar{X}^1)(X_j^2 - \bar{X}^2) R^{Y,1,\pi i} R^{Y,1,\pi j}\right) \\
&= \frac{1}{n^2} \sum_{i=1}^n (X_i^1 - \bar{X}^1)^2 \mathcal{E}(R^{Y,1,\pi i})^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i=1}^n (X_i^1 - \bar{X}^1)(X_j^2 - \bar{X}^2) \mathcal{E}(R^{Y,1,\pi i} R^{Y,1,\pi j}) \\
&= \frac{1}{n} S^2(X^1) [\mathcal{E}(R^{Y,1,\pi i})^2 - \mathcal{E}(R^{Y,1,\pi i} R^{Y,1,\pi j})] \\
&\text{et le lemme (3.6.3) assure que :}
\end{aligned}$$

$$\mathcal{E}(S(X^1, R^{Y,1,\pi}))^2 = \frac{S^2(X^1)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \quad (3.87)$$

2. De façon analogue, on peut montrer que :

$$\mathcal{E}(S(X^2, R^{Y,1,\pi}))^2 = \frac{S^2(X^2)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \quad (3.88)$$

3.

$$\begin{aligned}
& \mathcal{E}(S(X^1, R^{Y,1,\pi})S(X^2, R^{Y,1,\pi})) \\
&= \mathcal{E}\left(\frac{1}{n} \sum_{i=1}^n (X_i^1 - \bar{X}^1) R^{Y,1,\pi i} \frac{1}{n} \sum_{i=1}^n (X_i^2 - \bar{X}^2) R^{Y,1,\pi i}\right) \\
&= \mathcal{E}\left(\frac{1}{n^2} \sum_{i=1}^n (X_i^1 - \bar{X}^1)(X_i^2 - \bar{X}^2) (R^{Y,1,\pi i})^2\right) \\
&+ \mathcal{E}\left(\frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i=1}^n (X_i^1 - \bar{X}^1)(X_j^2 - \bar{X}^2) R^{Y,1,\pi i} R^{Y,1,\pi j}\right) \\
&= \frac{1}{n} S(X^1, X^2) (\mathcal{E}(R^{Y,1,\pi i})^2) - \frac{1}{n} S(X^1, X^2) \mathcal{E}(R^{Y,1,\pi i} R^{Y,1,\pi j}) \\
&= \frac{1}{n} S(X^1, X^2) (\mathcal{E}(R^{Y,1,\pi i})^2 - \mathcal{E}(R^{Y,1,\pi i} R^{Y,1,\pi j})) \quad (3.89)
\end{aligned}$$

En utilisant le lemme (3.6.4), on obtient ainsi :

$$\mathcal{E}(S(X^1, R^{Y.1,\pi})S(X^2, R^{Y.1,\pi})) = \frac{S(X^1, X^2)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \quad \square \quad (3.90)$$

Proposition 3.6.2. *Dans le cas non trivial ($\Delta \neq 0$), la variance permutationnelle de $\hat{\beta}_{2,\pi}$ est :*

$$\begin{aligned} \mathcal{V}ar(\hat{\beta}_{2,\pi}) &= \frac{S^2(X^1)(\mathcal{E}(R^{Y.1,\pi i})^2 - \mathcal{E}(R^{Y.1,\pi i} R^{Y.1,\pi j}))}{n\Delta} \\ &= \frac{S^2(X^1)S^2(Y) - [S(X^1, Y)]^2}{(n-1)\Delta} \end{aligned}$$

Preuve :

$$\mathcal{V}ar(\hat{\beta}_{2,\pi}) = \mathcal{E}(\hat{\beta}_{2,\pi})^2 - \mathcal{E}^2(\hat{\beta}_{2,\pi}) = \mathcal{E}(\hat{\beta}_{2,\pi})^2$$

par application de la proposition (3.6.1). Mais :

$$\begin{aligned} \Delta \cdot \mathcal{E}(\hat{\beta}_{2,\pi})^2 &= \mathcal{E} [S^2(X^1)S(X^2, R^{Y.1,\pi}) - S(X^1, R^{Y.1,\pi})S(X^1, X^2)]^2 \\ &= [S^2(X^1)]^2 \mathcal{E}[S(X^2, R^{Y.1,\pi})]^2 + S(X^1, X^2)]^2 \mathcal{E}[S(X^1, R^{Y.1,\pi})]^2 \\ &\quad - 2S^2(X^1)S(X^1, X^2)\mathcal{E}[S(X^1, R^{Y.1,\pi})S(X^2, R^{Y.1,\pi})] \end{aligned}$$

On utilise alors du lemme (3.6.4) pour écrire :

$$\mathcal{E}(\hat{\beta}_{2,\pi})^2 = \frac{S^2(X^1)[\mathcal{E}(R^{Y.1,\pi})^2 - \mathcal{E}(R^{Y.1,\pi i} R^{Y.1,\pi j})]}{n\Delta} \quad (3.91)$$

alors,

$$\mathcal{V}ar(\hat{\beta}_{2,\pi}) = \frac{S^2(X^1)S^2(Y) - [S(X^1, Y)]^2}{(n-1)\Delta} \quad \square \quad (3.92)$$

Cela montre que la variance permutationnelle de l'estimateur $\hat{\beta}_{2,\pi}$ est une valeurs fonction simple des moments empiriques du vecteur des observations (X^1, X^2, Y) invariant par permutation d'un ensemble des données. Nous souhaitons comparer cette valeur avec les estimateurs MCO de $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ dans le cas de la méthode de Freedman-Lane d'une part et pour la méthode de Kennedy d'autre part.

3.7 Espérances permutationnelle des Estimateurs de $\widehat{\text{VAR}}(\hat{\beta}_{2,\pi})$

Dans la proposition (3.6.2) nous avons montré que la variance permutationnelle de $\hat{\beta}_{2,\pi}$ est :

$$\mathcal{V}ar(\hat{\beta}_{2,\pi}) = \frac{S^2(X^1)S^2(Y) - [S(X^1, Y)]^2}{(n-1)\Delta}$$

Comme déjà dit, les estimateurs MCO de $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ obtenus par les méthodes de Freedman-Lane et de Kennedy sont différents. Pour étudier sa loi permutationnelle, nous calculons dans cette section les espérances permutationnelles des estimateurs de $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ fournis par les méthodes de Freedman-Lane, et de Kennedy. Puis, nous les comparons à la variance permutationnelle de $\hat{\beta}_{2,\pi}$.

3.7.1 Cas de la méthode de Freedman-Lane

Le modèle de Freedman-Lane, est caractérisé par l'équation de régression

$$Y_{\pi i}^F = \beta_{0,\pi}^F + \beta_{1,\pi}^F x^1 + \beta_{2,\pi}^F x_i^2 + \epsilon_i^F$$

et nous avons montré (cf. l'équation (3.19)) que :

$$\mathcal{V}ar(\hat{\beta}_{2,\pi})_F = \frac{S^2(X^1)\sigma^{2F}}{n\Delta}.$$

L'estimateur MCO de $\mathcal{V}ar(\hat{\beta}_{2,\pi})_F$ est alors :

$$\widehat{\mathcal{V}ar(\hat{\beta}_{2,\pi})_F} = S^2(\hat{\beta}_{2,\pi})_F = \frac{S^2(X^1)S^{2F}}{n\Delta}$$

où $S^{2F} = \frac{\sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n-3}$ est l'estimateur MCO de σ^{2F} . Nous nous proposons de calculer maintenant l'espérance permutationnelle de $S^2(\hat{\beta}_{2,\pi})_F$. Observons tout d'abord que :

$$\mathcal{E}(S^2(\hat{\beta}_{2,\pi})_F) = \mathcal{E}\left(\frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n(n-3)\Delta}\right) = \frac{S^2(X^1)}{n(n-3)\Delta} \cdot \sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^F)^2 \quad (3.93)$$

et donc, pour calculer l'espérance permutationnelle de $S^2(\hat{\beta}_{2,\pi})_F$, il faut d'abord connaître $\sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^F)^2$. D'après l'équation (3.24), nous avons :

$$\begin{aligned} \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 &= \sum_{i=1}^n (R^{Y.1,\pi i})^2 + n \frac{\Delta_2^2}{\Delta^2} S^2(X^1) + n \frac{\Delta_1^2}{\Delta^2} S^2(X^2) \\ &\quad + 2n \left(\frac{\Delta_1 \Delta_2}{\Delta^2}\right) S(X^1, X^2) - 2n \left(\frac{\Delta_1}{\Delta}\right) S(X^2, R^{Y.1,\pi i}) + \frac{\Delta_2}{\Delta} S(X^1, R^{Y.1,\pi i}) \end{aligned}$$

où par définition, on a :

$$\Delta_1 = S^2(X^2)S(X^1, R^{Y.1,\pi}) - S(X^1, X^2)S(X^2, R^{Y.1,\pi})$$

$$\Delta_2 = S^2(X^1)S(X^2, R^{Y.1,\pi}) - S(X^1, X^2)S(X^1, R^{Y.1,\pi})$$

$$\Delta = S^2(X^1)S^2(X^2) - [S(X^1, X^2)]^2$$

Alors, pour calculer l'espérance $\sum_{i=1}^n (\hat{\epsilon}_i^F)^2$ nous avons besoin de calculer les espérances de Δ_1 , Δ_2 , $\Delta_1\Delta_2$, $\Delta_1S(X^2, R^{Y.1,\pi})$ et $\Delta_2S(X^1, R^{Y.1,\pi})$

Lemme 3.7.1. *Pour tout jeu de données tel que $S^2(X^1) \neq 0$, si $R^{Y,1,\pi}$ est le vecteur résidu permuté on a :*

1. $\mathcal{E}(\Delta_1)^2 = \frac{S^2(X^2)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \cdot \Delta$
2. $\mathcal{E}(\Delta_2)^2 = \frac{S^2(X^1)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \cdot \Delta$
3. $\mathcal{E}(\Delta_1 \Delta_2) = -\frac{S(X^1 X^2)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \cdot \Delta$
4. $\mathcal{E}(\Delta_1 S(X^2, R^{Y,1,\pi})) = \frac{1}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \cdot \Delta$
5. $\mathcal{E}(\Delta_2 S(X^1, R^{Y,1,\pi})) = \frac{1}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \cdot \Delta$

Preuve : 1.

$$\begin{aligned} \mathcal{E}(\Delta_1)^2 &= \mathcal{E}[S^2(X^2)S(X^1, R^{Y,1,\pi}) - S(X^1, X^2)S(X^2, R^{Y,1,\pi})]^2 \\ &= [S^2(X^2)]^2 \mathcal{E}[S(X^1, R^{Y,1,\pi})]^2 + [S(X^1, X^2)]^2 \mathcal{E}[S(X^2, R^{Y,1,\pi})]^2 \\ &\quad - 2S^2(X^2)S(X^1, X^2) \mathcal{E}[S(X^1, R^{Y,1,\pi})S(X^2, R^{Y,1,\pi})] \end{aligned}$$

mais le lemme (3.6.4) permet d'écrire :

$$\begin{aligned} &= \frac{1}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) ([S^2(X^2)]^2 S^2(X^1) \\ &\quad + [S(X^1, X^2)]^2 S^2(X^2) - 2S^2(X^2)[S(X^1, X^2)]^2) \\ &= \frac{S^2(X^2)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \Delta \end{aligned} \tag{3.94}$$

2. Par une argumentation identique on peut montrer que :

$$\mathcal{E}(\Delta_2)^2 = \frac{S^2(X^1)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \Delta \tag{3.95}$$

3. Il reste à connaître l'espérance conditionnelle des produits. Ainsi,

$$\begin{aligned} \mathcal{E}(\Delta_1 \Delta_2) &= \mathcal{E}[S^2(X^1)S(X^2, R^{Y,1,\pi}) - S(X^1, X^2)S(X^1, R^{Y,1,\pi})] \\ &\quad \times [(S^2(X^2)S(X^1, R^{Y,1,\pi}) - S(X^1, X^2)S(X^2, R^{Y,1,\pi}))] \\ &= S^2(X^1)S^2(X^2) \mathcal{E}[S(X^1, R^{Y,1,\pi})S(X^2, R^{Y,1,\pi})] \\ &\quad - S(X^1, X^2)S^2(X^2) \mathcal{E}[S(X^1, R^{Y,1,\pi})]^2 - S^2(X^1)S(X^1, X^2) \mathcal{E}[S(X^2, R^{Y,1,\pi})]^2 \\ &\quad + [S(X^1, X^2)]^2 \mathcal{E}[S(X^1, R^{Y,1,\pi})S(X^2, R^{Y,1,\pi})] \end{aligned}$$

À l'aide du lemme (3.6.4) on obtient par substitution et réorganisation algébrique :

$$\mathcal{E}(\Delta_1 \Delta_2) = -\frac{S^2(X^1 X^2)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \Delta \quad (3.96)$$

4. Maintenant

$$\begin{aligned} \mathcal{E}((\Delta_1 S(X^2, R^{Y.1, \pi}))) &= \mathcal{E}[S^2(X^1) S(X^2, R^{Y.1, \pi}) \\ &\quad - S(X^1, X^2) S(X^1, R^{Y.1, \pi}) S(X^2, R^{Y.1, \pi})] \\ &= S^2(X^1) \mathcal{E}[S(X^2, R^{Y.1, \pi})]^2 \\ &\quad - S(X^1 X^2) \mathcal{E}[S(X^1, R^{Y.1, \pi}) S(X^2, R^{Y.1, \pi})] \end{aligned}$$

À partir du lemme (3.6.4), on obtient par substitution :

$$\mathcal{E}((\Delta_1 S(X^2, R^{Y.1, \pi}))) = \frac{\Delta}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \quad (3.97)$$

5. On montre de façon analogue que :

$$\mathcal{E}((\Delta_2 S(X^1, R^{Y.1, \pi}))) = \frac{\Delta}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \quad \square \quad (3.98)$$

Proposition 3.7.1. *Dans le cas non trivial $\Delta \neq 0$, $S^2(X^1) \neq 0$, l'espérance permutationnelle de $S^2(\hat{\beta}_{2, \pi})_F$ est égale à la variance permutationnelle de $\hat{\beta}_{2, \pi}$:*

$$\mathcal{E}(S^2(\hat{\beta}_{2, \pi})_F) = \frac{(S^2(X^1) S^2(Y) - [S(X^1, Y)]^2)}{(n-1)\Delta} = \text{Var}(\hat{\beta}_{2, \pi})$$

Par suite, l'estimateur de $\text{Var}(\hat{\beta}_{2, \pi})$ par la méthode de Freedman-Lane est un estimateur sans biais sous loi permutationnelle .

$$\text{Preuve : } \mathcal{E}(S^2(\hat{\beta}_{2, \pi})_F) = \mathcal{E}\left(\frac{S(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{(n-3)\Delta}\right) = \frac{S(X^1)}{(n-3)\Delta} \cdot \sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^F)^2$$

Mais

$$\begin{aligned} \sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^F)^2 &= \sum_{i=1}^n \mathcal{E}(R^{Y,1,\pi i})^2 + n \frac{S^2(X^1)}{\Delta^2} \mathcal{E}(\Delta_2^2) + n \frac{S^2(X^2)}{\Delta^2} \mathcal{E}(\Delta_1^2) \\ &+ 2n \frac{S(X^1, X^2)}{\Delta^2} \mathcal{E}(\Delta_1 \Delta_2) - 2 \frac{n}{\Delta} (\mathcal{E}(\Delta_2 S(X^1, R^{Y,1,\pi i})) + \mathcal{E}(\Delta_1 S(X^2, R^{Y,1,\pi i}))) \end{aligned}$$

On utilise alors les résultats du lemme (3.7.1) pour obtenir par substitution :

$$\begin{aligned} &= n \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) - 4 \frac{n\Delta}{(n-1)\Delta} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \\ &+ 2n\Delta \frac{S^2(X^1)S^2(X^2) - [S(X^1, X^2)]^2}{(n-1)\Delta^2} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \\ &= n \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) - 4 \frac{n}{(n-1)} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \\ &+ 2 \frac{n}{(n-1)} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \end{aligned}$$

Soit après simplification :

$$= \frac{n(n-3)}{n-1} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \quad (3.99)$$

Par suite :

$$\begin{aligned} \mathcal{E}(S^2(\hat{\beta}_{2,\pi})_F) &= \mathcal{E}\left(\frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{n(n-3)\Delta}\right) = \frac{S^2(X^1)}{n(n-3)\Delta} \cdot \sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^F)^2 \\ &= \frac{n(n-3)S^2(X^1)}{n(n-1)(n-3)\Delta} \left(S^2(Y) - \frac{[S(X^1, Y)]^2}{S^2(X^1)} \right) \\ &= \frac{(S^2(X^1)S^2(Y) - [S(X^1, Y)]^2)}{(n-1)\Delta} \\ &= \mathcal{V}ar(\hat{\beta}_{2,\pi}) \quad \square \end{aligned} \quad (3.100)$$

On déduit de cette identité le fait que l'estimateur de $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ par la méthode de Freedman-Lane est un estimateur sans biais pour $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ sous la loi permutationnelle .

3.7.2 Cas de la méthode de Kennedy

Le modèle de Kennedy, est déterminée par l'équation :

$$R^{Y,1,\pi i} = \beta_{2,\pi} R_i^{2,1} + \epsilon_i^K$$

et en définissant $\sigma^{2K} = \text{Var}(\epsilon^K)$ on a montré que :

$$\text{Var}(\hat{\beta}_{2,\pi})_K = \frac{S^2(X^1)\sigma^{2K}}{n\Delta}.$$

L'estimateur MCO de $\text{Var}(\hat{\beta}_{2,\pi})_K$ est donc :

$$\widehat{\text{Var}}(\hat{\beta}_{2,\pi})_K = S^2(\hat{\beta}_{2,\pi}) = \frac{S^2(X^1)S^{2K}}{n\Delta}$$

où $S^{2K} = \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n-3}$ est l'estimateur MCO de σ^{2K} . Ici comme dans la section (3.7.1), nous calculons l'espérance permutationnelle de $\widehat{\text{Var}}(\hat{\beta}_{2,\pi})_K$, en utilisant l'identité :

$$\mathcal{E}(S^2(\hat{\beta}_{2,\pi})_K) = \mathcal{E}\left(\frac{S(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-3)\Delta}\right) = \frac{S(X^1) \sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^K)^2}{n(n-3)\Delta} \cdot \sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^K)^2 \quad (3.101)$$

Donc, pour calculer l'espérance permutationnelle de $S^2(\hat{\beta}_{2,\pi})_K$ nous devons calculer $\sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^K)^2$. Mais selon le lemme (3.3.2), nous avons :

$$\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 = \sum_{i=1}^n (R^{Y,1,\pi i})^2 - n \frac{\Delta_1^2}{\Delta S^2(X^1)}$$

On est donc en situation de démontrer le résultat suivant :

Proposition 3.7.2. *Dans le cas non trivial $\Delta \neq 0$, on a :*

$$\begin{aligned} \mathcal{E}(S^2(\hat{\beta}_{2,\pi})_K) &= \frac{n-2}{n-3} \cdot \frac{(S^2(X^1)S^2(Y) - S(X^1, Y))}{(n-1)\Delta} \\ &= \frac{n-2}{n-3} \cdot \mathcal{V}ar(\hat{\beta}_{2,\pi}) \end{aligned} \quad (3.102)$$

Preuve :

$$\mathcal{E}(S^2(\hat{\beta}_{2,\pi})_K) = \frac{S^2(X^1)}{n(n-3)\Delta} \cdot \sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^K)^2$$

Le lemme (3.3.2) permet de plus d'écrire :

$$\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 = \sum_{i=1}^n (R^{Y,1,\pi i})^2 - n \frac{\Delta_1^2}{\Delta S^2(X^1)}$$

et donc

$$\sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^K)^2 = \sum_{i=1}^n \mathcal{E}(R^{Y,1,\pi i})^2 - n \frac{\mathcal{E}(\Delta_1^2)}{\Delta S^2(X^1)}.$$

On obtient alors à partir des lemmes (3.6.3) et (3.7.1) :

$$\begin{aligned} \sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^K)^2 &= n \left(S^2(Y) - \frac{S(X^1, Y)}{S^2(X^1)} \right) - \frac{n \Delta S^2(X^1)}{(n-1) \Delta S^2(X^1)} \left(S^2(Y) - \frac{S(X^1, Y)}{S^2(X^1)} \right) \\ &= \frac{n(n-2)}{(n-1)} \left(S^2(Y) - \frac{S(X^1, Y)}{S^2(X^1)} \right) \end{aligned} \quad (3.103)$$

et donc :

$$\begin{aligned} \mathcal{E}(S^2(\hat{\beta}_{2,\pi})_K) &= \frac{S^2(X^1)}{n(n-3)\Delta} \cdot \sum_{i=1}^n \mathcal{E}(\hat{\epsilon}_i^K)^2 \\ &= \frac{n(n-2)S^2(X^1)}{n(n-1)(n-3)\Delta} \left(S^2(Y) - \frac{S(X^1, Y)}{S^2(X^1)} \right) \\ &= \frac{n-2}{n-3} \cdot \frac{(S^2(X^1)S^2(Y) - S(X^1, Y))}{(n-1)\Delta} \\ &= \frac{n-2}{n-3} \cdot \mathcal{V}ar(\hat{\beta}_{2,\pi}) \quad \square \end{aligned} \quad (3.104)$$

Ce résultat montre que, l'estimateur de $\text{Var}(\beta_{2,\pi})$ par la méthode de Kennedy ($S^2(\hat{\beta}_{2,\pi})_K$) est un estimateur biaisé pour $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ sous la loi permutacionnelle, contrairement à l'estimateur donné par la méthode de Freedman-Lane ($S^2(\hat{\beta}_{2,\pi})_F$)

Proposition 3.7.3.

$$\mathcal{E}(S^2(\beta_{2,\pi})_K) > \mathcal{E}(S^2(\beta_{2,\pi})_F)$$

Preuve : En utilisant les propositions (3.7.1) et (3.7.2), on montre que :

$$\frac{\mathcal{E}(S^2(\beta_{2,\pi})_F)}{\mathcal{E}(S^2(\beta_{2,\pi})_K)} = \frac{\frac{(S^2(X^1)S^2(Y) - [S(X^1, Y)]^2)}{(n-1)\Delta}}{\frac{(n-2)(S^2(X^1)S^2(Y) - S(X^1, Y))}{(n-1)(n-3)\Delta}} = \frac{n-3}{n-2} < 1$$

alors :

$$\mathcal{E}(S^2(\beta_{2,\pi})_K) > \mathcal{E}(S^2(\beta_{2,\pi})_F) \quad \square \quad (3.105)$$

Mais il apparaît clairement que la différence devient négligeable quand la taille de l'échantillon est assez grande. D'autre part, d'après la proposition (3.4.1), on a pour chaque permutation π :

$$S^2(\beta_{2,\pi})_K \geq S^2(\beta_{2,\pi})_F$$

Par ailleurs, lorsque nous avons un échantillon de taille n , il existe $n!$ permutations et la probabilité permutacionnelle de chaque permutation est $1/n!$. Alors, la définition de l'espérance donne :

$$\mathcal{E}(S^2(\beta_{2,\pi})_K) = \sum_{i=1}^{n!} \frac{1}{n!} S^2(\beta_{2,\pi i})_K > \sum_{i=1}^{n!} \frac{1}{n!} S^2(\beta_{2,\pi i})_F = \mathcal{E}(S^2(\beta_{2,\pi})_F)$$

et la proposition (3.4.1) dit que :

$$S^2(\beta_{2,\pi})_K \geq S^2(\beta_{2,\pi})_F$$

Le rôle du biais permutational observé pour la méthode de Kennedy peut être contrôlé. C'est ce que confirme la proposition ci-dessous.

Proposition 3.7.4. *L'espérance permutational de $S^2(\hat{\beta}_{2,\pi})_{K^*}$ est égale à la variance permutational de $\hat{\beta}_{2,\pi}$:*

$$\mathcal{E}(S^2(\hat{\beta}_{2,\pi})_{K^*}) = \frac{(S^2(X^1)S^2(Y) - [S(X^1, Y)]^2)}{(n-1)\Delta} = \mathcal{V}ar(\hat{\beta}_{2,\pi})$$

Par suite, l'estimateur de $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ par la méthode de Kennedy corrigée en $(n-2)$ degrés de liberté est un estimateur sans biais pour $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ sous loi permutational.

Preuve : En utilisant la proposition (3.7.2), on a :

$$\mathcal{E}(S^2(\hat{\beta}_{2,\pi})_K) = \frac{n-2}{n-3} \cdot \frac{(S^2(X^1)S^2(Y) - S(X^1, Y))}{(n-1)\Delta} = \frac{n-2}{n-3} \mathcal{V}ar(\hat{\beta}_{2,\pi})$$

Donc, en multipliant $S^2(\hat{\beta}_{2,\pi})_K$ par le coefficient $(n-3)/(n-2)$, on obtient un estimateur sans biais de $\mathcal{V}ar(\hat{\beta}_{2,\pi})$, c'est-à-dire tel que :

$$\mathcal{E}\left(\frac{n-3}{n-2} \cdot S^2(\hat{\beta}_{2,\pi})_K\right) = \mathcal{V}ar(\hat{\beta}_{2,\pi})$$

Mais

$$\begin{aligned} \frac{n-3}{n-2} \cdot S^2(\hat{\beta}_{2,\pi})_K &= \frac{n-3}{n-2} \cdot \frac{S(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-3)\Delta} \\ &= \frac{S(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-2)\Delta} \\ &= S^2(\hat{\beta}_{2,\pi})_{K^*} \end{aligned} \quad (3.106)$$

C'est dire que l'estimateur de $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ par la méthode de Kennedy avec $(n-2)$ degrés de liberté est un estimateur sans biais de $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ pour la loi permutational :

$$\mathcal{E}(S^2(\hat{\beta}_{2,\pi})_{K^*}) = \frac{(S^2(X^1)S^2(Y) - [S(X^1, Y)]^2)}{(n-1)\Delta} = \mathcal{V}ar(\hat{\beta}_{2,\pi}) \quad \square \quad (3.107)$$

Dans les propositions (3.7.1) et (3.7.4), nous avons montré que l'espérance permutationnelle de $S^2(\hat{\beta}_{2,\pi})_F$ et $S^2(\hat{\beta}_{2,\pi})_{K^*}$ sont égales à $\mathcal{V}ar(\hat{\beta}_{2,\pi})$. C'est-à-dire qu'on dispose d'estimateurs sans biais de $\mathcal{V}ar(\hat{\beta}_{2,\pi})$ pour la loi permutationnelle. Cela a pour première conséquence d'indiquer qu'une juste comparaison des deux méthodes de test de permutation, en termes d'efficacité doit intégrer la correction de la statistique de Kennedy en $(n-2)$. Pour autant, on ne peut pas déduire de cette proposition que les méthodes de Freedman-Lane et de Kennedy corrigée avec $(n-2)$ degrés de liberté, sont équivalentes pour effectuer le test d'hypothèse $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$. En effet, même si $\mathcal{E}(S^2(\hat{\beta}_{2,\pi})_F) = \mathcal{E}(S^2(\hat{\beta}_{2,\pi})_{K^*}) = \mathcal{V}ar(\hat{\beta}_{2,\pi})$, il reste que nous avons montré dans la proposition (3.4.4) le résultat suivant :

$$S^2(\hat{\beta}_{2,\pi})_{K^*} - S^2(\hat{\beta}_{2,\pi})_F \begin{cases} \geq 0 & \text{si } [S(X^1, R^{Y.1, \pi^i})]^2 \geq \Delta S^2(\hat{\beta}_{2,\pi})_F \\ < 0 & \text{si } [S(X^1, R^{Y.1, \pi^i})]^2 < \Delta S^2(\hat{\beta}_{2,\pi})_F \end{cases}$$

Cela signifie qu'il existe des cas où $S^2(\hat{\beta}_{2,\pi})_{K^*} > S^2(\hat{\beta}_{2,\pi})_F$ et d'autres où $S^2(\hat{\beta}_{2,\pi})_{K^*} < S^2(\hat{\beta}_{2,\pi})_F$. Mais les résultats des simulations présentées au chapitre (2) montrent que dans plupart des cas $S^2(\hat{\beta}_{2,\pi})_{K^*} > S^2(\hat{\beta}_{2,\pi})_F$, puisque dans la plupart des cas, le risque d'erreur de première espèce de la méthode de Kennedy corrigée est supérieur à celui de la méthode de Freedman-Lane.

3.8 Étude comparative pour p variables explicatives

On considère le $(p+1)$ -uplet fourni par une variable expliquée, Y et un nombre ($p \geq 1$) de variables explicatives X^1, X^2, \dots, X^p . Une façon d'écrire la relation liant la variable expliquée aux variables explicatives consiste à spécifier un modèle linéaire stochastique du type :

$$Y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \dots + \beta_p x_i^p + \epsilon_i \quad (i = 0, 1, \dots) \quad (3.108)$$

où ϵ mesure l'écart du modèle aux données, lorsque le modèle (3.108) est appliqué à un échantillon de n observations $\{y_1, x_1^1, x_1^2, \dots, x_1^p\}$. Pour faciliter les écritures, on peut ré-écrire le modèle (3.108) sous forme matricielle. Posons :

$$\mathbf{x}_i = \begin{bmatrix} 1 \\ x_i^1 \\ \vdots \\ x_i^p \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

Alors, on a :

$${}^t \mathbf{x}_i \boldsymbol{\beta} = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \dots + \beta_p x_i^p \quad (3.109)$$

de sorte que l'équation (3.108) s'écrit en termes de vecteurs :

$$Y_i = {}^t \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i \quad (3.110)$$

On complète cette écriture en posant :

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \in \mathbb{R}^n \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \in \mathbb{R}^n \quad \mathbf{X} = \begin{bmatrix} {}^t\mathbf{X}_1 \\ {}^t\mathbf{X}_2 \\ \vdots \\ {}^t\mathbf{X}_n \end{bmatrix} = \begin{bmatrix} 1 & x_1^1 & \cdots & x_1^p \\ 1 & x_2^1 & \cdots & x_2^p \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_n^1 & \cdots & x_n^p \end{bmatrix}$$

et l'on obtient l'écriture matricielle d'un modèle de régression multiple

$$\underset{(n \times 1)}{\mathbf{Y}} = \underset{(n \times k)}{\mathbf{X}} \underset{(k \times 1)}{\boldsymbol{\beta}} + \underset{(n \times 1)}{\boldsymbol{\epsilon}} \quad \text{où } k = p + 1 \quad (3.111)$$

Ce modèle peut alors être regardé comme un ensemble de restrictions imposées à la loi conjointe du vecteur aléatoire $\mathbf{Z} = (\mathbf{Y}, \mathbf{X})$. Dit autrement, un modèle est un ensemble de lois de probabilités conjointes qui satisfont à un ensemble de postulats. D'un point de vue probabiliste, le modèle classique de régression linéaire multiple ci-dessus est caractérisé par les postulats présentés dans le chapitre (1) dans le cas multiple.

Afin de simplifier la présentation, nous supposons dans la suite que la matrice \mathbf{X} est de plein rang. Nous savons que l'estimateur MCO de $\boldsymbol{\beta}$ est :

$$\hat{\boldsymbol{\beta}} = ({}^t\mathbf{X}\mathbf{X})^{-1}{}^t\mathbf{X}\mathbf{Y} \quad (3.112)$$

et la variance de $\hat{\boldsymbol{\beta}}$ est dans ce cadre classique (*cf.* Neter et al [46]) :

$$\text{Var}(\hat{\boldsymbol{\beta}}) = ({}^t\mathbf{X}\mathbf{X})^{-1}\sigma^2 \quad (3.113)$$

où $\sigma^2 = \text{Var}(\epsilon)$. Lorsqu'on se fixe pour objectif d'effectuer un test d'hypothèse portant sur la nullité d'un coefficient de régression partielle par exemple : $H_0 : \beta_p = 0$ contre $H_1 : \beta_p \neq 0$, une statistique de test souvent utilisée pour évaluer cette hypothèse est la statistique de Student :

$$T_p = \frac{\hat{\beta}_p}{S(\hat{\beta}_p)} \sim t(n - k) \quad (3.114)$$

où $\hat{\beta}_p$ est l'estimateur MCO de β_p et $S^2(\hat{\beta}_p)$ est l'estimateur MCO de $\text{Var}(\hat{\beta}_p)$. Le choix du coefficient d'ordre p est arbitraire, car il est lié au choix de l'indexation des prédicteurs. Il ne retire donc rien à la généralité de notre étude. Remarquons que lorsque le terme d'erreur ϵ ne vérifie pas les postulats P1-P5 imposés dans le cas multiple (*cf.* chapitre (1)), la statistique T ne suit plus nécessairement une loi de Student.

Les résultats précédents conduisent à étendre à cette situation la comparaison des méthodes de test permutational proposées par Freedman-Lane et par Kennedy.

3.8.1 Cas de la méthode de Freedman-Lane

Afin d'obtenir le modèle de Freedman-Lane pour effectuer un test d'hypothèse $H_0 : \beta_p = 0$ contre $H_0 : \beta_p \neq 0$, il faut d'abord réaliser une régression linéaire multiple de la variable Y restreinte aux $(p - 1)$ prédicteurs X^1, X^2, \dots, X^{p-1} , au sens MCO, d'après le modèle

$$\mathbf{Y} = \mathcal{X}\boldsymbol{\alpha} + \mathbf{U} \tag{3.115}$$

où ${}^t\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_{p-1})$, $\mathcal{X} = (\mathbf{1}, \mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^{p-1})$ et $\mathbf{U} = (u_1, u_2, \dots, u_n)$. L'équation ci-dessus conduit à calculer la prévision :

$$\hat{\mathbf{Y}}^p = \mathcal{X}\hat{\boldsymbol{\alpha}} \tag{3.116}$$

où ${}^t\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_{p-1})$ est estimateur MCO de $\boldsymbol{\alpha}$ du modèle (3.116). On obtient le vecteur des résidus observés par ajustement de ce modèle restreint :

$$\mathbf{u} = \mathbf{y} - \hat{\mathbf{y}}^p = \mathbf{y} - \mathcal{X}\mathbf{a} \tag{3.117}$$

où ${}^t\mathbf{a} = (a_0, a_1, \dots, a_{p-1})$ est l'estimation MCO de $\boldsymbol{\alpha}$. C'est-à-dire : $\mathbf{a} = ({}^t\mathcal{X}\mathcal{X})^{-1}{}^t\mathcal{X}\mathbf{y}$, on obtient alors les résidus observés :

$$\mathbf{u} = \mathbf{y} - \hat{\mathbf{y}}^p = \mathbf{y} - \mathcal{X}({}^t\mathcal{X}\mathcal{X})^{-1}{}^t\mathcal{X}\mathbf{y} = (\mathbf{I} - \mathbf{H}_{\mathcal{X}})\mathbf{y} \tag{3.118}$$

où $\mathbf{H}_{\mathcal{X}} = \mathcal{X}({}^t\mathcal{X}\mathcal{X})^{-1}{}^t\mathcal{X}$.

Supposons maintenant qu'on réalise un ensemble de permutations de ces résidus observés. A chaque permutation π , on associe alors le vecteur des valeurs permutées $u^{\pi i}$ ($i = 1, 2, \dots, n$), défini comme :

$$\mathbf{u} = (u_1, u_2, \dots, u_n) \xrightarrow{\pi} \mathbf{u}^\pi = (u^{\pi 1}, u^{\pi 2}, \dots, u^{\pi n}) \tag{3.119}$$

Remarquons qu'on peut obtenir le vecteur \mathbf{u}^π par application de l'opérateur de permutation de matrice \mathbf{P} comme suit :

$$\mathbf{u}^\pi = \mathbf{P}\mathbf{u} \tag{3.120}$$

où \mathbf{P} est un matrice de terme général $p_{ij} \in \{0, 1\}$ telle que :

$${}^t\mathbf{P}\mathbf{P} = \mathbf{P}{}^t\mathbf{P} = \mathbf{I} \tag{3.121}$$

et d'après (3.118),

$$\mathbf{u}^\pi = \mathbf{P}(\mathbf{I} - \mathbf{H}_{\mathcal{X}})\mathbf{y} \tag{3.122}$$

On constitue ensuite le vecteur des pseudo-valeurs notées \mathbf{Y}_π^F , en ajoutant le vecteur des résidus permutés \mathbf{u}^π au vecteur des valeurs retenues dans le modèle d'équation (3.115) :

$$\mathbf{Y}_\pi^F = \hat{\mathbf{Y}}^p + \mathbf{u}^\pi = \mathcal{X}\hat{\boldsymbol{\alpha}} + \mathbf{u}^\pi \implies \begin{pmatrix} Y_{\pi 1}^F \\ Y_{\pi 2}^F \\ \vdots \\ Y_{\pi n}^F \end{pmatrix} = \begin{pmatrix} \hat{Y}_1^p + u^{\pi 1} \\ \hat{Y}_2^p + u^{\pi 2} \\ \vdots \\ \hat{Y}_n^p + u^{\pi n} \end{pmatrix} \tag{3.123}$$

Enfin, le modèle ajusté par Freedman-Lane s'écrit sous la forme du modèle général suivant :

$$\mathbf{Y}_\pi^F = \mathbf{X}\boldsymbol{\beta}_\pi^F + \boldsymbol{\epsilon}^F \quad (3.124)$$

où ${}^t\boldsymbol{\beta}_\pi^F = (\beta_{0,\pi}^F, \beta_{1,\pi}^F, \dots, \beta_{p,\pi}^F)$ et $\boldsymbol{\epsilon}^F$ est le vecteur d'erreurs et est tel que $\mathbb{E}(\boldsymbol{\epsilon}^F) = \mathbf{0}$ et $\text{Var}(\boldsymbol{\epsilon}^F) = \sigma^{2F}\mathbf{I}$.

Remarque 3.8.1. *Comme dans la proposition (3.3.1) (où on se limitait à deux variables explicatives), on peut montrer ici que les statistiques de test de permutation de l'hypothèse $H_0 : \beta_p = 0$ contre $H_1 : \beta_p \neq 0$ dans le modèle de Freedman-Lane (3.124) et dans le modèle :*

$$\mathbf{u}^\pi = \mathbf{X}\boldsymbol{\beta}'_\pi + \boldsymbol{\epsilon}' \quad (3.125)$$

où ${}^t\boldsymbol{\beta}'_\pi = (\beta'_{0,\pi}, \beta'_{1,\pi}, \dots, \beta'_{p,\pi})$, ${}^t\boldsymbol{\epsilon}' = (\epsilon'_1, \epsilon'_2, \dots, \epsilon'_p)$ et $\text{Var}(\boldsymbol{\epsilon}') = \sigma^{2'}\mathbf{I}$, sont équivalents. C'est-à-dire que :

- a) $\hat{\beta}'_{p,\pi} = \hat{\beta}_{p,\pi}^F$
- b) $\text{Var}(\hat{\beta}'_{p,\pi}) = \text{Var}(\hat{\beta}_{p,\pi}^F)$
- c) $\hat{\epsilon}'_i = \hat{\epsilon}_i^F$
- d) $T_{p,\pi}^F = \frac{\hat{\beta}_{p,\pi}^F}{S(\hat{\beta}_{p,\pi}^F)} = \frac{\hat{\beta}'_{p,\pi}}{S(\hat{\beta}'_{p,\pi})} = T_{p,\pi}'$

L'équation (3.125) permet donc de calculer la prévision :

$$\hat{\mathbf{u}}^\pi = \mathbf{X}\hat{\boldsymbol{\beta}}'_\pi \quad (3.126)$$

où $\hat{\boldsymbol{\beta}}'_\pi = {}^t(\hat{\beta}'_{0,\pi}, \hat{\beta}'_{1,\pi}, \dots, \hat{\beta}'_{p,\pi}) = ({}^t\mathbf{X}\mathbf{X})^{-1}{}^t\mathbf{X}\mathbf{u}^\pi$ est l'estimateur MCO de $\boldsymbol{\beta}'_\pi$ dans le modèle (3.125) et $\text{Var}(\hat{\boldsymbol{\beta}}'_\pi) = ({}^t\mathbf{X}\mathbf{X})^{-1}\sigma^{2'}$.

Lemme 3.8.1. *Les résidus du modèle de Freedman-Lane : $\mathbf{Y}_\pi^F = \mathbf{X}\boldsymbol{\beta}_\pi^F + \boldsymbol{\epsilon}^F$ vérifient :*

$$\sum_{i=1}^n (\hat{\epsilon}_i^F)^2 = {}^t\hat{\boldsymbol{\epsilon}}^F \hat{\boldsymbol{\epsilon}}^F = {}^t\mathbf{U}^\pi (\mathbf{I} - \mathbf{H}_\mathbf{X}) \mathbf{U}^\pi = {}^t\mathbf{U}^\pi \mathbf{U}^\pi - {}^t\mathbf{U}^\pi \mathbf{H}_\mathbf{X} \mathbf{U}^\pi$$

où $\mathbf{H}_\mathbf{X} = \mathbf{X}({}^t\mathbf{X}\mathbf{X})^{-1}{}^t\mathbf{X}$

Preuve : D'après la remarque (3.8.1), on sait que $\hat{\epsilon}_i^F = \epsilon'_i$. Alors, nous utilisons le modèle (3.125) : $\mathbf{U}^\pi = \mathbf{X}\boldsymbol{\beta}'_\pi + \boldsymbol{\epsilon}'$. Ce modèle permet d'écrire :

$$\hat{\boldsymbol{\epsilon}}^F = \hat{\boldsymbol{\epsilon}}' = \mathbf{U}^\pi - \mathbf{X}\hat{\boldsymbol{\beta}}'_\pi = \mathbf{U}^\pi - \mathbf{X}({}^t\mathbf{X}\mathbf{X})^{-1}{}^t\mathbf{X}\mathbf{U}^\pi = (\mathbf{I} - \mathbf{H}_\mathbf{X})\mathbf{U}^\pi \quad (3.127)$$

et alors,

$$\begin{aligned} {}^t\hat{\boldsymbol{\epsilon}}^F \hat{\boldsymbol{\epsilon}}^F &= {}^t\hat{\boldsymbol{\epsilon}}' \hat{\boldsymbol{\epsilon}}' &= {}^t\mathbf{U}^\pi (\mathbf{I} - \mathbf{H}_\mathbf{X}) (\mathbf{I} - \mathbf{H}_\mathbf{X}) \mathbf{U}^\pi \\ &= {}^t\mathbf{U}^\pi (\mathbf{I} - \mathbf{H}_\mathbf{X}) \mathbf{U}^\pi \\ &= {}^t\mathbf{U}^\pi \mathbf{U}^\pi - {}^t\mathbf{U}^\pi \mathbf{H}_\mathbf{X} \mathbf{U}^\pi \quad \square \end{aligned} \quad (3.128)$$

Afin de comparer les estimateurs de $\text{Var}(\hat{\beta}_2)$ par les méthodes de Freedman-Lane et de Kennedy, nous avons besoin de rappeler quelques propriétés et de démontrer quelques lemmes et propositions. Nous notons $\tilde{\mathbf{x}}^p = (\mathbf{I} - \mathbf{H}_{\mathcal{X}})\mathbf{x}^p$ la projection de \mathbf{x}^p sur le sous espace orthogonal à $\mathbf{H}_{\mathcal{X}}$

Lemme 3.8.2. (c.f. Srivastava et Khatri 1979 Corollaire 1.4.2 (ii) [54])
Soient \mathbf{P} et \mathbf{Q} deux matrices carrées et non singulières alors :

$$\begin{bmatrix} \mathbf{P} & \mathbf{S} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{P} - \mathbf{S}\mathbf{Q}^{-1}\mathbf{R})^{-1} & -\mathbf{P}^{-1}\mathbf{S}(\mathbf{Q} - \mathbf{R}\mathbf{P}^{-1}\mathbf{S})^{-1} \\ -(\mathbf{Q} - \mathbf{R}\mathbf{P}^{-1}\mathbf{S})^{-1}\mathbf{R}\mathbf{P}^{-1} & (\mathbf{Q} - \mathbf{R}\mathbf{P}^{-1}\mathbf{S})^{-1} \end{bmatrix} \quad \square \quad (3.129)$$

Le lemme suivant évalue $(\mathbf{P} - \mathbf{S}\mathbf{Q}^{-1}\mathbf{R})^{-1}$ dans le cas symétrique $\mathbf{R} = {}^t\mathbf{S}$

Lemme 3.8.3. (Woodbury matrix identity (cf. [52]))

$$(\mathbf{P} - \mathbf{S}\mathbf{Q}^{-1}{}^t\mathbf{S})^{-1} = \mathbf{P}^{-1} + \mathbf{P}^{-1}\mathbf{S}(\mathbf{Q}^{-1} + {}^t\mathbf{S}\mathbf{P}^{-1}\mathbf{S})^{-1}{}^t\mathbf{S}\mathbf{P}^{-1} \quad \square \quad (3.130)$$

Partitionnons maintenant la matrice $\mathbf{X} = (\mathbf{1}, \mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^p)$ en deux parties $\mathbf{X} = (\mathbf{1}, \mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^{p-1} \mid \mathbf{x}^p) = (\mathcal{X} \mid \mathbf{x}^p)$ de façon à accorder un rôle particulier au prédicteur d'intérêt. Dans la proposition suivante, nous décomposons la matrice de projection $\mathbf{H}_{\mathbf{X}}$ en deux matrices de projections, respectant cette partition et traduisant l'orthogonalité des sous espaces de l'espace des prédicteurs, qui sont engendrés par chaque bloc de variables.

Proposition 3.8.1. La matrice de projection sur $\mathbf{H}_{\mathbf{X}}$ associée à une matrice \mathbf{X} , bipartitionnée en colonnes, se décompose en :

$$\mathbf{H}_{\mathbf{X}} = \mathbf{X}({}^t\mathbf{X}\mathbf{X})^{-1}{}^t\mathbf{X} = \mathbf{H}_{\mathcal{X}} + \mathbf{H}_{\tilde{\mathbf{x}}^p}$$

où $\mathbf{H}_{\tilde{\mathbf{x}}^p}$ est la matrice de projection sur espace engendré par $\tilde{\mathbf{x}}^p$ et $\mathbf{H}_{\mathcal{X}}$ le projecteur sur l'espace engendré par X^1, \dots, X^{p-1} .

Preuve : D'après $\mathbf{X} = (\mathcal{X} \mid \mathbf{x}^p)$ on a :

$${}^t\mathbf{X}\mathbf{X} = \begin{bmatrix} {}^t\mathcal{X}\mathcal{X} & {}^t\mathcal{X}\mathbf{x}^p \\ \hline {}^t\mathbf{x}^p\mathcal{X} & {}^t\mathbf{x}^p\mathbf{x}^p \end{bmatrix} \quad (3.131)$$

En appliquant le lemme (3.8.2) pour $\mathbf{P} = {}^t\mathcal{X}\mathcal{X}$, $\mathbf{R} = {}^t\mathbf{S} = {}^t\mathcal{X}\mathbf{x}^p$, $\mathbf{Q} = {}^t\mathbf{x}^p\mathbf{x}^p$ on obtient :

$$\mathbf{R}\mathbf{P}^{-1}\mathbf{S} = {}^t\mathbf{x}^p\mathbf{H}_{\mathcal{X}}\mathbf{x}^p \quad (3.132)$$

et donc

$$({}^t\mathbf{X}\mathbf{X})^{-1} = \left[\begin{array}{c|c} [{}^t\mathcal{X}\mathcal{X} - {}^t\mathcal{X}\mathbf{x}^p({}^t\mathbf{x}^p\mathbf{x}^p)^{-1}{}^t\mathbf{x}^p{}^t\mathcal{X}]^{-1} & -({}^t\mathcal{X}\mathcal{X})^{-1}{}^t\mathcal{X}\mathbf{x}^p(a^{-1}) \\ \hline -(a^{-1}){}^t\mathbf{x}^p{}^t\mathcal{X}({}^t\mathcal{X}\mathcal{X})^{-1} & a^{-1} \end{array} \right] \quad (3.133)$$

où $a = {}^t\mathbf{x}^p[\mathbf{I} - \mathbf{H}_\mathcal{X}]\mathbf{x}^p$. Le bloc (1,1) peut être calculé à l'aide du lemme (3.8.3) :

$$[{}^t\mathcal{X}\mathcal{X} - {}^t\mathcal{X}\mathbf{x}^p({}^t\mathbf{x}^p\mathbf{x}^p)^{-1}{}^t\mathbf{x}^p{}^t\mathcal{X}]^{-1} = ({}^t\mathcal{X}\mathcal{X})^{-1} + ({}^t\mathcal{X}\mathcal{X})^{-1}{}^t\mathcal{X}\mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p{}^t\mathcal{X}({}^t\mathcal{X}\mathcal{X})^{-1} \quad (3.134)$$

et donc :

$$\begin{aligned} \mathbf{H}_\mathbf{X} &= \mathbf{X}({}^t\mathbf{X}\mathbf{X})^{-1}{}^t\mathbf{X} \\ &= (\mathcal{X} \mid \mathbf{x}^p)({}^t\mathbf{X}\mathbf{X})^{-1} \begin{pmatrix} {}^t\mathcal{X} \\ - \\ {}^t\mathbf{x}^p \end{pmatrix} \\ &= \mathcal{X}({}^t\mathcal{X}\mathcal{X})^{-1}{}^t\mathcal{X} + \mathcal{X}({}^t\mathcal{X}\mathcal{X})^{-1}{}^t\mathcal{X}\mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p{}^t\mathcal{X}({}^t\mathcal{X}\mathcal{X})^{-1}{}^t\mathcal{X} \\ &\quad - \mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p{}^t\mathcal{X}({}^t\mathcal{X}\mathcal{X})^{-1}{}^t\mathcal{X} - \mathcal{X}({}^t\mathcal{X}\mathcal{X})^{-1}{}^t\mathcal{X}\mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p + \mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p \\ &= \mathbf{H}_\mathcal{X} + \mathbf{H}_\mathcal{X}\mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p\mathbf{H}_\mathcal{X} - \mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p\mathbf{H}_\mathcal{X} - \mathbf{H}_\mathcal{X}\mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p + \mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p \\ &= \mathbf{H}_\mathcal{X} - (\mathbf{I} - \mathbf{H}_\mathcal{X})\mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p\mathbf{H}_\mathcal{X} + (\mathbf{I} - \mathbf{H}_\mathcal{X})\mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p \\ &= \mathbf{H}_\mathcal{X} + (\mathbf{I} - \mathbf{H}_\mathcal{X})(-\mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p\mathbf{H}_\mathcal{X} + \mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p) \\ &= \mathbf{H}_\mathcal{X} + (\mathbf{I} - \mathbf{H}_\mathcal{X})(\mathbf{x}^p(a^{-1}){}^t\mathbf{x}^p)(\mathbf{I} - \mathbf{H}_\mathcal{X}) \\ &= \mathbf{H}_\mathcal{X} + \tilde{\mathbf{x}}^p({}^t\tilde{\mathbf{x}}^p\tilde{\mathbf{x}}^p)^{-1}{}^t\tilde{\mathbf{x}}^p \\ &= \mathbf{H}_\mathcal{X} + \mathbf{H}_{\tilde{\mathbf{x}}^p} \end{aligned} \quad (3.135)$$

parce que

$$a = {}^t\mathbf{x}^p(\mathbf{I} - \mathbf{H}_\mathcal{X})\mathbf{x}^p = {}^t\mathbf{x}^p(\mathbf{I} - \mathbf{H}_\mathcal{X})(\mathbf{I} - \mathbf{H}_\mathcal{X})\mathbf{x}^p = {}^t\tilde{\mathbf{x}}^p\tilde{\mathbf{x}}^p \quad \square \quad (3.136)$$

Proposition 3.8.2. *Considérons le modèle (3.125). $\mathbf{U}^\pi = \hat{\boldsymbol{\beta}}' \mathbf{X} + \boldsymbol{\epsilon}'$, les estimateurs MCO de $\text{Var}(\hat{\beta}_{p,\pi})$ par la méthode de Freedman-Lane et le modèle ci-dessus sont égaux. C'est-à-dire que :*

$$\widehat{\text{Var}}(\hat{\beta}_{p,\pi}^F) = S^2(\hat{\beta}_{p,\pi}^F) = \frac{S^{2F}}{a} = \frac{S^{2'}}{a} = S^2(\hat{\beta}_{p,\pi}^{\hat{F}}) = \widehat{\text{Var}}(\hat{\beta}_{p,\pi}^{\hat{F}})$$

où

$$a = {}^t\mathbf{x}^p(\mathbf{I} - \mathbf{H}_\mathcal{X})\mathbf{x}^p = {}^t\mathbf{x}^p(\mathbf{I} - \mathbf{H}_\mathcal{X})(\mathbf{I} - \mathbf{H}_\mathcal{X})\mathbf{x}^p = {}^t\tilde{\mathbf{x}}^p\tilde{\mathbf{x}}^p$$

et

$$S^{2F} = \frac{{}^t\boldsymbol{\epsilon}^F \boldsymbol{\epsilon}^F}{n - k} = \frac{{}^t\boldsymbol{\epsilon}' \boldsymbol{\epsilon}'}{n - k} = S^{2'}$$

Preuve : Nous savons que $\text{Var}(\hat{\beta}') = ({}^t\mathbf{X}\mathbf{X})^{-1}\sigma^{2'}$ et $\text{Var}(\hat{\beta}^F) = ({}^t\mathbf{X}\mathbf{X})^{-1}\sigma^{2F}$ (cf. Draper *et al.* [11]). Donc :

$$\text{Var}(\hat{\beta}_{p,\pi}^F) = \sigma^{2F} [({}^t\mathbf{X}\mathbf{X})^{-1}]_{k \times k} \quad (3.137)$$

et

$$\text{Var}(\hat{\beta}'_{p,\pi}) = \sigma^{2'} [({}^t\mathbf{X}\mathbf{X})^{-1}]_{k \times k} \quad (3.138)$$

où $[({}^t\mathbf{X}\mathbf{X})^{-1}]_{k \times k}$ est l'élément $(k \times k)$ de la matrice $({}^t\mathbf{X}\mathbf{X})^{-1}$.

Par ailleurs, la proposition (3.7.4) indique que :

$$[({}^t\mathbf{X}\mathbf{X})^{-1}]_{k \times k} = \frac{1}{a} \quad (3.139)$$

Par conséquent,

$$\text{Var}(\hat{\beta}_{p,\pi}^F) = \frac{\sigma^{2F}}{a}. \quad (3.140)$$

et donc, l'estimateur MCO de $\text{Var}(\hat{\beta}_{p,\pi}^F)$ s'écrit :

$$\widehat{\text{Var}}(\hat{\beta}_{p,\pi}^F) = S^2(\hat{\beta}_{p,\pi}^F) = \frac{S^{2F}}{a} \quad (3.141)$$

où $S^{2F} = \frac{{}^t\boldsymbol{\epsilon}^F \boldsymbol{\epsilon}^F}{n - k}$. Mais d'après la remarque (3.8.1)(c) permet de vérifier que :

$$S^{2F} = \frac{{}^t\boldsymbol{\epsilon}^F \boldsymbol{\epsilon}^F}{n - k} = \frac{{}^t\boldsymbol{\epsilon}' \boldsymbol{\epsilon}'}{n - k} = S^{2'}.$$

Enfin,

$$\widehat{\text{Var}}(\hat{\beta}_{p,\pi}^F) = S^2(\hat{\beta}_{p,\pi}^F) = \frac{S^{2F}}{a} = \frac{S^{2'}}{a} = S^2(\hat{\beta}'_{p,\pi}) = \widehat{\text{Var}}(\hat{\beta}'_{p,\pi}) \quad \square \quad (3.142)$$

3.8.2 Cas de la méthode de Kennedy

Pour obtenir la méthode de Kennedy avec p variables explicatives, il faut d'abord réaliser une régression de la variable X^p en x^1, x^2, \dots, x^{p-1} , au sens MCO et suivant l'équation matricielle :

$$\mathbf{X}^p = \mathcal{X}^t \boldsymbol{\Gamma} + \mathbf{T} \quad (3.143)$$

où $\mathcal{X} = (\mathbf{1}, \mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^{p-1})$, ${}^t\boldsymbol{\Gamma} = \{\gamma_0, \gamma_1, \dots, \gamma_{p-1}\}$ et $\mathbf{T} = \{\tau_1, \tau_2, \dots, \tau_n\}$. L'équation ci-dessus nous conduit à calculer :

$$\hat{\mathbf{X}}^p = \mathcal{X} \hat{\boldsymbol{\Gamma}} \quad (3.144)$$

où $\mathbf{G} = \hat{\boldsymbol{\Gamma}} = \{\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_{p-1}\} = ({}^t\mathcal{X}\mathcal{X})^{-1} {}^t\mathcal{X}\mathbf{x}^p$ est l'estimateur MCO de $\boldsymbol{\Gamma}$. On calcule alors des résidus observés \mathbf{t} comme suit :

$$\mathbf{t} = \mathbf{x}^P - \hat{\mathbf{x}}^P = \mathbf{x}^P - \mathcal{X}(\mathcal{X}\mathcal{X})^{-1}\mathcal{X}\mathbf{x}^P = (\mathbf{I} - \mathbf{H}_{\mathcal{X}})\mathbf{x}^P = \tilde{\mathbf{x}}^p \quad (3.145)$$

où $\mathbf{t} = \{t_1, t_2, \dots, t_n\}$. Enfin on réalise une régression au sens MCO de la variable des résidus permutés U^π en t d'après le modèle :

$$U_i^\pi = \beta_{p,\pi}^K t_i + \epsilon_i^K \quad (3.146)$$

où $\text{Var}(\epsilon^K) = \sigma^{2K}$. Remarquons que du fait que les deux variables U_i^π et t_i sont centrées, le modèle de régression est sans terme constant. L'équation ci-dessus conduit à calculer la prévision

$$\hat{U}^\pi = \hat{\beta}_{p,\pi}^K t_i \quad (3.147)$$

où

$$\hat{\beta}_{p,\pi}^K = \frac{{}^t\mathbf{t}\mathbf{U}^\pi}{{}^t\mathbf{t}\mathbf{t}} = \frac{{}^t\mathbf{x}^p[\mathbf{I} - \mathbf{H}_{\mathcal{X}}]\mathbf{U}}{{}^t\mathbf{x}^p[\mathbf{I} - \mathbf{H}_{\mathcal{X}}]\mathbf{x}^P} = ({}^t\tilde{\mathbf{x}}^p\tilde{\mathbf{x}}^p)^{-1}{}^t\tilde{\mathbf{x}}^p\mathbf{U}^\pi \quad (3.148)$$

Par ailleurs, la variance de $\hat{\beta}_{p,\pi}$ est :

$$\text{Var}(\hat{\beta}_{p,\pi}^K)_K = \frac{\sigma^{2K}}{\sum_{i=1}^n t_i^2} = \frac{\sigma^{2K}}{{}^t\mathbf{t}\mathbf{t}} = \frac{\sigma^{2K}}{{}^t\mathbf{x}^P(\mathbf{I} - \mathbf{H}_{\mathcal{X}})\mathbf{x}^P} = \frac{\sigma^{2K}}{{}^t\tilde{\mathbf{x}}^p\tilde{\mathbf{x}}^p} = \frac{\sigma^{2K}}{a} \quad (3.149)$$

et un estimateur MCO de cette variance est :

$$\widehat{\text{Var}(\hat{\beta}_{p,\pi}^K)_K} = S^2(\hat{\beta}_{p,\pi}^K)_K = \frac{S^{2K}}{a} \quad (3.150)$$

où $S^{2K} = \frac{{}^t\hat{\epsilon}^K\hat{\epsilon}^K}{n-k} = \frac{\sum(\epsilon_i^k)^2}{n-k}$. La proposition (3.2.2) considérait le cas de deux variables explicatives et elle montre que $\hat{\beta}_{2,\pi}^F = \hat{\beta}_{2,\pi}^k$. Sa démonstration se généralise de façon immédiate au cas $p \geq 3$ et l'on peut montrer ici que $\hat{\beta}_{p,\pi}^F = \hat{\beta}_{p,\pi}^k$. Nous allons désormais noter cet estimateur $\hat{\beta}_{p,\pi}$.

Lemme 3.8.4. *Considérons le modèle de Kennedy*

$$U_i^\pi = \beta_{p,\pi} t_i + \epsilon_i^K$$

Alors, la somme des carrés des résidus de ce modèle vérifie :

$$\sum_{i=1}^n (\epsilon_i^k)^2 = {}^t\hat{\epsilon}^K\hat{\epsilon}^K = {}^t\mathbf{U}^\pi\mathbf{U}^\pi - {}^t\mathbf{U}^\pi\mathbf{H}_{\tilde{\mathbf{x}}^p}\mathbf{U}^\pi$$

Preuve : On sait que :

$$\begin{aligned} \hat{\epsilon}^K &= \mathbf{U}^\pi - \mathbf{t}\hat{\beta}_{p,\pi} &= \mathbf{U}^\pi - \mathbf{t} \cdot ({}^t\tilde{\mathbf{x}}^p\tilde{\mathbf{x}}^p)^{-1}{}^t\tilde{\mathbf{x}}^p\mathbf{U}^\pi \\ & &= \mathbf{U}^\pi - \tilde{\mathbf{x}}^p \cdot ({}^t\tilde{\mathbf{x}}^p\tilde{\mathbf{x}}^p)^{-1}{}^t\tilde{\mathbf{x}}^p\mathbf{U}^\pi \\ & &= \mathbf{U}^\pi - \mathbf{H}_{\tilde{\mathbf{x}}^p}\mathbf{U}^\pi \\ & &= (\mathbf{I} - \mathbf{H}_{\tilde{\mathbf{x}}^p})\mathbf{U}^\pi \end{aligned} \quad (3.151)$$

alors :

$$\begin{aligned}
 {}^t\hat{\epsilon}^K \hat{\epsilon}^K &= {}^t\mathbf{U}^\pi (\mathbf{I} - \mathbf{H}_{\tilde{\mathbf{x}}^p}) (\mathbf{I} - \mathbf{H}_{\tilde{\mathbf{x}}^p}) \mathbf{U}^\pi \\
 &= {}^t\mathbf{U}^\pi (\mathbf{I} - \mathbf{H}_{\tilde{\mathbf{x}}^p}) \mathbf{U}^\pi \\
 &= {}^t\mathbf{U}^\pi \mathbf{U}^\pi - {}^t\mathbf{U}^\pi \mathbf{H}_{\tilde{\mathbf{x}}^p} \mathbf{U}^\pi \quad \square
 \end{aligned} \tag{3.152}$$

Lemme 3.8.5.

$$\sum_{i=1}^n (\hat{\epsilon}_i^k)^2 - \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 = {}^t\hat{\epsilon}^K \hat{\epsilon}^K - {}^t\hat{\epsilon}^F \hat{\epsilon}^F = {}^t\mathbf{U}^\pi \mathbf{H}_{\mathcal{X}} \mathbf{U}^\pi \geq 0$$

Preuve : Les lemmes (3.8.1) et (3.8.4) ont pour corollaire :

$$\begin{aligned}
 {}^t\hat{\epsilon}^F \hat{\epsilon}^F &= {}^t\mathbf{U}^\pi \mathbf{U}^\pi - {}^t\mathbf{U}^\pi \mathbf{H}_{\mathbf{X}} \mathbf{U}^\pi \\
 {}^t\hat{\epsilon}^K \hat{\epsilon}^K &= {}^t\mathbf{U}^\pi \mathbf{U}^\pi - {}^t\mathbf{U}^\pi \mathbf{H}_{\tilde{\mathbf{x}}^p} \mathbf{U}^\pi
 \end{aligned}$$

En appliquant la proposition (3.8.1), on obtient :

$${}^t\hat{\epsilon}^F \hat{\epsilon}^F = {}^t\mathbf{U}^\pi \mathbf{U}^\pi - {}^t\mathbf{U}^\pi \mathbf{H}_{\mathcal{X}} \mathbf{U}^\pi = {}^t\mathbf{U}^\pi \mathbf{U}^\pi - {}^t\mathbf{U}^\pi \mathbf{H}_{\mathcal{X}} \mathbf{U}^\pi - {}^t\mathbf{U}^\pi \mathbf{H}_{\tilde{\mathbf{x}}^p} \mathbf{U}^\pi$$

et donc,

$$\begin{aligned}
 \sum_{i=1}^n (\hat{\epsilon}_i^k)^2 - \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 &= {}^t\hat{\epsilon}^K \hat{\epsilon}^K - {}^t\hat{\epsilon}^F \hat{\epsilon}^F \\
 &= {}^t\mathbf{U}^\pi \mathbf{U}^\pi - {}^t\mathbf{U}^\pi \mathbf{H}_{\tilde{\mathbf{x}}^p} \mathbf{U}^\pi - {}^t\mathbf{U}^\pi \mathbf{U}^\pi + {}^t\mathbf{U}^\pi \mathbf{H}_{\mathcal{X}} \mathbf{U}^\pi + {}^t\mathbf{U}^\pi \mathbf{H}_{\tilde{\mathbf{x}}^p} \mathbf{U}^\pi \\
 &= {}^t\mathbf{U}^\pi \mathbf{H}_{\mathcal{X}} \mathbf{U}^\pi \geq 0 \quad \square
 \end{aligned} \tag{3.153}$$

qui est égal au produit scalaire entre le vecteur \mathbf{U}^π et sa projection sur l'espace engendré par les vecteur $\mathbf{1}, \mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^{p-1}$.

Proposition 3.8.3. *La différence entre les estimateurs de $\text{Var}(\hat{\beta}_{p,\pi})$ par les méthodes de Freedman-Lane et de Kennedy est supérieure ou égale à zéro. C'est-à-dire que :*

$$\widehat{\text{Var}}(\hat{\beta}_{p,\pi})_K - \widehat{\text{Var}}(\hat{\beta}_{p,\pi})_F = S^2(\hat{\beta}_{p,\pi})_K - S^2(\hat{\beta}_{p,\pi})_F = \frac{{}^t\mathbf{U}^\pi \mathbf{H}_{\mathcal{X}} \mathbf{U}^\pi}{(n-k)a} \geq 0$$

où $a = {}^t\mathbf{x}^p [\mathbf{I} - \mathbf{H}_{\mathcal{X}}] \mathbf{x}^p$.

Preuve : D'après les équations (3.141) et (3.150), on a :

$$\begin{aligned}
S^2(\hat{\beta}_{p,\pi})_K - S^2(\hat{\beta}_{p,\pi})_F &= \frac{S^{2K}}{a} - \frac{S^{2F}}{a} \\
&= \frac{S^{2K} - S^{2F}}{a} \\
&= \frac{\sum_{i=1}^n (\hat{\epsilon}_i^k)^2 - \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{(n-k)a} \tag{3.154}
\end{aligned}$$

$$= \frac{{}^t\mathbf{U}^\pi \mathbf{H} \chi \mathbf{U}^\pi}{(n-k)a} \geq 0 \tag{3.155}$$

par application du lemme (3.8.5). \square

La proposition ci-dessus dit que :

$$S^2(\hat{\beta}_{p,\pi})_K \geq S^2(\hat{\beta}_{p,\pi})_F \tag{3.156}$$

On peut donc comparer les qualités des deux méthodes pour effectuer un test de l'hypothèse $H_0 : \beta_p = 0$ contre $H_1 : \beta_p \neq 0$:

1. Pour chaque permutation, la valeur absolue de la statistique de test de la méthode de Freedman-Lane est toujours supérieure ou égale à celle de la méthode de Kennedy :

$$|T_{p,\pi}^K| = \frac{|\hat{\beta}_{p,\pi}|}{S(\hat{\beta}_{p,\pi})_K} \leq \frac{|\hat{\beta}_{p,\pi}|}{S(\hat{\beta}_{p,\pi})_F} = |T_{p,\pi}^F|; \tag{3.157}$$

2. La distribution permutationnelle de $|T_{2,\pi}^K|$ a une dispersion supérieure ou égale à celle de $|T_{2,\pi}^F|$;

3. La p-valeur permutationnelle de la méthode de Freedman-Lane est supérieure à celle de la méthode de Kennedy :

$$p_{p,\pi}^F \geq p_{p,\pi}^k \tag{3.158}$$

Corollaire 3.8.1. *L'inégalité (3.158) induit deux conséquences :*

1. *Le risque d'erreur de première espèce encouru par la méthode de Freedman-Lane est inférieur ou égal à celui de la méthode de Kennedy ;*
2. *La puissance du test par la méthode de Freedman-Lane est inférieure ou égale à celle de la méthode de Kennedy.*

Preuve : D'après l'inégalité (3.158) ci-dessus, on a :

$$p_{p,\pi}^F \geq p_{p,\pi}^k$$

Par ailleurs, nous savons que lorsque la p-valeur observée est plus petite que le niveau α du test, on rejette l'hypothèse $H_0 : \beta_p = 0$. Par conséquent, comme dans le

corollaire (3.4.2) la probabilité de rejeter $H_0 : \beta_p = 0$ par la méthode de Freedman-Lane est inférieure ou égale à celle qui est obtenue par la méthode de Kennedy, c'est-à-dire :

$$Pr(P_{p,\pi}^F < \alpha) \leq Pr(P_{p,\pi}^K < \alpha) \tag{3.159}$$

Donc, la fonction puissance permutationnelle de la méthode de Freedman-Lane est inférieure ou égale à celle de la méthode de Kennedy. Si C est la région critique du test, on a :

$$\pi_\pi^F(\beta_p) = Pr_\pi^F(C) \leq Pr_\pi^K(C) = \pi_\pi^K(\beta_p) \tag{3.160}$$

1. Par conséquent, pour des données générées sous l'hypothèse $H_0 : \beta_p = 0$, la relation ci-dessus dit que le risque d'erreur de première espèce de la méthode de Freedman-Lane est inférieur à celui de Kennedy.

$$Pr_{H_0}(P_{p,\pi}^F < \alpha) \leq Pr_{H_0}(P_{p,\pi}^K < \alpha) \tag{3.161}$$

Ces résultats donnent un support mathématique aux résultats des simulations d'Andersson et Legendre [1] (Voir page 289).

2. Par contre, lorsque les données sont générées sous l'hypothèse alternative $H_1 : \beta_p \neq 0$ la relation ci-dessus dit que la puissance du test de la méthode de Freedman-Lane est inférieure ou égale à celle de la méthode de Kennedy.

$$Pr_{H_1}(P_{p,\pi}^F < \alpha) \leq Pr_{H_1}(P_{p,\pi}^K < \alpha) \quad \square \tag{3.162}$$

Cependant, lors de la vérification des puissances de tests par des méthodes permutationnelles pour p variables la méthode de Kennedy n'a été considérée ni dans les simulations d'Anderson et Legendre [1], ni celle de Legendre (2000). D'autre part, d'après la proposition (3.8.3), on a :

$$S^2(\hat{\beta}_{p,\pi})_K - S^2(\hat{\beta}_{p,\pi})_F = \frac{{}^t\mathbf{U}^\pi \mathbf{H}_X \mathbf{U}^\pi}{(n-k)a} \geq 0$$

où $a = {}^t\mathbf{x}^p [\mathbf{I} - \mathbf{H}_X] \mathbf{x}^p$.

Les valeurs a et n sont fixées par les données. Par suite, pour chaque permutation, plus la valeur ${}^t\mathbf{U}^\pi \mathbf{H}_X \mathbf{U}^\pi$ est grande et plus la différence des deux estimateurs est élevée. Par ailleurs, les deux estimateurs sont égaux si ${}^t\mathbf{U}^\pi \mathbf{H}_X \mathbf{U}^\pi = 0$. Géométriquement, on peut dire que lorsque le vecteur aléatoire \mathbf{U}^π est orthogonal à sa projection sur l'espace déterminé par $\mathbf{1}, \mathbf{X}^1, \dots, \mathbf{X}^{p-1}$, c'est-à-dire, lorsque X^p est non corrélé aux autres prédicteurs. Alors, les deux méthodes donnent des estimateurs identiques.

Proposition 3.8.4. *Supposons le nombre de prédicteurs k fixé. Si $\frac{{}^t\mathbf{U}^\pi \mathbf{H}_X \mathbf{U}^\pi}{a}$ (où $a = {}^t\mathbf{x}^p [\mathbf{I} - \mathbf{H}_X] \mathbf{x}^p \neq 0$) admet une limite finie l quand n tend vers l'infini, la différence entre les deux estimateurs de $\mathbb{V}\text{ar}(\hat{\beta}_{2,\pi})$ tend vers zéro.*

Preuve : D'après la proposition (3.8.3) on a :

$$S^2(\hat{\beta}_{p,\pi})_K - S^2(\hat{\beta}_{p,\pi})_F = \frac{{}^t\mathbf{U}^\pi \mathbf{H}_X \mathbf{U}^\pi}{(n-k)a} \geq 0$$

alors :

$$\begin{aligned} \lim_{n \rightarrow +\infty} S^2(\hat{\beta}_{p,\pi})_K - S^2(\hat{\beta}_{p,\pi})_F &= \lim_{n \rightarrow +\infty} \frac{{}^t\mathbf{U}^\pi \mathbf{H}_X \mathbf{U}^\pi}{(n-k)a} \\ &= \lim_{n \rightarrow +\infty} \frac{{}^t\mathbf{U}^\pi \mathbf{H}_X \mathbf{U}^\pi}{a} \times \lim_{n \rightarrow +\infty} \frac{1}{n-k} \\ &= l \times 0 \\ &= 0 \end{aligned} \tag{3.163}$$

c'est-à-dire que les deux estimateurs sont asymptotiquement égaux. Lorsque la taille de l'échantillon est grande, on en déduit que les statistiques de test par les méthodes de Freedman-Lane et de Kennedy deviennent équivalentes, c'est-à-dire :

$$T_{p,\pi}^F = \frac{\hat{\beta}_{p,\pi}}{S(\hat{\beta}_{p,\pi})_F} \approx \frac{\hat{\beta}_{p,\pi}}{S(\hat{\beta}_{p,\pi})_K} = T_{p,\pi}^K$$

Par conséquent, lorsque la taille de l'échantillon est grande, on encourt le même risque d'erreur de première espèce, ce que confirme la simulation d'Anderson et Legendre page (289) où $n = 90$. On obtient aussi la même puissance de test.

3.8.3 Influence de la normalisation

Considérons à nouveau le modèle de Kennedy :

$$U_i^\pi = \beta_{p,\pi} t_i + \epsilon_i^K$$

où $\text{Var}(\epsilon^K) = \sigma^{2K}$. Comme dans la section (4.4) où on avait deux variables explicatives, $p = 2$, ici à chaque permutation π , on peut considérer qu'on dispose de $(n-2)$ degrés de liberté pour la statistique de test permutationnelle de la méthode de Kennedy ($T_{p,\pi}^K = \hat{\beta}_{p,\pi}/S(\hat{\beta}_{p,\pi})_K$, puisque pour chaque permutation on réalise une régression de la variable U^π en T). On ajuste donc un modèle linéaire à deux paramètres au lieu des k paramètres considérés par la méthode de Freedman-Lane. Cela revient à considérer U^π et T comme deux nouvelles variables (Anderson et Legendre [1]) et non par des variables déduites de traitements antérieurs. Dans ce cas, un estimateur MCO de σ^{2K} dans le modèle de Kennedy est $S^{2K*} = \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 / (n-2)$, c'est-à-dire que la somme des carrés des résidus est divisée par $(n-2)$ au lieu de $(n-k)$. On dit aussi qu'elle est normalisée par $(n-2)$. L'équation (3.150) permet d'obtenir un estimateur MCO pour la variance de $\hat{\beta}_{p,\pi}$ sous le modèle de Kennedy s'écrit donc :

$$\widehat{\text{Var}}(\hat{\beta}_{p,\pi})_{K^*} = S^2(\beta_{p,\pi})_{K^*} = \frac{S^{2K}}{a} = \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{(n-2)a} \quad (3.164)$$

Contrairement à la proposition (3.8.3) (où nous avons considéré $(n-k)$ degrés de liberté pour la statistique de test permutatonnelle suivant la méthode de Kennedy) nous avons montré que l'estimateur MCO de $\text{Var}(\hat{\beta}_{p,\pi})$ par la méthode de Freedman-Lane est inférieur à celui de la méthode de Kennedy, dans la proposition suivante nous montrerons lorsque les résidus de la méthode de Kennedy est normalisé en $(n-2)$ degrés de liberté, il existe des cas où l'estimateur de $\text{Var}(\hat{\beta}_{p,\pi})$ obtenu par la méthode de Freedman-Lane est supérieur à celui de la méthode de Kennedy corrigée.

Proposition 3.8.5. *Si $a \neq 0$, pour la méthode de Kennedy, en retenant $(n-2)$ degrés de liberté et non plus $(n-k)$ degrés de liberté, nous avons :*

$$S^2(\hat{\beta}_{p,\pi})_{K^*} - S^2(\hat{\beta}_{p,\pi})_F \begin{cases} \geq 0 & \text{si } {}^t\mathbf{U}^\pi \mathbf{H}_X \mathbf{U}^\pi \geq (k-2)S^{2F} \\ < 0 & \text{si } {}^t\mathbf{U}^\pi \mathbf{H}_X \mathbf{U}^\pi < (k-2)S^{2F} \end{cases}$$

$$\text{où, } a = {}^t\mathbf{x}^p[\mathbf{I} - \mathbf{H}_X]\mathbf{x}^p, \quad S^2(\hat{\beta}_{p,\pi})_{K^*} = \frac{S^{2K^*}}{a} \quad \text{et} \quad S^2(\hat{\beta}_{p,\pi})_F = \frac{S^{2F}}{a}$$

Preuve : Nous savons que : $S^2(\beta_{p,\pi})_{K^*} = \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{(n-2)a}$ $S^2(\beta_{p,\pi})_F = \frac{\sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{(n-k)a}$
donc,

$$\begin{aligned} S^2(\hat{\beta}_{p,\pi})_{K^*} - S^2(\hat{\beta}_{p,\pi})_F &= \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{(n-2)a} - \frac{\sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{(n-k)a} \\ &= \frac{[(n-k) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - (n-2) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2]}{(n-2)(n-k)a} \end{aligned}$$

les valeurs $(n-2)$, $(n-k)$ et a sont positives. Par suite,

$$S^2(\hat{\beta}_{p,\pi})_{K^*} - S^2(\hat{\beta}_{p,\pi})_F \begin{cases} \geq 0 & \text{si } [(n-k) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - (n-2) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2] \geq 0 \\ < 0 & \text{si } [(n-k) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - (n-2) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2] < 0 \end{cases}$$

Mais,

$$\begin{aligned}
\left[(n-k) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - (n-2) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 \right] &\geq 0 \quad \text{ce qui confirme} \\
\frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n} &\geq \frac{n-2}{n-k} \quad \text{et donc} \\
\frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\epsilon}_i^F)^2}{\sum_{i=1}^n (\hat{\epsilon}_i^F)^2} &\geq \frac{k-2}{n-k} \quad \text{mais d'après le lemme (3.8.5)} \\
\frac{{}^t\mathbf{U}^\pi \mathbf{H}_\mathcal{X} \mathbf{U}^\pi}{\sum_{i=1}^n (\hat{\epsilon}_i^F)^2} &\geq \frac{k-2}{n-k} \quad \text{et d'après la définition de } S^{2F} \\
{}^t\mathbf{U}^\pi \mathbf{H}_\mathcal{X} \mathbf{U}^\pi &\geq (k-2)S^{2F} \quad (3.165)
\end{aligned}$$

Par conséquent :

$$S^2(\hat{\beta}_{p,\pi})_{K^*} - S^2(\hat{\beta}_{p,\pi})_F \begin{cases} \geq 0 & \text{si } {}^t\mathbf{U}^\pi \mathbf{H}_\mathcal{X} \mathbf{U}^\pi \geq (k-2)S^{2F} \\ < 0 & \text{si } {}^t\mathbf{U}^\pi \mathbf{H}_\mathcal{X} \mathbf{U}^\pi < (k-2)S^{2F} \end{cases} \quad (3.166)$$

C'est-à-dire que contrairement à la proposition (3.8.3) où l'on avait toujours $S^2(\hat{\beta}_{p,\pi})_K \geq S^2(\hat{\beta}_{p,\pi})_F$, ici, en retenant $(n-2)$ degrés de liberté pour la méthode de Kennedy, pour chaque permutation où la valeur ${}^t\mathbf{U}^\pi \mathbf{H}_\mathcal{X} \mathbf{U}^\pi$ est inférieure à $(k-2)S^{2F}$, on sait que

$$S^2(\hat{\beta}_{p,\pi})_{K^*} < S^2(\hat{\beta}_{p,\pi})_F \quad \square$$

Corollaire 3.8.2. *Le test d'hypothèse $H_0 : \beta_p = 0$ contre $H_1 : \beta_p \neq 0$, en considérant $(n-2)$ degrés de liberté pour la méthode de Kennedy et non plus $(n-k)$, vérifie les deux propriétés suivantes :*

1. *il existe des cas où : $Pr_{H_0}(P_{p,\pi}^F < \alpha) > Pr_{H_0}(P_{p,\pi}^{K^*} < \alpha)$*
2. *il existe des cas où : $Pr_{H_1}(P_{p,\pi}^F < \alpha) > Pr_{H_1}(P_{p,\pi}^{K^*} < \alpha)$*

Preuve : D'après la proposition (3.8.5) on a :

$$S^2(\hat{\beta}_{p,\pi})_{K^*} - S^2(\hat{\beta}_{p,\pi})_F \begin{cases} \geq 0 & \text{si } {}^t\mathbf{U}^\pi \mathbf{H}_\mathcal{X} \mathbf{U}^\pi \geq (k-2)S^{2F} \\ < 0 & \text{si } {}^t\mathbf{U}^\pi \mathbf{H}_\mathcal{X} \mathbf{U}^\pi < (k-2)S^{2F} \end{cases}$$

De ce fait, pour la méthode de Kennedy corrigée il existe des cas où $S^2(\hat{\beta}_{p,\pi})_{K^*} < S^2(\hat{\beta}_{p,\pi})_F$. On en déduit qu'il existe des cas où $T_{p,\pi}^{K^*} = \hat{\beta}_{p,\pi}/S(\hat{\beta}_{p,\pi})_{K^*} > T_{p,\pi}^F = \hat{\beta}_{p,\pi}/S(\hat{\beta}_{p,\pi})_F$.

Ainsi, contrairement au corollaire (3.4.1) on ne peut pas toujours dire que la p-valeur permutationnelle de la méthode de Freedman-Lane est supérieure ou égale à celle de la méthode de Kennedy ($p_{p,\pi}^F \geq p_{p,\pi}^K$), et la relation

$$Pr(P_{p,\pi}^F < \alpha) \leq Pr(P_{p,\pi}^K < \alpha)$$

n'est pas toujours vraie si l'on retient la définition de $S^2(\hat{\beta}_{2,\pi})_{K^*}$ avec $(n - 2)$ degrés de liberté. Il existe alors des cas où

$$Pr(P_{p,\pi}^F < \alpha) > Pr(P_{p,\pi}^{K^*} < \alpha) \tag{3.167}$$

Par conséquent, pour des données générées sous l'hypothèse $H_0 : \beta_2 = 0$, la relation ci-dessus peut être précisée :

$$\text{il existe des cas où : } Pr_{H_0}(P_{p,\pi}^F < \alpha) > Pr_{H_0}(P_{p,\pi}^{K^*} < \alpha) \tag{3.168}$$

Elle dit qu'il existe des cas où le risque d'erreur de première espèce de la méthode de Kennedy avec $(n - 2)$ degrés de liberté est inférieur à celui de Freedman-Lane. Mais dans la simulation d'Anderson et Legendre (cf [1]page 289) du fait de la faible étendue du domaine des valeur des facteurs, on n'observait pas de cas où le risque d'erreur de première espèce de la méthode de Kennedy avec $(n - 2)$ degrés de liberté est inférieur à celui de la méthode de Freedman-Lane.

Par contre, lorsque les données sont générées sous l'hypothèse $H_1 : \beta_2 \neq 0$ la relation (3.168) s'écrit précisément :

$$\text{il existe des cas où : } Pr_{H_0}(P_{p,\pi}^F < \alpha) > Pr_{H_0}(P_{p,\pi}^{K^*} < \alpha) \tag{3.169}$$

Cela dit qu'il existe des cas où la puissance de test de la méthode de Freedman-Lane est supérieure à celle de la méthode de Kennedy. Remarquons que dans la simulation d'Anderson et Legendre la puissance n'est pas contrôlée. \square

Au bilan, comme nous l'avons montré pour $p = 2$ dans le corollaire (3.4.3), dans le cas général pour $p > 1$ il existe des cas où le risque d'erreur de première espèce de la méthode de Kennedy avec $(n - 2)$ degrés de liberté est inférieur à celui de la méthode de Freedman-Lane, et par contre il existe des cas où la puissance de la méthode de Kennedy avec $(n - 2)$ degrés de liberté est inférieur à celui de la méthode de Freedman-Lane.

Proposition 3.8.6. *Quand k est fixé, si ${}^t\mathbf{U}^\pi \mathbf{H}_X \mathbf{U}^\pi / a$ (où $a = {}^t\mathbf{x}^p [\mathbf{I} - \mathbf{H}_X] \mathbf{x}^p \neq 0$) admet une limite finie l , la différence entre les deux estimateurs de $\text{Var}(\hat{\beta}_{p,\pi})$ tend vers zéro quand n tend vers l'infini.*

Preuve : D'après la proposition ci-dessus on peut écrire :

$$\begin{aligned}
\lim_{n \rightarrow +\infty} [S^2(\hat{\beta}_{p,\pi})_{K*} - S^2(\hat{\beta}_{p,\pi})_F] &= \lim_{n \rightarrow +\infty} \left[\frac{\left[(n-k) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - (n-2) \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 \right]}{(n-2)(n-k)a} \right] \\
&\approx \lim_{n \rightarrow +\infty} \left[\frac{n \left(\sum_{i=1}^n (\hat{\epsilon}_i^K)^2 - \sum_{i=1}^n (\hat{\epsilon}_i^F)^2 \right)}{n^2 a} \right] \\
&\approx \lim_{n \rightarrow +\infty} \frac{{}^t \mathbf{U}^\pi \mathbf{H}_X \mathbf{U}^\pi}{a} \times \lim_{n \rightarrow +\infty} \frac{1}{n} \\
&\approx l \times 0 \\
&\approx 0
\end{aligned} \tag{3.170}$$

c'est-à-dire que les deux estimateurs sont asymptotiquement égaux. Par suite, lorsque la taille de l'échantillon de données est grande, les statistiques de test par les méthodes de Freedman-Lane et de Kennedy avec $(n-2)$ degrés de liberté sont approximativement égales :

$$T_{p,\pi}^F = \frac{\hat{\beta}_{p,\pi}}{S(\hat{\beta}_{p,\pi})_F} \approx \frac{\hat{\beta}_{p,\pi}}{S(\hat{\beta}_{p,\pi})_{K*}} = T_{p,\pi}^{K*}$$

Par conséquent, on doit observer dans ce cas le même risque d'erreur de première espèce, ce que confirme les simulations d'Anderson et Legendre (*cf.* [1] page 289 pour $n \geq 36$). On doit observer aussi la même puissance, même si cette étude n'entrait pas dans les simulations d'Anderson et Legendre. Nous avons traité ce sujet dans nos simulations pour $p = 2$ en section 2.5.3, pour $n > 36$.

Comme dans la proposition (3.4.6) où nous avons montré que l'estimateur de $\text{Var}(\hat{\beta}_2)$ par la méthode Kennedy avec $(n-2)$ degrés de liberté est plus petit que celle de la même méthode avec $(n-3)$ degrés de liberté. Nous démontrons dans la proposition suivante le résultat correspondant, dans un cadre général : il reste vrai dans ce cas que l'estimateur de $\text{Var}(\hat{\beta}_p)$ par la méthode Kennedy avec $(n-2)$ degrés de liberté est plus petit que celui de la même méthode avec $(n-k)$ degrés de liberté.

Proposition 3.8.7. *L'estimateur de $\text{Var}(\hat{\beta}_{p,\pi})$ par la méthode de Kennedy avec $(n-2)$ degrés de liberté, est toujours inférieur à celui de la même méthode avec $(n-k)$ degrés de liberté :*

$$S^2(\beta_{p,\pi})_K - S^2(\beta_{p,\pi})_{K*} = \frac{(k-2) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{(n-2)(n-k)a} \geq 0 \tag{3.171}$$

où $a = {}^t \mathbf{x}^p [\mathbf{I} - \mathbf{H}_X] \mathbf{x}^p$ est supposé fini et non nulle.

Preuve : Nous savons que l'estimateurs MCO de $\text{Var}(\hat{\beta}_{p,\pi})$ par la méthode de Kennedy avec $(n-2)$ et $(n-k)$ degrés de liberté sont :

$$S^2(\beta_{p,\pi})_K = \frac{S^{2K}}{a} = \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{(n-k)a}$$

$$S^2(\beta_{p,\pi})_{K^*} = \frac{S^{2K^*}}{a} = \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{(n-2)a}$$

donc :

$$\begin{aligned} S^2(\beta_{2,\pi})_K - S^2(\beta_{2,\pi})_{K^*} &= \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{(n-k)a} - \frac{\sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{(n-2)a} \\ &= \frac{(k-2) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{(n-2)(n-k)a} \\ &= \frac{(k-2)S^{2K}}{(n-2)a} \geq 0 \end{aligned} \quad (3.172)$$

Une conséquence de la proposition ci-dessus est que :

$$S^2(\beta_{2,\pi})_K - S^2(\beta_{2,\pi})_{K^*} = \frac{(k-2) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{(n-2)(n-k)a} = \frac{(k-2)S^{2K}}{(n-2)a}$$

Si S^{2K}/a (où $a = {}^t\mathbf{x}^p[\mathbf{I} - \mathbf{H}_{\mathcal{X}}]\mathbf{x}^p$) admet une limite finie et non nulle quand n est fixé et k tend vers l'infini, la différence entre les deux estimateur de $\text{Var}(\hat{\beta}_{p,\pi})$, augmente et on en déduit que la différence entre le risque d'erreur de première espèce de la méthode de Kennedy avec $(n-k)$ et celui avec $(n-2)$ degrés de liberté augmente, ce que confirme la simulation d'Anderson et Legendre (*cf.* [1] page 289).

D'après l'équation (3.172), on a :

$$S^2(\hat{\beta}_{p,\pi})_K \geq S^2(\hat{\beta}_{p,\pi})_{K^*} \quad (3.173)$$

et par suite :

1. Dans chaque permutation la valeur absolue de la statistique de test dans la méthode de Kennedy avec $(n-2)$ degrés de liberté est toujours supérieure à celle de la même méthode avec $(n-k)$ degrés de liberté, c'est-à-dire que :

$$|T_{p,\pi}^{K^*}| = \frac{|\hat{\beta}_{p,\pi}|}{S^2(\hat{\beta}_{p,\pi})_{K^*}} \geq \frac{|\hat{\beta}_{p,\pi}|}{S^2(\hat{\beta}_{p,\pi})_K} = |T_{p,\pi}^K| \quad (3.174)$$

2. La distribution permutationnelle de $|T_{p,\pi}^{K^*}|$ a une dispersion supérieure ou égale à celle de $|T_{p,\pi}^K|$.

3. La p-valeur permutationnelle de la méthode de Kennedy avec $(n-2)$ degrés de liberté est inférieure ou égale à celle de la même méthode avec $(n-K)$ degrés de liberté, c'est-à-dire que :

$$p_{p,\pi}^{K^*} \geq p_{p,\pi}^K \quad (3.175)$$

Corollaire 3.8.3. *Pour effectuer un test d'hypothèse $H_0 : \beta_p = 0$ contre $H_1 : \beta_p \neq 0$, à l'aide de la méthode de Kennedy :*

1. *un risque d'erreur de première espèce de la méthode avec $(n - 2)$ degrés de liberté inférieur ou égal à celui obtenu avec $(n - k)$ degrés de liberté.*
2. *une puissance de test de la méthode avec $(n - 2)$ degrés de liberté inférieure ou égale à celle obtenue avec $(n - k)$ degrés de liberté.*

Preuve : D'après la remarque ci-dessus, on a :

$$p_{2,\pi}^{K*} \geq p_{2,\pi}^K.$$

On conclue donc que la probabilité de rejeter $H_0 : \beta_p = 0$ par la méthode de Kennedy avec $(n - 2)$ degrés de liberté est inférieure ou égale à celle obtenue par la même méthode avec $(n - k)$ degrés de liberté, c'est-à-dire que :

$$Pr(P_{p,\pi}^{K*} < \alpha) \leq Pr(P_{p,\pi}^K < \alpha) \quad (3.176)$$

Donc, la fonction de puissance permutationnelle de la méthode de Kennedy avec $(n - 2)$ degrés de liberté est inférieure ou égale à celle de la même méthode avec $(n - k)$ degrés de liberté. Si C est la région critique du test, on a :

$$\eta_{\pi}^{K*}(\beta_p) = Pr_{\pi}^{K*}(C) \leq Pr_{\pi}^K(C) = \eta_{\pi}^K(\beta_p) \quad (3.177)$$

1. Par conséquent, pour des données générées sous l'hypothèse $H_0 : \beta_p = 0$, la relation ci-dessus dit que le risque d'erreur de première espèce de la méthode de Kennedy avec $(n - 2)$ degrés de liberté est inférieur à celui de la même méthode avec $(n - k)$ degrés de liberté.

$$Pr_{H_0}(P_{p,\pi}^{K*} < \alpha) \leq Pr_{H_0}(P_{p,\pi}^K < \alpha) \quad (3.178)$$

Ces résultats donnent un support mathématique aux résultats des simulations d'Andersson et Legendre (*cf.* [1]page 289).

2. Par contre, lorsque les données sont générées sous l'hypothèse $H_1 : \beta_p \neq 0$ la relation ci-dessus (4.63) dit que la puissance du test par la méthode de Kennedy avec $(n - 2)$ degrés de liberté est inférieure ou égale à celle de la même méthode avec $(n - k)$ degrés de liberté.

$$Pr_{H_1}(P_{p,\pi}^{K*} < \alpha) \leq Pr_{H_1}(P_{p,\pi}^K < \alpha) \quad (3.179)$$

Ce qui n'est pas contrôlé dans la simulation d'Anderson et Legendre.

Proposition 3.8.8. *Si k est fixé et si S^{2k}/a admet une limite finie et non nulle, notée g quand n tend vers l'infini, la différence entre les deux estimateurs de $\text{Var}(\hat{\beta}_{p,\pi})$ par la méthode de Kennedy avec $(n - 2)$ et $(n - k)$ degrés de liberté respectivement, sont asymptotiquement égaux.*

Preuve : D'après la proposition (3.8.7), on a :

$$S^2(\beta_{p,\pi})_K - S^2(\beta_{p,\pi})_{K*} = \frac{(k-2) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{(n-2)(n-k)a}$$

alors,

$$\begin{aligned} \lim_{n \rightarrow +\infty} [S^2(\hat{\beta}_{p,\pi})_K - S^2(\hat{\beta}_{p,\pi})_{K*}] &= \lim_{n \rightarrow +\infty} \left[\frac{(k-2)S^{2k}}{(n-2)a} \right] \\ &\approx \lim_{n \rightarrow +\infty} \frac{(k-2)S^{2k}}{a} \lim_{n \rightarrow +\infty} \frac{1}{n} \\ &= g \times 0 = 0 \quad \square \end{aligned} \quad (3.180)$$

Ainsi les statistiques $|T_{p,\pi}^K|$ et $|T_{p,\pi}^{K*}|$ sont asymptotiquement égales et pour effectuer un test sur l'hypothèse $H_0 : \beta_p = 0$ contre $H_0 : \beta_p \neq 0$, elles donnent les mêmes risques d'erreurs de première espèce et les mêmes puissances. C'est que confirme la simulation d'Anderson et Legendre (cf.†89 pour $n = 90$)

Remarque 3.8.2. D'après la proposition ci-dessus et en considérant la proposition (3.4.6), où on avait :

$$\begin{aligned} \lim_{n \rightarrow +\infty} [S^2(\hat{\beta}_{2,\pi})_K - S^2(\hat{\beta}_{2,\pi})_{K*}] &= \lim_{n \rightarrow +\infty} \left[\frac{S^2(X^1) \sum_{i=1}^n (\hat{\epsilon}_i^K)^2}{n(n-2)(n-3)\Delta} \right] \\ &\approx \lim_{n \rightarrow +\infty} \frac{S^2(X^1)S^{2K}}{\Delta} \lim_{n \rightarrow +\infty} \frac{1}{n^2} \\ &= \frac{\sigma_1^2 \sigma^{2K}}{\Delta*} \times 0 = 0 \quad \square \end{aligned}$$

Cela dit que la vitesse de convergence est en $O(n^{-2})$ si $k = 3$, est en $O(n^{-1})$ si $k > 3$ (cf. proposition ci-dessus). C'est ce que confirme la simulation d'Anderson et Legendre ([1], page 289, où le nombre de variables est 2, 5 et 10).

Chapitre 4

Tests d'association entre matrices de distance

4.1 Introduction

L'objectif de ce chapitre est à l'origine des recherches décrites dans cette thèse. Notre ambition initiale était en effet l'étude des tests de permutations utilisables pour étudier la dépendance partielle entre tableaux de distances interpoints.

On se place alors dans un cadre multivarié de dimension $p > 1$. Chaque unité statistique est alors observée à l'aide d'un vecteur aléatoire \mathbf{X} à valeurs dans un espace vectoriel normé de dimension $p > 1$ (p entier non nul) et l'on suppose donnée une distance entre unités statistiques déduit des valeurs de \mathbf{X} . L'exemple le plus classique est :

$$d_{ij} = d_{ij}(\mathbf{X}) = \|\mathbf{x}^i - \mathbf{x}^j\| = \left(\sum_{k=1}^p (x_k^i - x_k^j)^2 \right)^{1/2} \quad (4.1)$$

où \mathbb{R}^p est muni de la norme euclidienne standard

$$\|\mathbf{x}\|^2 = \sum_{j=1}^p (x_j)^2 = {}^t\mathbf{x}\mathbf{x}. \quad (4.2)$$

Lorsqu'on observe \mathbf{X} sur un échantillon de taille n , on peut associer à \mathbf{X} une matrice carrée d'ordre n , de terme général $d_{ij} = d_{ij}(\mathbf{X})$:

$$\mathbf{D} = D(\mathbf{X}) = \begin{pmatrix} 0 & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{12} & 0 & d_{23} & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{1n} & d_{2n} & \cdots & \cdots & 0 \end{pmatrix} \quad (4.3)$$

La formulation d'un modèle statistique en terme de matrices de distances offre une grande souplesse et par suite, a beaucoup d'applications dans les sciences empiriques telles que en économie, en biologie, en écologie etc. Elle permet d'unifier

les approches adaptées au cas de variables aléatoires ($p = 1$), de vecteurs aléatoires ($p > 1$) ou de fonctions aléatoires.

Lorsqu'on dispose d'une information empirique exprimée au travers de tableaux de distances, il devient naturel pour de nombreux chercheurs de rechercher des liaisons (linéaires) entre des matrices de distances. Un cas particulier de modèle appartenant à cette classe et très couramment utilisé formule le problème en termes de modèle de régression sur des matrices de distances.

Dans le cas de données dyadiques, comme dans la situation usuelle opérant sur des données vectorielles, l'analyse de régression classique se heurte à des difficultés importantes lorsque les postulats fondant la méthode sont violés par les données de distance disponibles.

Dans les deux cas, il est alors prudent d'opter pour des méthodes d'inférence libre (au sens distribution free) et en particulier pour des méthodes permutationnelles. Dans le cas de données disponibles sous la forme de tableaux de distances, deux obstacles demeurent :

1. les distances inter points sont des nombres positifs et d'échelle de rapport, le zéro représentant une situation de référence.

2. il n'est pas réaliste d'admettre que les distances d_{ik} et d_{jl} sont des variables aléatoires indépendantes dès que (i, k) et (j, l) ont au moins un élément en commun.

Les modèles de régression utilisés ne font donc pas intervenir de terme constant et sont donc d'un emploi plus délicat que les modèles usuels.

En 2000, Legendre ([35]) a étudié par simulation l'application des quatre méthodes de test de permutation présentées dans le chapitre 2 au cas de matrices des distances. Dans ce cas, les variables X^1 , X^2 et Y définissant un modèle de régression sont formées respectivement par les éléments de la partie triangulaire inférieure des matrices de distances $\mathbf{D}^1 = D(\mathbf{X}^1)$, $\mathbf{D}^2 = D(\mathbf{X}^2)$ et $\mathbf{D}^Y = D(\mathbf{Y})$. Ces trois vecteurs doivent vérifier les contraintes imposées par les propriétés d'une matrice de distance.

Dans ce chapitre, nous présentons tout d'abord le test de Mantel ([39]) proposé dans le cas de deux matrices de distance donc dans un cadre de régression simple. Ensuite, nous expliquons la méthode et les résultats obtenus par Legendre. Enfin nous complétons et comparons son travail grâce à une simulation plus large.

4.2 Le test de Mantel

Le test de Mantel [39] permet de déterminer si une variable est spatialement structurée ou plus généralement s'il y a une liaison entre deux variables. Pour cela, ce test a pour caractéristique de tester la liaison entre deux matrices de distances (ou de similarités) évaluées sur les mêmes objets.

Le test de Mantel suppose qu'on a observé les $n(n - 1)/2$ éléments de deux ma-

trices de distances entre les éléments d'un même ensemble¹ : une matrice A contenant les distances entre les objets caractérisés par la première variable et une matrice B contenant les distances entre les objets caractérisés par la deuxième variable.

$$\mathbf{A} = \begin{bmatrix} 0 & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & 0 & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & 0 & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & 0 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 0 & b_{12} & b_{13} & \cdots & b_{1n} \\ b_{21} & 0 & b_{23} & \cdots & b_{2n} \\ b_{31} & b_{32} & 0 & \cdots & b_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots & 0 \end{bmatrix}$$

Afin d'évaluer l'intensité de la relation linéaire entre les deux variables, on mesure alors l'association entre éléments de ces deux matrices A et B par une statistique appropriée telle que la corrélation entre les éléments correspondants.

Du fait que les matrices de distance sont symétriques, la corrélation entre tous les éléments hors diagonaux des deux matrices est la même que la corrélation entre les $N = n(n-1)/2$ éléments des parties triangulaires inférieures (ou supérieures), cette corrélation est donnée par l'équation :

$$R = \frac{\sum_* a_{ij} b_{ij} - \sum_* a_{ij} \sum_* b_{ij} / N}{\sqrt{[\sum_* a_{ij}^2 - (\sum_* a_{ij})^2 / N][\sum_* b_{ij}^2 - (\sum_* b_{ij})^2 / N]}} \quad (4.4)$$

où $\sum_* = \sum_{i < j}, i, j = 1, \dots, n$.

Par suite, on détermine la signification de la statistique de test par comparaison avec la distribution permutationnelle obtenue pour cette statistique en permutant aléatoirement l'ordre des éléments dans une des matrices. Cette approche opère simplement comme le test de permutation portant sur le coefficient de régression dans un modèle de régression simple présenté dans le chapitre 1. Nous avons seulement remplacer les deux variables quantitatives par deux matrices de distances.

On peut alors reprendre l'analyse présentée au chapitre 1, en adoptant comme statistique de test : la statistique de Student T , R le coefficient de corrélation ou encore la somme des produits $\sum a_{ij} b_{ij}$ puisque les tests de permutation utilisant ces trois indicateurs sont équivalents.

4.2.1 Simulation de Legendre

Afin de comparer dans le cas de matrices de distances, les quatre méthodes de test de permutation présentées dans le chapitre 2, Legendre ([35]) a réalisé une simulation. Sa démarche consiste en les étapes suivantes.

¹mais à partir de mesures différentes par exemple une matrice de distances écologiques basées sur les abondances d'espèces, et une matrice de distance géographiques entre les objets

1. L'auteur génère trois matrices de données $\mathbf{Z}^j, j = 1, 2, 3$ d'ordre $(n \times 10)$ à l'aide d'un vecteur aléatoire de dimension $p = 10$ qui suit une loi $LG_{10}(\mathbf{0}, \mathbf{I})$ pour une simulation et une loi très dissymétrique $Exp(1)^3$ pour l'autre :

$$\mathbf{Z}^j = \begin{bmatrix} z_{11}^j & \cdots & z_{110}^j \\ z_{21}^j & \cdots & z_{210}^j \\ \vdots & \ddots & \vdots \\ z_{n1}^j & \cdots & z_{n10}^j \end{bmatrix}$$

2. A partir de $\mathbf{Z}^1, \mathbf{Z}^2$ et \mathbf{Z}^3 il calcule les matrices carrées d'ordre n de terme général les distances entre paires d'unités statistiques. La matrice de distances associée à \mathbf{Z}^j est notée $\mathbf{D}^j, j = 1, 2, 3$:

$$\mathbf{D}^j = \begin{bmatrix} 0 & d_{12}^j & \cdots & d_{1n}^j \\ d_{21}^j & 0 & \cdots & d_{2n}^j \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1}^j & d_{n2}^j & \cdots & d_{nn}^j \end{bmatrix}$$

3. Il définit alors les vecteurs $\mathbf{d}^j, j = 1, 2, 3$ de dimension $N = n(n-1)/2$ respectivement formés par les éléments de la partie triangulaire inférieure (ou supérieure) des matrices de distances $\mathbf{D}^1, \mathbf{D}^2$ et \mathbf{D}^3 . Pour simplifier les notations, on note ici les composantes de \mathbf{d}^j par un seul indice $e = \{1, \dots, N\}$

$$\mathbf{d}^j = \begin{bmatrix} d_1^j \\ d_2^j \\ \vdots \\ d_N^j \end{bmatrix}$$

4. Il définit alors la matrice \mathbf{D} formée par concaténation des matrices de distances :

$$\mathbf{D} = [\mathbf{d}^1 | \mathbf{d}^2 | \mathbf{d}^3] = \begin{bmatrix} d_1^1 & d_1^2 & d_1^3 \\ d_2^1 & d_2^2 & d_2^3 \\ \vdots & \vdots & \vdots \\ d_N^1 & d_N^2 & d_N^3 \end{bmatrix} \quad (4.5)$$

A partir des $\mathbf{d}^1, \mathbf{d}^2$ et \mathbf{d}^3 , Legendre construit des vecteurs $\mathbf{X}^1, \mathbf{X}^2$ et \mathbf{Y} dont la structure de corrélation est contrôlée à l'aide d'une matrice de corrélation \mathbf{R} de la façon suivante :

$$\mathbf{D}_{\mathbf{R}} = [\mathbf{X}^1 | \mathbf{X}^2 | \mathbf{Y}] = \mathbf{D} \mathbf{R}^{1/2} \quad \text{où} \quad \mathbf{R} = \begin{bmatrix} 1 & \rho(X^1, X^2) & \rho(X^1, Y) \\ \rho(X^1, X^2) & 1 & \rho(X^2, Y) \\ \rho(X^1, Y) & \rho(X^2, Y) & 1 \end{bmatrix} \quad (4.6)$$

On est alors en état de réaliser à partir des trois vecteurs $\mathbf{X}^1, \mathbf{X}^2$ et \mathbf{Y} les tests de permutation présentés au chapitre 2. Notons que Legendre a utilisé le coefficient

de corrélation partielle comme statistique de test. Les principaux résultats de ses simulations sont les suivants :

1. La méthode de Kennedy augmente le risque d'erreur de première espèce.
2. En augmentant la corrélation $\rho(X^1, X^2)$, la méthode Manly augmente le risque d'erreur de première espèce.
3. Lorsque les valeurs de Z^1 , Z^2 et Z^3 sont générées suivant une loi fortement disymétrique, telle que la loi $Exp(1)^3$, la méthode de Ter Brack a la valeur maximale du risque d'erreur de première espèce.
4. Les risques d'erreur de première espèce encourus par ces méthodes dans les cas de test bilatéral et de test unilatéral sont différentes.
5. Dans la plupart des cas, la puissance de la méthode de Ter Brack est la plus forte.

Notre étude par simulation conduit à des observations différentes des résultats obtenus par Legendre. Nous devons donc admettre que dans le cas des matrices de distances il est nécessaire d'effectuer une étude plus complète des facteurs agissant sur la puissance des tests de permutation appliqués à la dépendance (partielle).

4.3 Notre simulation

Tout d'abord, considérons le cadre de la régression linéaire double :

$$Y_e = \beta_0 + \beta_1 x_e^1 + \beta_2 x_e^2 + \epsilon_e \quad (e = 0, 1, \dots, N), \quad (4.7)$$

où Y, x^1, x^2 et ϵ sont désormais les éléments de la partie triangulaire inférieure de quatre matrices des distances, $N = n(n-1)/2$ et e désigne la paire d'unités statistique (i, j) avec, $i < j$, $i, j = 1, \dots, n$. On effectue donc ici une régression linéaire sur données dyadiques. Il reste à montrer que les composantes de ces variables vérifient les propriétés des matrices de distances. L'objectif est d'effectuer le test de l'hypothèse $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$.

Nous étudions en particulier l'influence sur le risque d'erreur de première espèce et sur la puissance du test des quatre facteurs suivants :

- a) le nombre d'unités statistiques n ;
- b) le degré de corrélation entre les variables explicatives $\rho(X^1, X^2)$;
- c) la valeur du paramètre β_1 associé à la covariable X^1 ;
- d) la distribution du terme d'erreur aléatoire ϵ_e .

4.3.1 Analyse du risque d'erreur de première espèce du test

A) Le cadre de la simulation

Nous avons étudié le risque d'erreur de première espèce pour les spécifications suivantes :

- a) le nombre d'unités statistiques : n appartient à $\{6, 9, 12\}$;
- b) le coefficient associé à la covariable X^1 : β_1 appartient à l'ensemble $\{1, 2.5\}$
- c) le coefficient de corrélation entre les variables explicatives X^1 et X^2 : $\rho(X^1, X^2)$ est à valeurs dans $\{0.1, 0.9\}$;
- d) la distribution du terme d'erreur aléatoire ϵ : $\{Wishart, Exp(2), Beta(2, 0.3)\}$.

B) Algorithme de simulation

1. Nous avons généré deux matrices $\mathbf{Z}^j, j = 1, 2$ d'ordre $(n \times 10)$ à l'aide d'un vecteur aléatoire de dimension $p = 10$ qui suit une loi $LG_{10}(\mathbf{0}, \mathbf{I})$:

$$\mathbf{Z}^j = \begin{bmatrix} z_{11}^j & z_{12}^j & z_{13}^j & \cdots & z_{110}^j \\ z_{21}^j & z_{22}^j & z_{23}^j & \cdots & z_{210}^j \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1}^j & z_{n2}^j & z_{n3}^j & \cdots & z_{n10}^j \end{bmatrix}$$

2. A partir de \mathbf{Z}^1 et \mathbf{Z}^2 , nous avons calculé les matrices carrées d'ordre n de terme général la distance euclidienne entre paires d'unités. On note $\mathbf{D}^j, j = 1, 2$ ces deux matrices :

$$\mathbf{D}^j = D(\mathbf{Z}^j) = \begin{bmatrix} 0 & d_{12}^j & d_{13}^j & \cdots & d_{1n}^j \\ d_{21}^j & 0 & d_{23}^j & \cdots & d_{2n}^j \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{n1}^j & d_{n2}^j & d_{n3}^j & \cdots & d_{nn}^j \end{bmatrix}$$

3. Nous avons défini alors les vecteurs \mathbf{d}^1 et \mathbf{d}^2 de taille $N = n(n-1)/2$ respectivement formés par les éléments de la partie triangulaire inférieure (ou supérieure) des matrices de distances \mathbf{D}^1 et \mathbf{D}^2 . Pour simplifier les notations, on repère les composantes de $\mathbf{d}^j, j = 1, 2$ à l'aide d'un seul indice $e = 1, \dots, N$.

$$\mathbf{d}^j = \begin{bmatrix} d_1^j \\ d_2^j \\ \vdots \\ d_N^j \end{bmatrix}$$

Contrairement au chapitre 2 où nous avons étudié les tests de permutation dans le cas des variables ordinaires, ici les composantes du vecteur \mathbf{d}^1 ou \mathbf{d}^2 ne sont pas indépendantes, puisqu'elles sont des éléments d'une matrice de distances déduite d'une configuration aléatoire de n points de \mathbb{R}^p et qu'elles vérifient donc des contraintes liées à la définition d'une distance, en particulier le postulat d'inégalité triangulaire.

4. Nous avons voulu étudier l'effet d'une transformation des distances sur le seuil et sur la puissance du test. Afin de limiter les calculs, nous avons limité notre comparaison à une transformation logarithmique et une transformation quadratique. A partir des vecteurs \mathbf{d}^1 et \mathbf{d}^2 nous avons donc calculé :

- a) les vecteurs des carrés des distances, notés \mathbf{d}_c^1 et \mathbf{d}_c^2 respectivement ;
- b) le vecteur des logarithmes des distances notés respectivement \mathbf{d}_l^1 et \mathbf{d}_l^2 .

5. Nous sommes maintenant en mesure d'explicitier le modèle retenu pour simuler des données. On définit tout d'abord la matrice \mathbf{D} :

$$\mathbf{D} = [\mathbf{d}^1 | \mathbf{d}^2] = \begin{bmatrix} d_1^1 & d_1^2 \\ d_2^1 & d_2^2 \\ \vdots & \vdots \\ d_N^1 & d_N^2 \end{bmatrix} \quad (4.8)$$

Puisque les deux matrices de distances \mathbf{D}^1 et \mathbf{D}^2 sont générées de manière indépendante, on a $\rho(\mathbf{d}^1, \mathbf{d}^2) = 0$. A partir de \mathbf{D} , nous avons construit les vecteurs \mathbf{X}^1 et \mathbf{X}^2 de corrélation $\rho(X^1, X^2)$ de la façon suivante :

$$\mathbf{D}_{\mathbf{R}} = [\mathbf{X}^1 | \mathbf{X}^2] = \mathbf{D}\mathbf{R}^{1/2} \quad \text{où} \quad \mathbf{R} = \begin{bmatrix} 1 & \rho(X^1, X^2) \\ \rho(X^1, X^2) & 1 \end{bmatrix} \quad (4.9)$$

où $\rho(X^1, X^2)$ est fixé dans chaque simulation. Ensuite, nous calculons le terme signal comme : $\mathbf{D}_{\mathbf{R}}\boldsymbol{\beta}$ où $\boldsymbol{\beta} = \{\beta_1, \beta_2\}$. Remarquons que pour calculer le risque d'erreur de première espèce, on doit poser $\beta_2 = 0$.

6. Pour simuler le terme d'erreur, nous avons généré un N -échantillon aléatoire $\boldsymbol{\epsilon} = {}^t(\epsilon_1, \epsilon_2, \dots, \epsilon_N)$ suivant l'une des trois lois suivantes $\{Wishart, Exp(2), Beta(2, 0.3)\}$. On distingue ainsi le cas d'une loi symétrique (loi de Wishart) et le cas de lois dissymétriques à dissymétrie droite ($Beta(2, 0.3)$) et gauche ($Exp(2)$).

La procédure de génération de nombres aléatoires nécessite ici une étape de rejet de façon à s'assurer que les données générées vérifient bien les propriétés des matrices de distances. Aussi, si les données simulées ne vérifient pas les dites propriétés, on répète cette étape jusqu'à ce qu'elles y satisfassent.

7. Nous avons de plus observé, bien que de façon grossière, l'effet des variances du terme signal ($\mathbf{D}_{\mathbf{R}}\boldsymbol{\beta}$), et du terme d'erreur ($\boldsymbol{\epsilon}$). Nous avons donc imposé une condition au rapport signal sur bruit :

$$\text{Var}(\mathbf{D}_{\mathbf{R}}\boldsymbol{\beta}) / \text{Var}(\boldsymbol{\epsilon}) \geq 4 \quad (4.10)$$

Si les données générées pour ϵ ne vérifiaient pas cette dernière condition, on divisait les valeurs du terme ϵ par 2 pour obtenir la relation (4.10). Cette simulation a été comparée à la situation opposée où le rapport vérifie :

$$\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) \leq 1/4 \quad (4.11)$$

mais cette dernière situation est étudiée dans le seul cas où les valeurs du terme d'erreur sont générées selon une loi Wishart.

8. Enfin, les valeurs de Y ont été générées par l'équation :

$$\mathbf{Y} = \mathbf{D}_R\boldsymbol{\beta} + \epsilon \quad (4.12)$$

les données ainsi générées devraient elles aussi vérifier les propriétés des matrices de distances. Aussi, si ces données ne vérifiaient pas les dites propriétés, on répétait les étapes 6 et 7 jusqu'à ce qu'elles y satisfassent.

9. On a répété les étapes 5-8 pour les deux autres ensembles de données générées en étape 4-a et 4-b, c'est-à-dire pour le vecteur des carrés des distances et le vecteur des logarithmes des distances. Au bilan, nous avons ainsi généré trois ensembles de données.

10. Afin de comparer des statistiques de test pour chacune des méthodes nous avons réalisé les tests de permutation avec trois statistiques : le T de Student, le coefficient de corrélation partielle noté par $r_{Y2.1\pi}$ et le coefficient de corrélation partielle au carré, noté $r_{Y2.1\pi}^2$. Ainsi, chaque méthode de test de permutation a conduit au calcul des ces trois statistiques de test sur les données permutées obtenues selon le principe caractérisant chaque méthode.

Pour chaque combinaison de niveaux des facteurs, 3000 ensembles de données ont été générés. Puis, pour chacune des quatre méthodes de test, chaque ensemble de données a été permuté 999 fois. Pour chaque méthode, les statistiques de tests de permutation $t_{2,\pi}$, $r_{Y2.1\pi}$ et $r_{Y2.1\pi}^2$ ont été calculées pour chaque ensemble de données permutées.

11. Pour comparer les tests d'hypothèses dans le cas bilatéral et le cas unilatéral, et cela à l'aide de chacune des méthodes, nous avons calculé les p-valeurs bilatérales et unilatérales correspondant aux trois statistiques de test. Cela signifie qu'après 999 permutations (pour chaque ensemble de données), la p-valeur permutationnelle a été calculée pour les trois statistiques de test et par chacune des quatre méthodes. Par exemple dans le cas d'une hypothèse bilatérale et pour la statistique de test t , la p-valeur permutationnelle est calculée par la formule :

$$p_t^B - \text{valeur}_\pi = \frac{\#(|t_\pi| \geq |t_{obs}|)}{1000} \quad (4.13)$$

et pour l'hypothèse unilatérale $H_0 : \beta_2 \leq 0$ contre $H_1 : \beta_2 > 0$ la p-valeur permutationnelle s'obtient par :

$$p_t^U - \text{valeur}_\pi = \frac{\#(t_\pi \geq t_{obs})}{1000} \quad (4.14)$$

Dans le cas du coefficient de corrélation partielle $r_{Y2.1}$, les p-valeurs respectivement bilatérale et unilatérale sont :

$$p_r^B - \text{valeur}_\pi = \frac{\#(|r_{Y2.1,\pi}| \geq |r_{Y2.1(obs)}|)}{1000} \quad (4.15)$$

$$p_r^U - \text{valeur}_\pi = \frac{\#(r_{Y2.1,\pi} \geq r_{Y2.1(obs)})}{1000} \quad (4.16)$$

de plus pour la statistique de test $r_{Y2.1}^2$:

$$p_{r^2}^B - \text{valeur}_\pi = \frac{\#(r_{Y2.1,\pi}^2 \geq r_{Y2.1(obs)}^2)}{1000} \quad (4.17)$$

$$p_{r^2}^U - \text{valeur}_\pi = \frac{\#(r_{Y2.1,\pi}^2 \geq |r_{Y2.1(obs)}|^2)}{1000} \quad (4.18)$$

où t_{obs} et $r_{Y2.1(obs)}$ sont les statistiques t et $r_{Y2.1}$ calculées sur les données originales (non permutées).

Le niveau de signification a été fixé à $\alpha = 0,05$ pour ces tests . Ainsi, pour chaque ensemble de données on rejette l'hypothèse $H_0 : \beta_2 = 0$ quand la $p - \text{valeur}_\pi$ est inférieure à 0.05. Sur les 3000 $p - \text{valeur}_\pi$ calculées le pourcentage de $P - \text{valeur}_\pi$ (bilatérales ou unilatérales) inférieures à 0.05 donne le risque d'erreur de première espèce atteint. En fait, ce pourcentage estime la probabilité de rejeter $H_0 : \beta_2 = 0$ lorsque H_0 est vraie (parce que les données sont générées sous $H_0 : \beta_2 = 0$). De plus, un intervalle de confiance à 95% a été calculé à partir du risque d'erreur de première espèce (calculé sur 3000 ensembles).

Enfin, dans chacune des méthodes, on a calculé le risque d'erreur de première espèce pour les trois statistiques t , $r_{Y2.1}$ et dans le cas des deux hypothèses : bilatérale et unilatérale.

Afin de compléter la comparaison pour chaque ensemble de données on a de plus utilisé les tests paramétriques classiques qui consistent à calculer une statistique de Student et sa p-valeur donnée par la théorie Gaussienne. De la même façon, sur les 3000 ensembles de données le risque d'erreur de première espèce et son intervalle de confiance sont calculés à l'aide des formules fournies par la théorie Gaussienne.

C) Les résultats des simulations

Dans cette section nous présentons les résultats relatifs aux deux situations : les termes d'erreur suivent (**C1**) une loi de Wishart et (**C2**) une loi $Beta(2,0.3)$.

Lorsque les termes d'erreur suivent une loi $Exp(2)$, les résultats des simulations sont les mêmes que le cas de données générées selon une loi $Beta(2, 0.3)$ et c'est pourquoi nous les avons consignés dans l'annexe 4.

C1) Cas des données générées selon une loi de Wishart :

Cette partie présente quatre comparaisons : la première concerne les transformations des distances simulées, la deuxième étudie l'influence du choix d'une statistique de test, la troisième oppose les alternatives bilatérales et unilatérales et la dernière comparaison contraste deux situations marquées par un rapport signal sur bruit fort ou faible.

C1-1 Comparaison des trois ensembles des données : distance interpoint, carré de distance et transformée logarithmique de distances

L'objet de cette partie est de comparer par des résultats de simulations le risque d'erreur de première espèce de la statistique de tests T de Student sur l'hypothèse alternative bilatérale, pour trois ensembles de données : distance interpoint, carré des distances et transformée logarithmique des distances. Les valeurs des variables Y , X^1 , et X^2 sont générées telles que $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$.

En considérant les figures 4.1 et 4.2 on peut faire les observations suivantes :

1. Quelles que soient les valeurs des facteurs et pour les trois ensembles de données, en augmentant le nombre d'unités (les valeurs des autres facteurs restant fixées), le risque d'erreur de première espèce augmente. Ceci est peut-être dû au fait que les valeurs des variables X^1 ou X^2 sont dépendantes.
2. Pour toutes les combinaisons des facteurs et pour les trois ensembles de données, l'augmentation du coefficient de corrélation de 0.1 à 0.9 (les valeurs des autres facteurs restant fixées) n'a pas d'effet sur le risque d'erreur de première espèce.
3. Les figures 4.1 et 4.2 ne montrent aucune différence majeure. En résumé, la valeur vraie du coefficient β_1 ne semble pas influencer sur les risques d'erreur de première espèce.
4. Les quatre méthodes de test permutation ainsi que la méthode classique, génèrent des niveaux d'erreur de première espèce très semblables quelles que soient les valeurs de n , β_1 et ρ .
5. On constate que lorsque les résidus sont générés selon la loi de Wishart, le risque d'erreur de première espèce est plus fort que dans le cas $Exp(2)$ et $Beta(2, 0.3)$. La symétrie de la loi des résidus semble donc dégrader la situation pour toutes les combinaisons des facteurs, par les quatre méthodes de test de permutation ainsi que la méthode classique.
6. Quelle que soit la combinaison des facteurs choisie, les transformations quadratique et logarithmique n'ont pas d'effet sur le risque d'erreur de première

espèce. C'est dire que les résultats obtenus par ces deux ensembles de données diffèrent très peu de ceux qu'on a établis sur l'ensemble initial des données de distance.

7. Nous avons présenté ici la statistique T de Student. Nous avons par ailleurs obtenu des résultats analogues pour les statistiques de test concurrentes : le coefficient de corrélation partielle ($r_{Y2.1}$) et bien évidemment aussi pour le carré du coefficient de corrélation partielle ($r_{Y2.1}^2$) (cf. LABSAD en ligne²).
8. Les résultats présentés dans les figures 4.1 et 4.2 correspondent aux hypothèses alternatives bilatérales. Nous avons obtenu des résultats équivalents pour les hypothèses alternatives unilatérales (cf. LABSAD en ligne).
9. Les résultats donnés dans ces figures correspondent au cas où la variance du terme de signal est quatre fois plus grande que celle du terme d'erreur $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$. Des résultats équivalents sont obtenus lorsque la variance du terme d'erreur est quatre fois plus grande que celle du terme signal $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) < 4$ (cf. LABSAD).

²tous les résultats de ces simulations seront accessibles et consultables en ligne sur le site internet du LABSAD

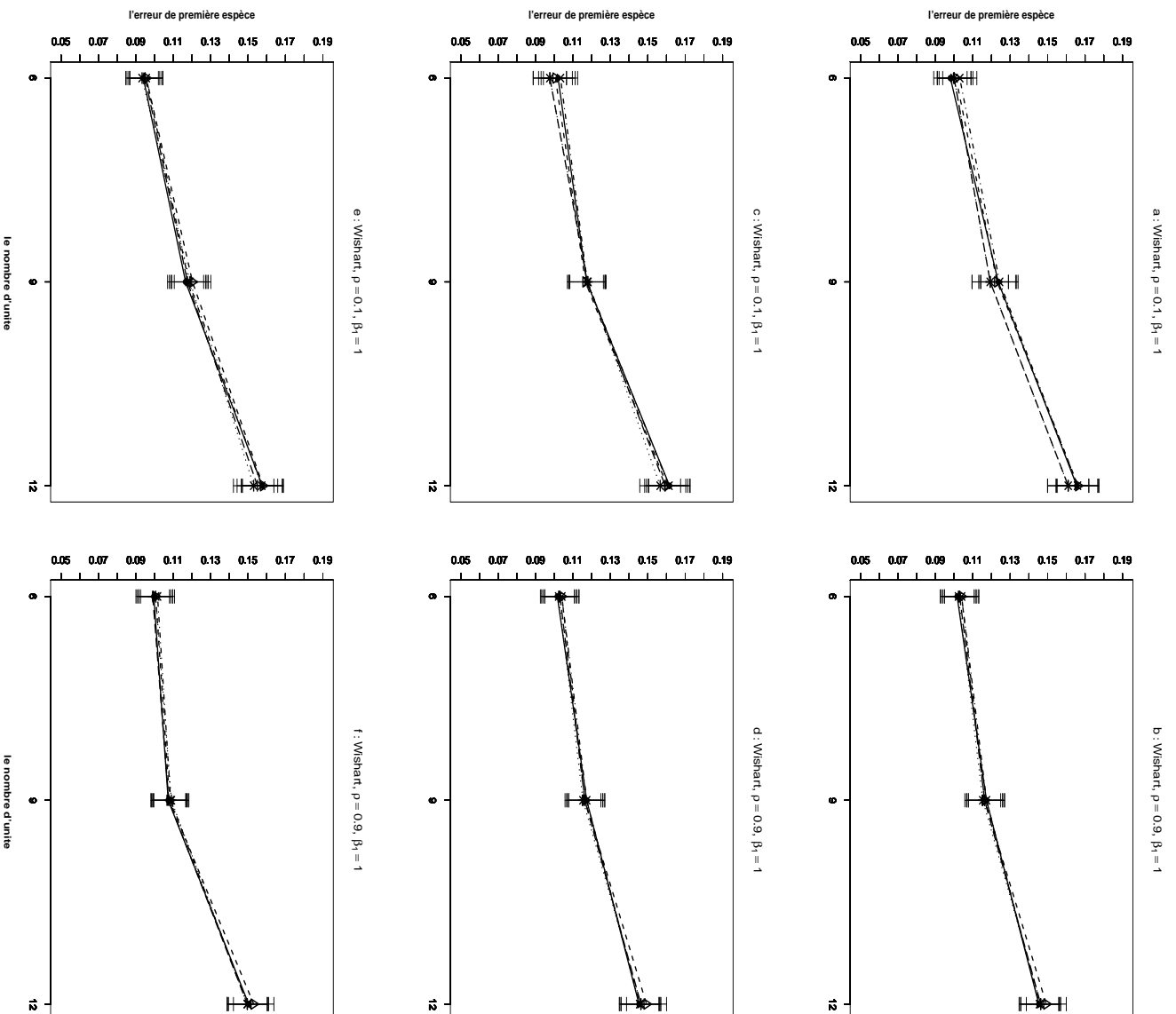


FIG. 4.1 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Brack— $+$ — et Manly— $*$ —... 30000 n-échantillons du résidu e ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. La statistique de test est le T de Student pour toutes les figures. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures a, c et e . $\rho = 0.9$ pour les figures b, d et f . X^1 et X^2 sont des distances interpoints pour les figures a et b ; des carrés des distances pour les figures c et d ; des transformées logarithmiques des distances pour les figures e et f .

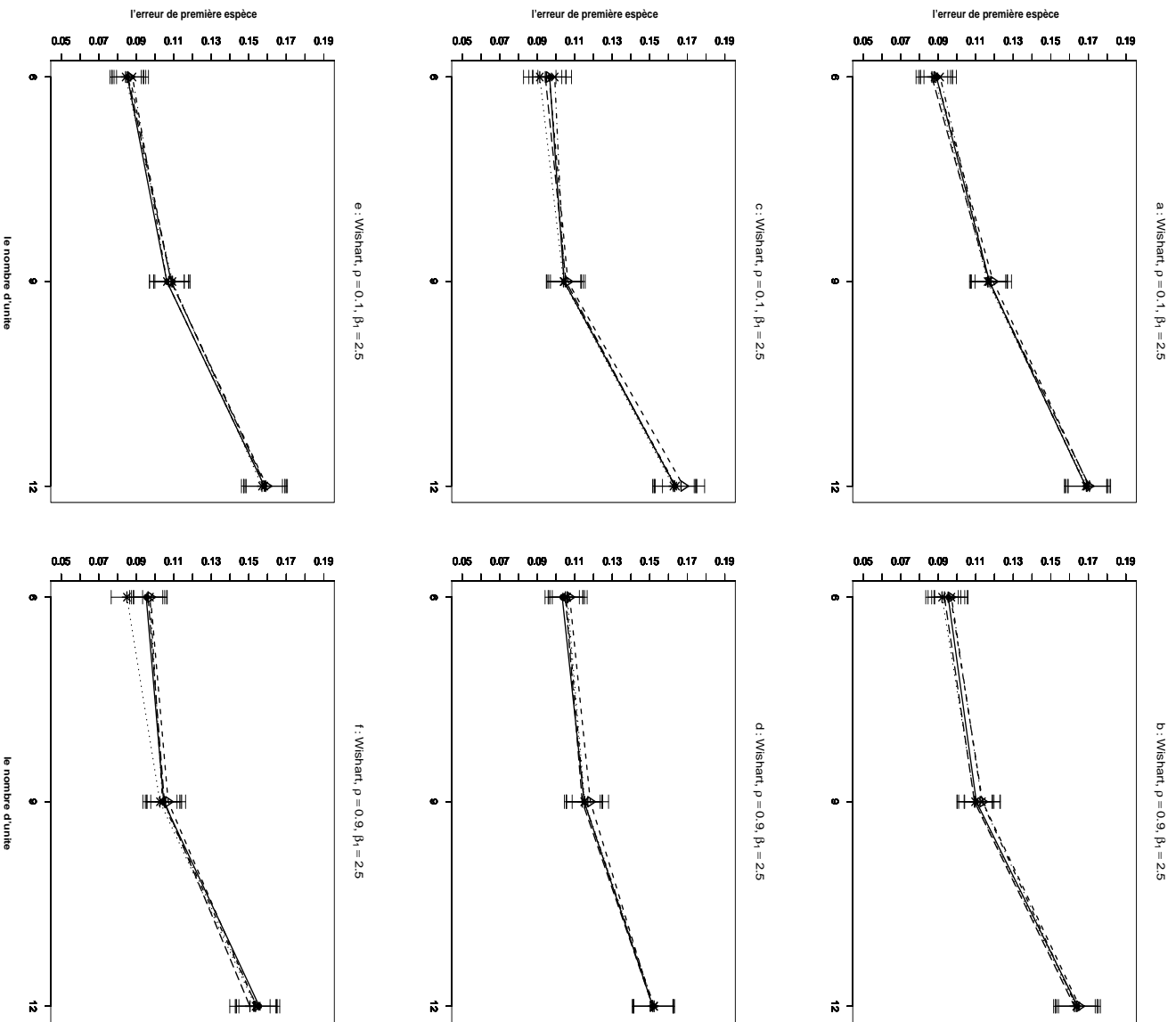


FIG. 4.2 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— \bullet —; Freedman et Lane— Δ —; Kennedy— \times —; Ter Brack— $+$ — et Manly— $*$ —, 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. La statistique de test est le T de Student pour toutes les figures. $\beta_1 = 2.5$ pour toutes les figures. $\rho = 0.1$ pour les figures a, c et e. $\rho = 0.9$ pour les figures b, d et f. X^1 et X^2 sont des distances interpoints pour les figures a et b; des carrés des distances pour les figures c et d; des transformées logarithmiques des distances pour les figures e et f.

C1-2 Comparaison des trois statistiques de tests

L'objet de cette partie est la comparaison des résultats des simulations pour le risque d'erreur de première espèce concernant les statistiques de tests T de Student, $r_{Y_{2,1}}$ et $r_{Y_{2,1}}^2$, dans le cas de l'hypothèse alternative bilatérale. Ici, X^1 et X^2 sont des distances interpoints telles que $\text{Var}(\mathbf{D}_{\mathbf{R}}\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$.

Des figures 4.3 et 4.4, on peut tirer les conclusions suivantes :

1. Les mêmes quatre premiers résultats obtenus sont identiques à ceux qui ont été présentés dans la partie (C1-1).
2. On constate que quelle que soit la combinaison des facteurs choisie, il n'y a pas de différences notables entre les résultats obtenus sur les trois statistiques de test : le T de Student, $r_{Y_{2,1}}$ et $r_{Y_{2,1}}^2$.
3. D'après ces simulations, dans le cas permutational, la différence entre les deux statistiques $r_{Y_{2,1}}$ et $r_{Y_{2,1}}^2$ est rigoureusement nulle.
4. Nos résultats ont été obtenus lorsque les variables X_1 et X_2 étaient les distances interpoints. Il est intéressant de signaler que la comparaison des statistiques de tests sur les données associées aux variables de carré de distance ou de transformée logarithmique de distances, indique des résultats identiques.
5. Les résultats présentés sur les figures 4.1 et 4.2 correspondant au hypothèses alternatives bilatérales. Nous avons obtenu des résultats équivalents pour les hypothèses alternatives unilatérales. (cf. LABSAD en ligne)
6. Les résultats donnés dans ces figures correspondent au cas où la variance du terme de signal était quatre fois plus grande que celle du terme d'erreur $\text{Var}(\mathbf{D}_{\mathbf{R}}\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$. Toute fois des résultats équivalents sont obtenus lorsque la variance du terme d'erreur est quatre fois plus grande que celle du terme signal $\text{Var}(\mathbf{D}_{\mathbf{R}}\boldsymbol{\beta})/\text{Var}(\epsilon) < 4$ (cf. LABSAD en ligne).

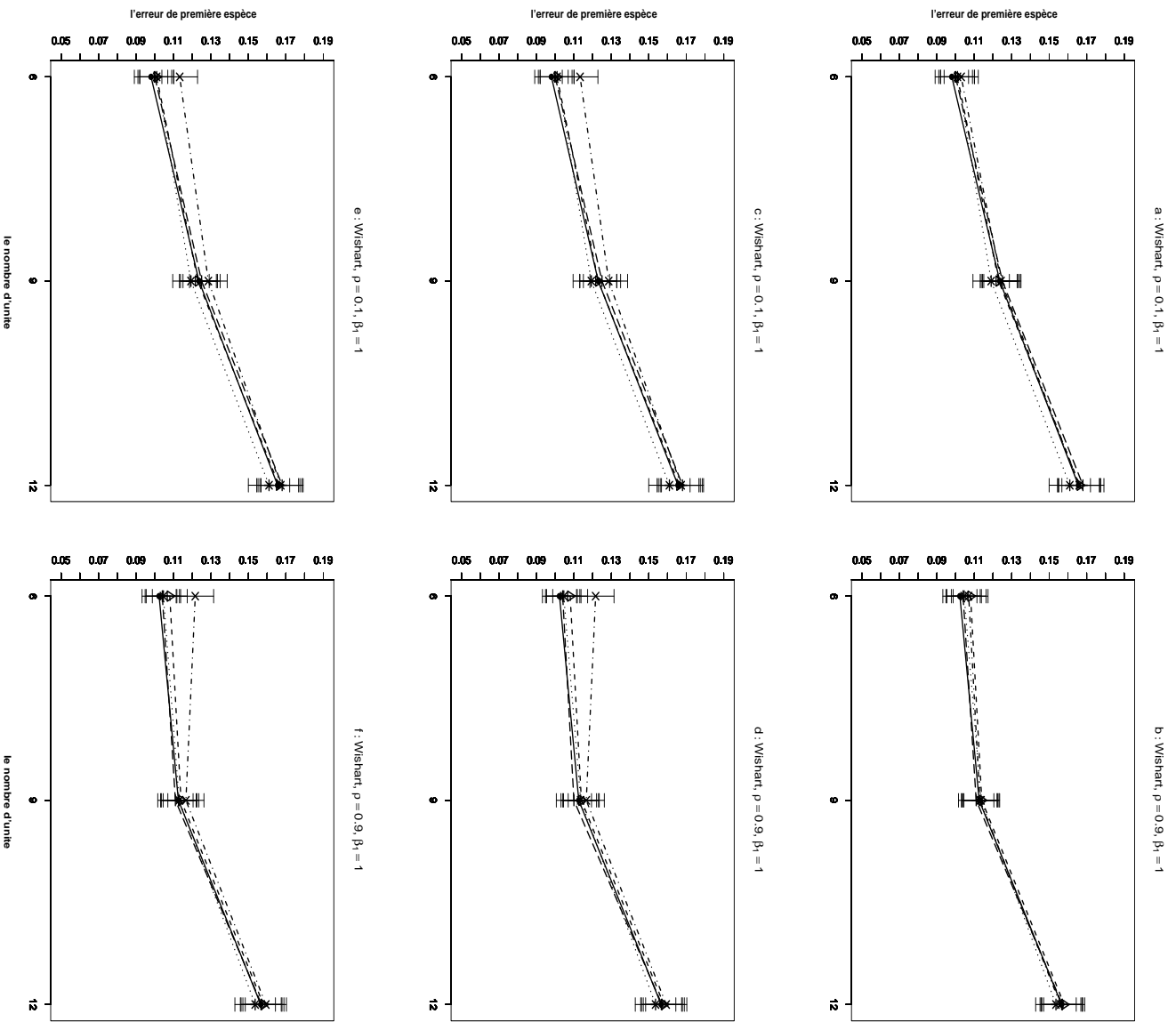


FIG. 4.3 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Braek— $+$ — et Manly— $*$ —... 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures a, c et e . $\rho = 0.9$ pour les figures b, d et f . X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est : le T de Student pour les figures a et b ; le coefficient de corrélation partielle pour les figures c et d ; carrée du coefficient de corrélation partielle pour les figures e et f .

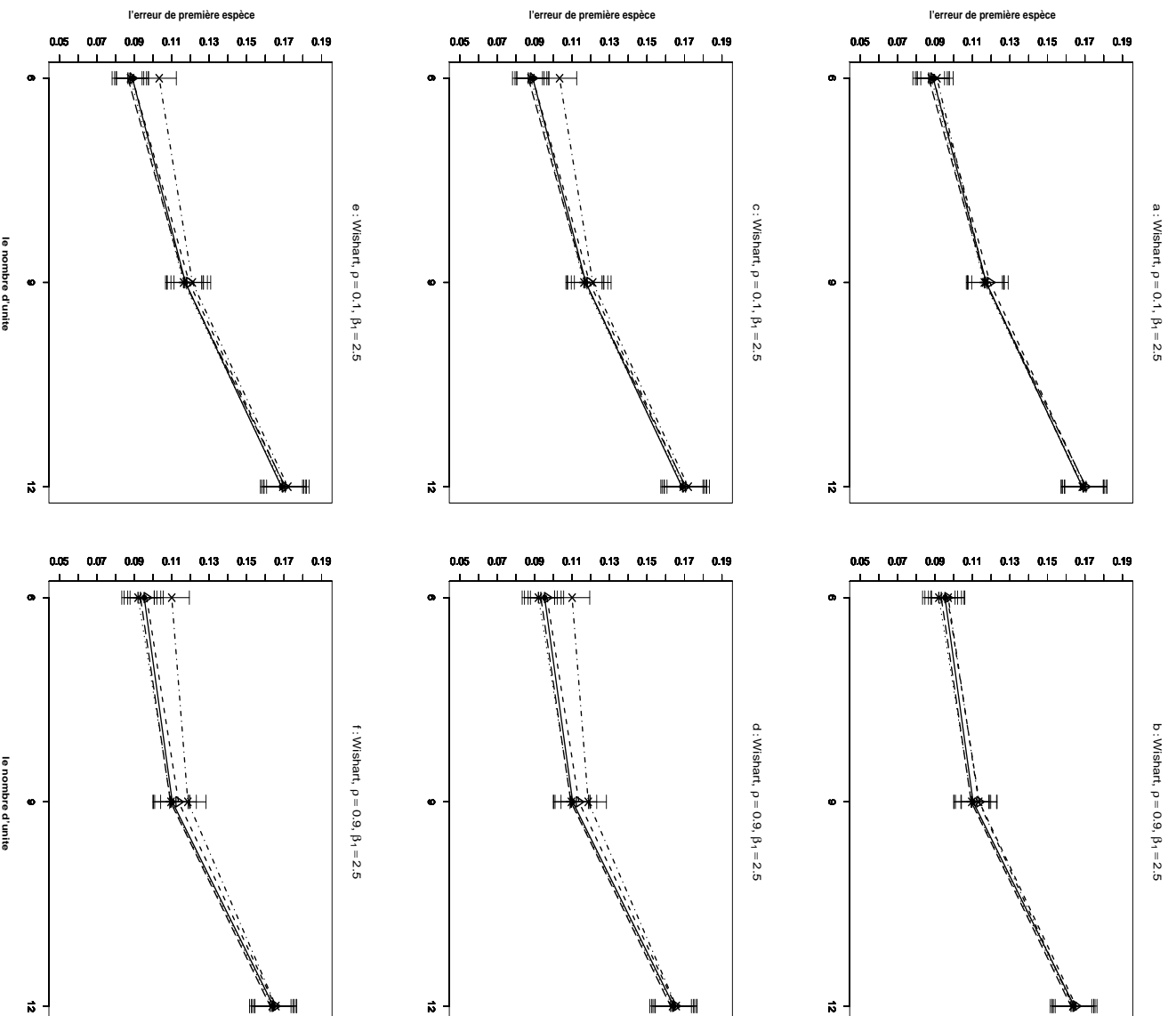


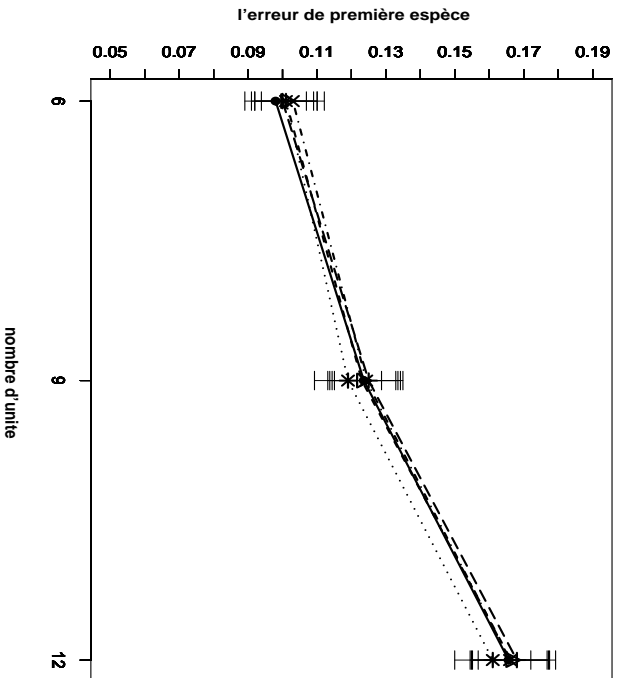
FIG. 4.4 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Brack— $+$ — et Manly $\dots*$. 30000 n-échantillons du résidu e ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures a , c et e . $\rho = 0.9$ pour les figures b , d et f . X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est : le T de Student pour les figures a et b ; le coefficient de corrélation partielle pour les figures c et d ; carrée du coefficient de corrélation partielle pour les figures e et f .

C1-3 Comparaison des hypothèses alternatives bilatérales et unilatérales

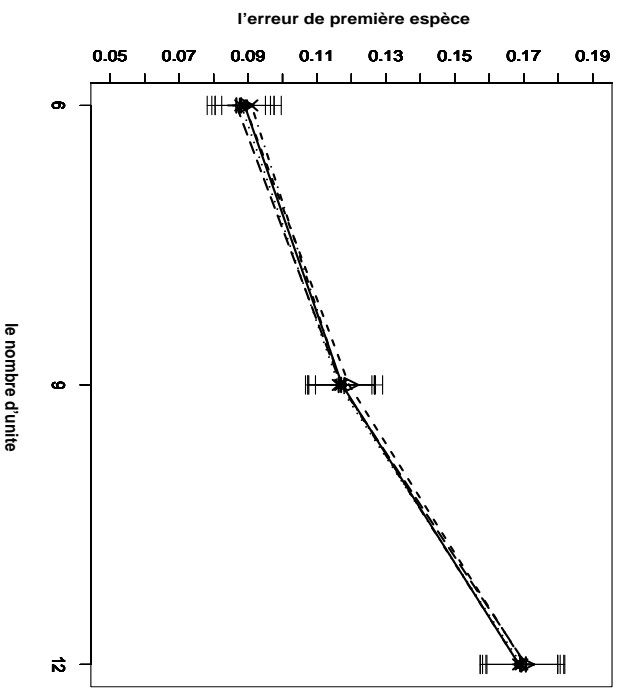
Dans cette partie nous voulons comparer par des résultats de simulations le risque d'erreur de première espèce de la statistique de test T de Student, sur les deux cas : l'hypothèse alternative bilatérale et hypothèse alternative unilatérale où X^1 et X^2 sont des distances interpoints telles que $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$. En considérant les figures 4.5 et 4.6 on peut faire les observations suivantes :

1. En comparant les figures 4.5-a,b (respectivement 4.6-a,b) avec les figures 4.5-c,d (respectivement 4.6-c,d) pour toutes les combinaisons de facteurs, le risque d'erreur de première espèce dans le cas unilatéral est un peu amoindri par rapport au cas bilatéral.
2. Les résultats ont été obtenus lorsque les variables X_1 et X_2 sont définies comme des distances interpoints. Nous avons obtenu des résultats très semblables pour les carrés des distances et pour les transformées logarithmiques de ces distances. (cf. LABSAD en ligne)
3. Ces résultats correspondent au cas de la statistique de test T de Student, mais nous avons obtenu des résultats très voisins pour les deux autres statistiques de test : coefficient de corrélation partielle ($r_{Y2.1}$) et bien évidemment aussi le carré du coefficient de corrélation partielle ($r_{Y2.1}^2$) (cf. LABSAD).
4. Ces résultats correspondent au cas où nous avons simulé des données de sorte que la variance du terme de signal soit quatre fois plus grande que celle du terme d'erreur $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$. Nous avons confronté cette situation à un contexte très différent à priori : celui d'une variance du terme d'erreur quatre fois plus grande que celle du terme signal $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) < 4$. (cf. LABSAD en ligne)

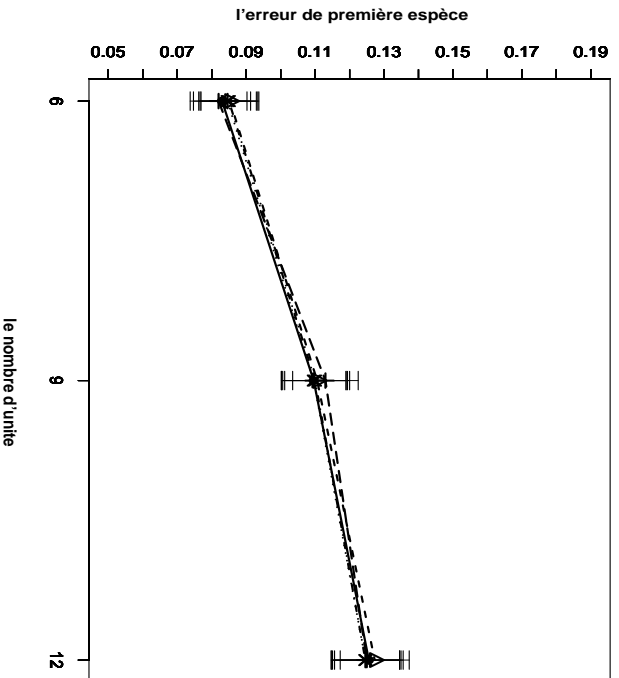
a : Wishart, $\rho = 0.1$, $\beta_1 = 1$



b : Wishart, $\rho = 0.1$, $\beta_1 = 2.5$



c : Wishart, $\rho = 0.1$, $\beta_1 = 1$



d : Wishart, $\rho = 0.1$, $\beta_1 = 2.5$

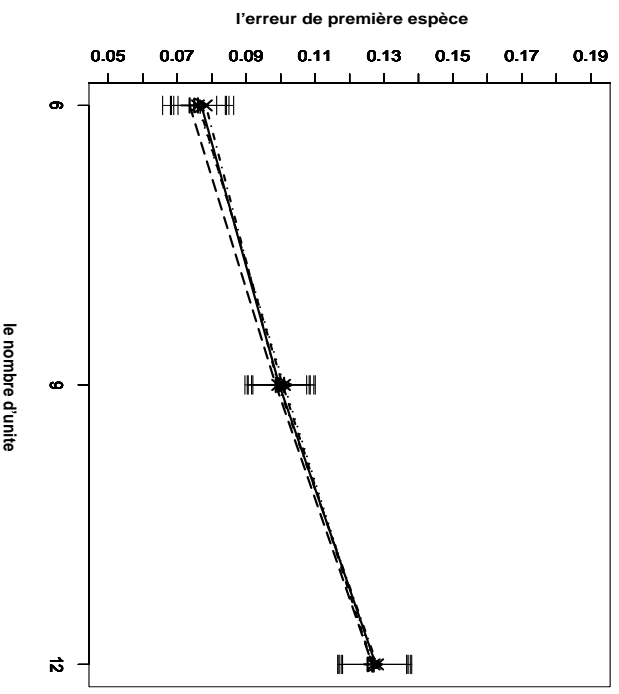


FIG. 4.5 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités pour les méthodes Gaussienne $-\triangle--$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot \times \cdot-$; Ter Braek $-\cdot + -$ et Manly $\dots * \dots$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est le T de Student pour toutes figures. $\rho_1 = 0.1$ pour toutes les figures. $\beta_1 = 1$ pour les figures a et c. $\beta_1 = 2.5$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

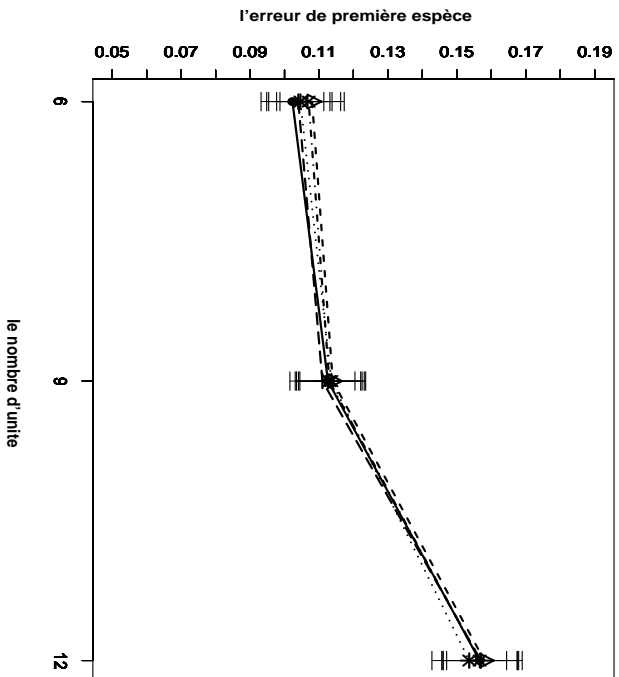
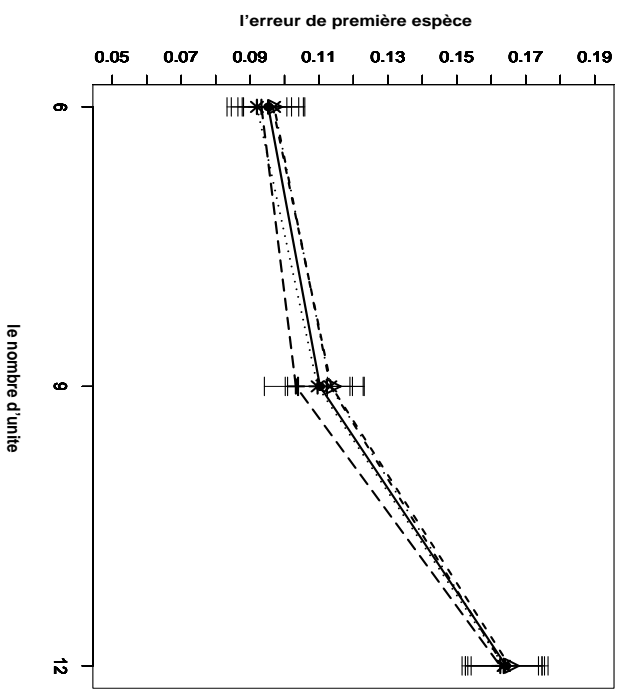
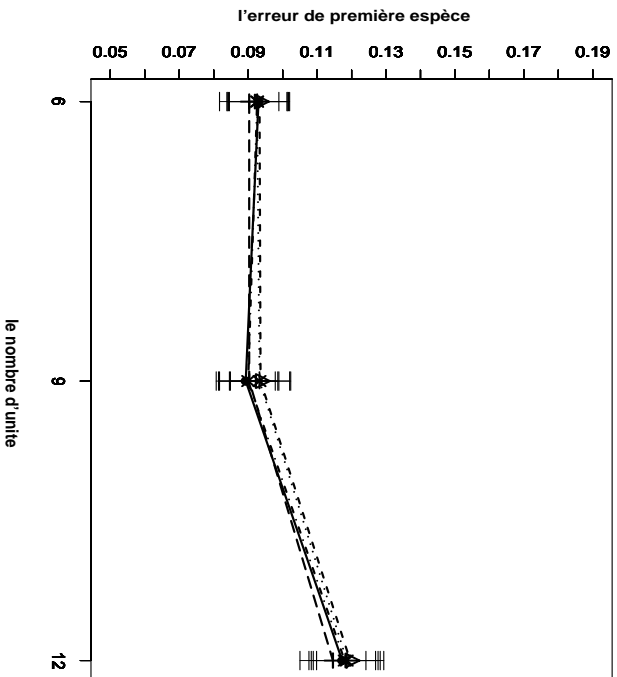
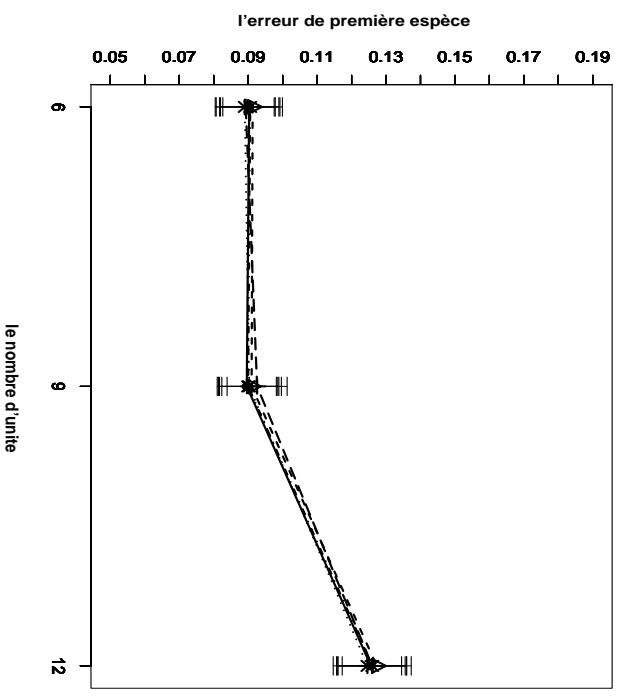
a : Wishart, $\rho = 0.9$, $\beta_1 = 1$ b : Wishart, $\rho = 0.9$, $\beta_1 = 2.5$ c : Wishart, $\rho = 0.9$, $\beta_1 = 1$ d : Wishart, $\rho = 0.9$, $\beta_1 = 2.5$ 

FIG. 4.6 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités pour les méthodes Gaussienne— \triangle —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Braek— $+$ — et Manly— \dots — \ast — \dots . 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est le T de Student pour toutes figures. $\rho_1 = 0.9$ pour toutes les figures. $\beta_1 = 1$ pour les figures a et c. $\beta_1 = 2.5$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

C1-4 Comparaison des données générés pour le terme d'erreur où $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$ et $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) < 4$

L'objet de cette partie est de comparer par des résultats de simulations le risque d'erreur de première espèce de la statistique de tests T de Student, sur l'hypothèse alternative bilatérale dans deux cas distincts : les données générées pour les variables X^1 , X^2 et Y sont des distances interpoints telles que $\text{Var}(\beta_1 X^1 + \beta_2 X^2)/\text{Var}(\epsilon) \geq 4$ puis $\text{Var}(\beta_1 X^1 + \beta_2 X^2)/\text{Var}(\epsilon) < 4$.

En considérant les figures 4.7 et 4.8 on peut tirer les conclusion suivantes :

1. Quelles que soient les valeurs des facteurs et pour les trois ensembles de données, en augmentant le nombre d'unités (lorsque les valeurs des autres facteurs restent fixés), le risque d'erreur de première espèce augmente.
2. Pour toutes les combinaisons des valeurs de n , β_1 et ρ , nous observons que lorsque $\text{Var}(\beta_1 X^1 + \beta_2 X^2)/\text{Var}(\epsilon) < 4$ ou au contraire $\text{Var}(\beta_1 X^1 + \beta_2 X^2)/\text{Var}(\epsilon) \geq 4$, les risques d'erreur de première espèce obtenus par les quatre méthodes de tests de permutation sont plus grands que le niveau considéré $\alpha = 0.05$.
3. Pour toutes les combinaisons des facteurs et pour les trois ensembles de données, l'augmentation du coefficient de corrélation de 0.1 à 0.9 (lorsque les valeurs des autres facteurs restent fixes) n'a pas d'effet sur le risque d'erreur de première espèce.
4. Les figures 4.7 et 4.8 ne montrent aucune différence majeure. En résumé, la valeur vraie du coefficient β_1 ne semble pas influencer sur les risques d'erreur de première espèce.
5. Les quatre méthodes de test permutation ainsi que la méthode classique, génèrent des niveaux d'erreur de première espèce très semblables, quelles que soient les valeurs de n , β_1 et ρ .
6. Ces résultats correspondent au cas de la statistiques de test T de Student, mais nous avons obtenu des résultats très voisins pour les deux statistiques de test : coefficient de corrélation partielle ($r_{Y2.1}$) et bien évidemment aussi le carré du coefficient de corrélation partielle ($r_{Y2.1}^2$) (cf. LABSAD).
7. Les résultats présentés dans les figure 4.7 et 4.8 correspondent aux hypothèses alternatives bilatérales. Nous avons obtenu des résultats équivalents pour les hypothèses alternatives unilatérales. (c.f LABSAD en ligne)

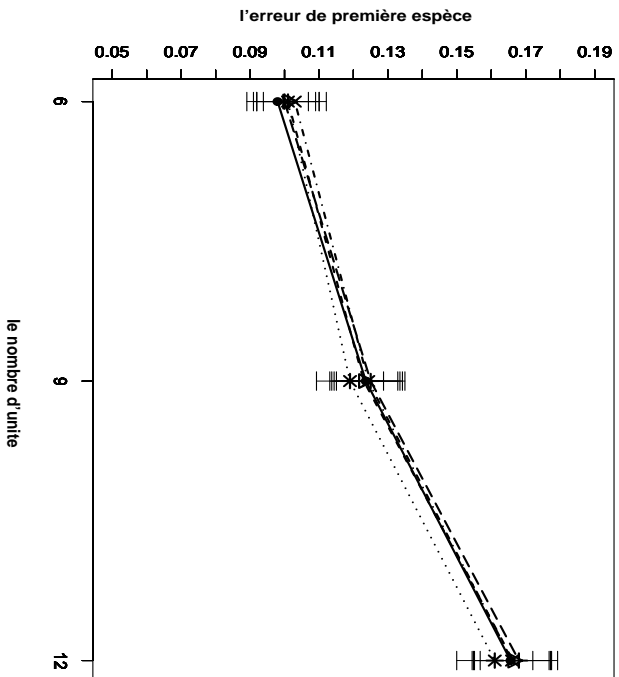
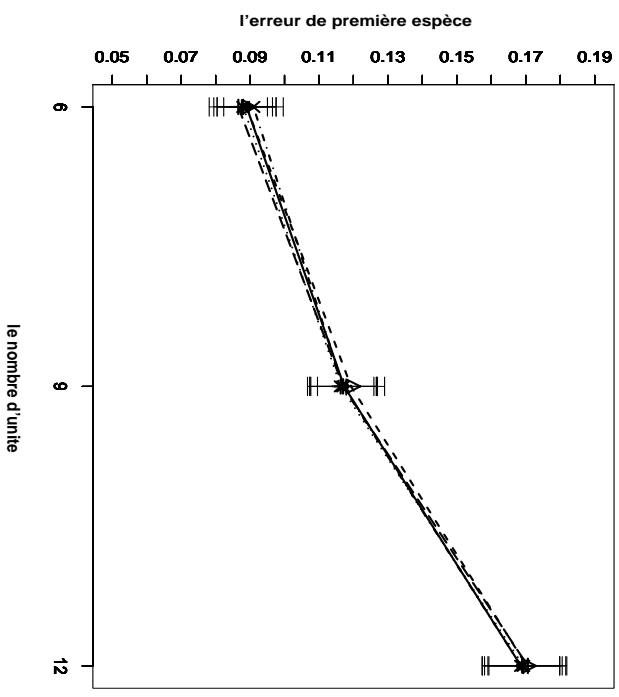
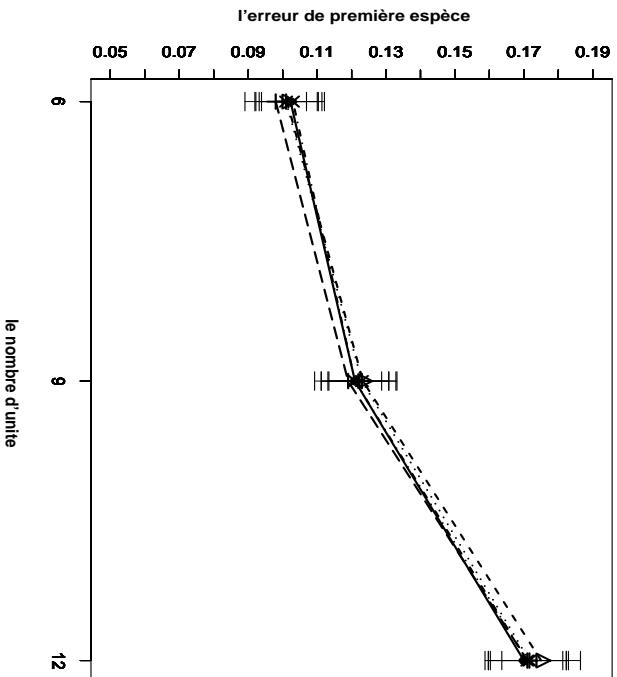
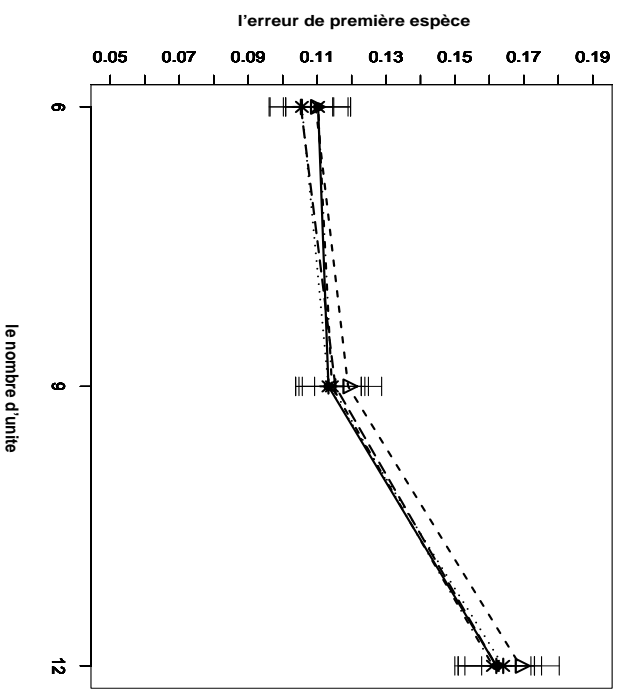
a : Wishart, $\rho = 0.1$, $\beta_1 = 1$ b : Wishart, $\rho = 0.1$, $\beta_1 = 2.5$ c : Wishart, $\rho = 0.1$, $\beta_1 = 1$ d : Wishart, $\rho = 0.1$, $\beta_1 = 2.5$ 

FIG. 4.7 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— \triangle —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Braek— $+$ — et Manly— \dots — \ast — \dots . 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. Pour toutes les figures, X^1 et X^2 sont des distances interpoints, la statistique de test est le T de Student et $\rho_1 = 0.1$. $\beta_1 = 1$ pour les figures a et c. $\beta_1 = 2.5$ pour les figures b et d. Les données générées pour les variables X^1 , X^2 et Y sont telles que $\mathbb{V}ar(\beta_1 X^1 + \beta_2 X^2) / \mathbb{V}ar(\epsilon) \geq 4$ dans les figures a et b, $\mathbb{V}ar(\beta_1 X^1 + \beta_2 X^2) / \mathbb{V}ar(\epsilon) < 4$, dans les figures c et d

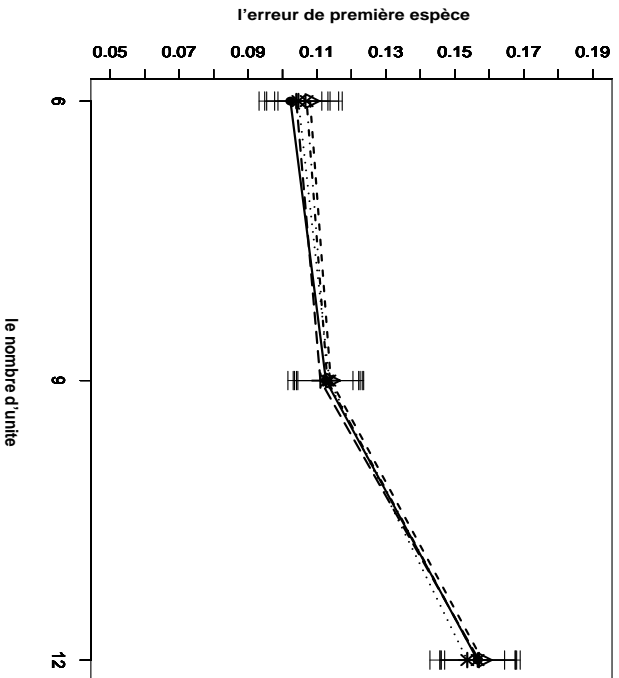
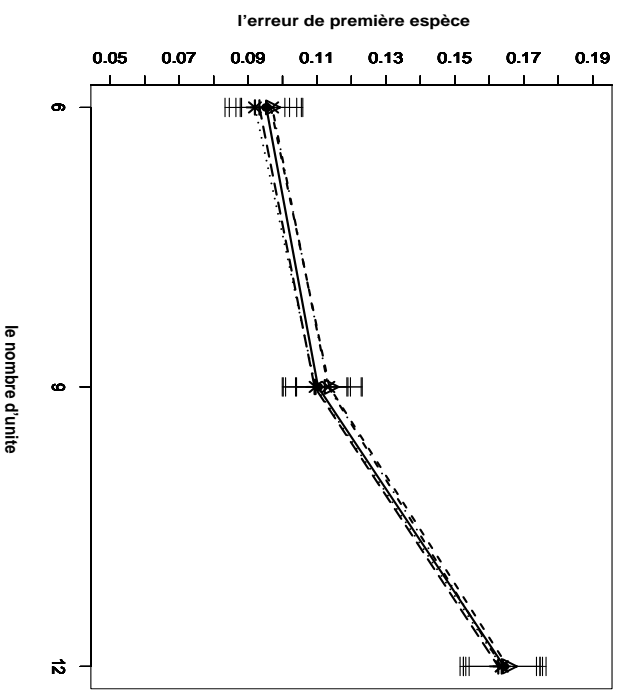
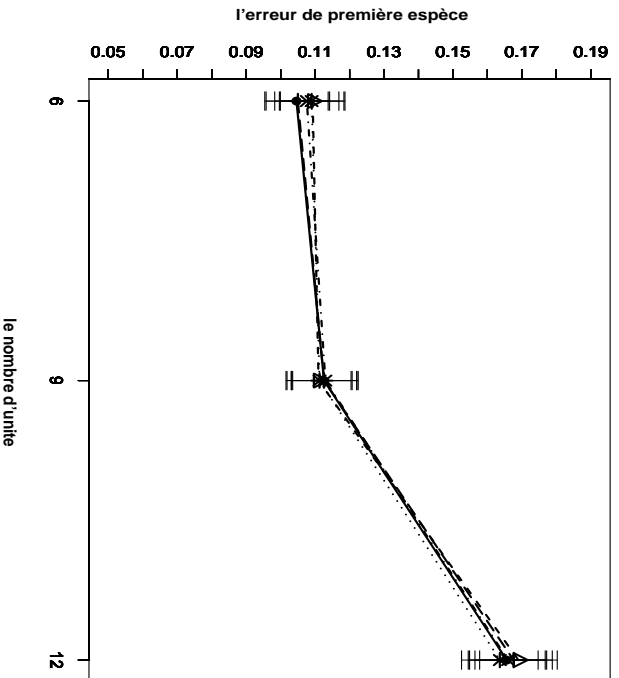
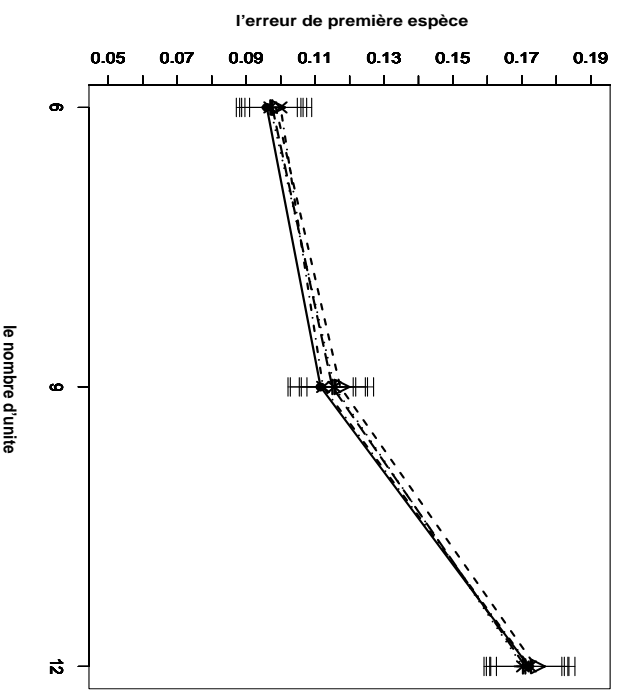
a : Wishart, $\rho = 0.9$, $\beta_1 = 1$ b : Wishart, $\rho = 0.9$, $\beta_1 = 2.5$ c : Wishart, $\rho = 0.9$, $\beta_1 = 1$ d : Wishart, $\rho = 0.9$, $\beta_1 = 2.5$ 

FIG. 4.8 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— \triangle —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Braek— $+$ — et Manly— \dots — \ast — \dots . 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. Pour toutes les figures, X^1 et X^2 sont des distances interpoints, la statistique de test est le T de Student et $\rho_1 = 0.9$. $\beta_1 = 1$ pour les figures a et c. $\beta_1 = 2.5$ pour les figures b et d. Les données générées pour les variables X^1 , X^2 et Y sont telles que $\text{Var}(\beta_1 X^1 + \beta_2 X^2) / \text{Var}(\epsilon) \geq 4$ dans les figures a et b, $\text{Var}(\beta_1 X^1 + \beta_2 X^2) / \text{Var}(\epsilon) < 4$, dans les figures c et d.

C2- Cas des données générées selon une loi $Beta(2, 0.3)$:

Cette partie présente trois comparaisons : la première concerne les transformations des distances simulées, la deuxième étudie l'influence du choix d'une statistique de test, la dernière oppose les alternatives bilatérales et unilatérales.

C2-1 Comparaison des trois ensembles des données : distance interpoint, carré des distances et transformée logarithmique des distances

Le cadre des simulations est ici identique à la section (C1 – 1), seule la loi Wishart est remplacée par loi $Beta(2, 0.3)$.

En considérant les figures 4.9 4.10 on peut faire les observations suivantes :

1. Contrairement à ce qu'on observait dans la section précédente où les erreurs de première espèce des quatre méthodes de test de permutation étaient plus grandes que le niveau nominal considéré $\alpha = 0.05$, nous observons ici des erreurs de première espèce proches de 0.05. C'est dire que les méthodes de test de permutation approchent bien le niveau considéré.
2. Bien que l'erreur de première espèce des méthodes de test de permutation soit proche de 0.05, ces dernières ne se superposent pas complètement les unes aux autres.
3. En comparant les figures 14.9-a,c,e (respectivement 14.10-a,b,c) aux figures 14.9-b,d,f (respectivement 14.10-b,d,f), nous constatons que l'augmentation du coefficient de corrélation n'a pas d'effet sur l'erreur de première espèce. Nous avons déjà observé des résultats analogues lors de l'étude des méthodes de test de permutation adaptées au cas de données ordinaires (cf. chapitre 2).
4. Pour toutes les combinaisons de facteurs, l'augmentation de n (de $n = 6$ à $n = 12$) (les valeurs des autres facteurs restant fixées), n'entraîne pas d'augmentation de l'erreur de première espèce.
5. Selon ces simulations, quelle que soit la combinaison des modalités des facteurs retenue, il ne semble pas exister de différence entre les résultats sur les données de distance interpoint initiales, les carrés des distances et les transformées logarithmiques des distances.
6. La consultation des figures 4.9 et 4.10 ne permet pas de déclarer qu'une de ces méthodes opère beaucoup mieux que les autres.
7. Les résultats présentés ici concernent la statistique de test T de Student. Mais, nous avons obtenu des résultats analogues pour les autres statistiques

de test : le coefficient de corrélation partielle ($r_{Y_{2.1}}$) et le carré du coefficient de corrélation partielle ($r_{Y_{2.1}}^2$) (cf. LABSAD en ligne).

8. Ces résultats concernent les hypothèses alternative bilatérales, nous avons obtenus des résultats analogues dans le cas d'hypothèses alternatives unilatérales (cf. LABSAD).

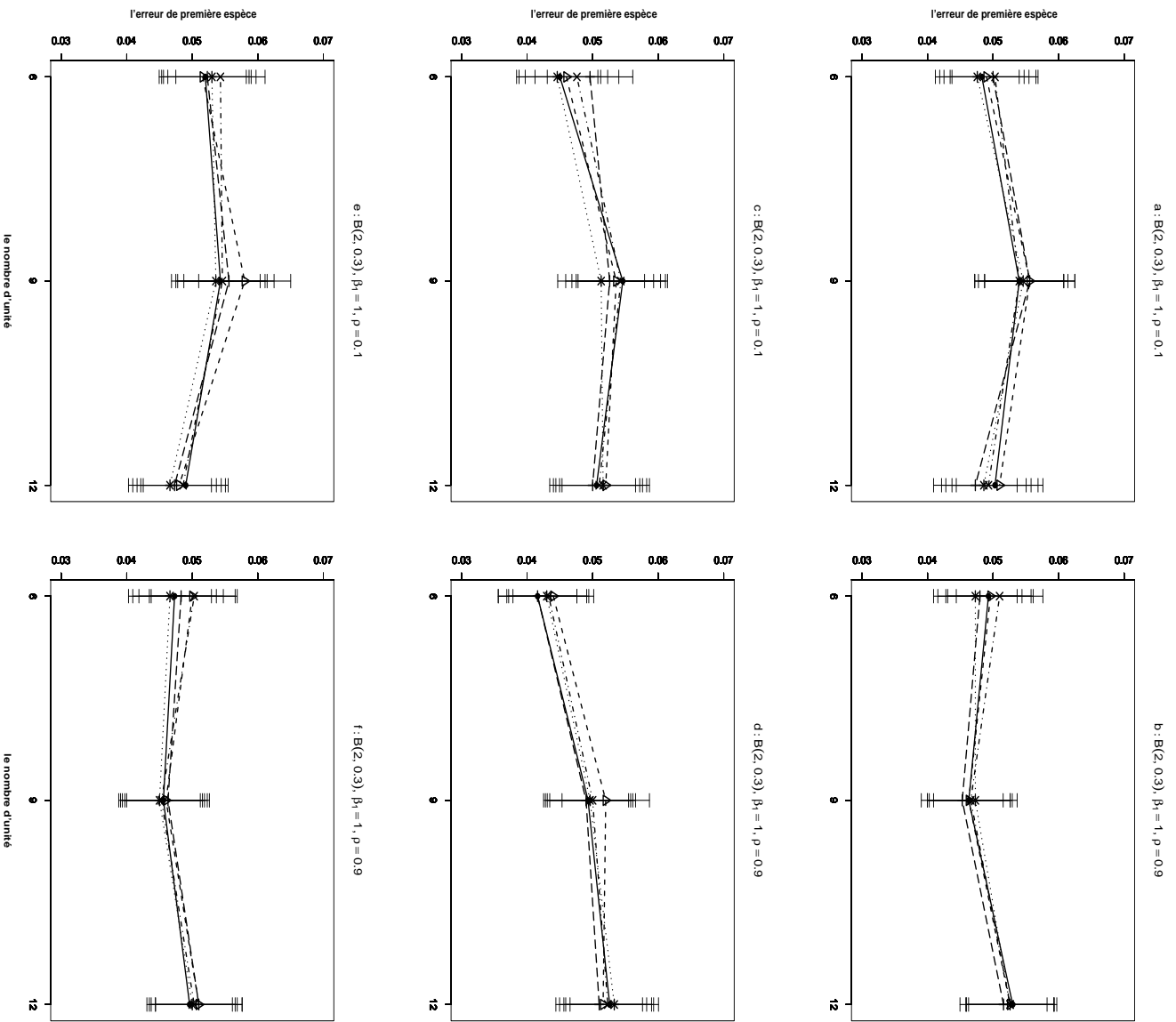


FIG. 4.9 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Brack— $+$ — et Manly— $*$ —... 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. La statistique de test est le T de Student pour toutes les figures. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures a, c et e . $\rho = 0.9$ pour les figures b, d et f . X^1 et X^2 sont des distances interpoints pour les figures a et b ; des carrés des distances pour les figures c et d ; des transformées logarithmiques des distances pour les figures e et f .

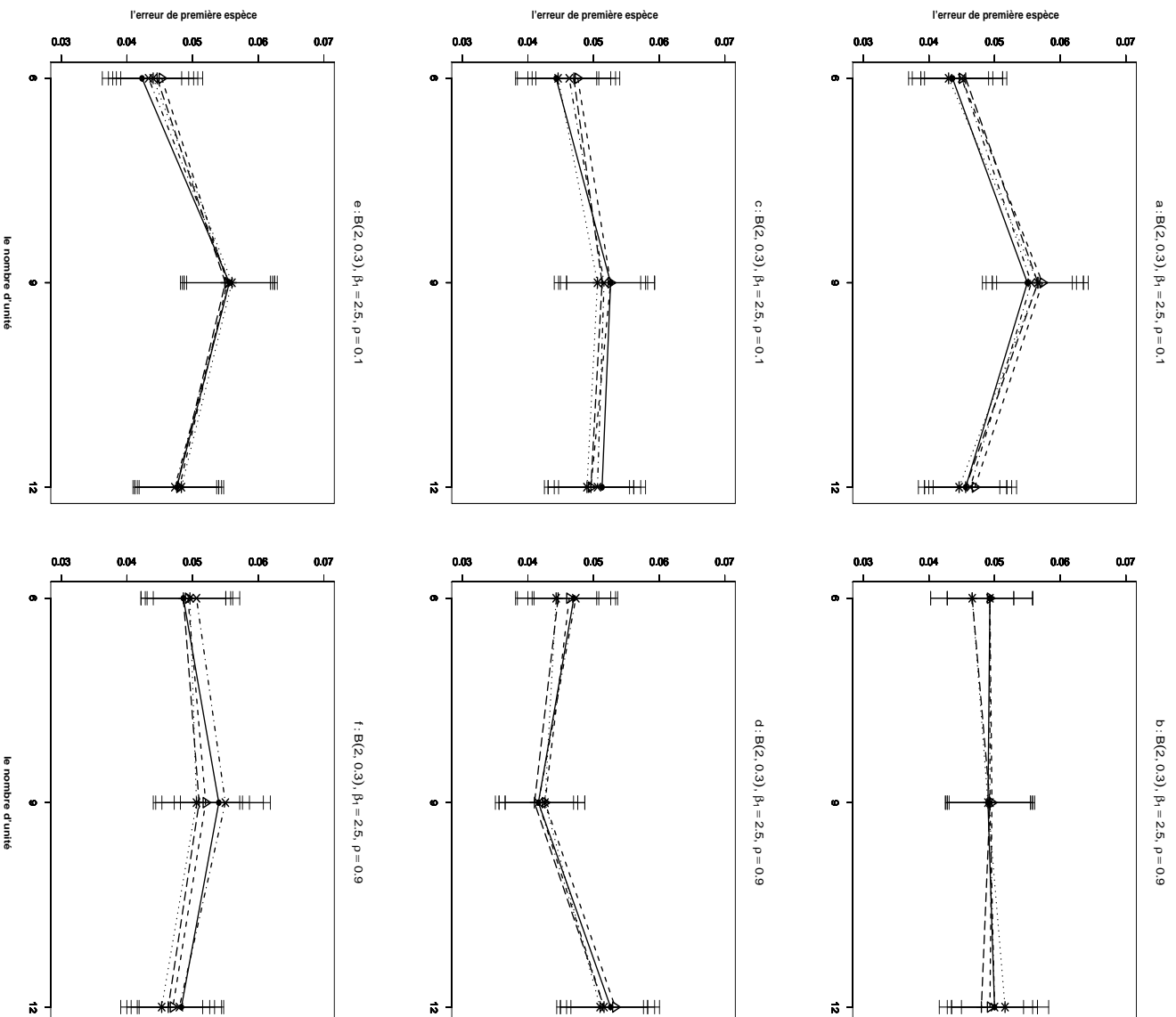


FIG. 4.10 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot\times\cdot-$; Ter Brack $-\cdot-$ et Manly $\dots*$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. La statistique de test est le T de Student pour toutes les figures. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures *a*, *c* et *e*. $\rho = 0.9$ pour les figures *b*, *d* et *f*. X^1 et X^2 sont des distances interpoints pour les figures *a* et *b*; des carrés des distances pour les figures *c* et *d*; des transformées logarithmiques des distances pour les figures *e* et *f*.

C2-2 Comparaison des trois statistiques de tests

Le cadre des simulations est ici identique à la section (C1 – 2), seule la loi Wishart est remplacée par loi $Beta(2, 0.3)$.

En considérant les figures 4.11 et 4.12 on peut tirer les conclusions suivantes

1. Pour toutes les combinaisons des facteurs n, β_1 et ρ , le risque d'erreur de première espèce obtenu par les méthodes de test de permutations est proche du niveau nominal $\alpha = 0.05$. Il est important de rappeler que ce niveau a été considéré comme le niveau de référence.
2. Pour toutes les combinaisons des facteurs n, ρ et β_1 , l'augmentation du nombre d'unités n ainsi que celle du coefficient de corrélation n'ont pas d'effet sur le risque d'erreur de première espèce.
3. Les statistiques de test du coefficient de corrélation et du coefficient de corrélation partielle fonctionnent de manière homogène pour toutes les combinaisons des facteurs.
4. Dans les figures 4.11-c,d (respectivement 4.12-c,d) où nous avons utilisé la statistique $r_{Y_{2,1}}$ ainsi que dans les figure 4.11-e,f (respectivement 4.12-e,f) où nous avons utilisé la statistique $r_{Y_{2,1}}^2$, le risque d'erreur de première espèce obtenu par la méthode de Kennedy (surtout lorsque $n = 6$) est plus grand que pour les autres méthodes. Mais dans les figures 4.11-a,b lorsque nous avons utilisé la statistique T de Student, le risque d'erreur de première espèce obtenu par la méthode de Kennedy est proche de celui obtenu par les autres méthodes.
5. Selon les résultats de ces simulations, les trois autres méthodes de test de permutation (Freedman-Lane, Ter Brack, Manly) fonctionnent de façon très semblable. Nous pouvons donc difficilement sélectionner l'une d'entre elles comme la meilleure méthode (cf. LABSAD en ligne).
6. Ces résultats ont été obtenus lorsque les variables X_1 et X_2 étaient les distances interpoints initiales. Il est intéressant de signaler que l'application des mêmes méthodes aux variables de carrés des distances ou aux transformées logarithmes de distances fournit des résultats analogues.
7. Ces résultats concernent les hypothèses bilatérales, mais nous avons obtenus les mêmes résultats pour les hypothèses unilatérales.

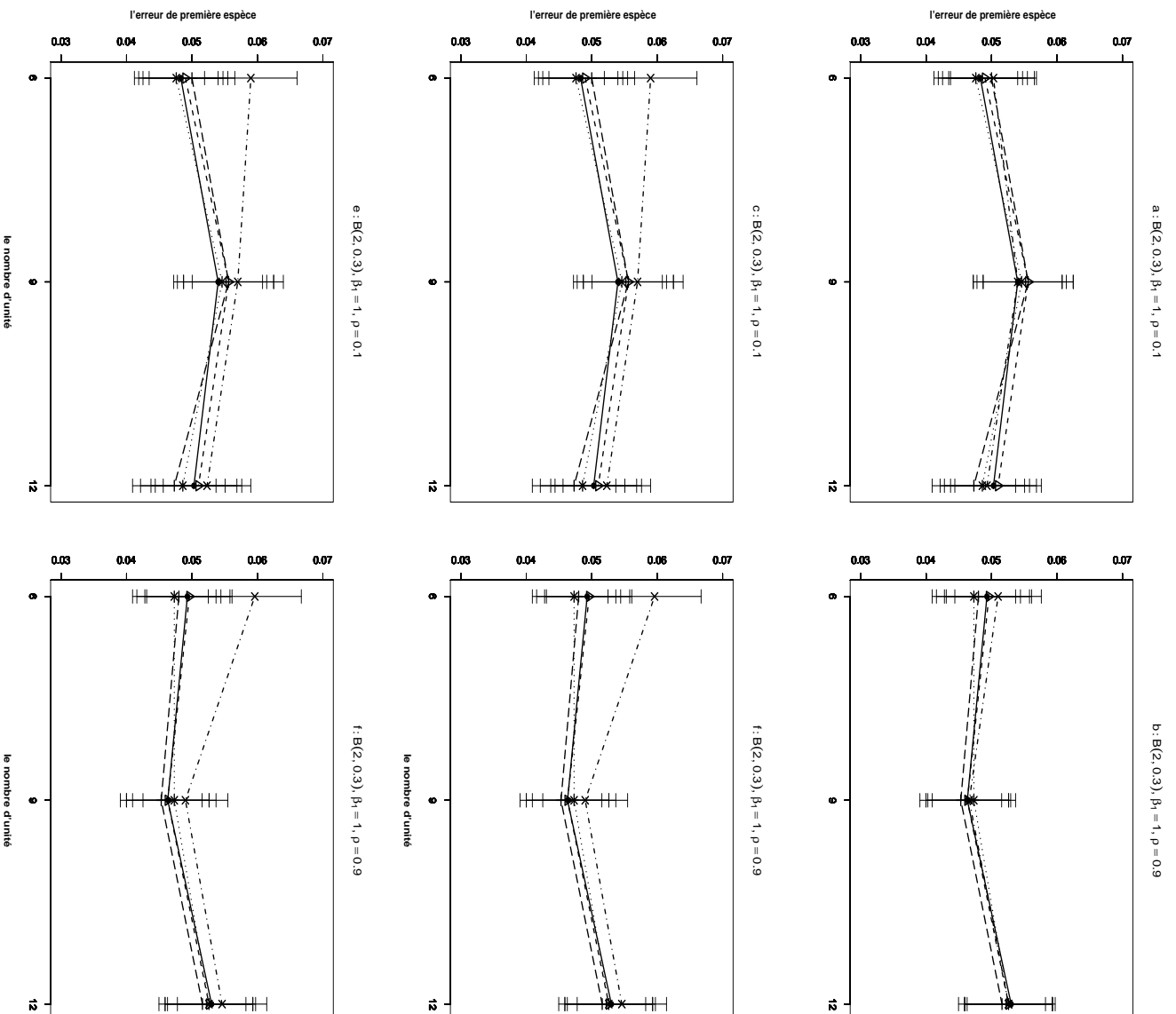


FIG. 4.11 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Brack— $+$ — et Manly— $*$ —... 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures a, c et e . $\rho = 0.9$ pour les figures b, d et f . X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est : le T de Student pour les figures a et b ; le coefficient de corrélation partielle pour les figures c et d ; carrée du coefficient de corrélation partielle pour les figures e et f .

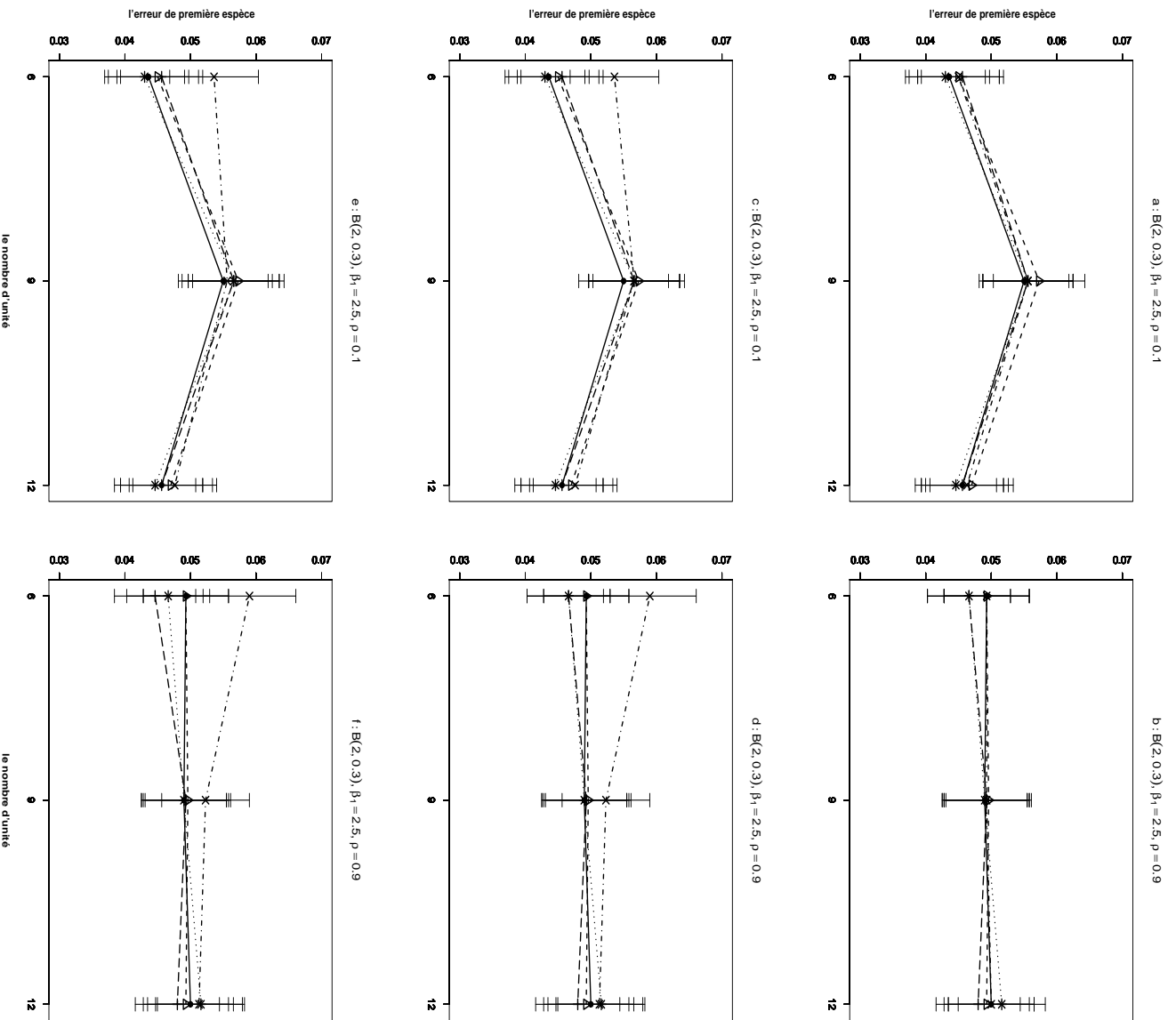


FIG. 4.12 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Braek— $+$ — et Manly— \ast —... 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures a , c et e . $\rho = 0.9$ pour les figures b , d et f . X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est : le T de Student pour les figures a et b ; le coefficient de corrélation partielle pour les figures c et d ; carrée du coefficient de corrélation partielle pour les figures e et f .

C2-3 Comparaison des hypothèses alternatives bilatérale et unilatérale

Le cadre des simulations est ici identique à la section (C1 – 3), seule la loi Wishart est remplacée par loi $Beta(2, 0.3)$.

En considérant les figures 4.13 et 4.14 on peut faire les observations suivantes :

1. La comparaison des figures 4.13-a,b (respectivement 4.14-a,b) avec 4.13-c,d (respectivement 4.14-c,d), nous permet de constater que les risques d'erreur de première espèce obtenus par des méthodes de test de permutation sur l'hypothèse alternative unilatérale et l'hypothèse alternative bilatérale sont proches.
2. Les quatre méthodes de test de permutation ainsi que la méthode classique ne présentent pas de grandes différences lorsqu'on compare le résultat du test correspondant aux hypothèses alternatives unilatérale et bilatérale.
3. Pour toutes les combinaisons de facteurs retenues dans cette étude, l'augmentation du nombre d'unité n ainsi que celle du coefficient de corrélation ρ (les autres facteurs restant fixés) semblent avoir très peu d'effet sur l'erreur de première espèce.
4. Ces résultats sont valables pour la statistiques de test T de Student. Mais, nous avons obtenu des résultats très semblables pour les deux autres statistiques de test : le coefficient de corrélation partielle ($r_{Y2.1}$) et le carré du coefficient de corrélation partielle ($r_{Y2.1}^2$).

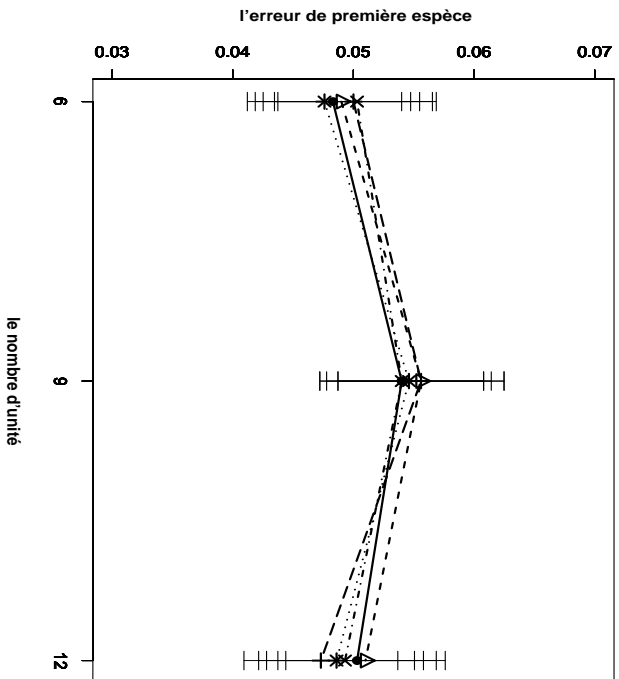
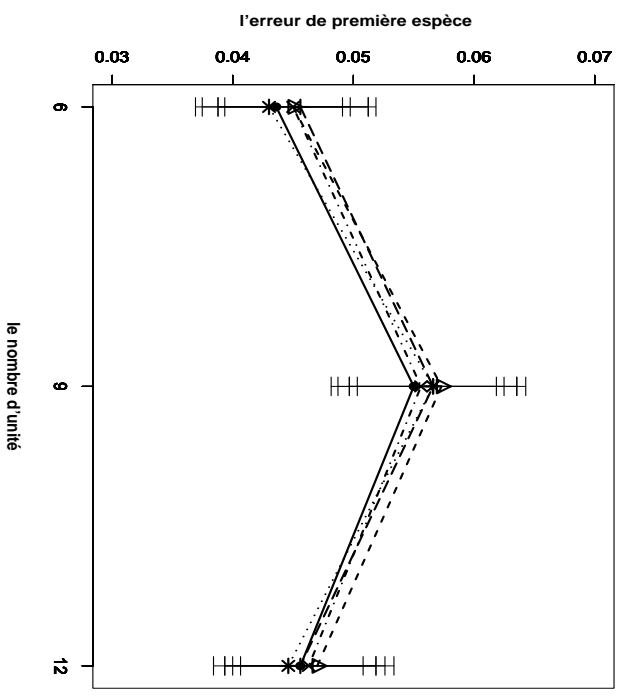
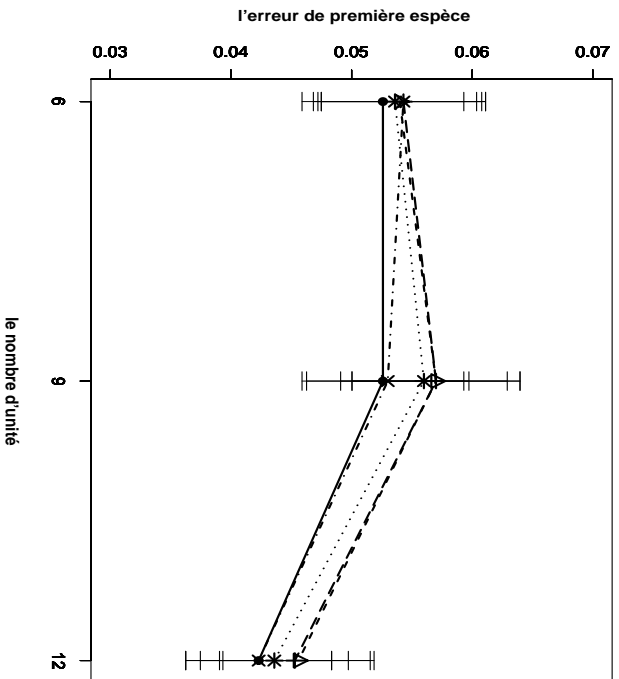
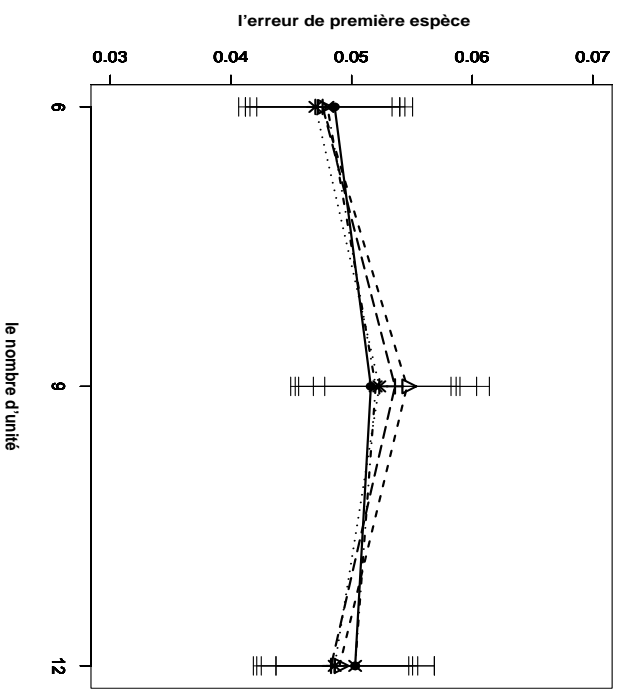
a : $B(2, 0.3), \rho = 0.1, \beta_1 = 1$ b : $B(2, 0.3), \rho = 0.1, \beta_1 = 2.5$ c : $B(2, 0.3), \rho = 0.1, \beta_1 = 1$ d : $B(2, 0.3), \rho = 0.1, \beta_1 = 2.5$ 

FIG. 4.13 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— Δ —; Freedman et Lane — \bullet —; Kennedy — \times —; Ter Braek — $+$ — et Manly $\dots * \dots$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est le T de Student pour toutes figures. $\rho_1 = 0.1$ pour toutes les figures. $\beta_1 = 1$ pour les figures a et c. $\beta_1 = 2.5$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

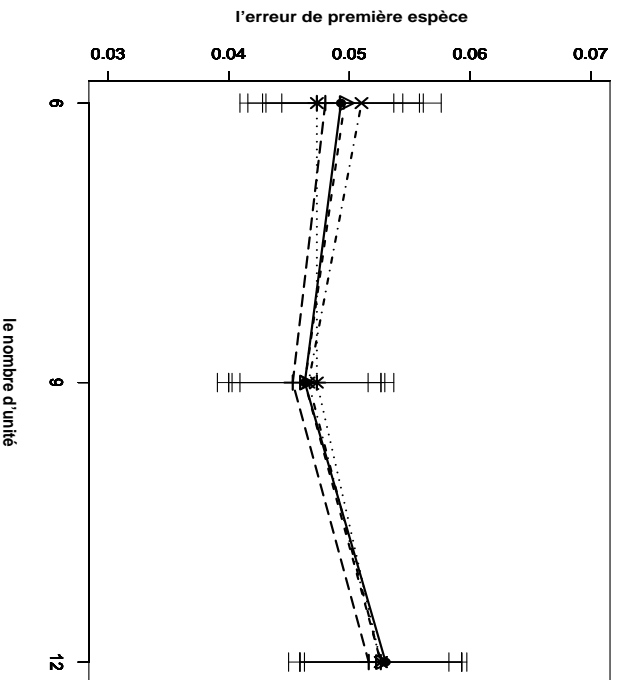
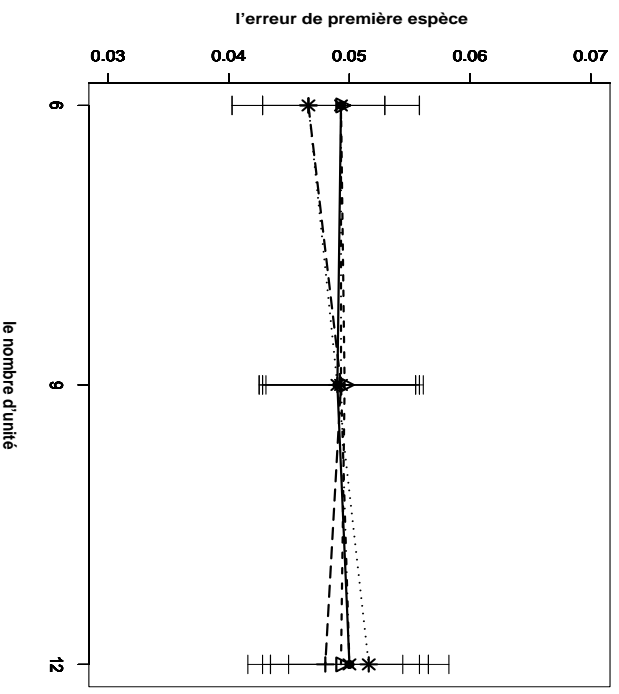
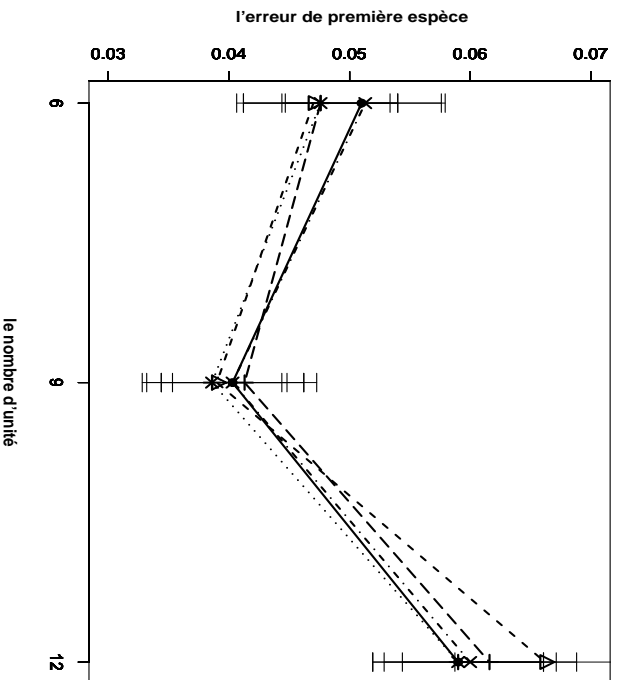
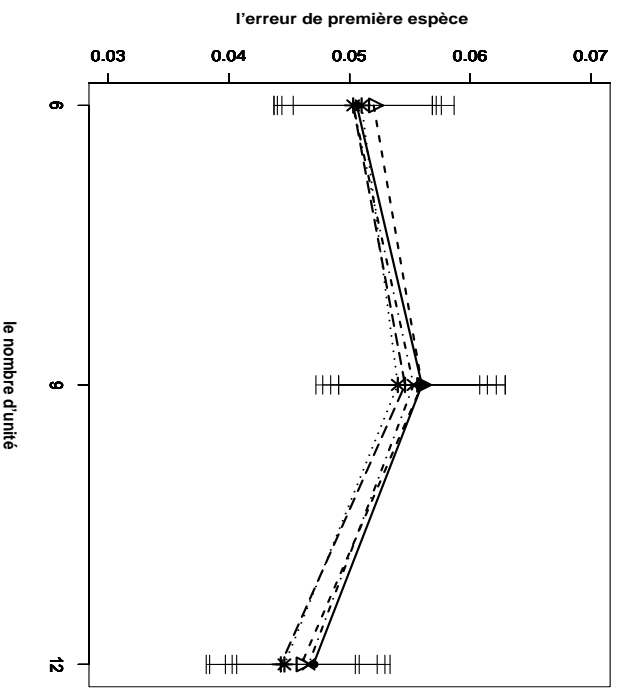
a : $B(2, 0.3), \rho = 0.9, \beta_1 = 1$ b : $B(2, 0.3), \rho = 0.9, \beta_1 = 2.5$ c : $B(2, 0.3), \rho = 0.9, \beta_1 = 1$ d : $B(2, 0.3), \rho = 0.9, \beta_1 = 2.5$ 

FIG. 4.14 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Braek— $+$ — et Manly— \dots — \ast — \dots . 30000 n-échantillons du résidu e ont été simulés suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est le T de Student pour toutes figures. $\rho_1 = 0.9$ pour toutes les figures. $\beta_1 = 1$ pour les figures a et c. $\beta_1 = 2.5$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

4.3.2 Analyse de la puissance du test

De façon complémentaire aux résultats de simulations présentés dans la section précédente pour le risque d'erreur de première espèce, nous voulons étudier maintenant l'influence des facteurs étudiés plus haut sur la fonction puissance.

A) Le cadre de la simulation

- a) Pour vérifier les effets du nombre d'unités statistiques n , nous avons tiré un n-échantillon de dimension 10, dans trois cas : $n \in \{6, 9, 12\}$;
- b) Pour vérifier les effets du degré de liaison avec la covariable nous avons considéré les coefficients de régression partielle $\beta_1 = \{0, 1, 2.5\}$.
- c) Afin d'étudier les effets du degré de corrélation entre les variable X^1 et X^2 , nous avons limité notre étude au cadre d'une corrélation faible, fixée ici à 0.1 et d'une corrélation forte, égale à 0.9. Nous avons donc retenu les valeurs de simulation $\rho(X^1, X^2) = \{0.1, 0.9\}$.
- d) La distribution du terme d'erreur aléatoire ϵ reprend ici le choix précédent de trois lois variant par leur caractère d'asymétrie : $\{Wishart, Exp(2), Beta(2, 0.3)\}$.

Les étapes suivies pour générer des données et calculer les puissances empiriques obtenues par les méthodes de test de permutation sur des données de distance et pour les différentes statistiques de test dans le cas où les hypothèses alternatives sont bilatérales ou unilatérales, sont les mêmes que celles qui ont été définies lors de l'étude du risque d'erreur de première espèce.

B) Les résultats des simulations pour la puissance

Dans cette section nous présentons les résultats relatifs aux deux situations : les termes d'erreur suivent (**B1**) une loi de Wishart et (**B2**) une loi $Beta(2, 0.3)$. Lorsque les termes d'erreur suivent une loi $Exp(2)$, les résultats des simulations sont les mêmes que le cas de données générées selon une loi de $Beta(2, 0.3)$, nous les avons consignés dans l'annexe 4.

B1) Cas des données générées selon une loi de Wishart :

Cette partie présente quatre comparaisons : la première concerne les transformations des distances simulées, la deuxième étudie l'influence du choix d'une statistique de test, la troisième oppose les alternatives bilatérale et unilatérale et la dernière comparaison contraste deux situations marquées par un rapport signal sur bruit fort ou faible.

B1-1 Comparaison des trois ensembles des données : distance interpoint, carré de distance et transformée logarithmiques de distances

L'objet de cette partie est de comparer par des résultats de simulations la puissance (de test) de la statistique de test T de Student sur l'hypothèse alternative bilatérale, pour trois ensembles de données : distance interpoint, carré des distances et transformée logarithmique des distances. Les valeurs des variables Y , X^1 , et X^2 sont générées telles que $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$.

En considérant les figures 4.15 et 4.16 on peut faire les observations suivantes :

1. Par consultation des figures 4.15 et 4.16, on observe que les quatre méthodes de test de permutation fonctionnent de façon très semblable pour toutes les combinaisons de modalités des facteurs retenues dans notre comparaison. C'est-à-dire que les puissances semblent être sensiblement égales.
2. Lorsque β_1 est égal à zéro, la puissance du test de permutation est proche de 1, quelles que soient les modalités des autres facteurs.
3. Pour toutes les combinaisons de facteurs, nous présentons les graphiques correspondant aux deux cas : 1) lorsque X^1 et X^2 sont des distances interpoints les figures 4.15-a,b (respectivement 4.16- a,b) ; 2) lorsque X^1 et X^2 sont des transformées logarithmiques les figures 4.15-e,f (4.16-e,f), leurs fonctions puissance sont presque identiques. C'est-à-dire que ces deux ensembles de données fonctionnent de façon indifférenciée dans la plage de valeurs obtenues dans ces simulations. De plus, pour ces deux ensembles de données, en augmentant β_1 (ainsi que ρ) lorsque les valeurs des autres facteurs sont fixées, la puissance diminue. On observe de façon complémentaire que lorsque le nombre d'unités augmente, les autres facteurs restant fixés, la puissance augmente.
4. Lorsque les valeurs des variables X_1 et X_2 sont des carrés de distances (cf. figures 4.15-c,d et 4.16-c,d), pour toutes les combinaisons de facteurs, la puissance est très proche de 1 et est plus grande que les puissances obtenues à partir des distances originelles ou des données transformées logarithmiques des distances. De plus lorsque β_1 ainsi que ρ augmentent, la puissance ne diminue pas.
5. Lorsque $\rho = 0.9$, la perte de puissance obtenue en augmentant β_1 est plus forte que dans le cas où $\rho = 0.1$.
6. Nous avons présenté ici les résultats correspondent au cas de la statistique T de Student. Il convient de préciser que nous avons obtenu les résultats analogues en utilisant les deux autres statistiques de test : le coefficient de corrélation partielle ($r_{Y2.1}$) et le carré de coefficient de corrélation partielle ($r_{Y2.1}^2$) (cf. LAD SAD en ligne).
7. Les résultats présentés correspondent au cas d'hypothèses alternatives bilatérales mais nous avons obtenu les résultats analogues, dans le cas d'hypothèses

alternatives unilatérales.

8. De plus, ces résultats concernent le cas où la variance du terme de signal est quatre fois plus grande que celle du terme d'erreur $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$. Ces résultats ont été comparés à ceux obtenus lors de simulations avec une variance du terme d'erreur quatre fois plus grande que celle du terme signal $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) < 4$.

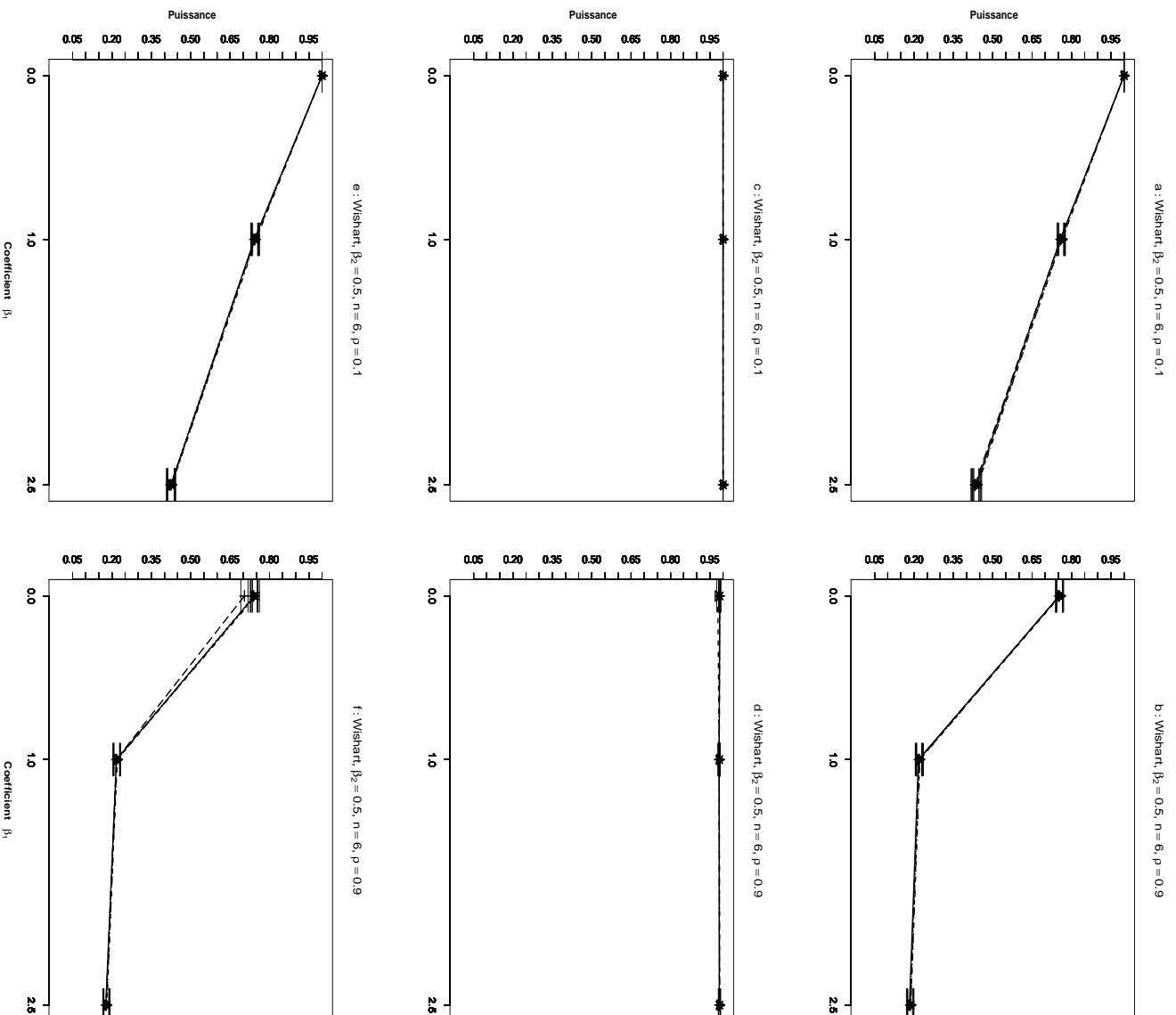


FIG. 4.15 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot \times \cdot-$; Ter Braek $-\text{+}-$ et Manly $\dots * \dots$. 30000 n-échantillons du résidu ϵ ont été simulées suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, $\beta_2 = 0.5$ et $n = 6$. $p = 0.1$ pour les figures a, c et e. $p = 0.9$ pour les figures b, d et f. X^1 et X^2 sont des distances interpoints pour les figures a et b; des carrés des distances pour les figures c et d; des transformées logarithmiques des distances pour les figures e et f.

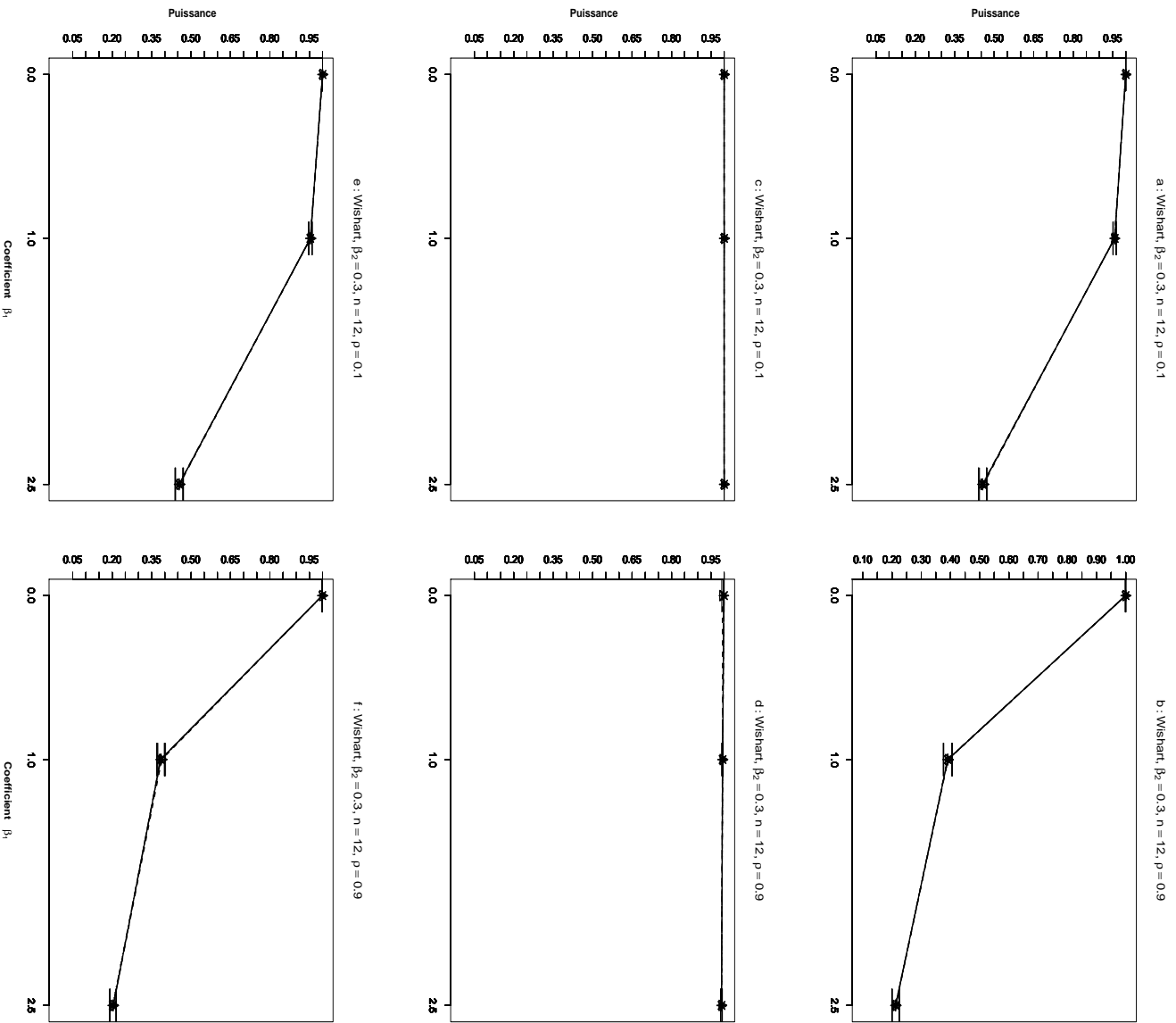


FIG. 4.16 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot\times\cdot-$; Ter Brack $-\text{+}-$ et Manly $\dots*\dots$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, $\beta_2 = 0.3$ et $n = 12$. $\rho = 0.1$ pour les figures a, c et e. $\rho = 0.9$ pour les figures b, d et f. X^1 et X^2 sont des distances interpoints pour les figures a et b; des carrés des distances pour les figures c et d; des transformées logarithmiques des distances pour les figures e et f.

B1-2 Comparaison des trois statistiques de test

L'objet de cette partie est la comparaison des résultats des simulations pour la puissance (de test) concernant les trois statistiques de tests T de Student, $r_{Y_{2.1}}$ et $r_{Y_{2.1}}^2$, dans le cas d'une hypothèse alternative bilatérale. Ici, X^1 et X^2 sont des distances interpoints telles que $\text{Var}(\mathbf{D}_{\mathbf{R}}\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$.

En considérant les figures 4.17 et 4.18 on peut faire les observations suivantes :

1. En comparant les figures 4.17-a,c,d où $n = 6$, nous constatons que les trois statistiques de test T , $r_{Y_{2.1}}$ et $r_{Y_{2.1}}^2$ fonctionnent presque de la même manière. On fait le même constat à partir des figures 4.18 pour lesquelles $n = 12$.
2. Les figures 4.17 et 4.18 nous permettent de constater que la puissance diminue lorsque nous augmentons le coefficient de corrélation. Nous avons déjà observé une telle liaison lorsque nous avons étudié les méthodes de test de permutation sur les variables ordinaires (dans le deuxième chapitre).
3. Pour toutes les combinaisons des facteurs, en augmentant le coefficient β_1 , nous constatons que la puissance diminue.
4. Selon les résultats obtenus, les quatre méthodes de test fonctionnent de manière très semblable pour toutes les combinaisons des facteurs.
5. Nous constatons qu'en augmentant le nombre d'unités n pour chacune des trois statistiques de test, la puissance augmente.
6. En conséquence, les résultats montrent qu'il n'existe pas de différences substantielles entre les trois statistiques de test.
7. Ces résultats sont obtenus lorsque X^1 et X^2 sont des distances interpoints. Mais nous avons obtenus des résultats très semblables à partir des carrés des distances et des transformées logarithmiques des distances.
8. Les résultats obtenus correspondant au cas d'hypothèses alternatives bilatérales, mais nous avons obtenu les résultats analogues pour les hypothèses alternatives unilatérales.
9. De plus, ces résultats traitent le cas où la variance du terme de signal est quatre fois plus grande que celle du terme d'erreur $\text{Var}(\mathbf{D}_{\mathbf{R}}\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$. Ces résultats sont confirmés lorsque la variance du terme d'erreur est quatre fois plus grand que celle du terme signal $\text{Var}(\mathbf{D}_{\mathbf{R}}\boldsymbol{\beta})/\text{Var}(\epsilon) < 4$.

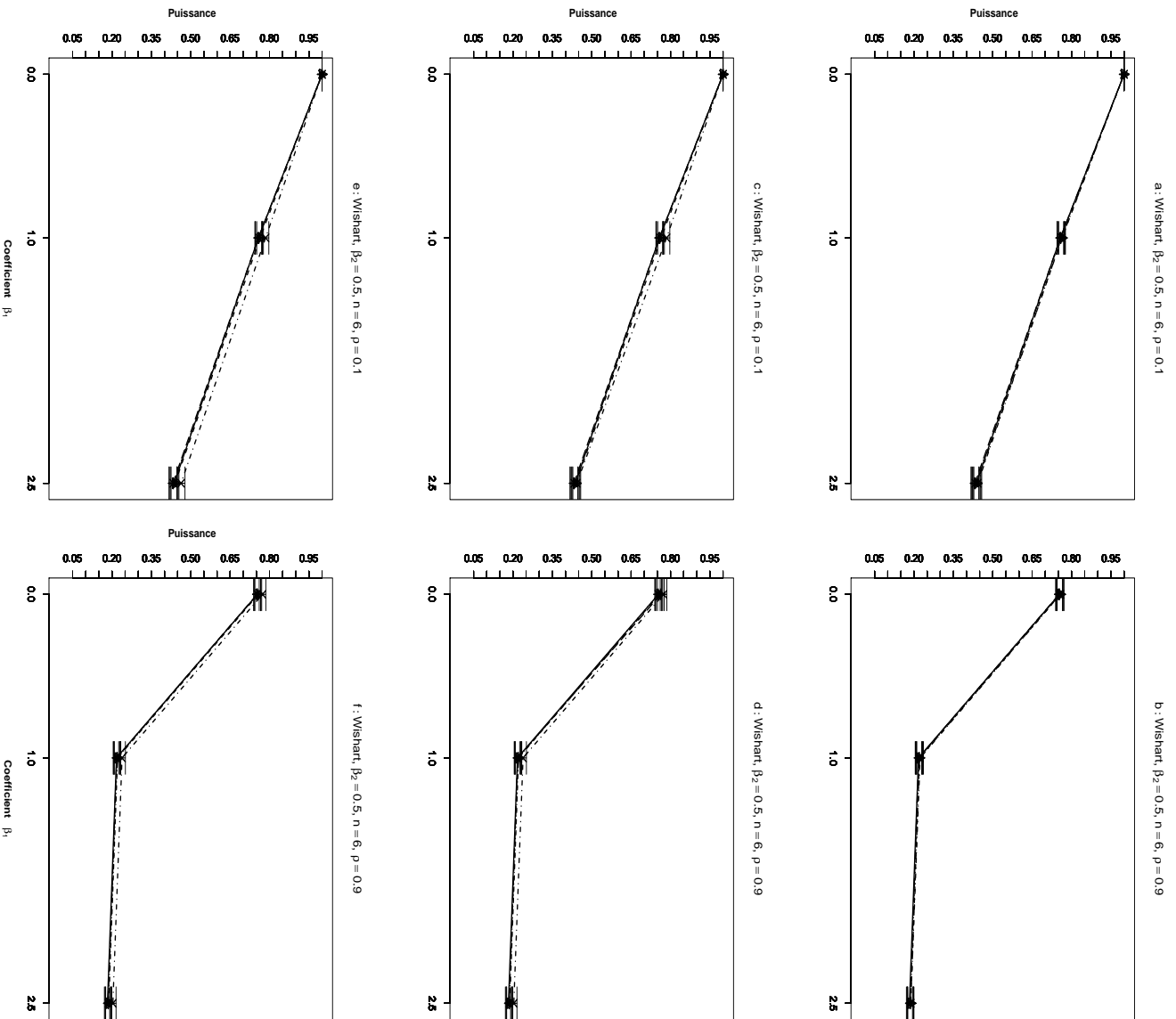


FIG. 4.17 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Braack— $+$ — et Manly— $*$ — \dots . 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures *a*, *c* et *e*. $\rho = 0.9$ pour les figures *b*, *d* et *f*. X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est : T de Student pour les figures *a* et *b*; le coefficient de corrélation partielle pour les figures *c* et *d*; carrée du coefficient de corrélation partielle pour les figures *e* et *f*.

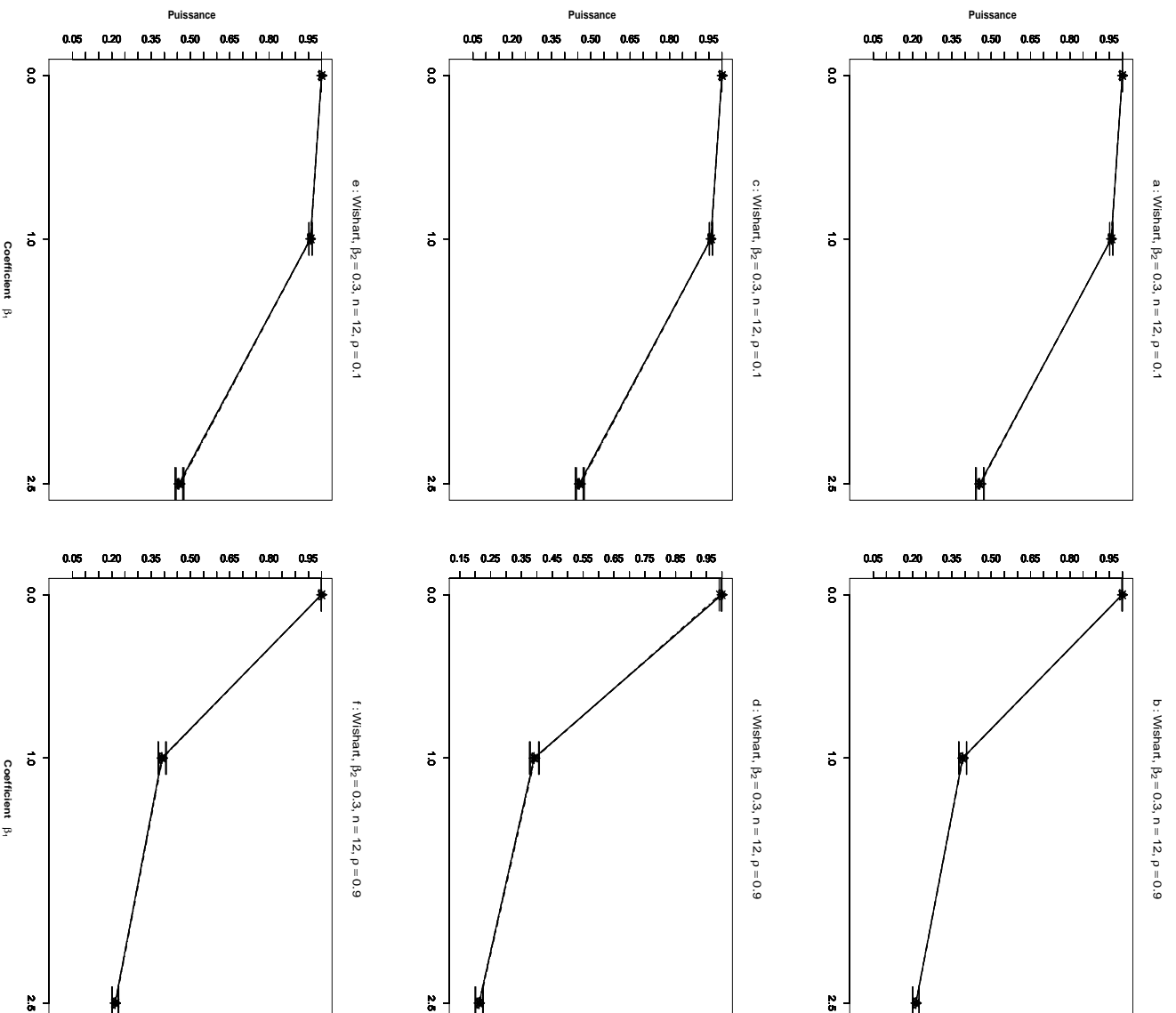


FIG. 4.18 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot \times \cdot-$; Ter Braek $-\text{+}-$ et Manly $\dots * \dots$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. Pour toutes les figures $\beta_2 = 0.3$ et $n = 12$. $\rho = 0.1$ pour les figures a, c et e. $\rho = 0.9$ pour les figures b, d et f. X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est : T de Student pour les figures a et b; le coefficient de corrélation partielle pour les figures c et d; carrée du coefficient de corrélation partielle pour les figures e et f.

B1-3 Comparaison des hypothèses alternatives bilatérale et unilatérale

Dans cette partie nous voulons comparer par des résultats de simulations la puissance (de test) de la statistique de test T de Student, sur les deux cas : l'hypothèse alternative bilatérale et hypothèse alternative unilatérale où X^1 et X^2 sont des distances interpoints tel que $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$

En considérant les figure 4.19 et 4.20 on peut tirer les conclusions suivantes :

1. Les quatre méthodes de test de permutation ainsi que la méthode classique ont les mêmes valeurs de la fonction puissance, pour toutes les combinaisons des facteurs .
2. En comparant les figures 4.19-a, (respectivement 4.19-b) et 4.19-c (respectivement 4.19-d), nous constatons que la puissance est un peu plus grande dans les cas unilatéraux que celle qui résulte de tests bilatéraux. Ces résultats sont également obtenus à partir des figures 4.20 où $\rho = 0.9$.
3. La puissance croît en augmentant n dans les deux cas.
4. Pour le test d'hypothèse unilatéral comme bilatéral, la puissance diminue lorsque nous augmentons β_1 . De plus nous confirmons que la diminution est moins rapide dans le cas du $\rho = 0.9$ (figure 4.20) que dans le cas $\rho = 0.1$ (figure 4.19)
5. En conséquence, les résultats montrent qu'il n'existe pas de différence majeure entre les méthodes appliquées dans le cas d'une alternative unilatérale et bilatérale.
6. De plus, les résultats présentés correspondent au cas où la variance du terme de signal était quatre fois plus grande que celle du terme d'erreur $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$, mais les résultats sont analogues lorsque la variance du terme d'erreur est quatre fois plus grande que celle du terme signal $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) < 4$.
7. Ces résultats sont obtenus lorsque X^1 et X^2 sont des distances interpoint, mais nous avons obtenus des résultats très semblables à partir des carrés de distances et des transformées logarithmiques de distances.
8. Nous avons présenté ici le cas de la statistique T de Student. Mais avons obtenu les résultats correspondants en utilisant les deux statistiques de test (cf. LABSAD en ligne).

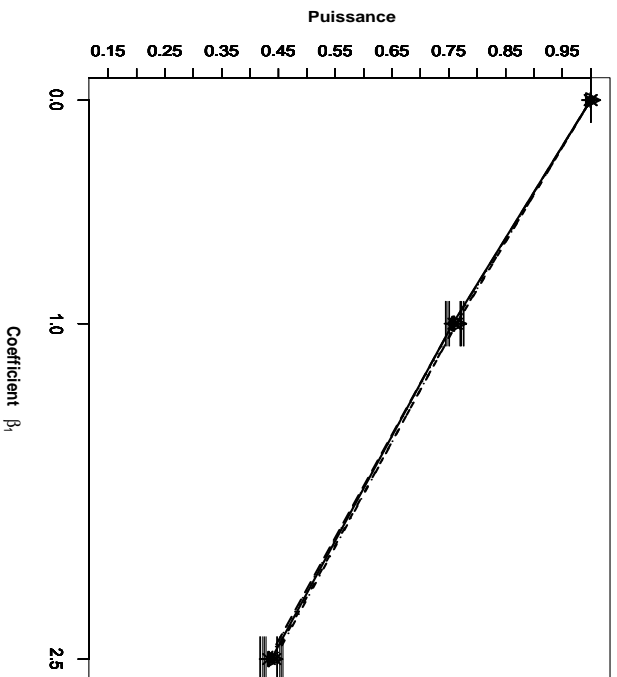
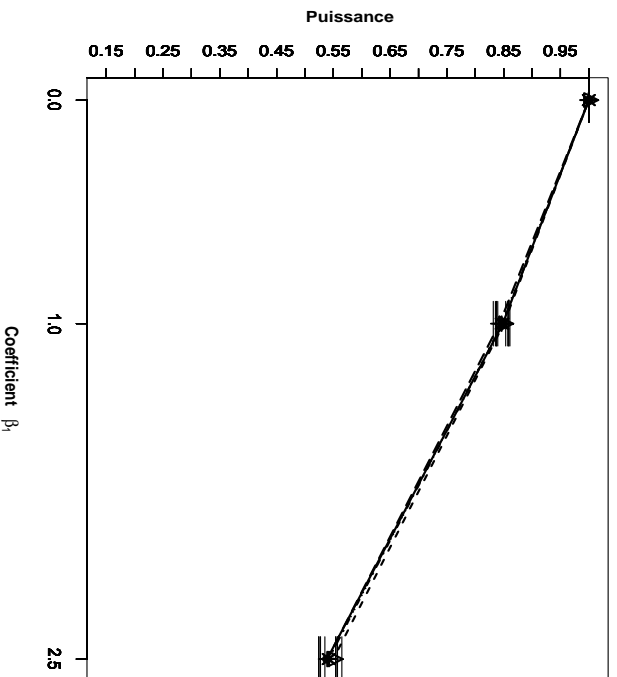
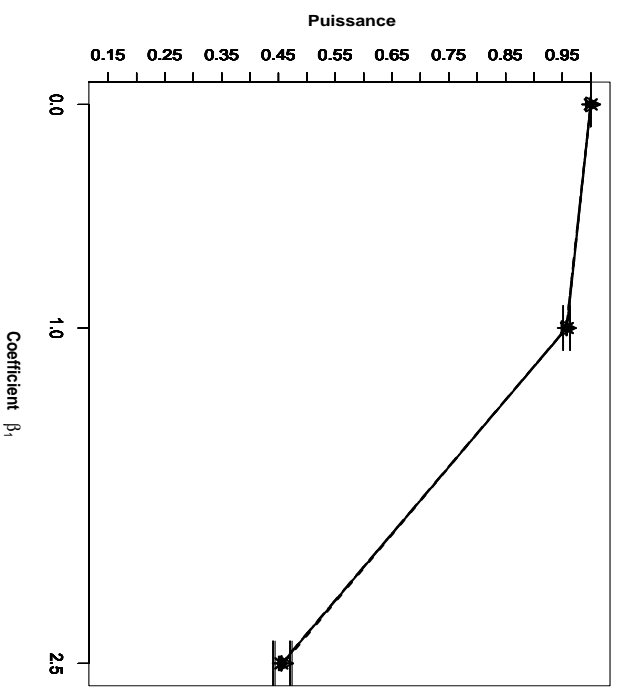
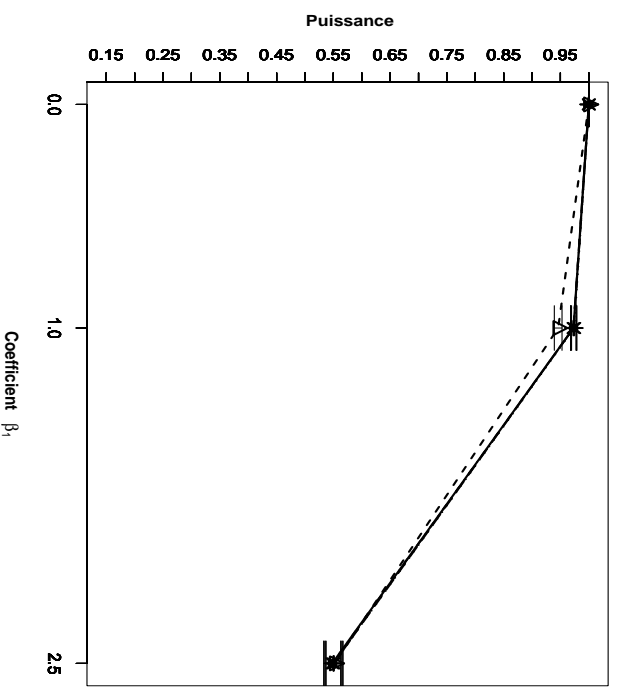
a : Wishart, $\rho = 0,1$, $\beta_2 = 0,5$, $n = 6$ c : Wishart, $\rho = 0,1$, $\beta_2 = 0,5$, $n = 6$ b : Wishart, $\rho = 0,1$, $\beta_2 = 0,3$, $n = 12$ d : Wishart, $\rho = 0,1$, $\beta_2 = 0,3$, $n = 12$ 

FIG. 4.19 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot \times \cdot-$; Ter Braek $-\text{+}-$ et Manly $\dots * \dots$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, X^1 et X^2 sont des distances interpoints et $\rho_1 = 0,1$. $n = 6$, $\beta_1 = 0,5$ pour les figures a et c. $n = 12$, $\beta_1 = 0,3$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

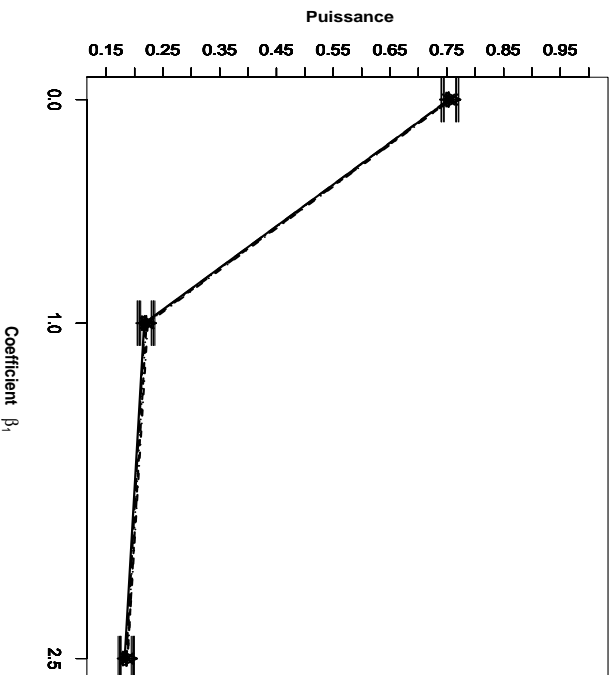
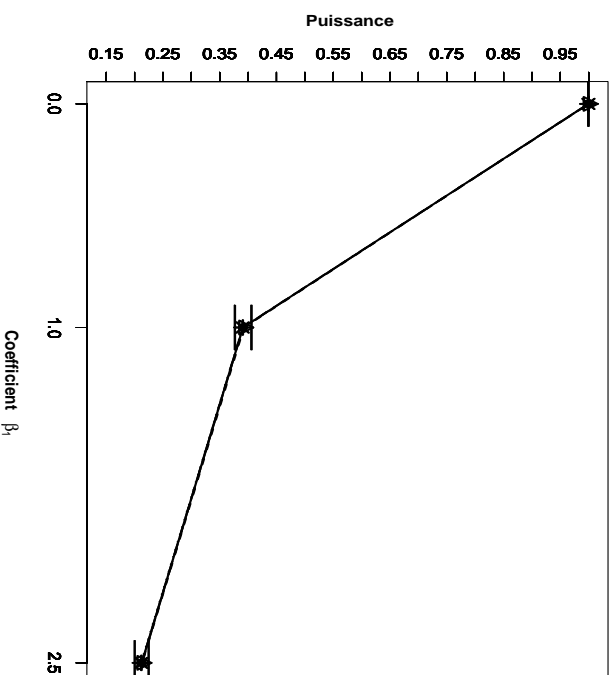
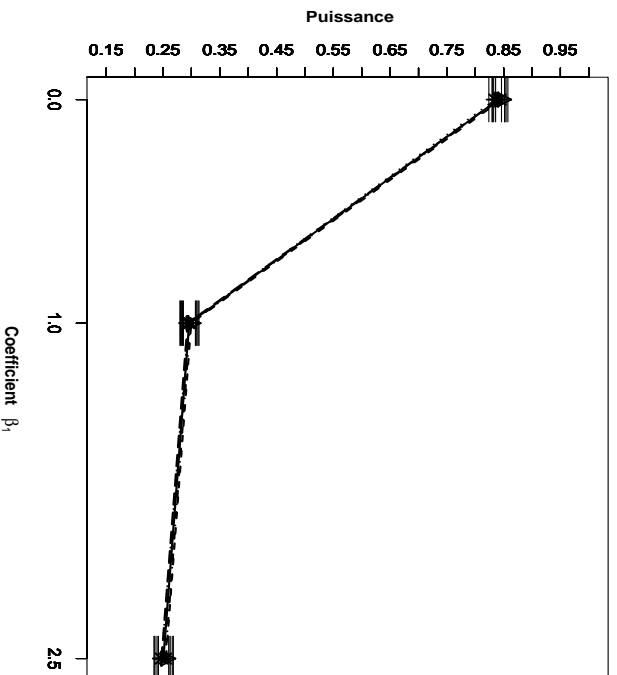
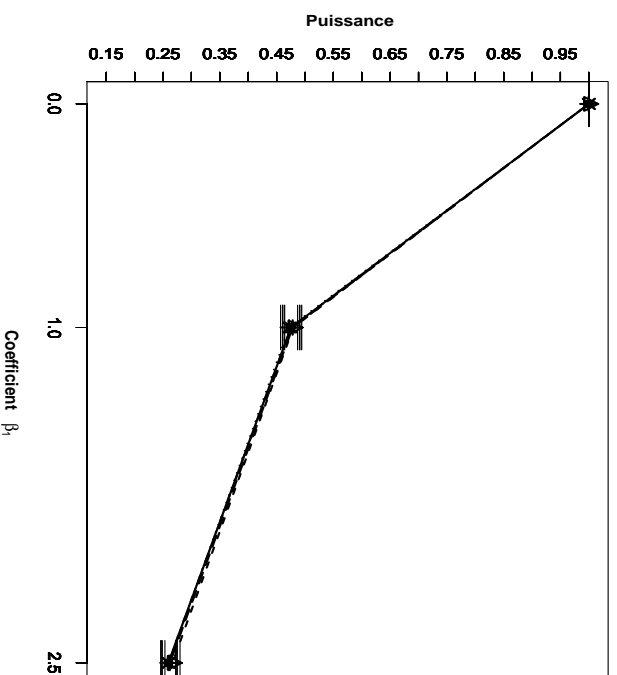
a : Wishart, $\rho = 0.9$, $\beta_2 = 0.5$, $n = 6$ b : Wishart, $\rho = 0.9$, $\beta_2 = 0.3$, $n = 12$ c : Wishart, $\rho = 0.9$, $\beta_2 = 0.5$, $n = 6$ d : Wishart, $\rho = 0.9$, $\beta_2 = 0.3$, $n = 12$ 

FIG. 4.20 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\times-$; Ter Braek $-\text{+}-$ et Manly $-\dots*$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, X^1 et X^2 sont des distances interpoints et $\rho_1 = 0.9$. $n = 6$, $\beta_1 = 0.5$ pour les figures a et c. $n = 12$, $\beta_1 = 0.3$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

B1-4 Comparaison des données générées pour le terme d'erreur où $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) \geq 4$ et $\text{Var}(\mathbf{D}_R\boldsymbol{\beta})/\text{Var}(\epsilon) < 4$

L'objet de cette partie est de comparer par des résultats de simulations la puissance de la statistique de test T de Student pour tester l'hypothèse alternative bilatérale dans deux cas distincts : les données générées pour les variables X^1 , X^2 et Y sont des distances interpoints telles que $\text{Var}(\beta_1 X^1 + \beta_2 X^2)/\text{Var}(\epsilon) \geq 4$ puis $\text{Var}(\beta_1 X^1 + \beta_2 X^2)/\text{Var}(\epsilon) < 4$.

En considérant les figures 4.21 et 4.22 on peut faire les observations suivantes :

1. Lorsque, la variance du terme d'erreur est quatre fois plus grande que celle du terme de signal 14.21-c,d (respectivement 14.22-c,d), la puissance diminue fortement par rapport au cas où la variance du terme de signal est quatre fois plus grande que celle du terme d'erreur 14.21-a,b (respectivement 14.22-a,b).
2. Pour toutes les combinaison des facteurs, nous observons qu'en augmentant le taux de dépendance entre prédicteurs ρ , lorsque les valeurs des autres facteurs sont fixées, la puissance du test diminue.
3. Nous observons également qu'en augmentant la valeur de coefficient de régression partielle β_1 , la puissance diminue.
4. Les résultats présentés dans cette section correspondent au cas d'hypothèses alternatives bilatérales. Mais nous avons obtenu des résultats analogues pour le cas d'hypothèses alternatives unilatérales.
5. Ces résultats sont obtenus lorsque X^1 et X^2 sont des distances interpoint, mais nous avons obtenus des résultats très semblable à partir des carrés de distances et des transformées logarithmiques de distances.
6. Nous avons présenté ici le cas de la statistique T de Student. Mais avons obtenu les résultats correspondants en utilisant les deux statistiques de test : le coefficient de corrélation partielle ($r_{Y2,1}$) et le carré du coefficient de corrélation partielle ($r_{Y2,1}^2$) (cf. LABSAD en ligne).

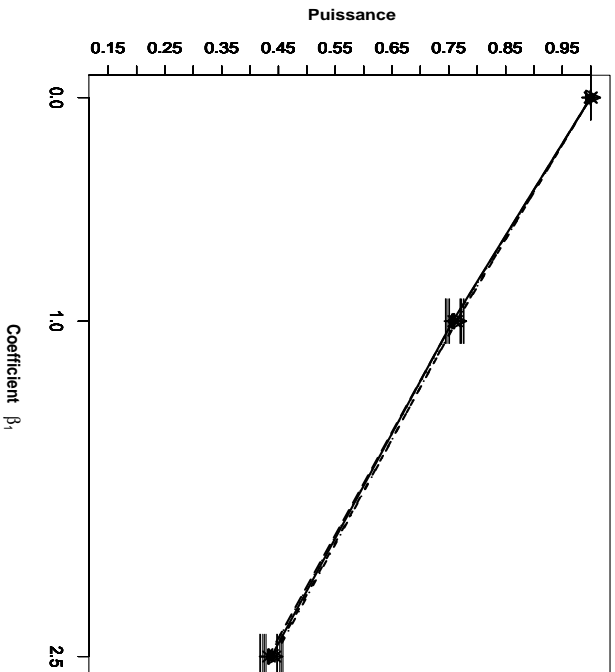
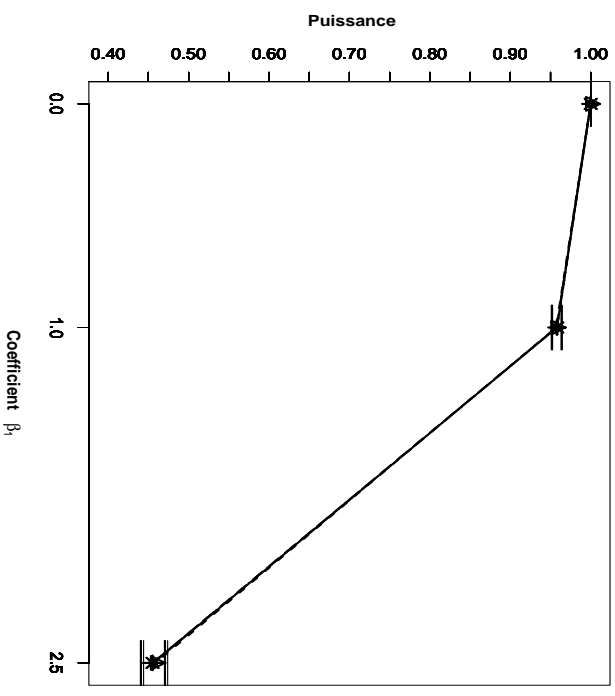
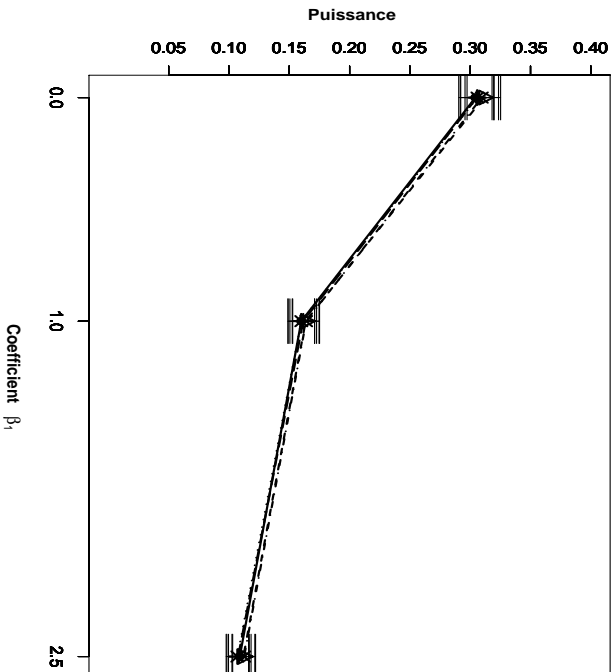
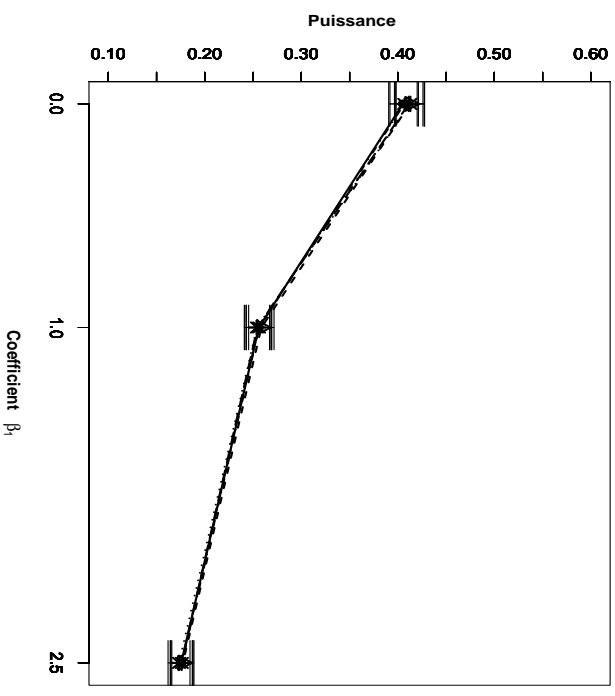
a : Wishart, $\rho = 0.1$, $\beta_2 = 0.5$, $n = 6$ b : Wishart, $\rho = 0.1$, $\beta_2 = 0.3$, $n = 12$ c : Wishart, $\rho = 0.1$, $\beta_2 = 0.5$, $n = 6$ d : Wishart, $\rho = 0.1$, $\beta_2 = 0.3$, $n = 12$ 

FIG. 4.21 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\triangle-$; Freedman et Lane $-\bullet-$; Kennedy $-\times-$; Ter Braek $-\dagger-$ et Manly $\dots*$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de test est le T de Student et $\rho_1 = 0.1$. $\beta_2 = 0.5$ pour les figures a et c. $\beta_2 = 0.3$ pour les figures b et d. Les données générées pour les variables X^1 , X^2 et Y sont telles que $\text{Var}(\beta_1 X^1 + \beta_2 X^2) / \text{Var}(\epsilon) \geq 4$ dans les figures a et b, $\text{Var}(\beta_1 X^1 + \beta_2 X^2) / \text{Var}(\epsilon) < 4$, dans les figures c et d

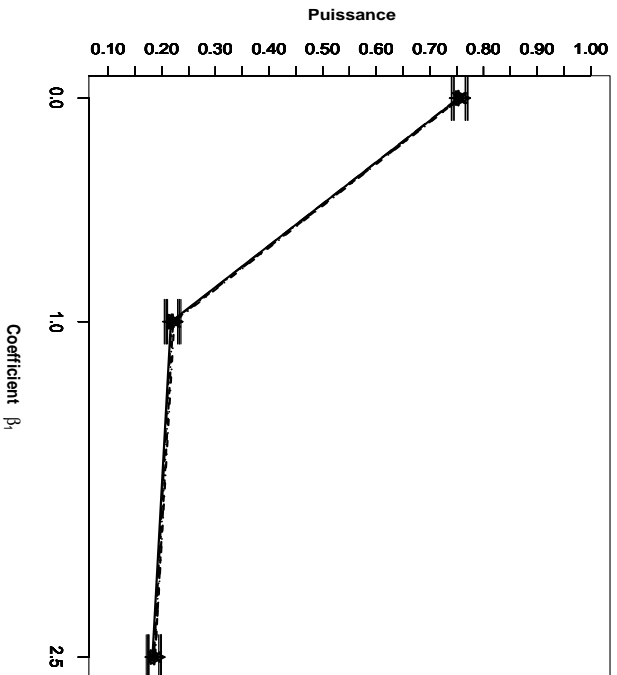
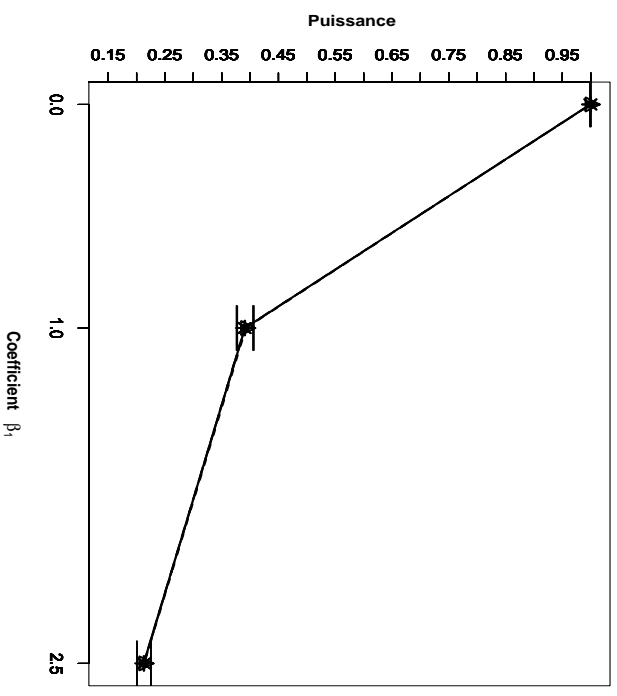
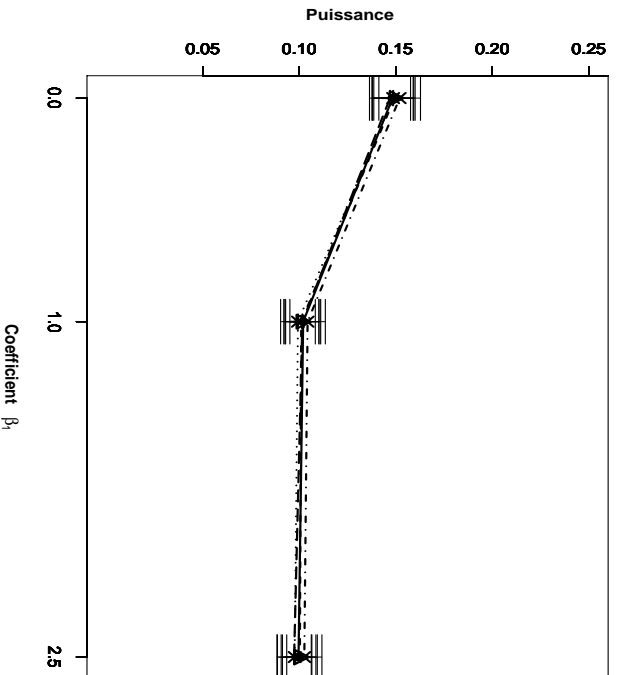
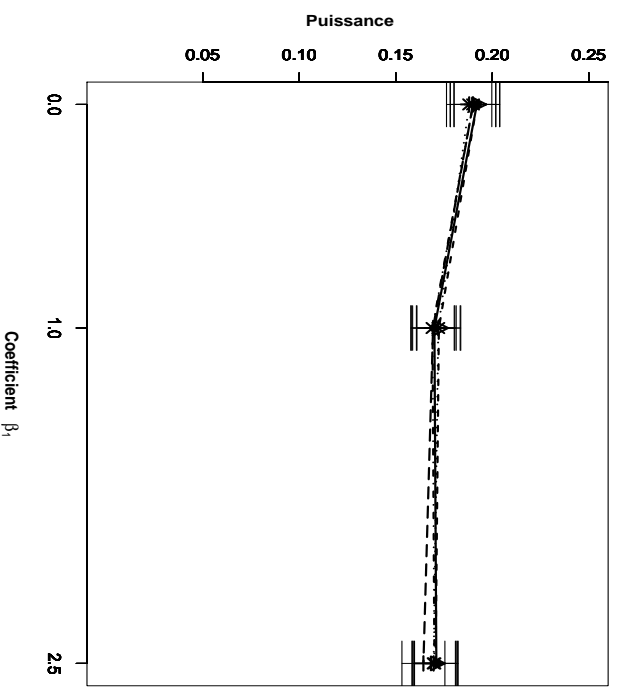
a : Wishart, $\rho = 0.9$, $\beta_2 = 0.5$, $n = 6$ b : Wishart, $\rho = 0.9$, $\beta_2 = 0.5$, $n = 12$ c : Wishart, $\rho = 0.9$, $\beta_2 = 0.5$, $n = 6$ d : Wishart, $\rho = 0.9$, $\beta_2 = 0.5$, $n = 12$ 

FIG. 4.22 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot \times -$; Ter Braek $-\text{+}-$ et Manly $\dots * \dots$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de Wishart et vérifient les propriétés d'une matrice de distance. Pour toutes les figures, X^1 et X^2 sont des distances interpoints, la statistique de test est le T de Student et $\rho_1 = 0.9$. $\beta_2 = 0.5$ pour les figures a et c. $\beta_2 = 0.3$ pour les figures b et d. Les données générées pour les variables X^1 , X^2 et Y sont telles que $\text{Var}(\beta_1 X^1 + \beta_2 X^2) / \text{Var}(\epsilon) \geq 4$ dans les figures a et b, $\text{Var}(\beta_1 X^1 + \beta_2 X^2) / \text{Var}(\epsilon) < 4$, dans les figures c et d.

B2) Cas des données générées pour le terme d'erreur selon la loi $Beta(2, 0.3)$

Cette partie présente trois comparaisons : la première concerne les transformations des distances simulées, la deuxième étudie l'influence du choix d'une statistique de test, la dernière oppose les alternatives bilatérales et unilatérales.

B2-1 Comparaison des trois ensembles des données : distance inter-point, carré de distances et transformées logarithmiques des distances

Le cadre des simulations est ici identique à la section (B1 – 1), seule la loi Wishart est remplacée par loi $Beta(2, 0.3)$.

En considérant les figures 4.23 et 4.24 on peut tirer les conclusions suivantes :

1. Contrairement au cas où les données du terme d'erreur sont générées à partir d'une distribution de Wishart, la puissance du test ne diminue pas lorsque le coefficient de régression partielle β_1 augmente.
2. Pour toutes les combinaisons des facteurs, la puissance est plus élevée lorsque les variables X_1 et X_2 sont les carrés de distances que lorsque les données X_1 et X_2 sont les distances interpoints initiales ou bien des transformées logarithmiques de distances.
3. Contrairement au cas précédent, dans la plupart des cas, en augmentant le coefficient ρ on n'observe pas de diminution de la puissance du test.
4. Les résultats obtenus montrent qu'ici encore la puissance des quatre méthodes de permutation ainsi que celle de la méthode classique sont très proches.
5. Nous avons présenté dans les figures suivantes la statistique du T de Student. Nous avons obtenu des résultats semblables en utilisant les statistiques $r_{Y2.1}$ et $r_{Y2.1}^2$.
6. Nous avons de plus illustré la situation dans le cas d'une hypothèse alternative bilatérale. Nous avons obtenu des résultats analogue en utilisant une hypothèse alternative unilatérale (cf. LABSAD en ligne).

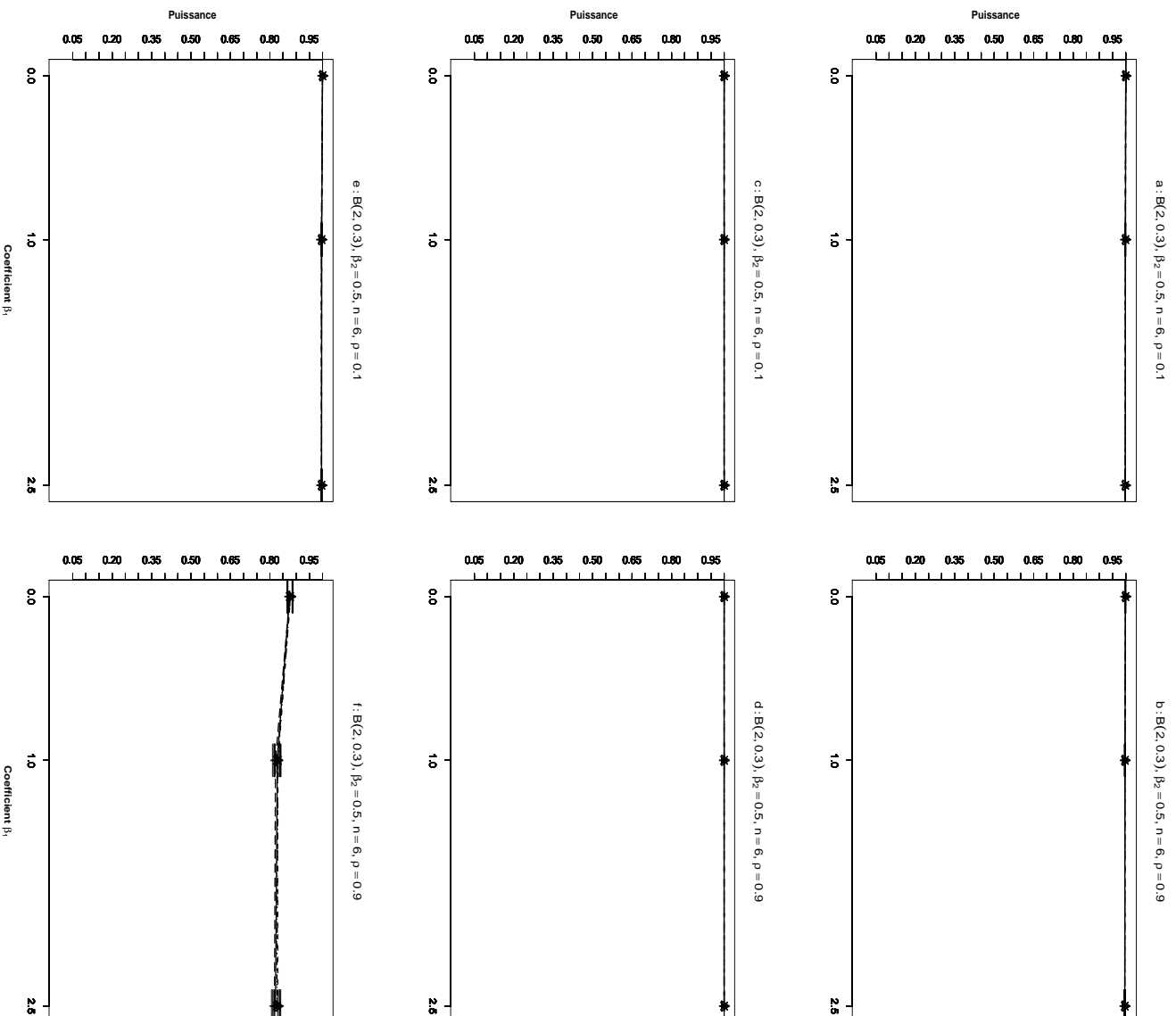


FIG. 4.23 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot \times \cdot-$; Ter Braek $-\text{+}-$ et Manly $\dots * \dots$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, $\beta_2 = 0.5$ et $n = 6$. $\rho = 0.1$ pour les figures *a*, *c* et *e*. $\rho = 0.9$ pour les figures *b*, *d* et *f*. X^1 et X^2 sont des distances interpoints pour les figures *a* et *b*; des carrés des distances pour les figures *c* et *d*; des transformées logarithmiques des distances pour les figures *e* et *f*.

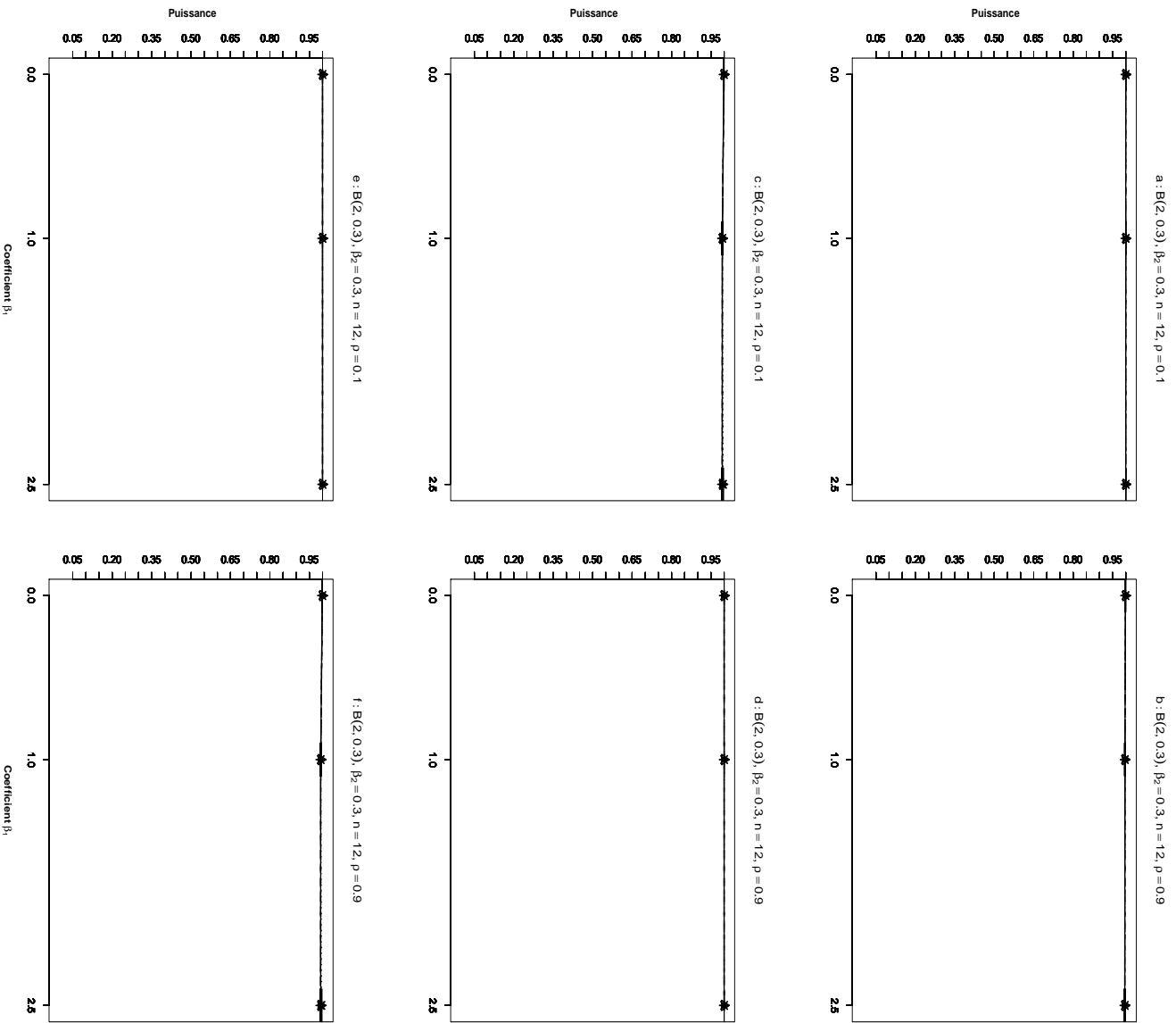


FIG. 4.24 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $--\Delta--$; Freedman et Lane $—\bullet—$; Kennedy $-\cdot\times\cdot-$; Ter Brack $—+—$ et Manly $\dots*\dots$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, $\beta_2 = 0.3$ et $n = 12$. $\rho = 0.1$ pour les figures a , c et e . $\rho = 0.9$ pour les figures b , d et f . X^1 et X^2 sont des distances interpolés pour les figures a et b ; des carrés des distances pour les figures c et d ; des transformées logarithmiques des distances pour les figures e et f .

B2-2 Comparaison des trois statistiques de tests

Le cadre des simulations est ici identique à la section (B1 – 2), seule la loi Wishart est remplacée par loi $Beta(2, 0.3)$.

En considérant les figures 4.25 et 4.26 on peut faire les observations suivantes :

1. La comparaison des figures 4.25 et 4.26, nous permet de constater que la puissance augmente lorsqu'on augmente le nombre d'unités n pour toutes les combinaisons des facteurs.
2. Les résultats confirment qu'il n'existe pas de différence entre les statistiques T de Student, $r_{Y_{2.1}}^2$ et $r_{Y_{2.1}}^2$.
3. Les figures présentées montrent qu'une augmentation de ρ diminue un peu la puissance.
4. Les variables X_1 et X_2 sont ici les distances interpoints. On note encore que nous avons obtenu des résultats très proches pour les carrés de distances et les transformées logarithmiques de distances.
5. Nous avons obtenu ces résultats dans le cas d'hypothèses bilatérales. Les simulations faites confirment également la validité de ces résultats dans le cas d'hypothèses alternatives bilatérales (cf. LABSAD en ligne).

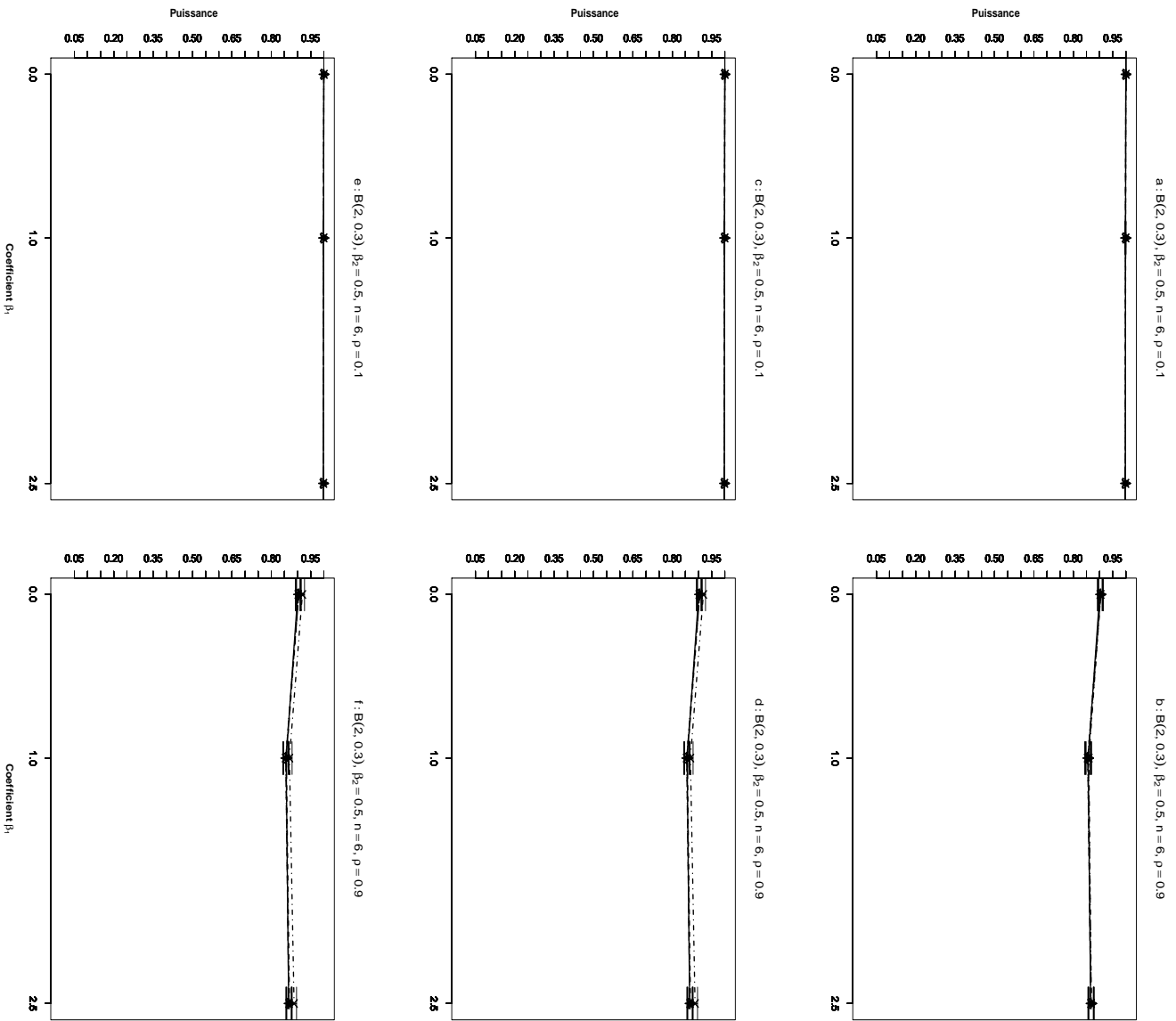


FIG. 4.25 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $--\Delta--$; Freedman et Lane $---\bullet---$; Kennedy $-\cdot\times\cdot-$; Ter Brack $---+$ et Manly $\dots*\dots$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. Pour toute les figures $\beta_2 = 0.5$ et $n = 6$. $\rho = 0.1$ pour les figures a, c et e. $\rho = 0.9$ pour les figures b, d et f. X^1 et X^2 sont des distances interpolés pour toutes les figures. La statistique de test est : T de Student pour les figures a et b; le coefficient de corrélation partielle pour les figures c et d; carrée du coefficient de corrélation partielle pour les figures e et f.

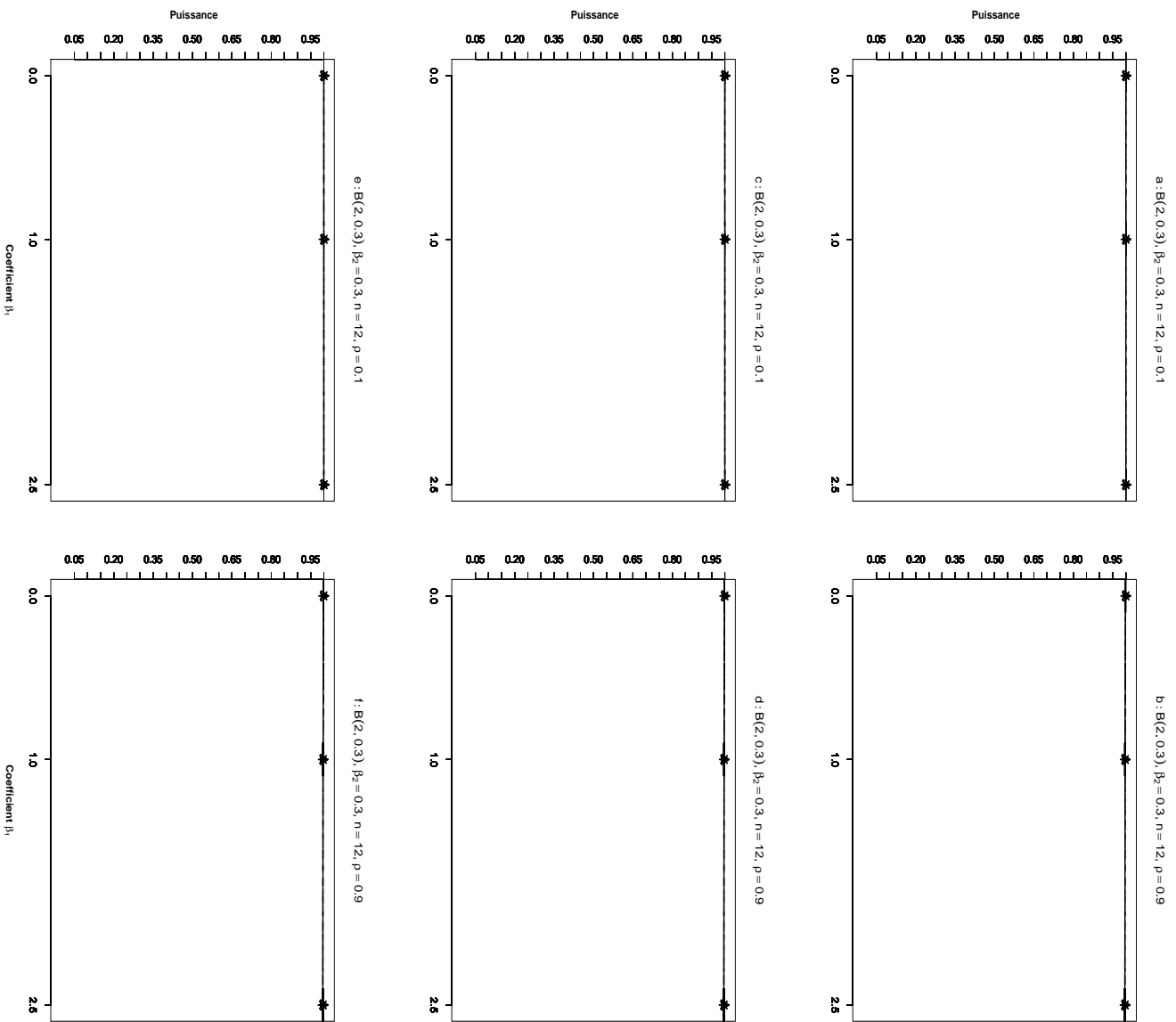


FIG. 4.26 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot \times \cdot-$; Ter Braek $-\text{+}-$ et Manly $\dots * \dots$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. Pour toute les figures $\beta_2 = 0.3$ et $n = 12$. $\rho = 0.1$ pour les figures a, c et e. $\rho = 0.9$ pour les figures b, d et f. X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est : T de Student pour les figures a et b; le coefficient de corrélation partielle pour les figures c et d; carrée du coefficient de corrélation partielle pour les figures e et f.

B2-3 Comparaison des hypothèses alternatives bilatérale et unilatérale

Le cadre des simulations est ici identique à la section (B1 – 3), seule la loi Wishart est remplacée par loi $Beta(2, 0.3)$.

En considérant les figures 4.27 et 4.28 on peut faire les observations suivantes :

1. Nous constatons par comparaison des figures 4.27-a,b (respectivement 4.28-a,b) avec 4.27-c,d (4.28c,d) qu'il n'existe pas de différence entre les puissances des tests unilatéral et bilatéral.
2. L'augmentation du nombre d'unités de $n = 6$ à $n = 12$ produit une augmentation de la puissance.
3. Ces résultats ont été obtenus lorsque X^1 et X^2 sont des distances interpoints. Mais nous avons obtenu des résultats analogues à partir des carrés des distances et des transformées logarithmiques des distances.
4. Nous avons présenté ici le cas de la statistique T de Student. Mais nous avons obtenu des résultats de simulation très voisins en utilisant les deux autres statistiques de test : le coefficient de corrélation partielle ($r_{Y2.1}$) et le carré du coefficient de corrélation partielle ($r_{Y2.1}^2$) (cf. LABSAD en ligne).

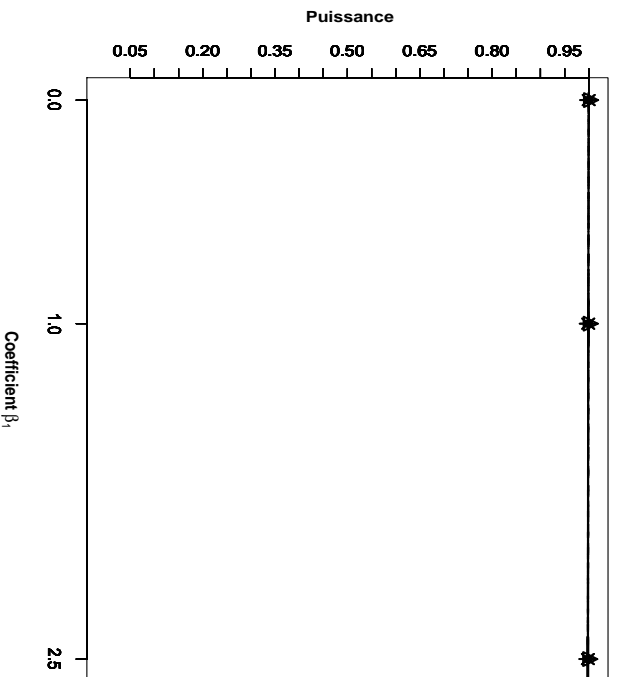
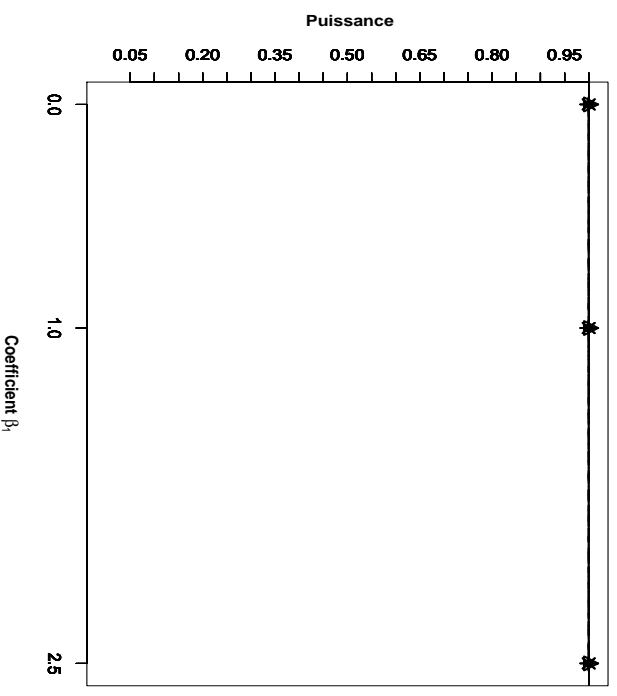
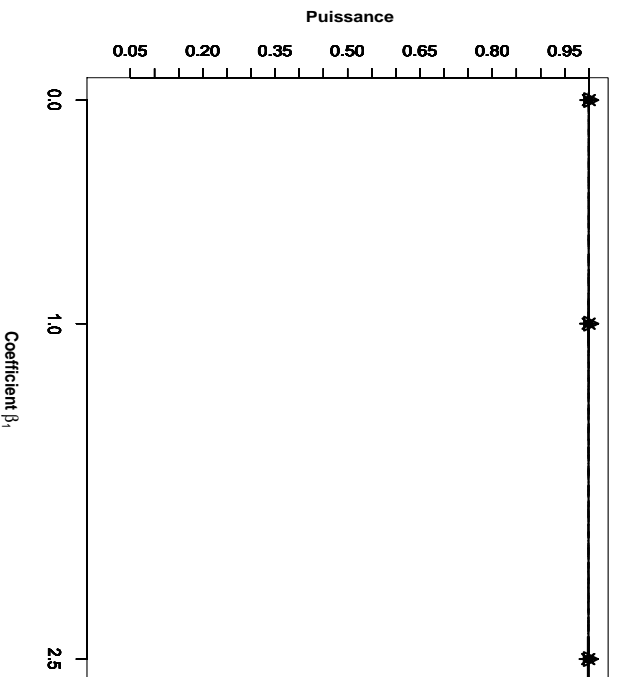
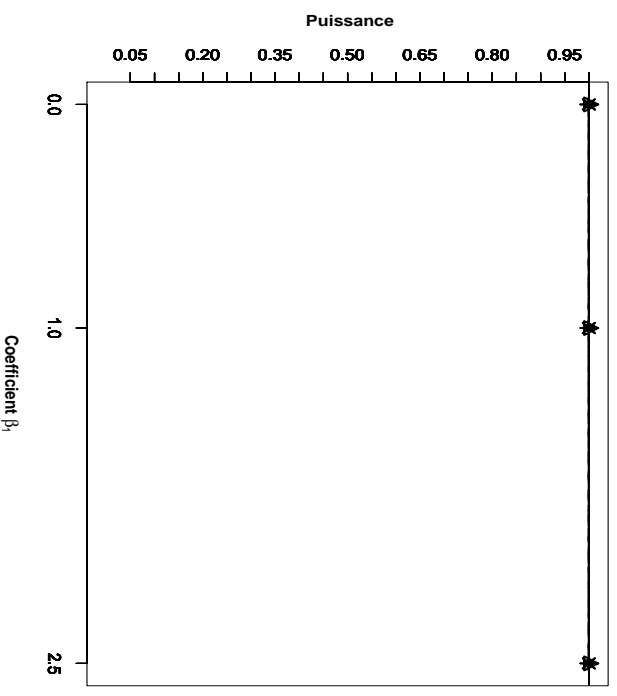
a : $B(2, 0.3)$, $\rho = 0.1$, $\beta_2 = 0.5$, $n = 6$ b : $B(2, 0.3)$, $\rho = 0.1$, $\beta_2 = 0.3$, $n = 12$ c : $B(2, 0.3)$, $\rho = 0.1$, $\beta_2 = 0.5$, $n = 6$ d : $B(2, 0.3)$, $\rho = 0.1$, $\beta_2 = 0.3$, $n = 12$ 

FIG. 4.27 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot \times \cdot-$; Ter Braek $-\text{+}-$ et Manly $\dots * \dots$. 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, X^1 et X^2 sont des distances interpoints et $\rho_1 = 0.1$. $n = 6$, $\beta_1 = 0.5$ pour les figures a et c. $n = 12$, $\beta_1 = 0.3$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

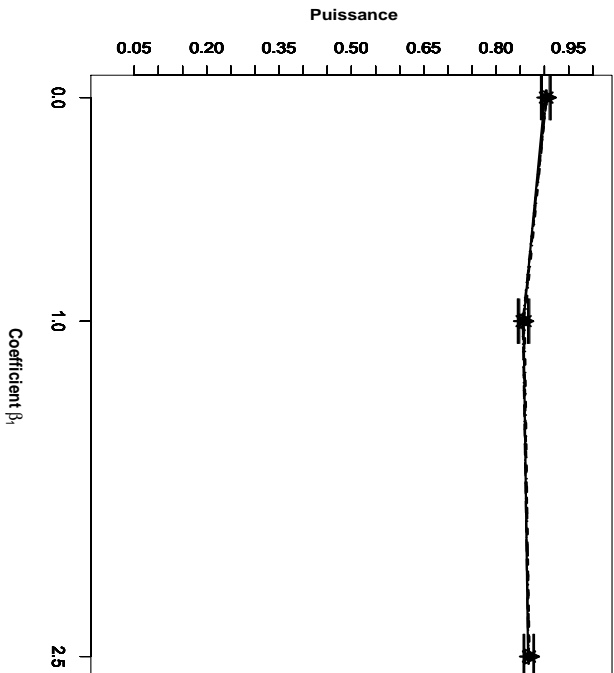
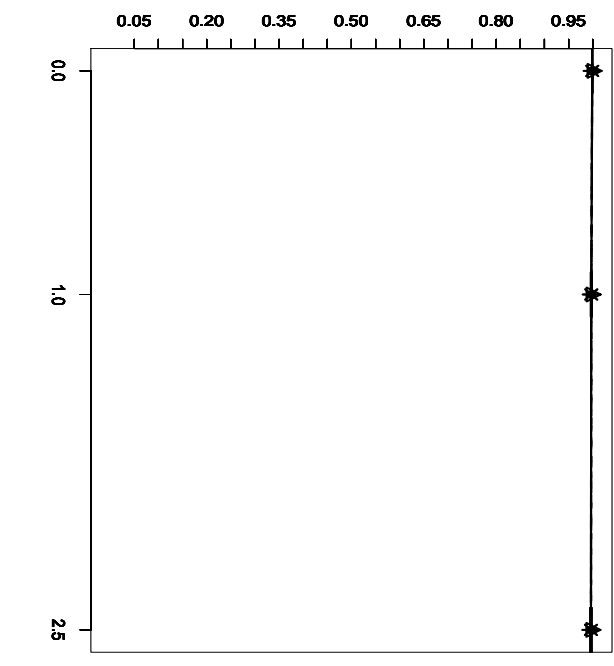
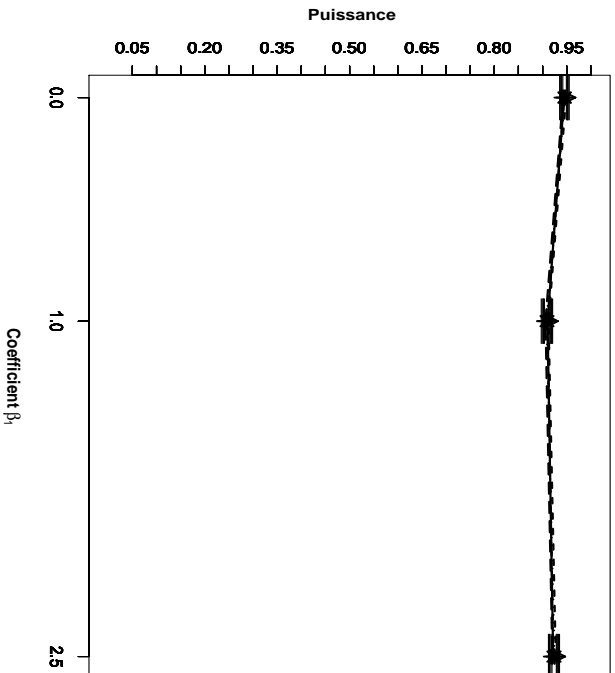
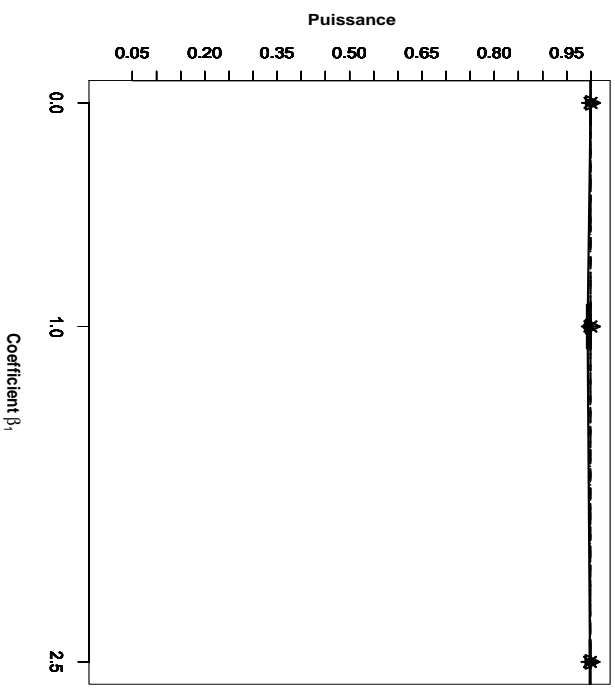
a: $B(2, 0.3)$, $\rho = 0.9$, $\beta_2 = 0.5$, $n = 6$ b: $B(2, 0.3)$, $\rho = 0.9$, $\beta_2 = 0.3$, $n = 12$ c: $B(2, 0.3)$, $\rho = 0.9$, $\beta_2 = 0.5$, $n = 6$ d: $B(2, 0.3)$, $\rho = 0.9$, $\beta_2 = 0.3$, $n = 12$ 

FIG. 4.28 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot \times \cdot-$; Ter Braek $-\cdot + \cdot-$ et Manly $\dots * \dots$. 30000 n-échantillons du résidu ϵ ont été simulées suivant une loi de $Beta(2, 0.3)$ et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, X^1 et X^2 sont des distances interpolés et $\rho_1 = 0.9$. $n = 6$, $\beta_1 = 0.5$ pour les figures a et c. $n = 12$, $\beta_1 = 0.3$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

Remarque 4.3.1. *Dans le cas de termes d'erreur générés à partir de la distribution exponentielle $Exp(2)$, nous avons obtenu des résultats de simulations très semblables à ceux qui résultent du choix de la distribution $Beta(2, 0.3)$. C'est pourquoi nous avons reporté la présentation de ces résultats en annexe 4. Cette analyse est un peu étonnante et déçoit nos a priori puisque les distributions $Exp(2)$ et $Beta(2, 0.3)$ présentent respectivement une dissymétrie gauche et une dissymétrie droite. Elles favorisent donc de façon opposée la présence de petites distances et de grandes dans l'échantillon. Nous attendions en conséquence des performances différentes des tests de permutation dans ces deux cas.*

C) Conclusion

Au bilan les quatre apports essentiels de ce chapitre nous semblent être :

1. Les résultats de nos simulations montrent de fait des différences avec les résultats présentés par Legendre [35]. Il nous paraît difficile de déterminer les raisons exactes de ces différences, parce que cet auteur ne dit pas clairement la méthode de simulation qu'il a retenue et en particulier s'il a bien simulé des tableaux de distance. Pour notre part, nous avons fait une première simulation qui ne garantissait pas que les données satisfassent les axiomes d'une distance. Cette simulation est présentée en partie (**D**) ci-dessous. C'est sur la deuxième simulation, respectant cette contrainte et donc plus complexe que repose les résultats présentés dans ce chapitre.

La comparaison des deux simulations, avec et sans contrainte distancielle, nous a permis de constater des différences notables, qui pourrait expliquer certaines des différences avec les simulations de Legendre.

2. Les résultats de nos simulations de tableaux de distances sont à l'évidence très différents de ceux que nous avons présentés au chapitre 2 et qui correspondent à une situation de régression multiple standard, car opérant sur des variables ordinaires, non dyadiques, et donc non soumises à une dépendance structurelle entre composantes des vecteurs observés.

Dans ce dernier cas, les résultats de nos simulations semblent plus en accord avec les constats rapportés par Legendre ([35]) et par Andersson-Legendre ([1]).

3. La situation nous paraît beaucoup plus difficile à maîtriser dans le cas de données dyadiques que dans le cas ordinaire. En nous appuyant sur nos simulations, nous ne pouvons pas déclarer quelle méthode du test de permutation est la meilleure. Nous ne pouvons pas non plus dire quelle méthode fonctionne mieux dans tel ou tel cas. En réalité, les résultats de nos simulations semblent indiquer que toutes les méthodes appliquées dans ce contexte fonctionnent de manière identique. Le seul point remarquable (et un peu étonnant) tient au cas des données du terme d'erreur générées suivant une loi de Wishart. Nous

avons en effet observé dans ce cas une augmentation importante de l'erreur de première espèce pour les quatre méthodes de test. D'autre part, un apport non anticipé de nos simulations de tableaux de distances tient au constat de la sensibilité des fonctions puissance obtenues par une procédure de test de permutation qui résulte tant d'une variation du coefficient de régression partielle β_1 que de celle du coefficient de corrélation entre matrices de distance prédicatrices.

4. Au bilan, l'étude comparative des méthodes de tests de permutation sur matrices de distances est beaucoup plus complexe que celle qui traite le cas des variables ordinaires. Il nous semble qu'une réelle compréhension des facteurs influents n'est pas encore acquise et nos simulations pèchent en conséquence par une imprécision certaine. Tant des études analytiques que de nouvelles simulations sont nécessaires pour lever toutes les zones d'ombre qui accompagnent ce problème actuellement.

D) La première simulation

Les résultats précédents ont été établis à partir de simulations générant des données qui respectent les axiomes de définition d'une distance. Dans une étude préliminaire, nous avons réalisé une simulation dans laquelle nous n'avons pas vérifié les axiomes de définition d'une distance pour la variable Y . En effet, les valeurs générées pour le terme d'erreur ne vérifient pas forcément les propriétés d'une matrice de distance et les valeurs de Y qui sont calculées par : $\mathbf{Y} = \mathbf{W}_R \boldsymbol{\beta} + \boldsymbol{\epsilon}$ ne vérifient donc pas non plus nécessairement ces propriétés. Le cadre de cette simulation est :

D1) Analyse du risque d'erreur de première espèce du test

Nous avons étudié le risque d'erreur de première espèce sur les valeurs des paramètres ci-dessous :

- a) le nombre d'unités statistiques : $n = \{9, 18, 36, 54\}$;
- b) le coefficient associé à la covariable X^1 : $\beta_1 = \{0, 2.5\}$
- c) le coefficient de corrélation entre les variables explicatives X^1 et X^2 :
 $\rho(X^1, X^2) = \{0.1, 0.9\}$;
- d) la distribution du terme d'erreur aléatoire ϵ : $\{Student(1), LG(0, 1), Exp(2)\}$.

D2) Analyse de la puissance du test

Nous avons étudié la puissance du test sur les valeurs des paramètres ci-dessous :

- a) le nombre d'unités statistiques : $n = \{6, 9, 18\}$;
- b) le coefficient associé à la covariable X^1 : $\beta_1 = \{0, 2.5\}$
- c) le coefficient de corrélation entre les variables explicatives X^1 et X^2 :
 $\rho(X^1, X^2) = \{0.1, 0.9\}$;
- d) la distribution du terme d'erreur aléatoire ϵ : $\{Student(1), LG(0, 1), Exp(2)\}$.

Remarquons que les données générées pour les variables X^1 et X^2 sont différentes de celles que nous avons générées à l'aide de l'algorithme 4.3.1 ci-dessus. Dans le cas précédent, en effet les valeurs des vecteurs \mathbf{d}^1 et \mathbf{d}^2 sont :

$$\mathbf{d}^1 = \begin{bmatrix} d_1^1/s_1 \\ d_2^1/s_1 \\ \vdots \\ d_N^1/s_1 \end{bmatrix}, \quad \mathbf{d}^2 = \begin{bmatrix} d_1^2/s_2 \\ d_2^2/s_2 \\ \vdots \\ d_N^2/s_2 \end{bmatrix}$$

où $s_1 = \sum_{i=1}^n d_i^1$ et $s_2 = \sum_{i=1}^n d_i^2$. De plus au cours de cette simulation, nous avons utilisé comme statistique de test le T de Student, et l'hypothèse alternative retenue était bilatérale. Les résultats de cette simulation sont disponibles sur le site internet du LABSAD.

Annexe 1

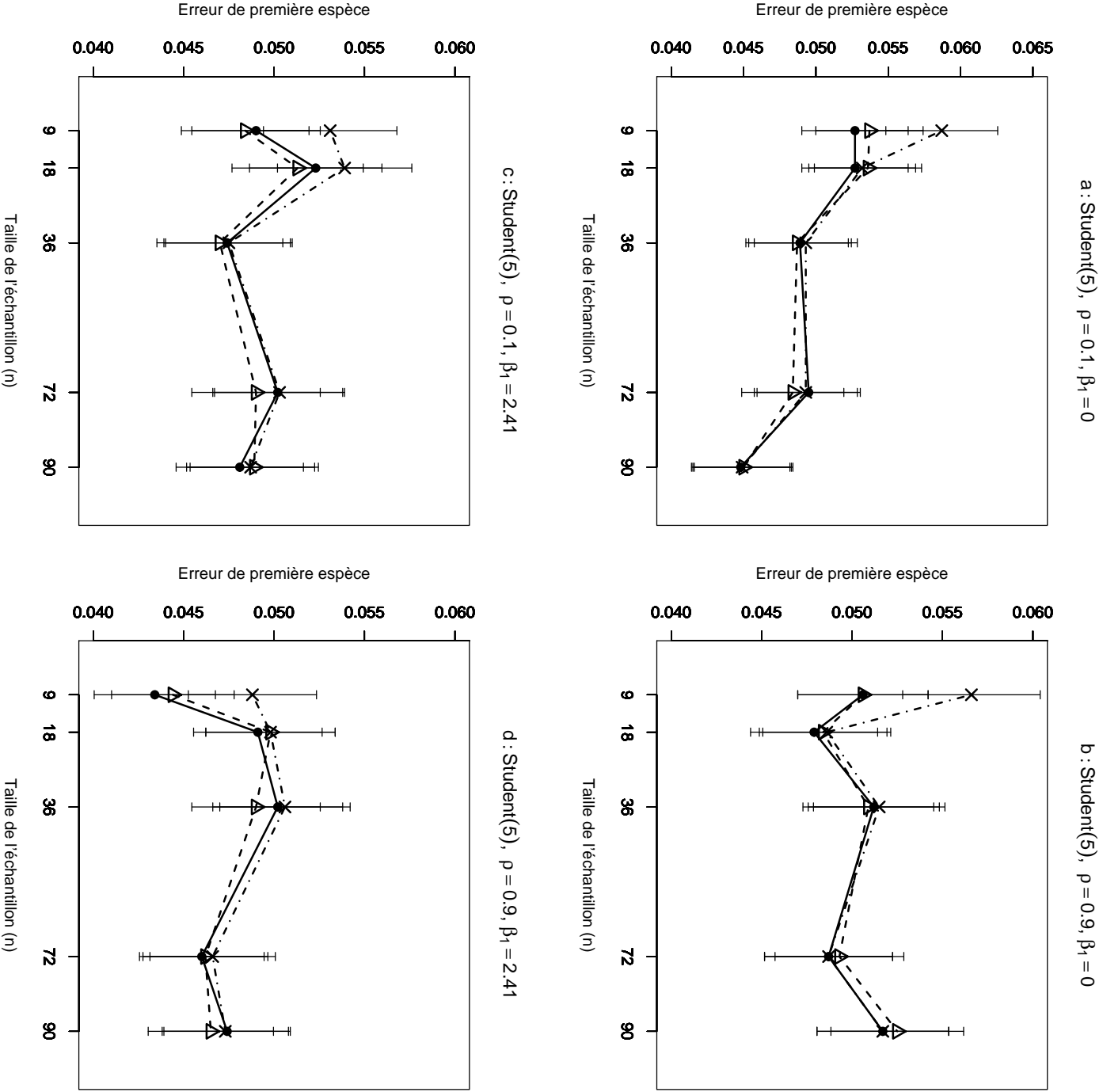


FIG. 4.29 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— Δ —, Freedman et Lane— \bullet — et Kennedy— \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Student(5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $n=9,18,36,72,90$. Les segments verticaux figurent l'intervalle de confiance à 95%.
 Figure a : $\rho = 0.1, \beta_1 = 0$; figure b : $\rho = 0.9, \beta_1 = 0$; figure c : $\rho = 0.1, \beta_1 = 2.41$, figure d : $\rho = 0.9, \beta_1 = 2.41$.

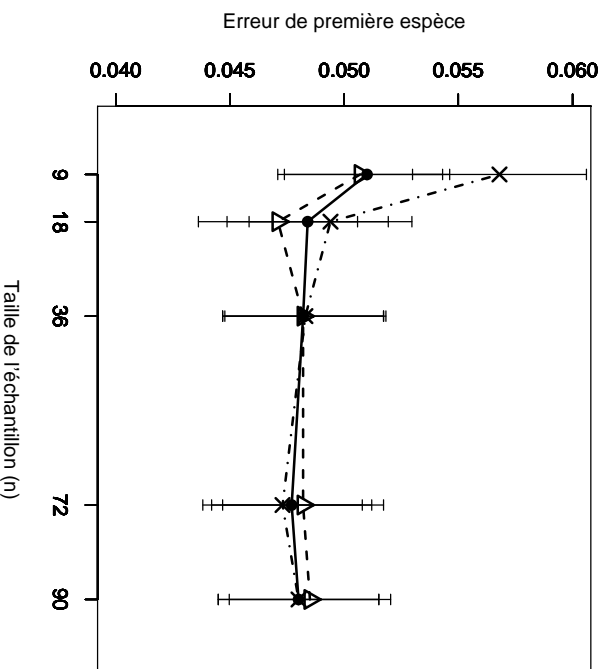
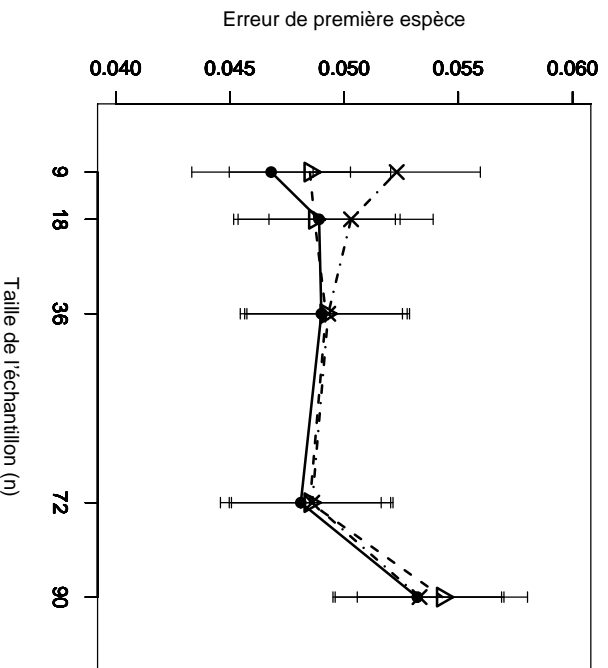
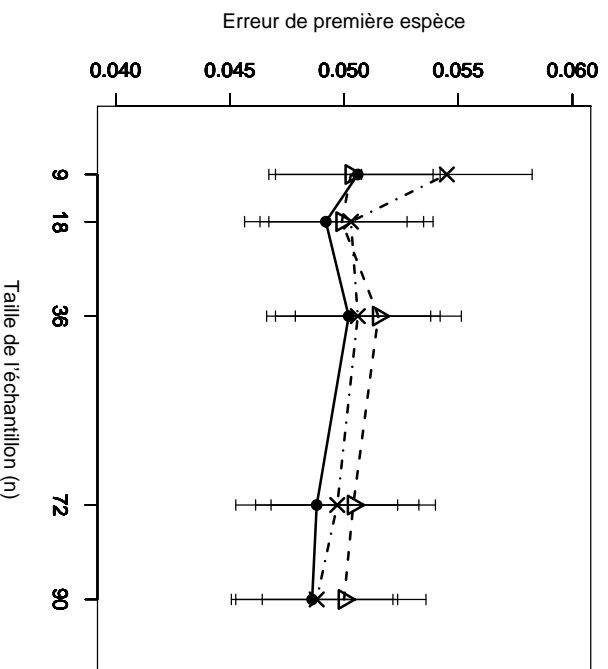
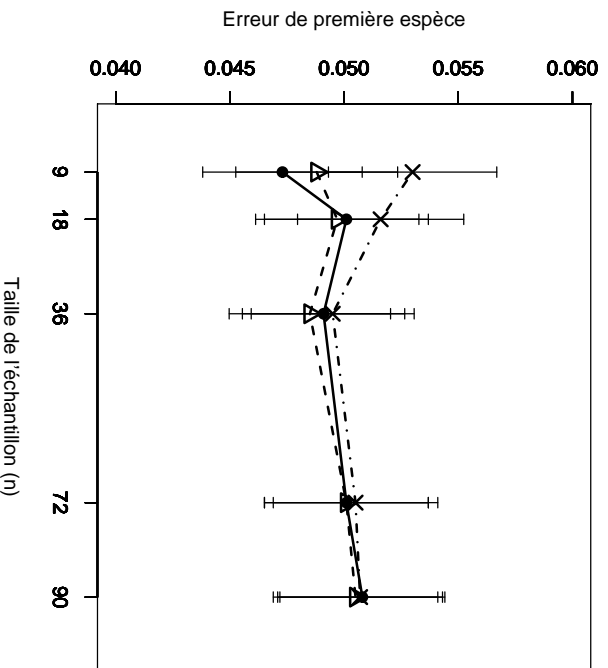


FIG. 4.30 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, Freedman et Lane— \bullet — et Kennedy— \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Exp(0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $n=9, 18, 36, 72, 90$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1, \beta_1 = 0$; figure b : $\rho = 0.9, \beta_1 = 0$; figure c : $\rho = 0.1, \beta_1 = 2.41$, figure d : $\rho = 0.9, \beta_1 = 2.41$.

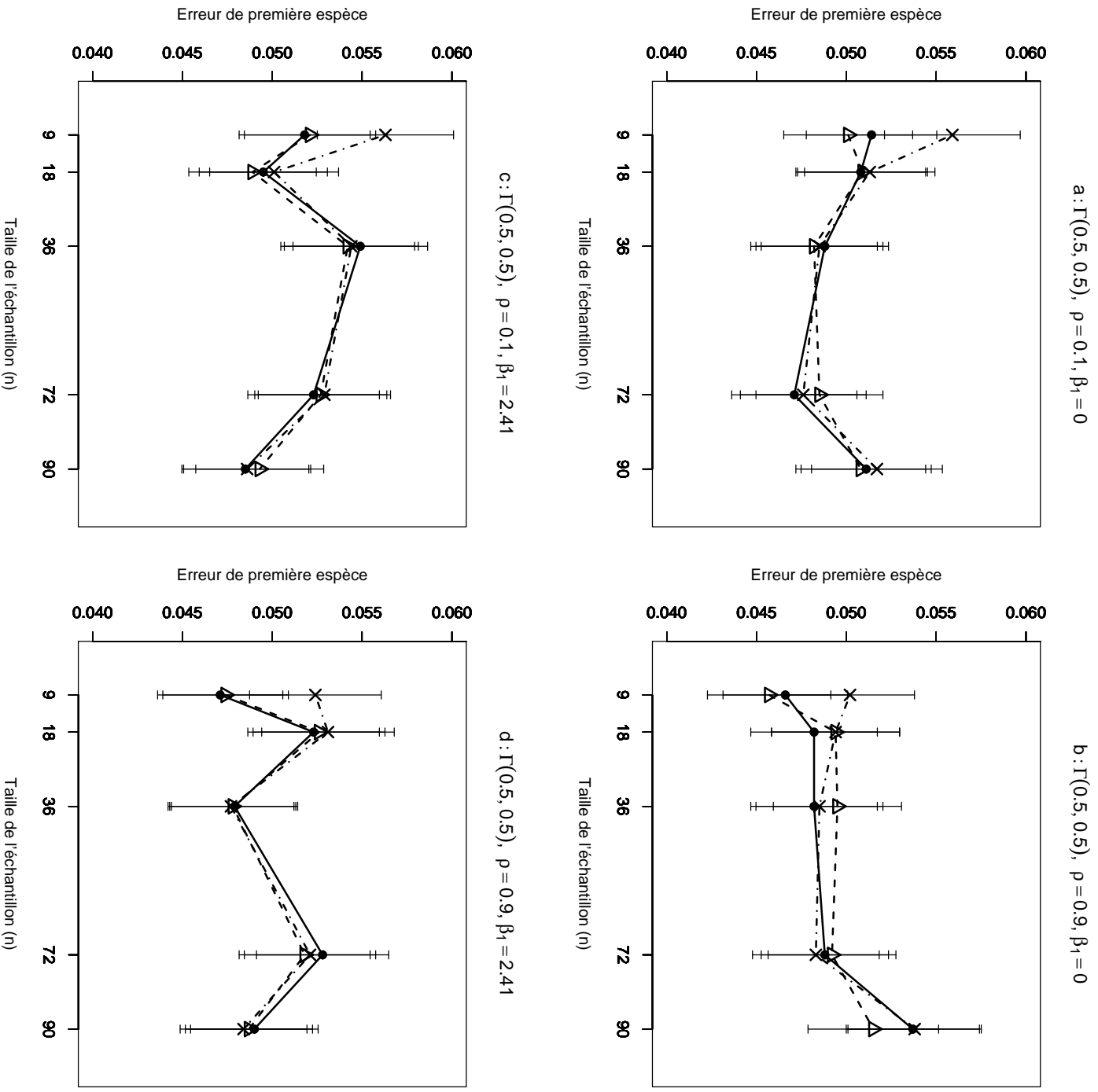


FIG. 4.31 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, Freedman et Lane— \bullet — et Kennedy— \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(0.5,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $n=9,18,36,72,90$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_1 = 0$; figure b : $\rho = 0.9$, $\beta_1 = 0$; figure c : $\rho = 0.1$, $\beta_1 = 2.41$, figure d : $\rho = 0.9$, $\beta_1 = 2.41$.

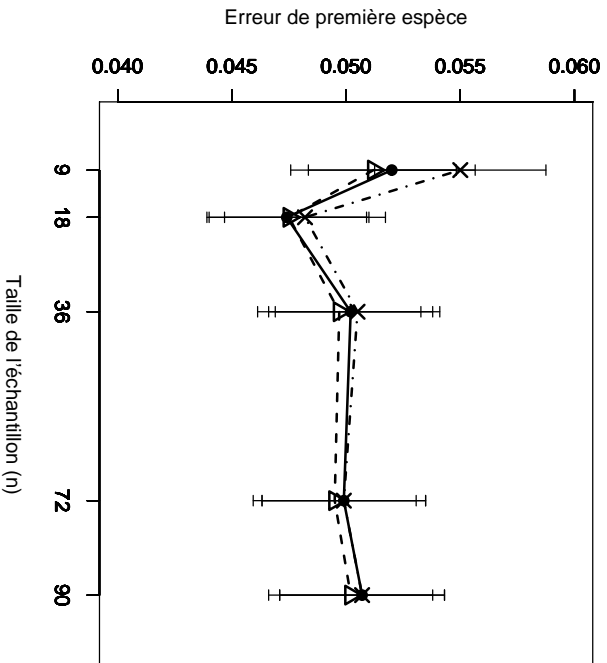
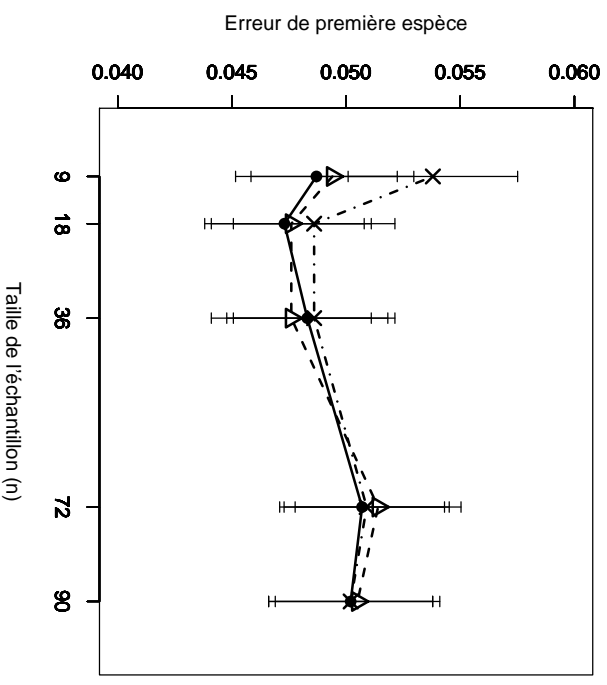
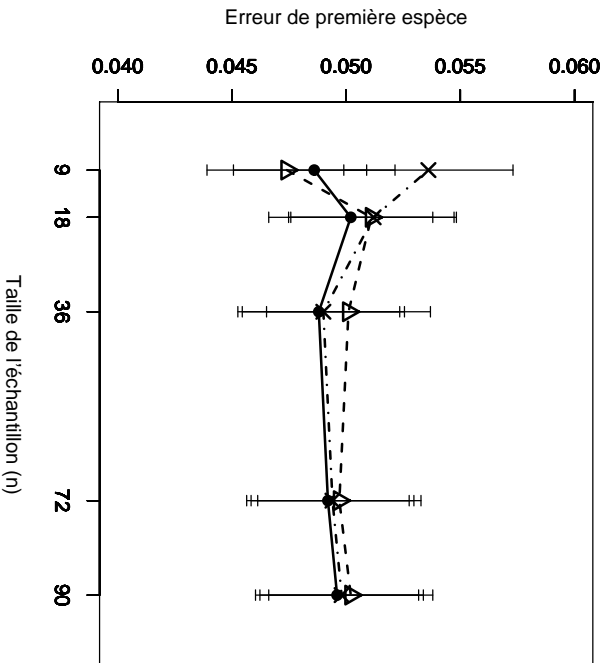
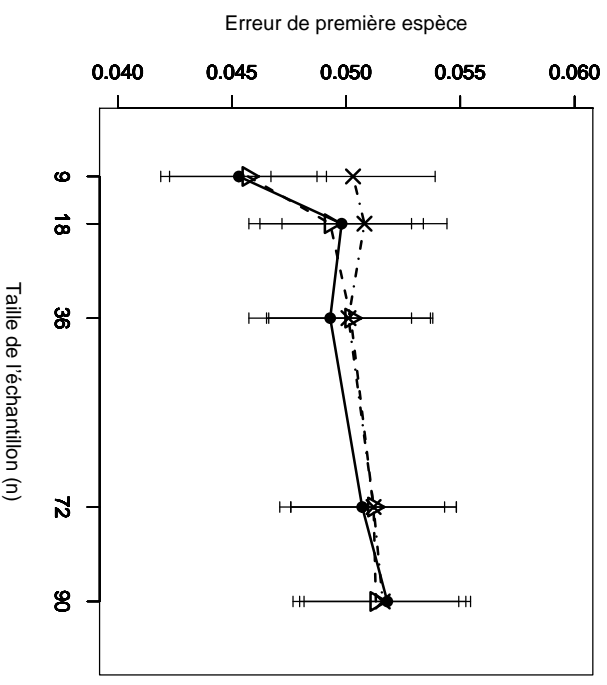
a : $\Gamma(0.5, 2)$, $\rho = 0.1$, $\beta_1 = 0$ b : $\Gamma(0.5, 2)$, $\rho = 0.9$, $\beta_1 = 0$ c : $\Gamma(0.5, 2)$, $\rho = 0.1$, $\beta_1 = 2.41$ d : $\Gamma(0.5, 2)$, $\rho = 0.9$, $\beta_1 = 2.41$ 

FIG. 4.32 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, Freedman et Lane— \bullet — et Kennedy— \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(0.5,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $n=9, 18, 36, 72, 90$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_1 = 0$; figure b : $\rho = 0.9$, $\beta_1 = 0$; figure c : $\rho = 0.1$, $\beta_1 = 2.41$, figure d : $\rho = 0.9$, $\beta_1 = 2.41$.

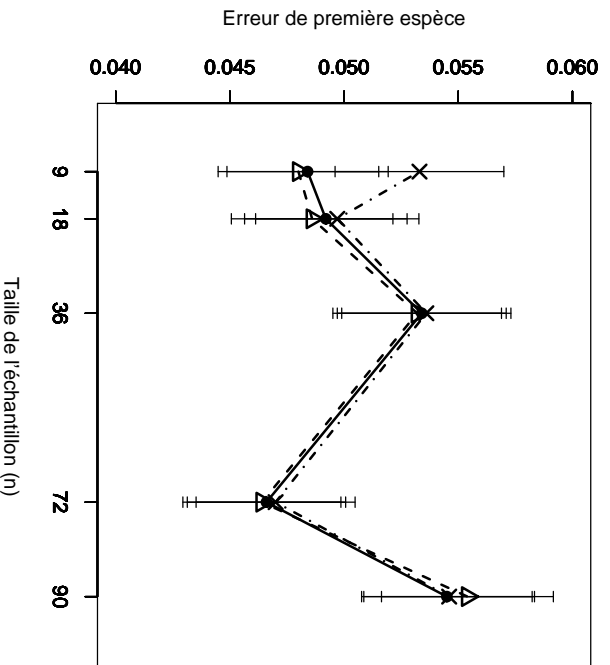
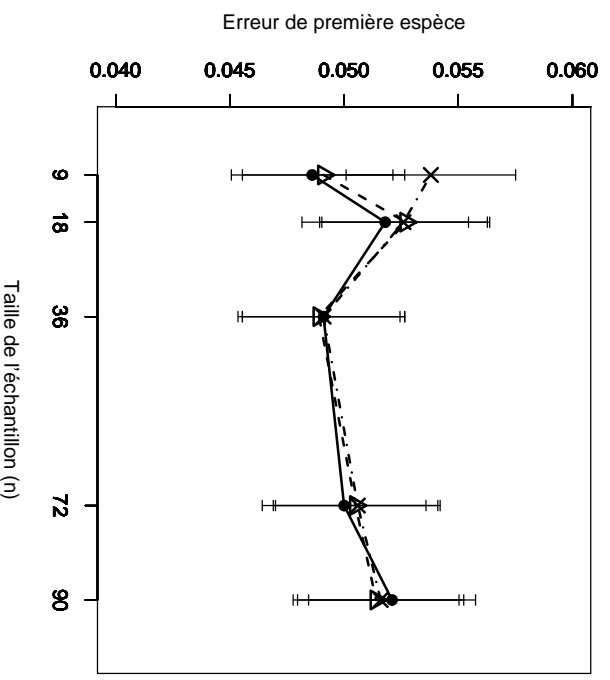
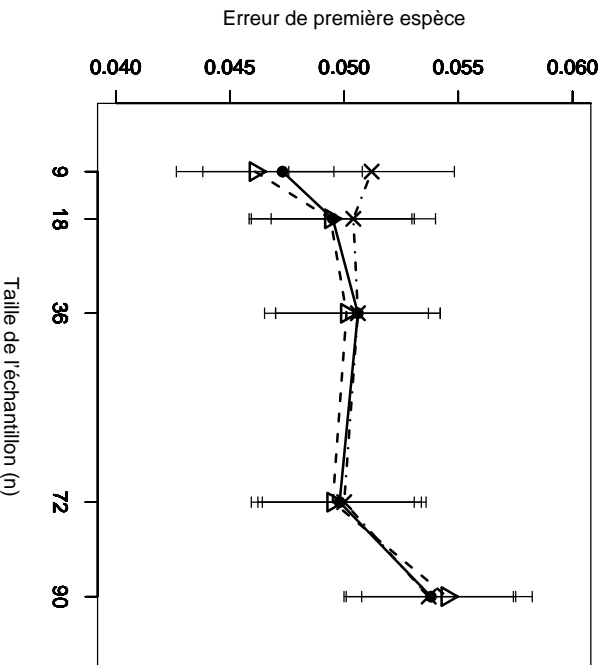
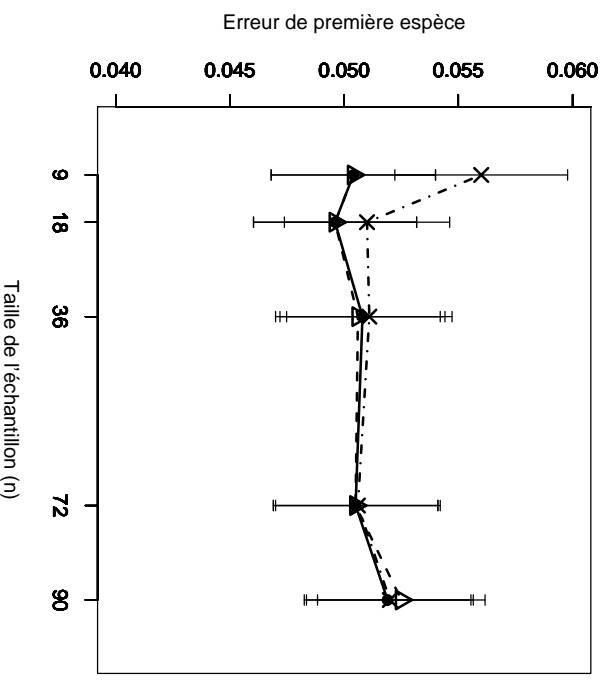
a : $\Gamma(1, 0.5)$, $\rho = 0.1$, $\beta_1 = 0$ b : $\Gamma(1, 0.5)$, $\rho = 0.9$, $\beta_1 = 0$ c : $\Gamma(1, 0.5)$, $\rho = 0.1$, $\beta_1 = 2.41$ d : $\Gamma(1, 0.5)$, $\rho = 0.9$, $\beta_1 = 2.41$ 

FIG. 4.33 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, Freedman et Lane— \bullet — et Kennedy— \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $n=9,18,36,72,90$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_1 = 0$; figure b : $\rho = 0.9$, $\beta_1 = 0$; figure c : $\rho = 0.1$, $\beta_1 = 2.41$, figure d : $\rho = 0.9$, $\beta_1 = 2.41$.

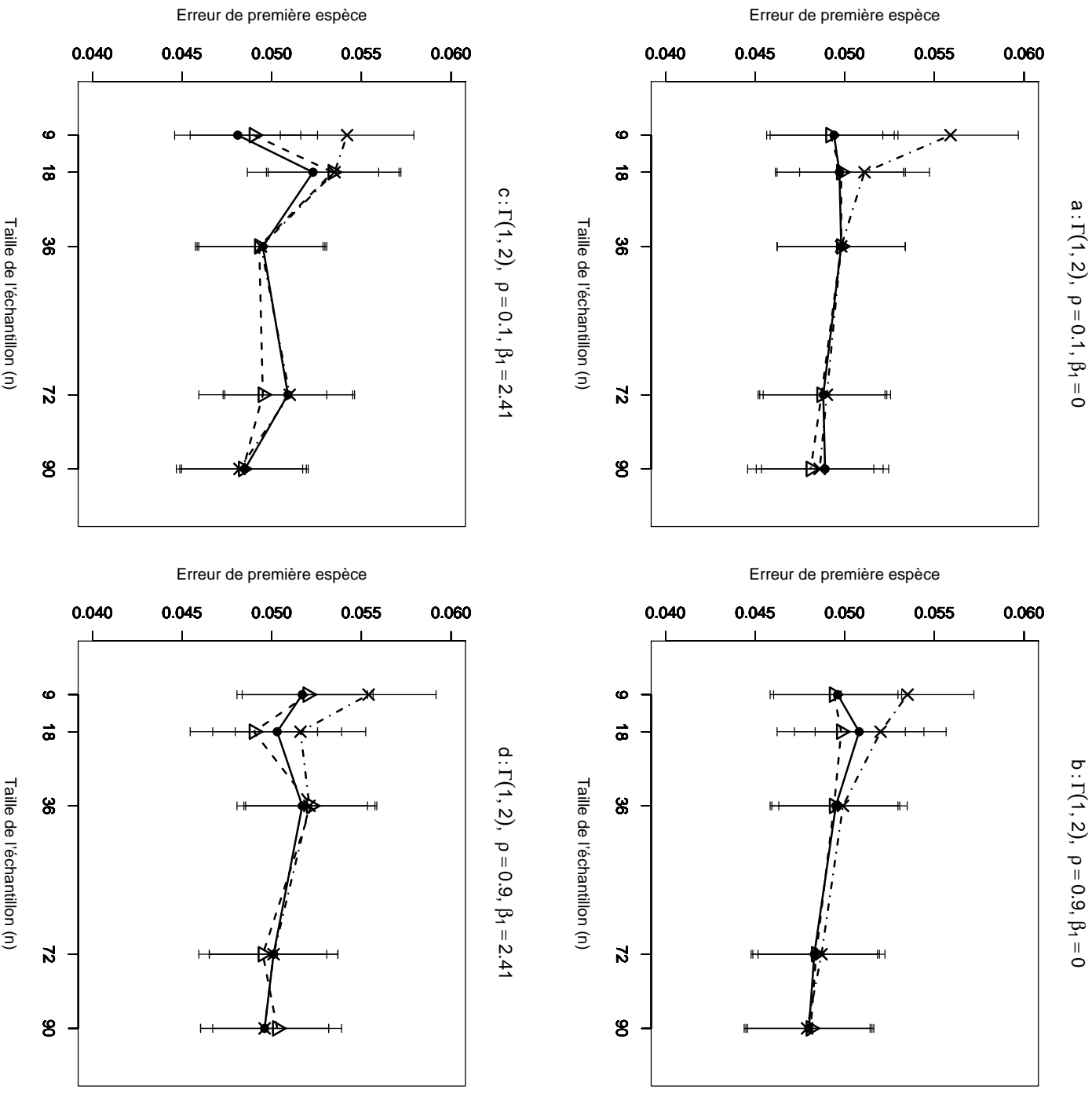


FIG. 4.34 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— \triangle —, Freedman et Lane— \bullet — et Kennedy— \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $n=9, 18, 36, 72, 90$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_1 = 0$; figure b : $\rho = 0.9$, $\beta_1 = 0$; figure c : $\rho = 0.1$, $\beta_1 = 2.41$, figure d : $\rho = 0.9$, $\beta_1 = 2.41$.

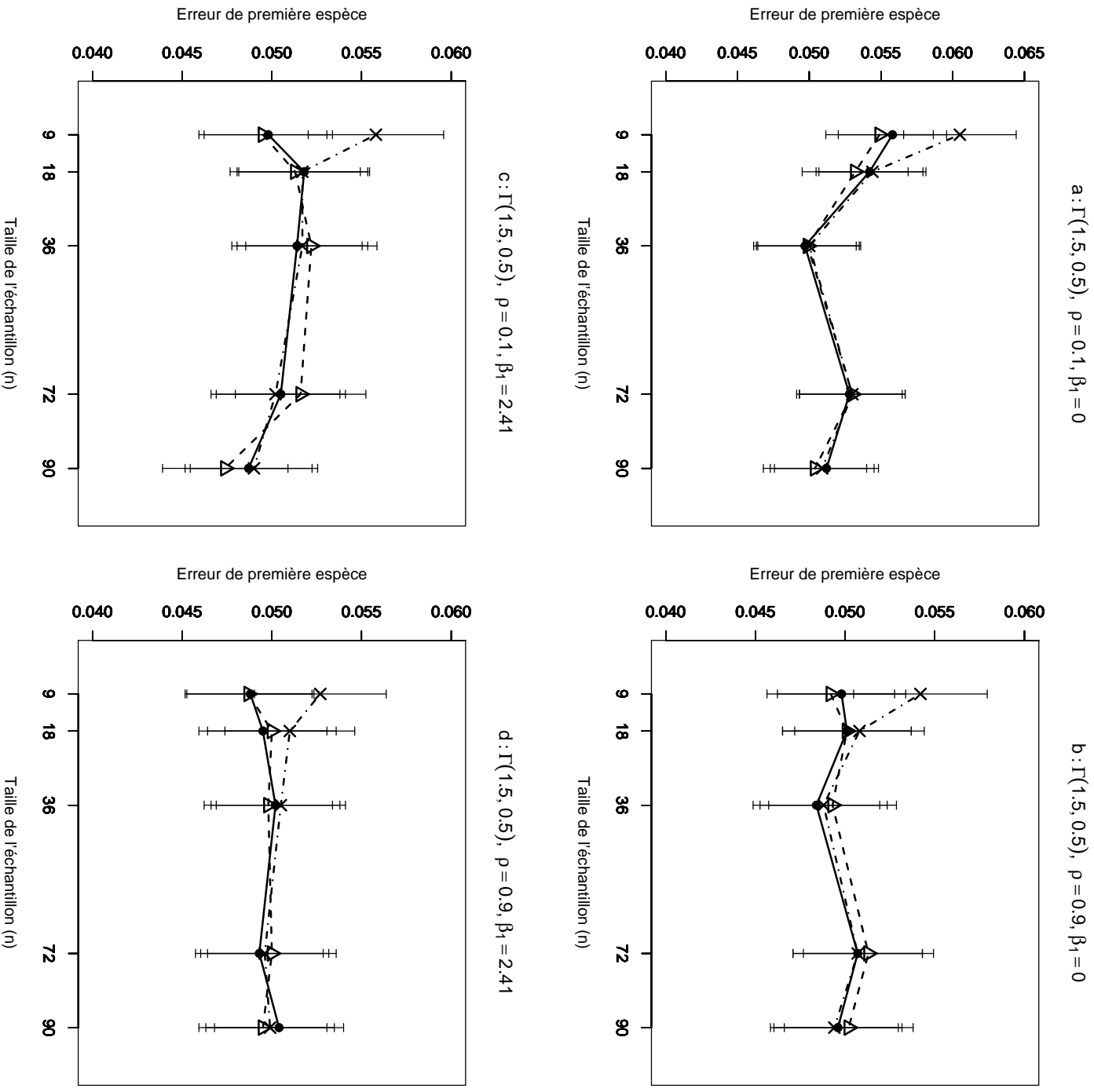


FIG. 4.35 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— Δ —, Freedman et Lane — \bullet — et Kennedy — \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1.5,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $n=9,18,36,72,90$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_1 = 0$; figure b : $\rho = 0.9$, $\beta_1 = 0$; figure c : $\rho = 0.1$, $\beta_1 = 2.41$, figure d : $\rho = 0.9$, $\beta_1 = 2.41$.

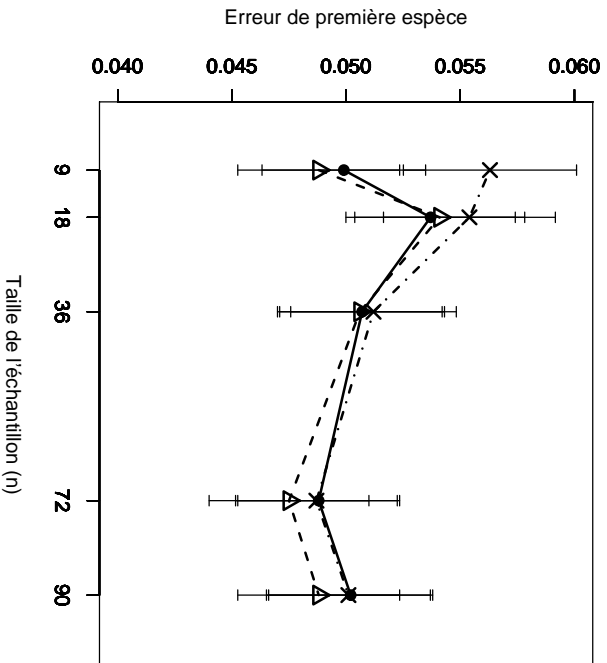
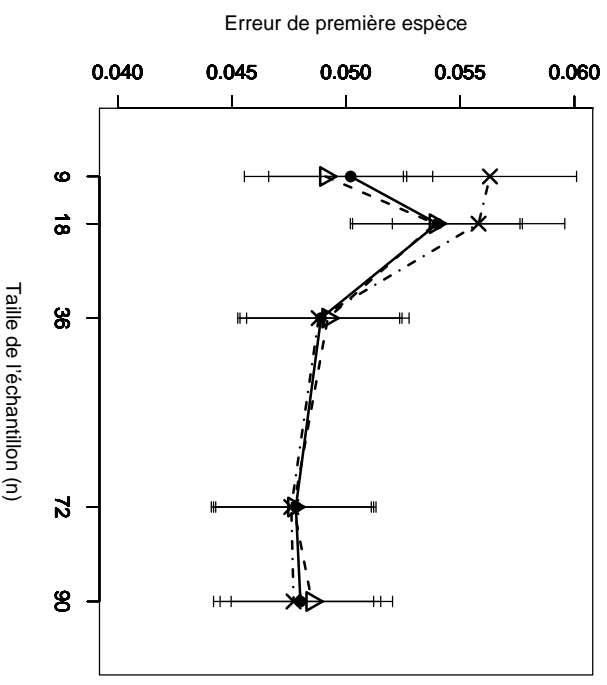
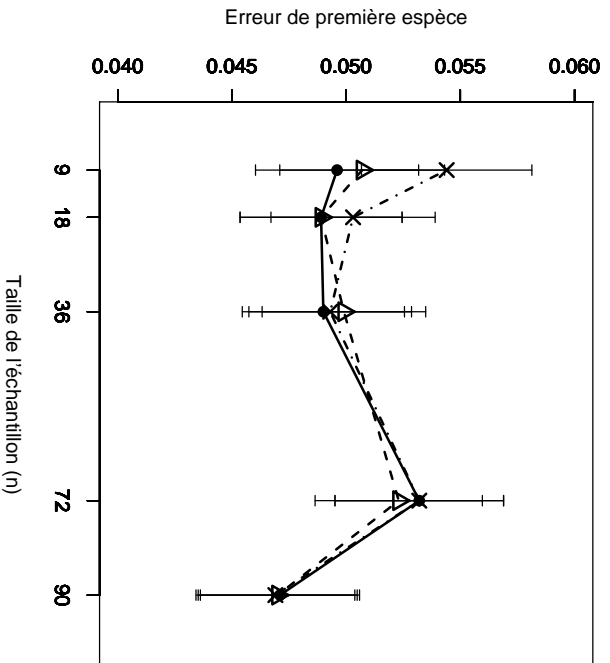
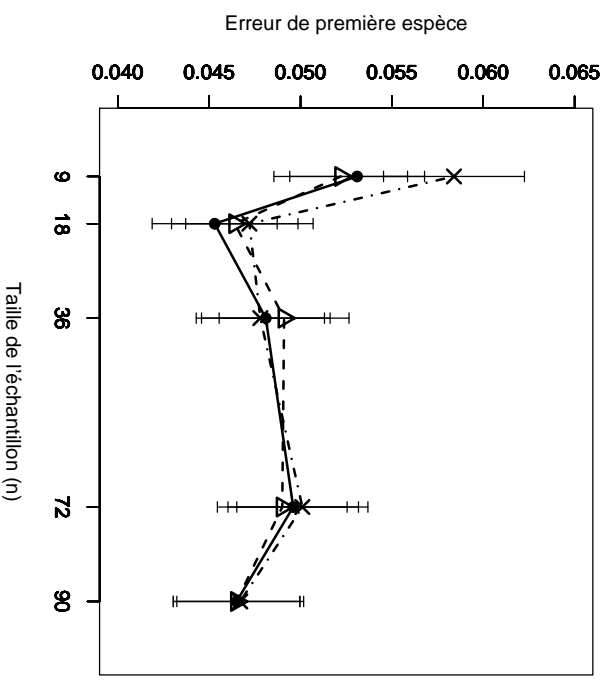
a: $\Gamma(1.5, 2)$, $\rho = 0.1$, $\beta_1 = 0$ b: $\Gamma(1.5, 2)$, $\rho = 0.9$, $\beta_1 = 0$ c: $\Gamma(1.5, 2)$, $\rho = 0.1$, $\beta_1 = 2.41$ d: $\Gamma(1.5, 2)$, $\rho = 0.9$, $\beta_1 = 2.41$ 

FIG. 4.36 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— Δ —, Freedman et Lane— \bullet — et Kennedy— \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1.5,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $n=9, 18, 36, 72, 90$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_1 = 0$; figure b : $\rho = 0.9$, $\beta_1 = 0$; figure c : $\rho = 0.1$, $\beta_1 = 2.41$, figure d : $\rho = 0.9$, $\beta_1 = 2.41$.

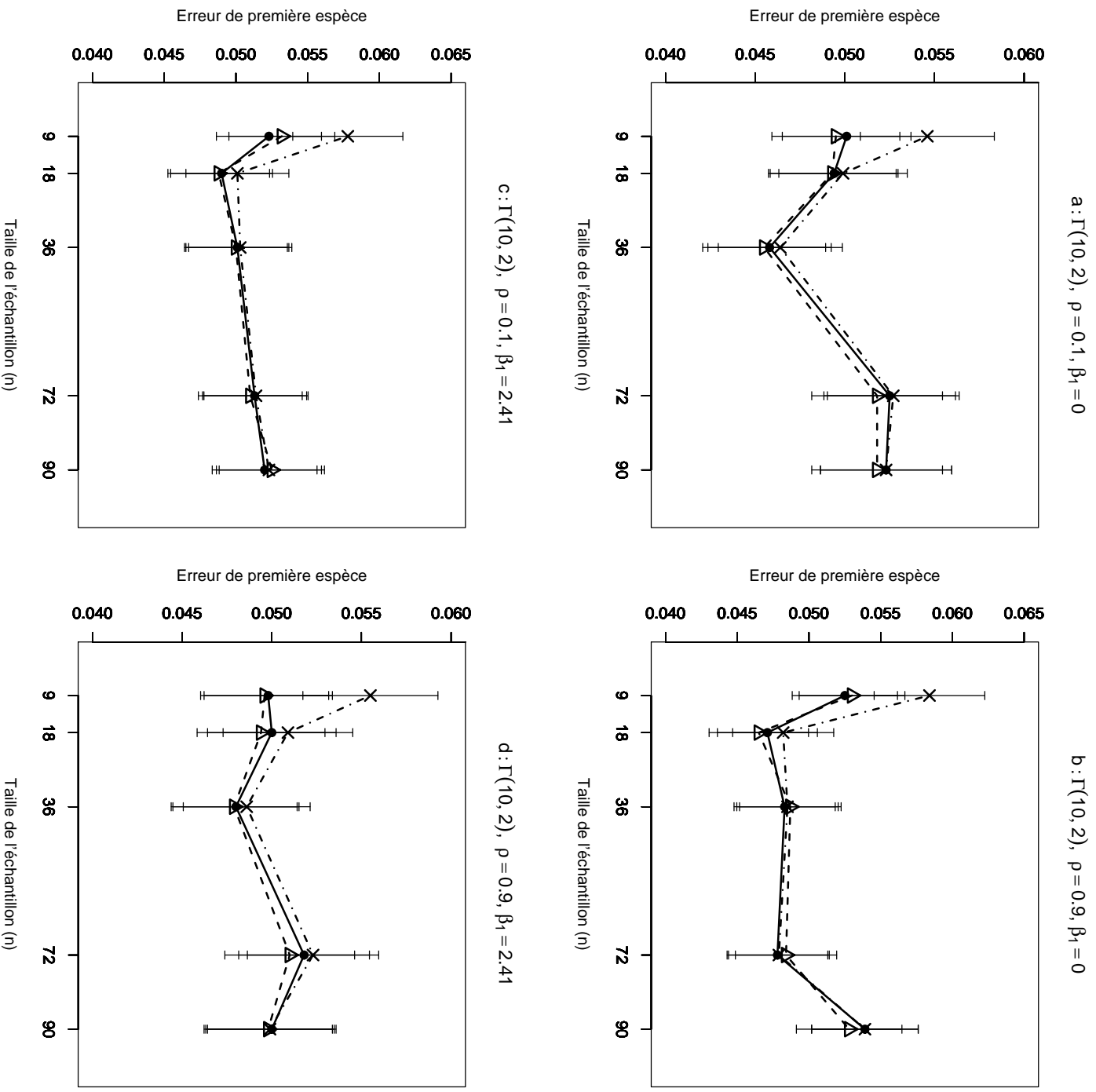


FIG. 4.37 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne— Δ —, Freedman et Lane— \bullet — et Kennedy— \times —, en fonction de la taille de l'échantillon. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(10,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $n=9,18,36,72,90$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_1 = 0$; figure b : $\rho = 0.9$, $\beta_1 = 0$; figure c : $\rho = 0.1$, $\beta_1 = 2.41$, figure d : $\rho = 0.9$, $\beta_1 = 2.41$.

Annexe 2

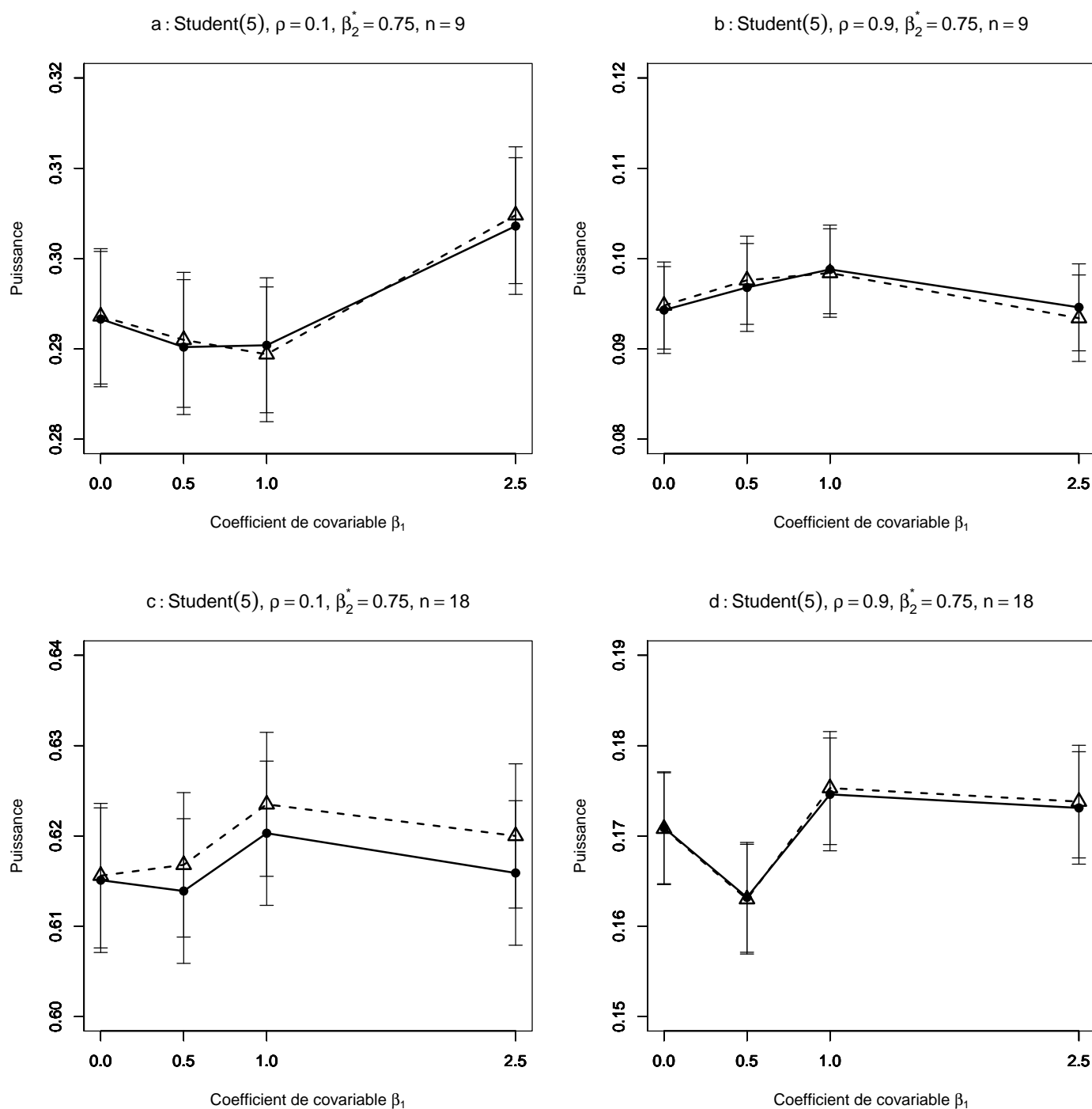


FIG. 4.38 – Puissance estimée du T-test de $H_0: \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Student(5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

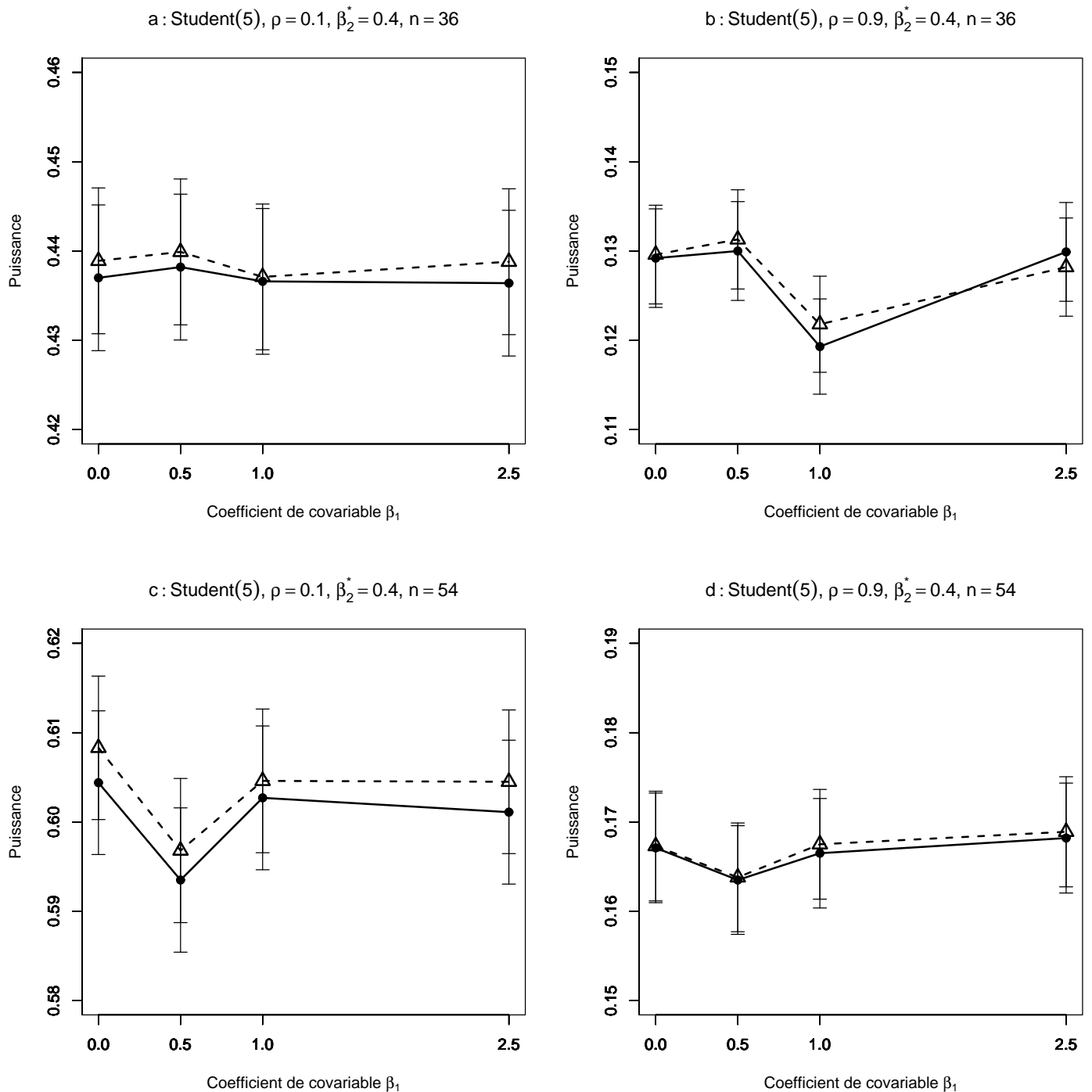


FIG. 4.39 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Student(5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

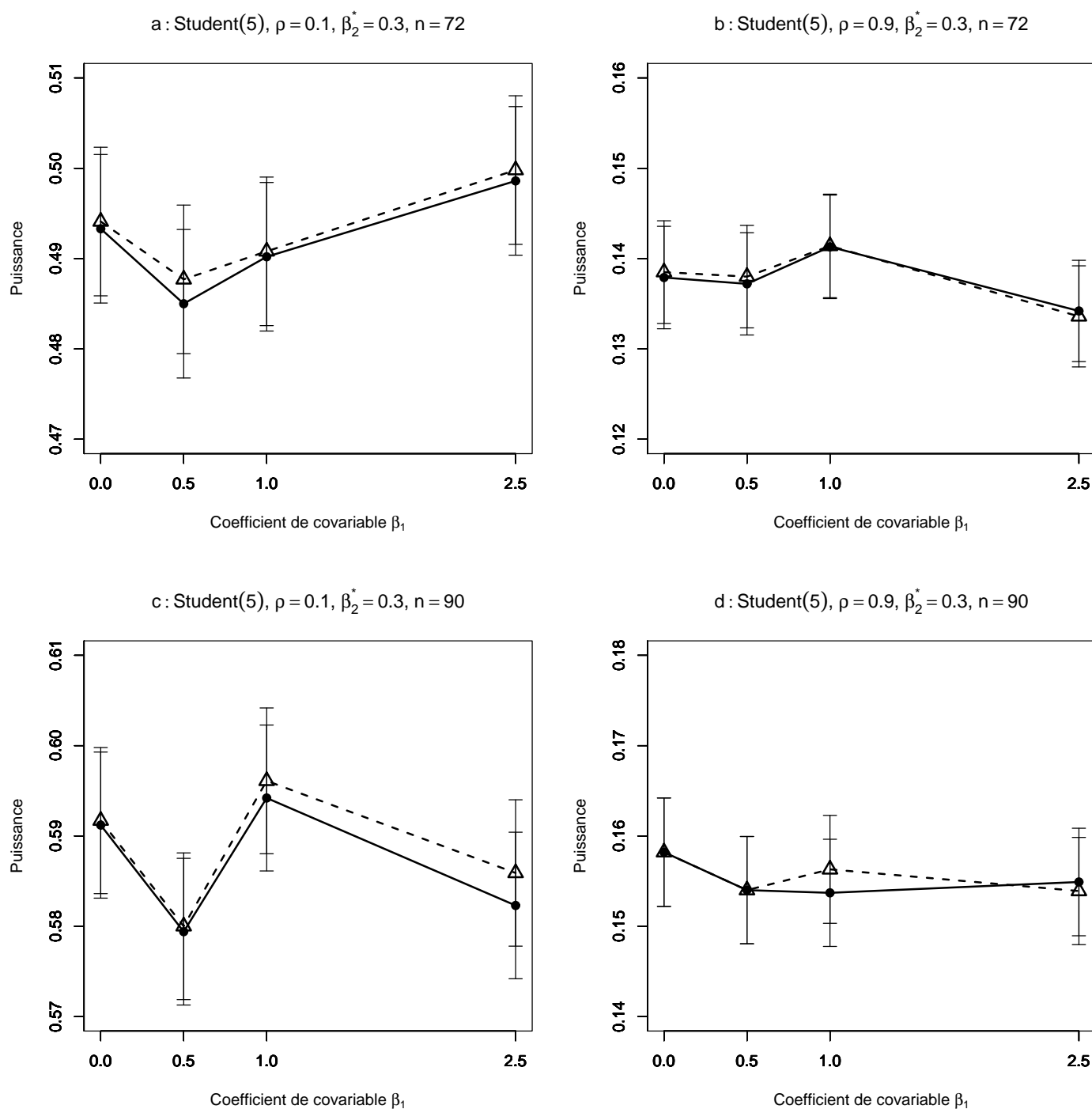


FIG. 4.40 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Student(5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

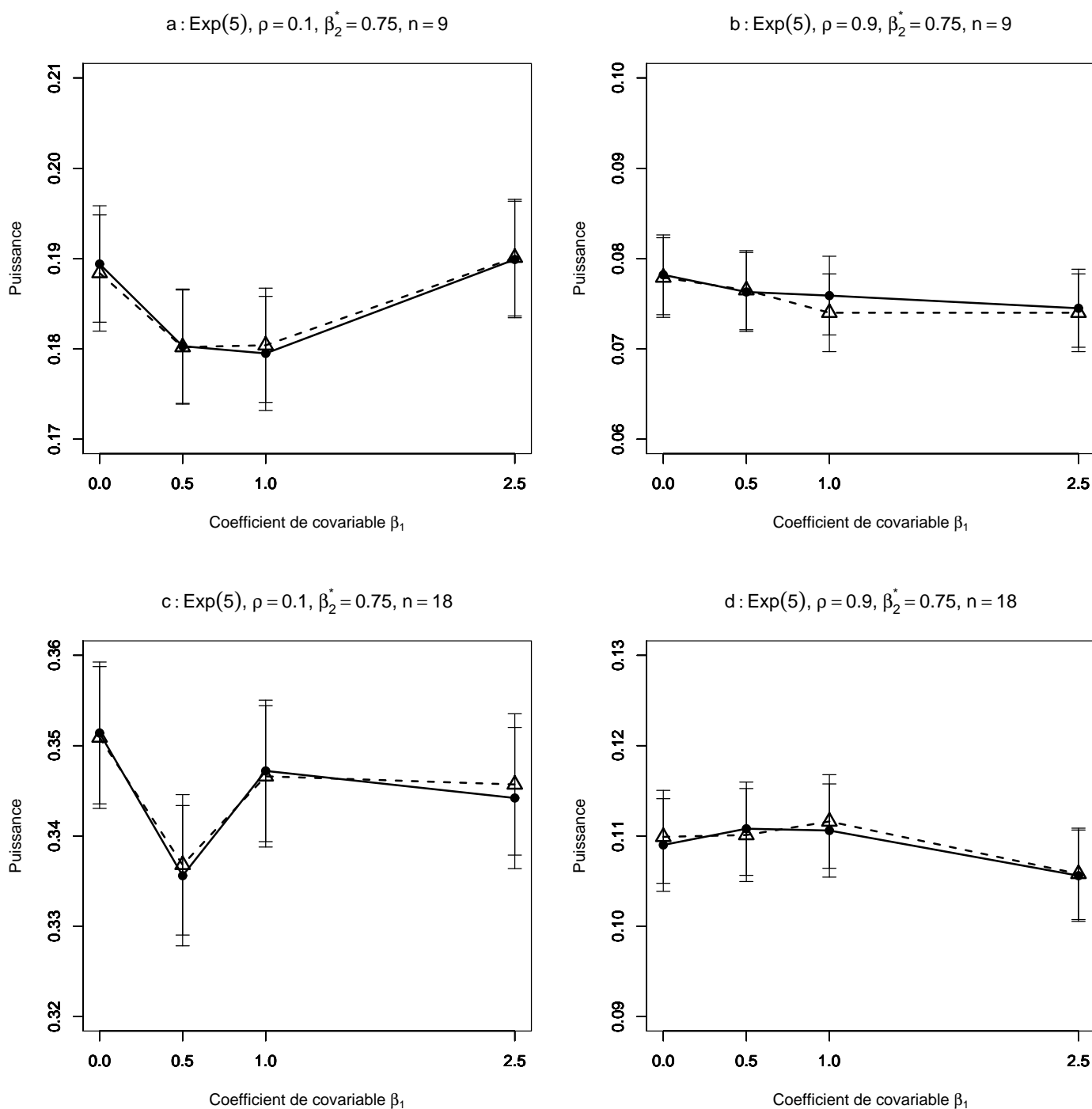


FIG. 4.41 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Exponentielle(0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1 = 0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%.
 Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

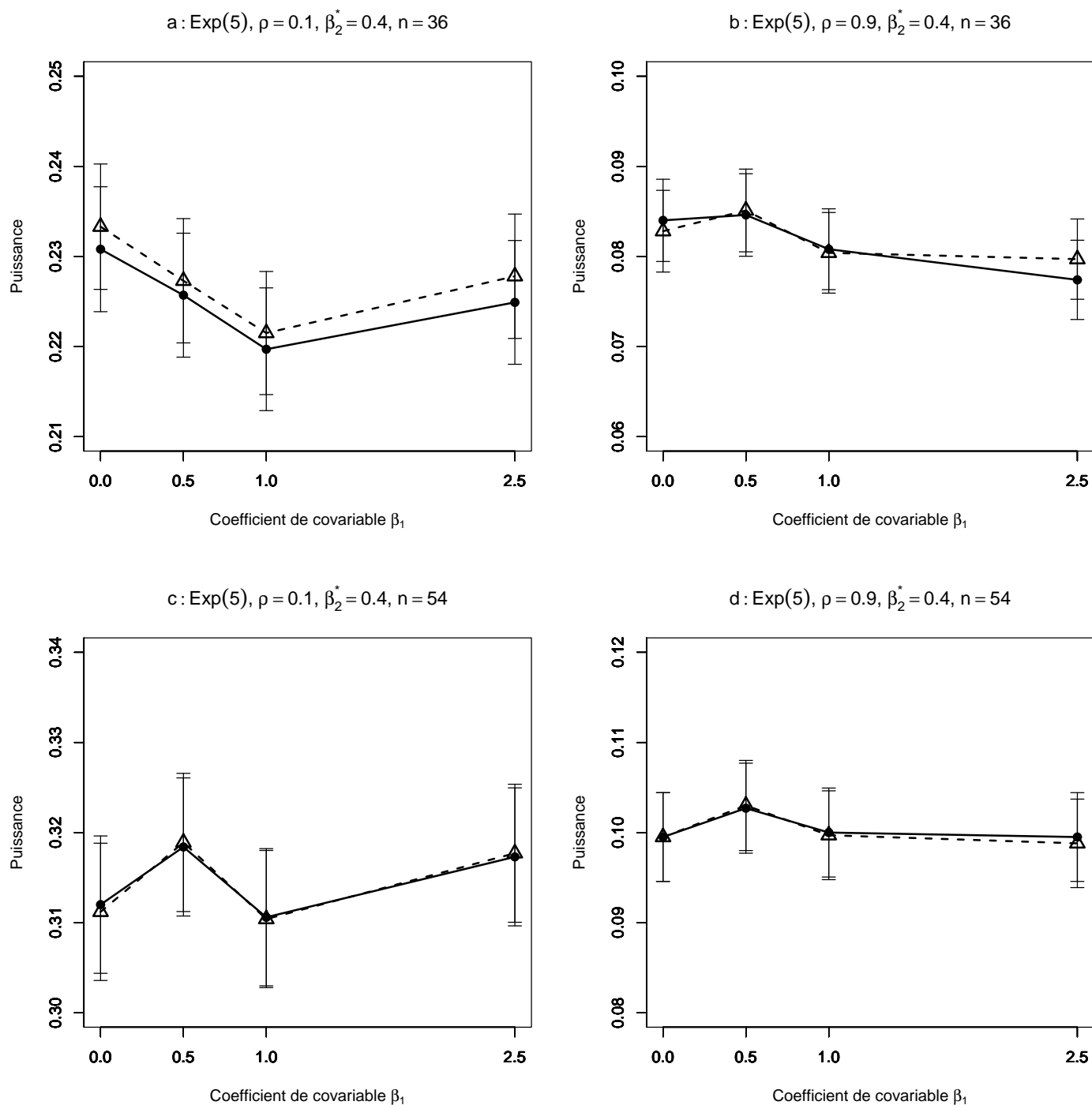


FIG. 4.42 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Exponentielle(0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%.
 Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

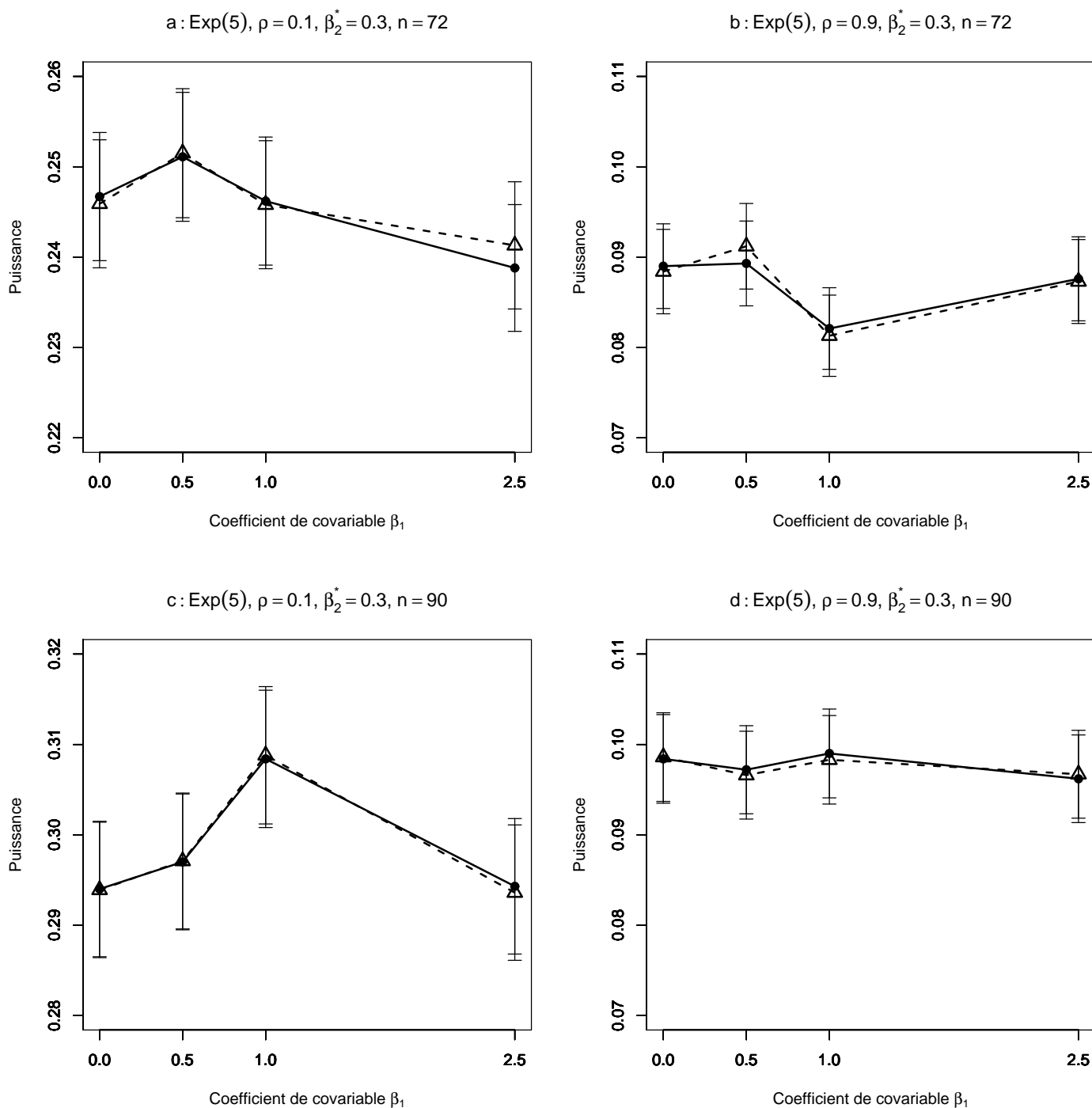


FIG. 4.43 – Puissance estimée du T-test de $H_0: \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Exponentielle(0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

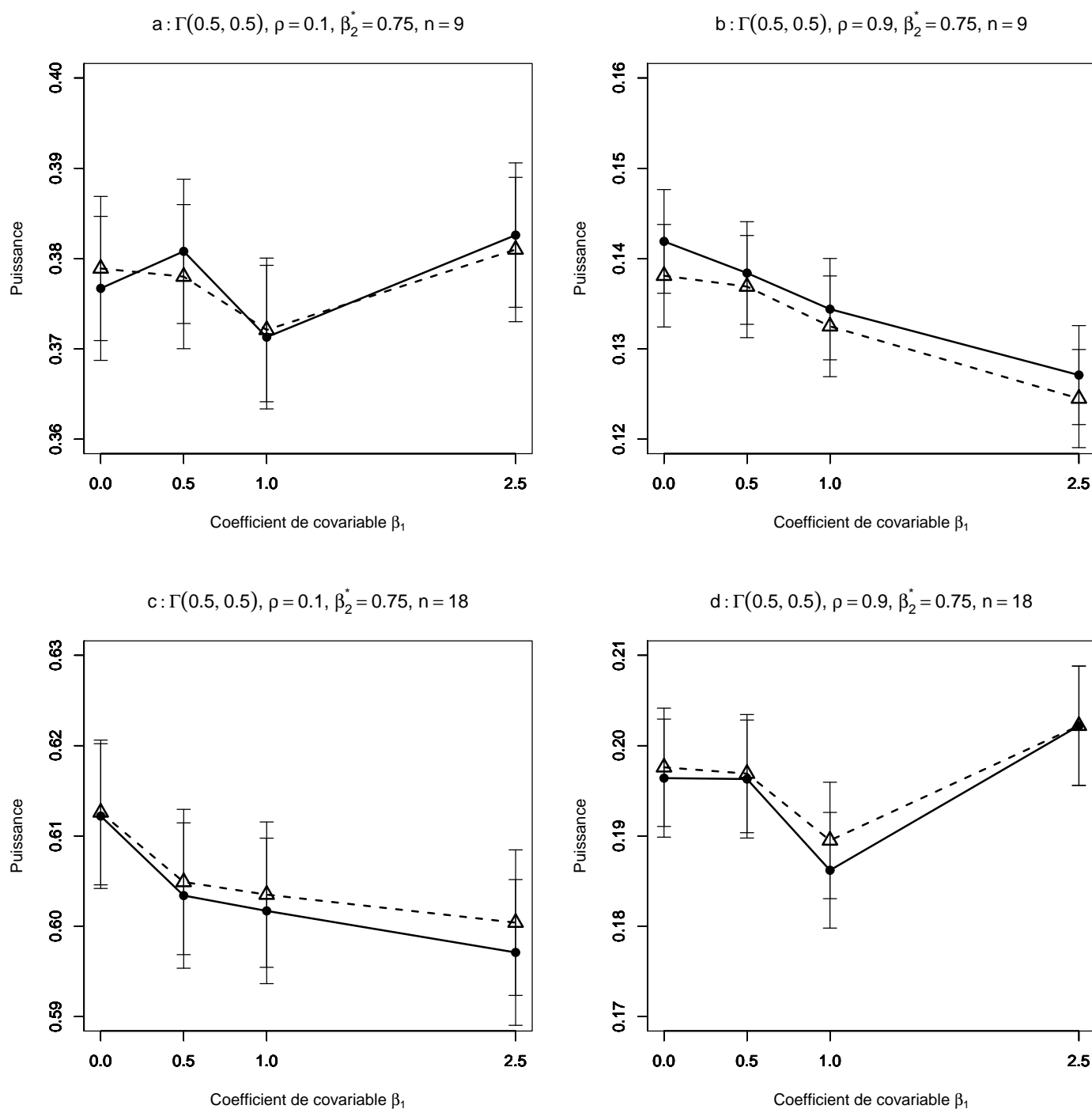


FIG. 4.44 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(0.5,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

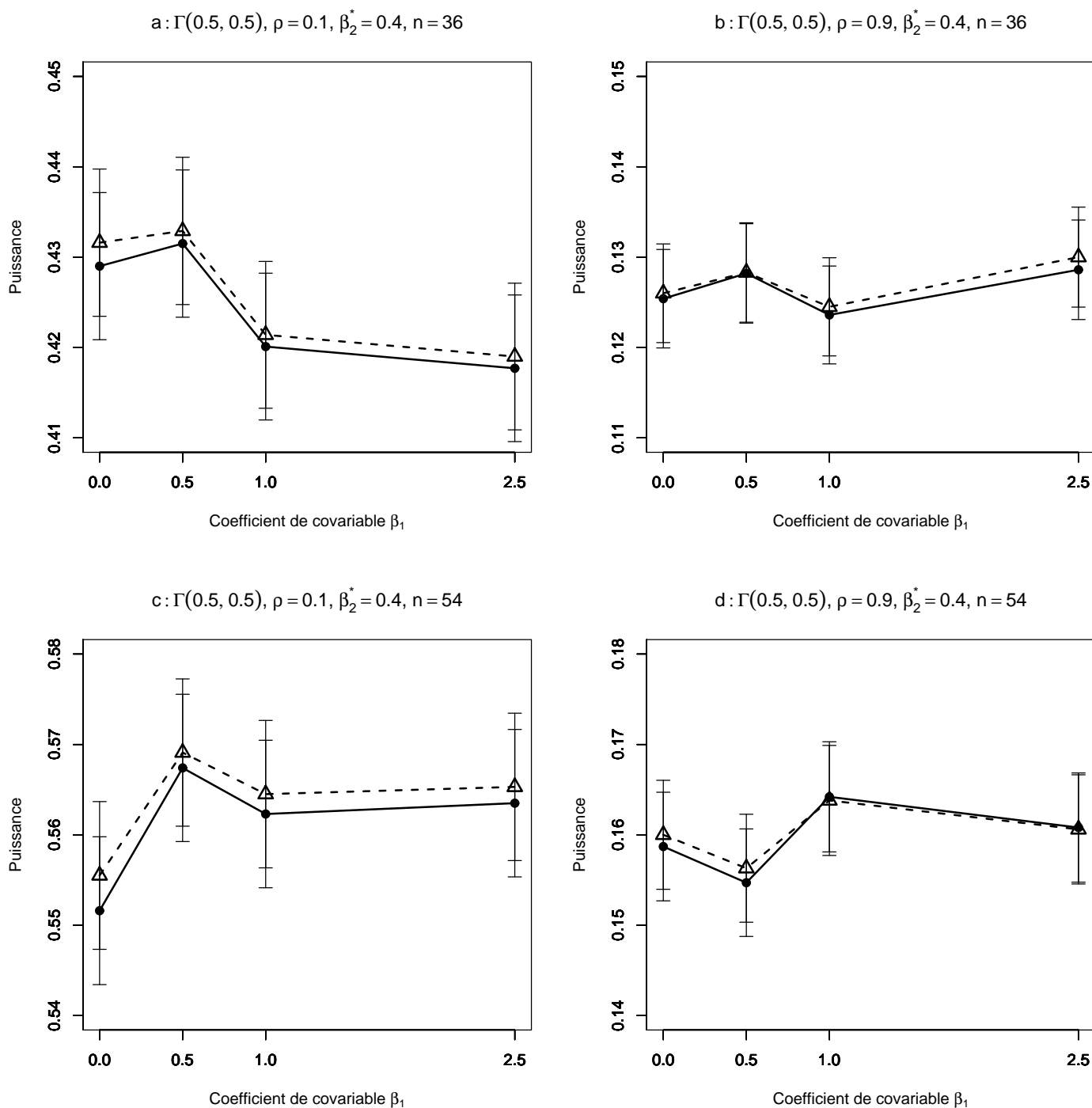


FIG. 4.45 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(0.5,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

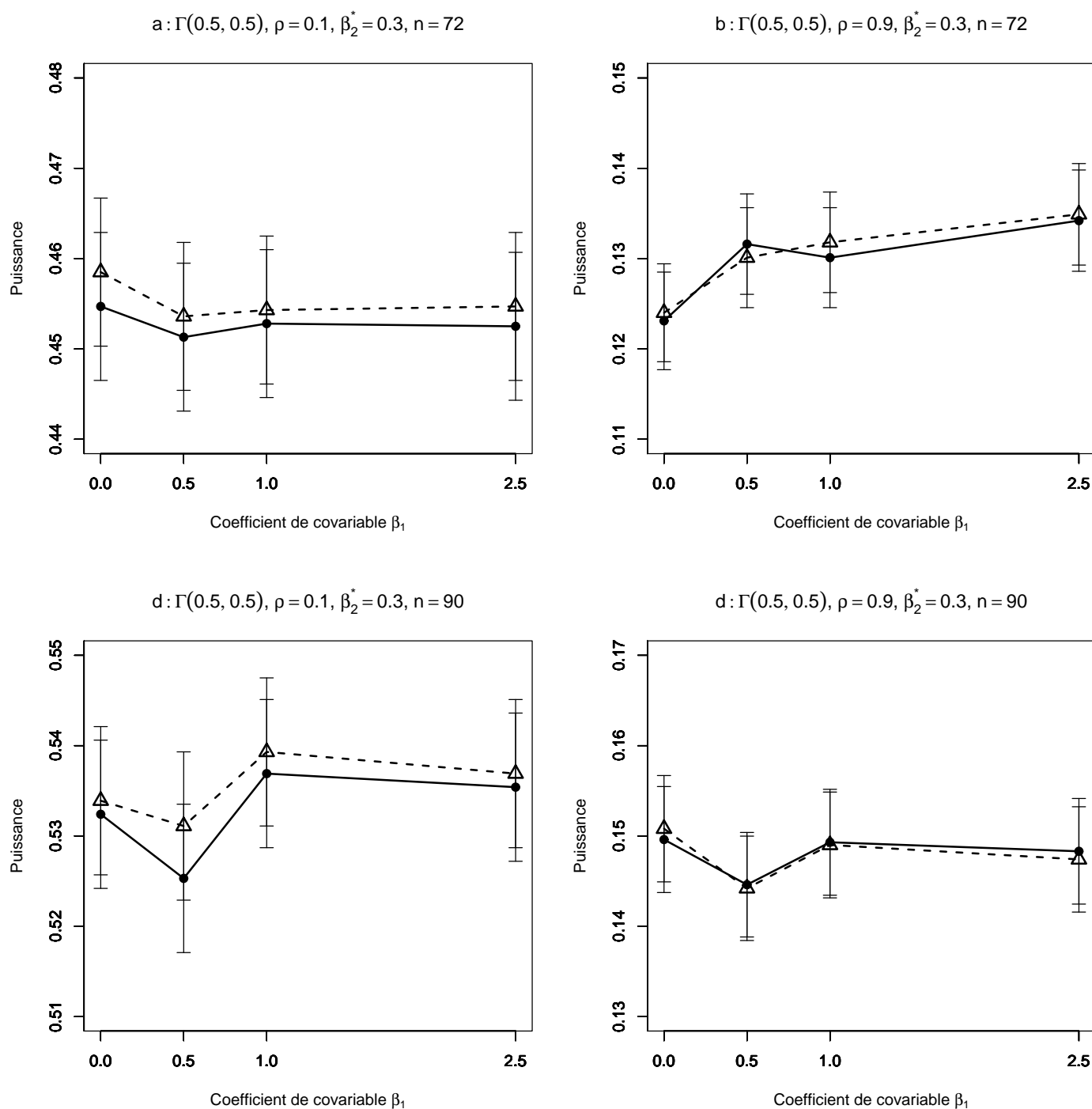


FIG. 4.46 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(0.5,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

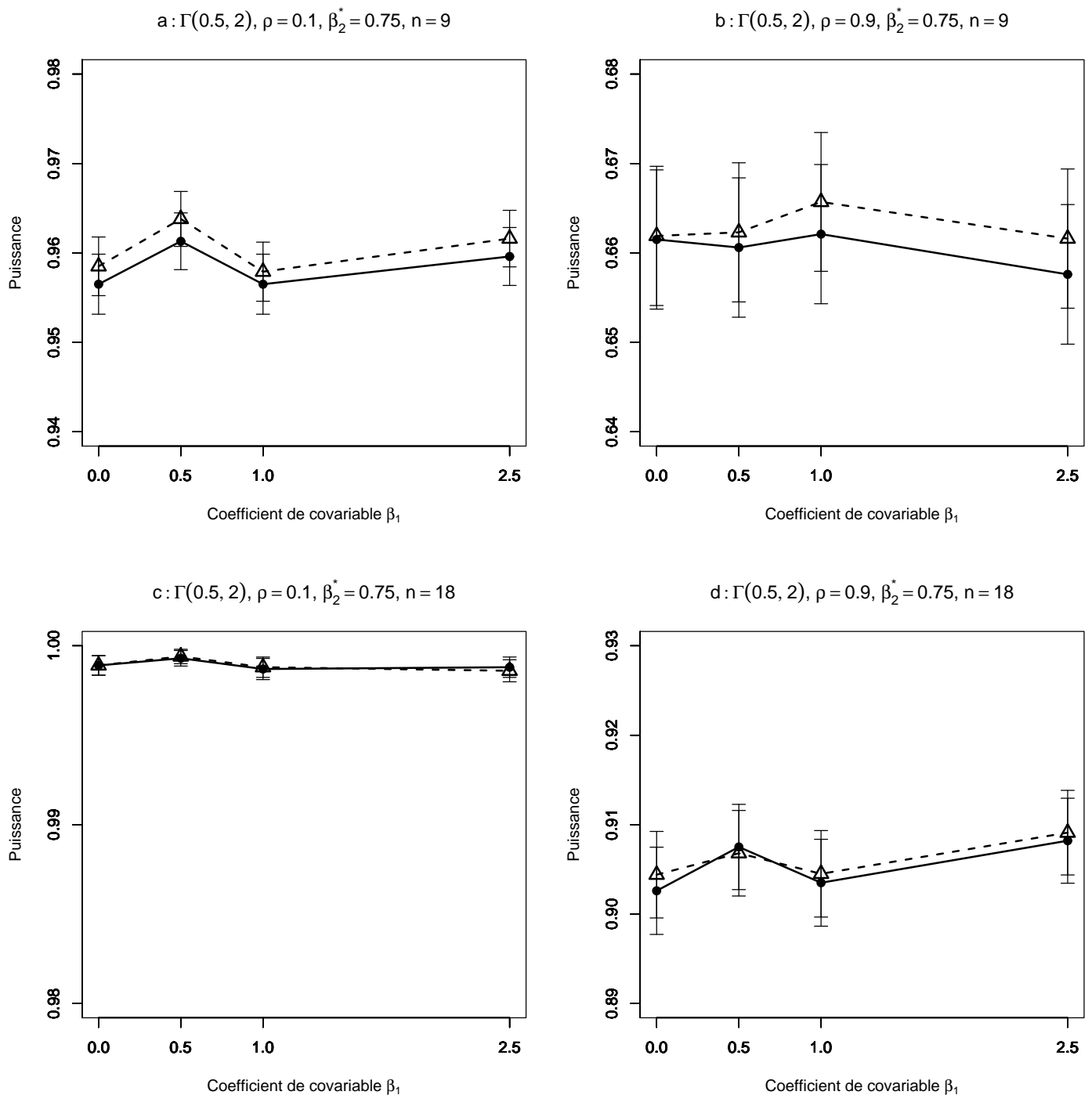


FIG. 4.47 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(0.5,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

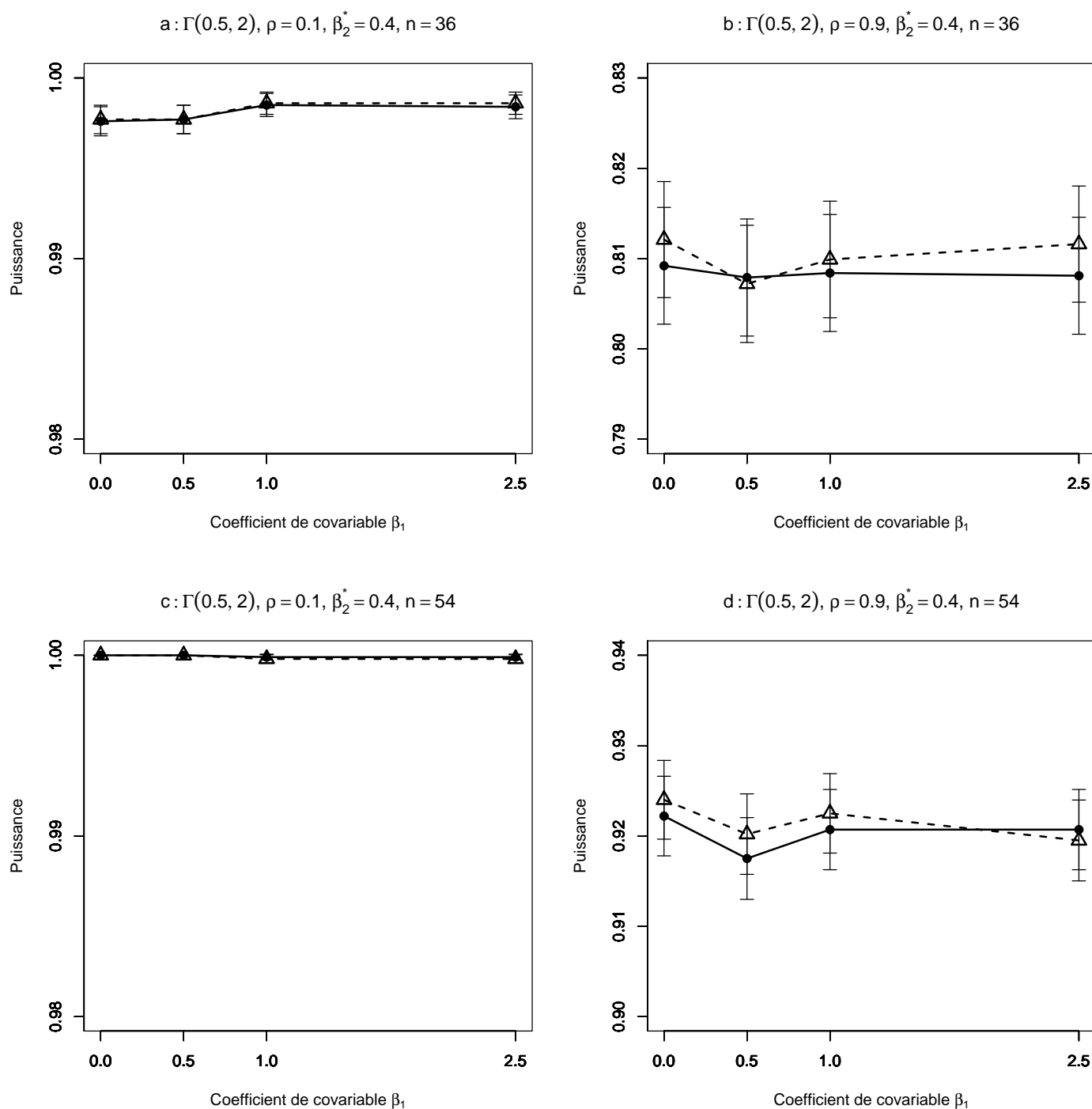


FIG. 4.48 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(0.5,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

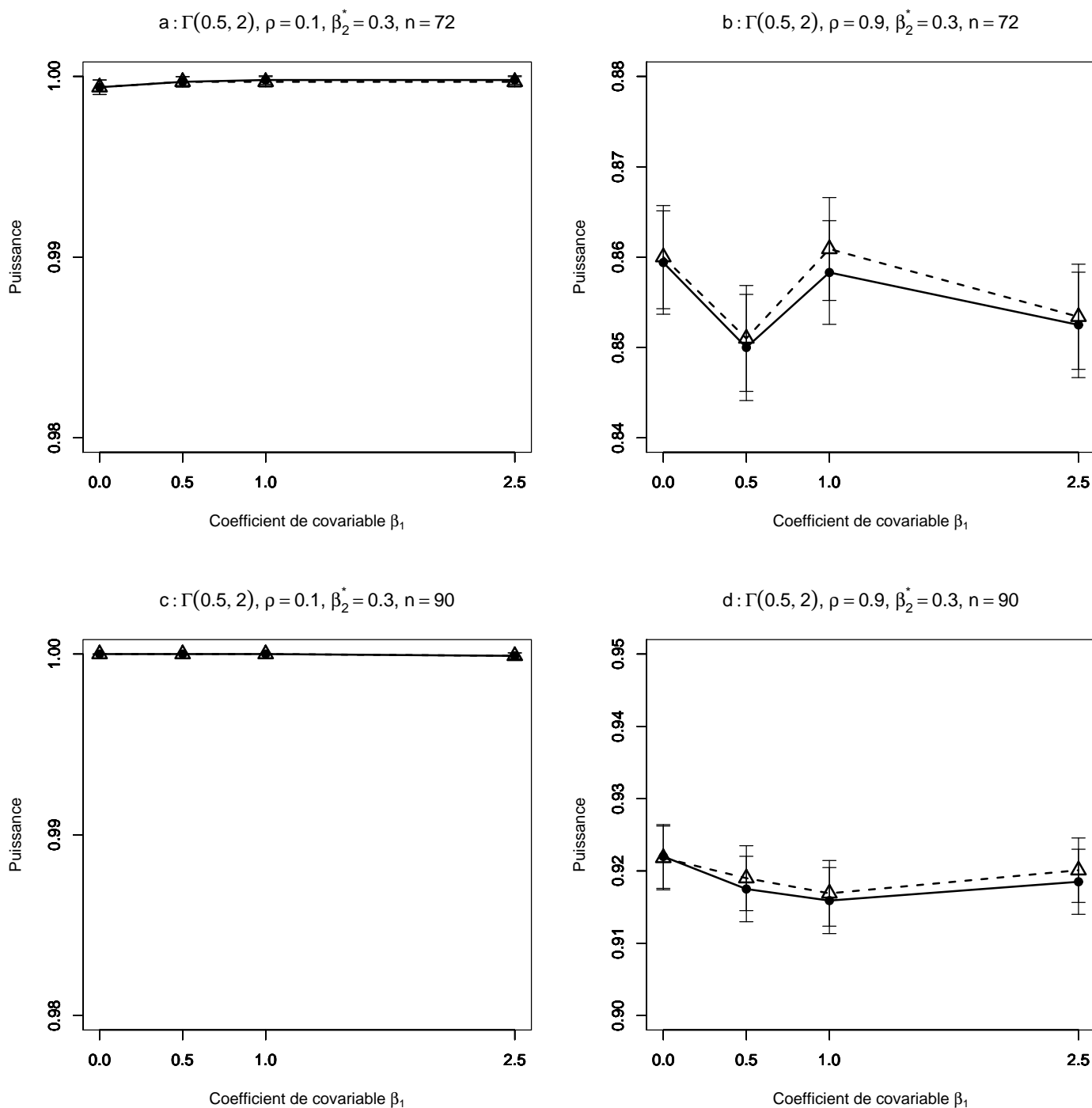


FIG. 4.49 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(0.5,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

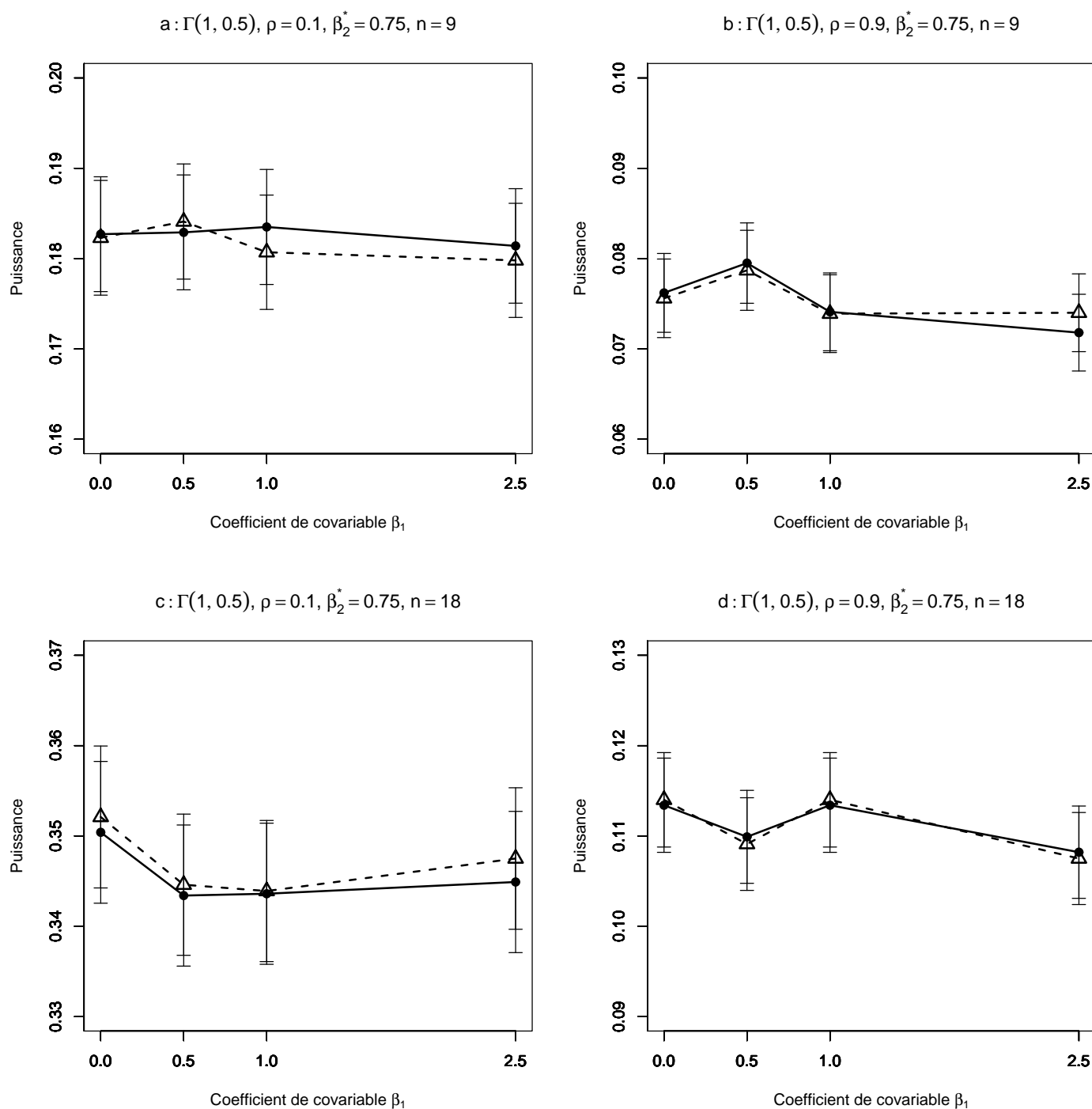


FIG. 4.50 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

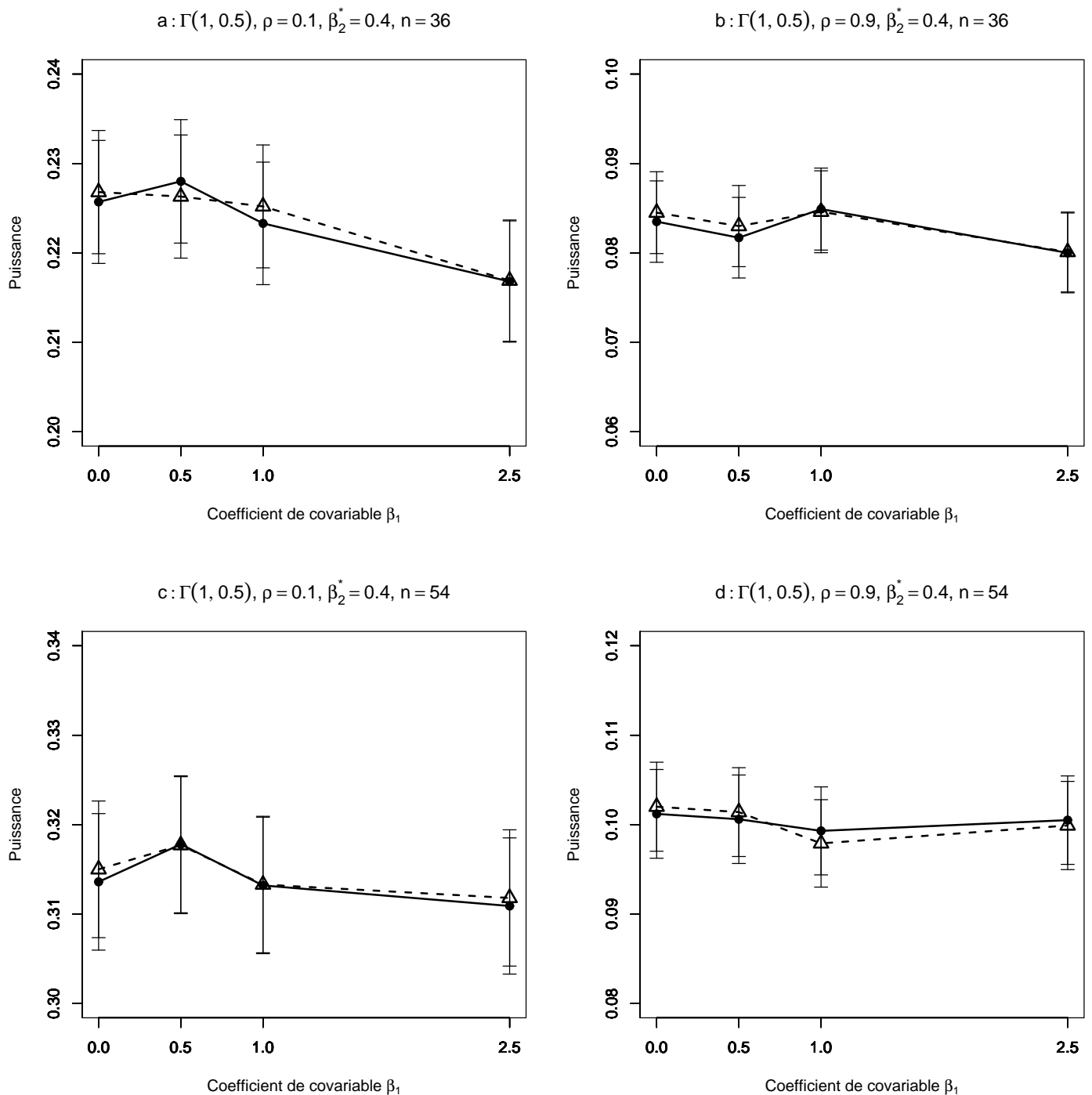


FIG. 4.51 – Puissance estimée du T-test de $H_0: \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

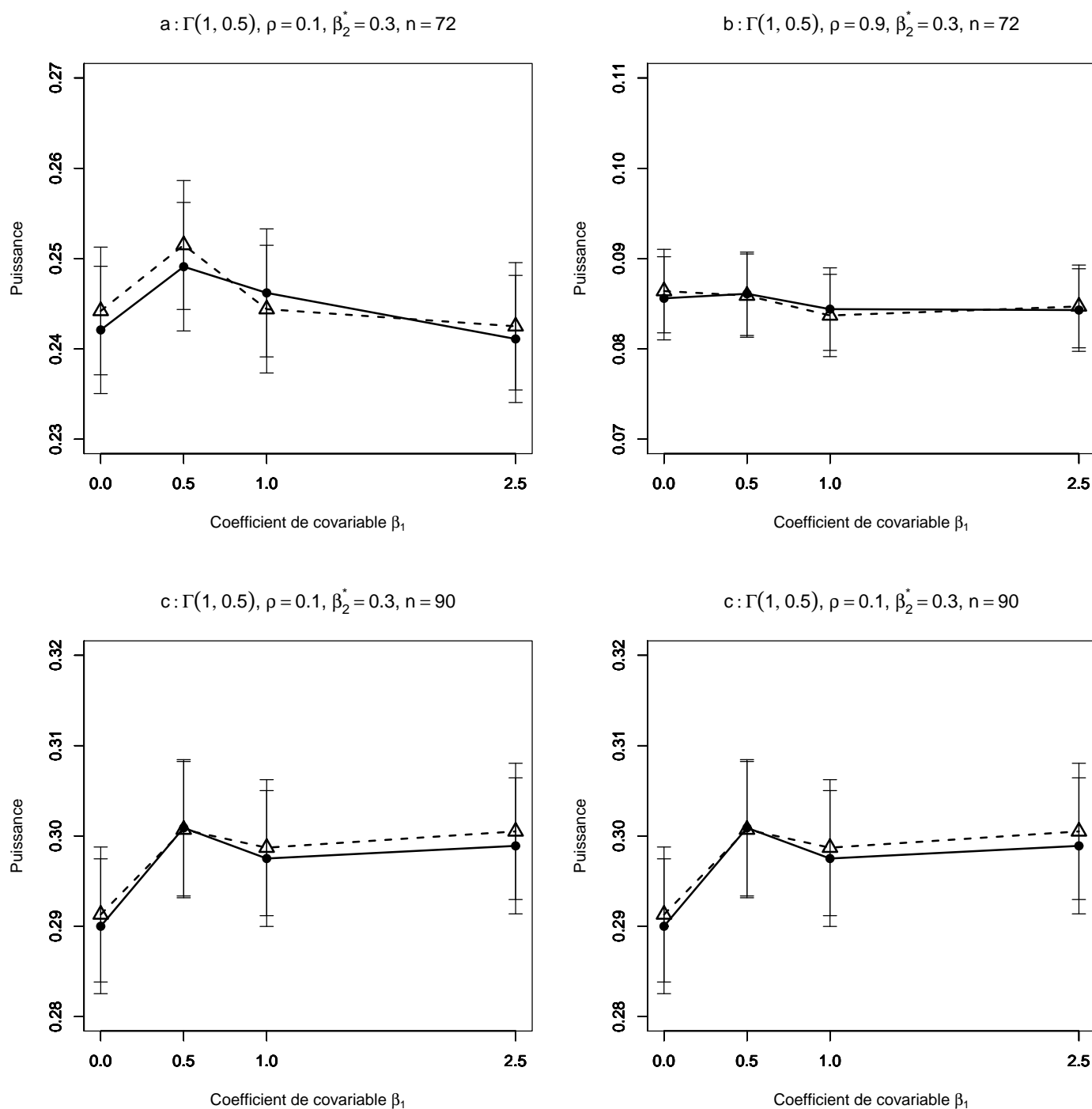


FIG. 4.52 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

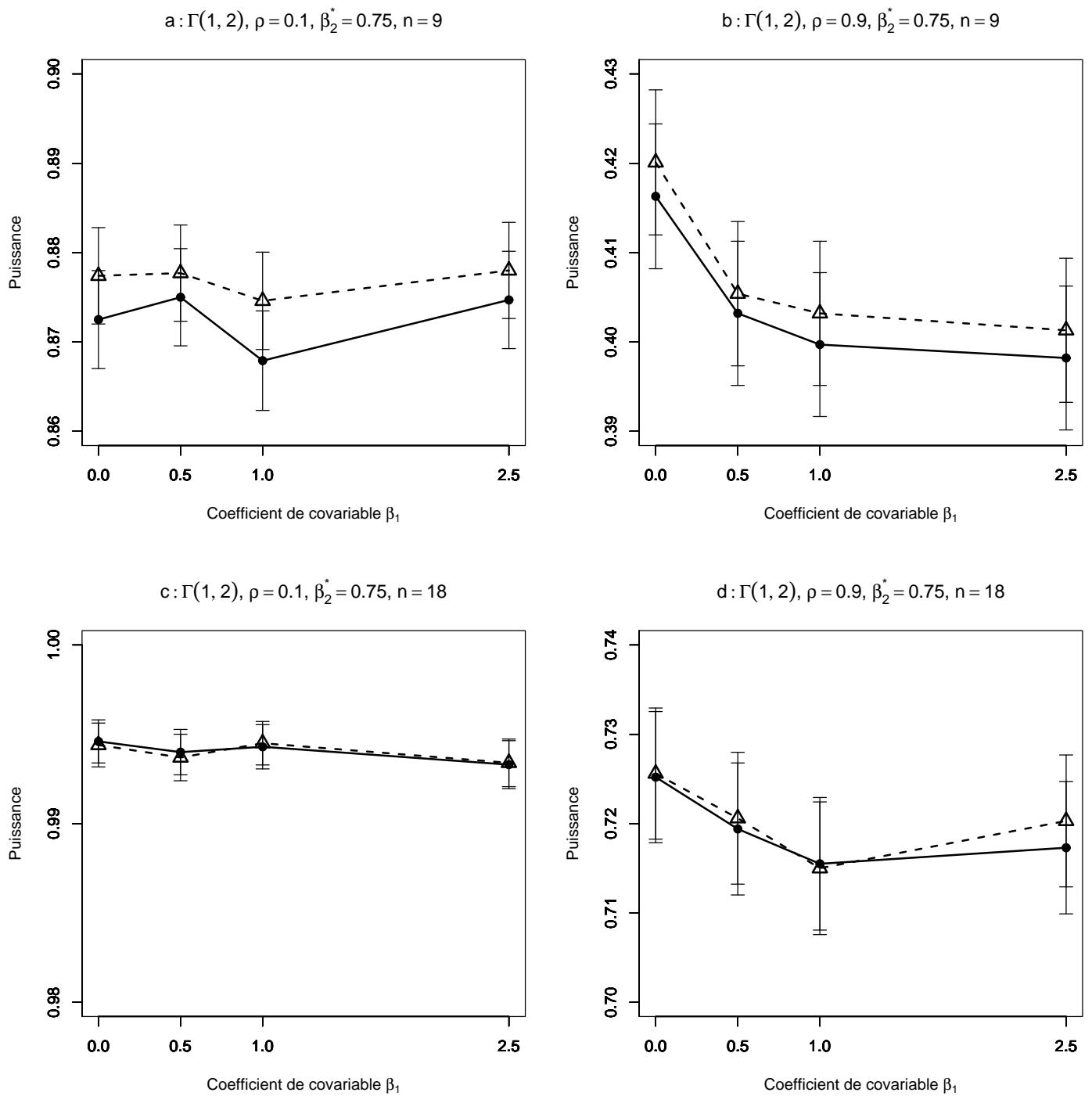


FIG. 4.53 – Puissance estimée du T-test de $H_0: \beta_2 = \beta_2^*$ par les méthodes Gaussienne— \triangle — et Freedman et Lane— \bullet —, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

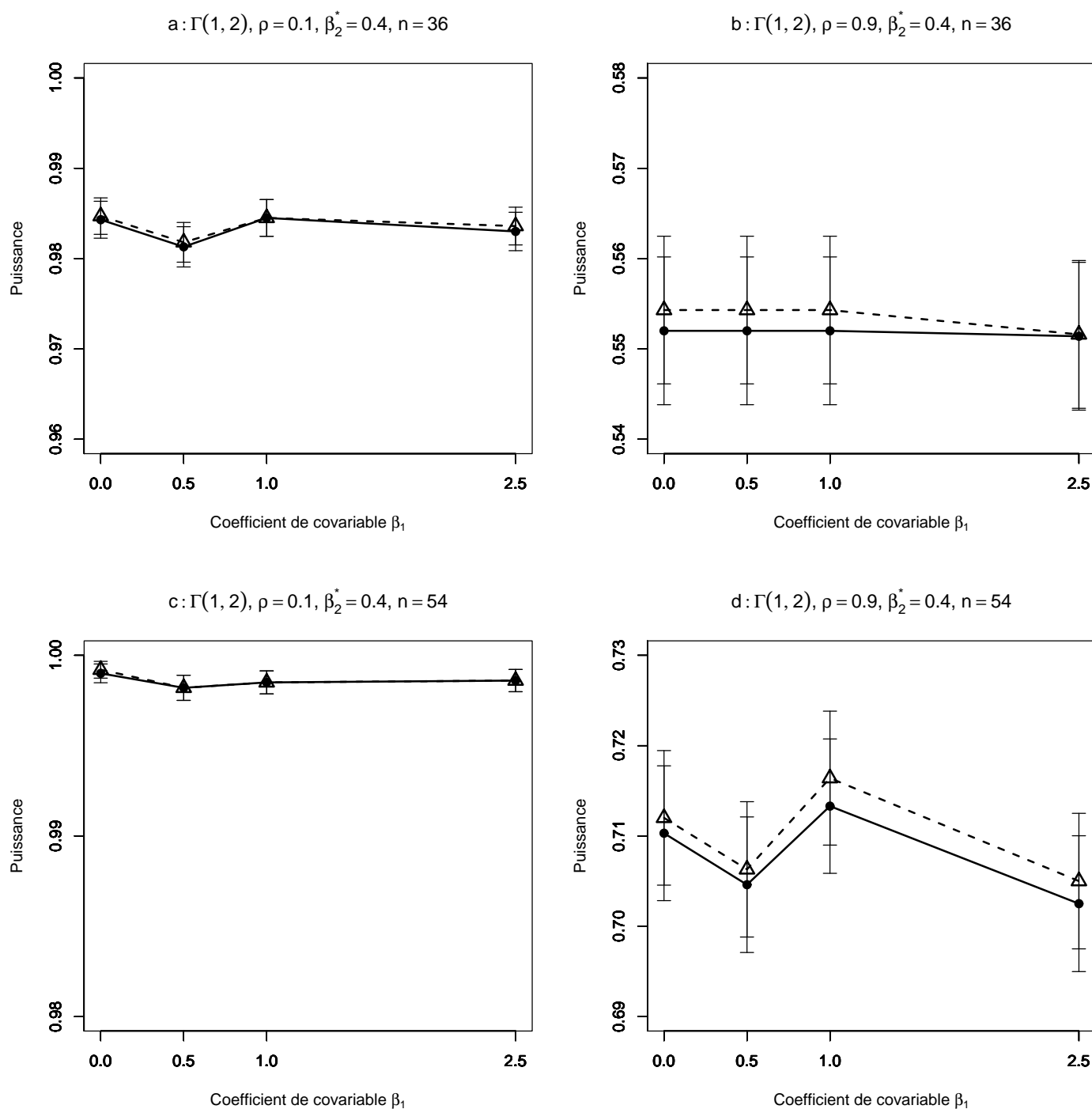


FIG. 4.54 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

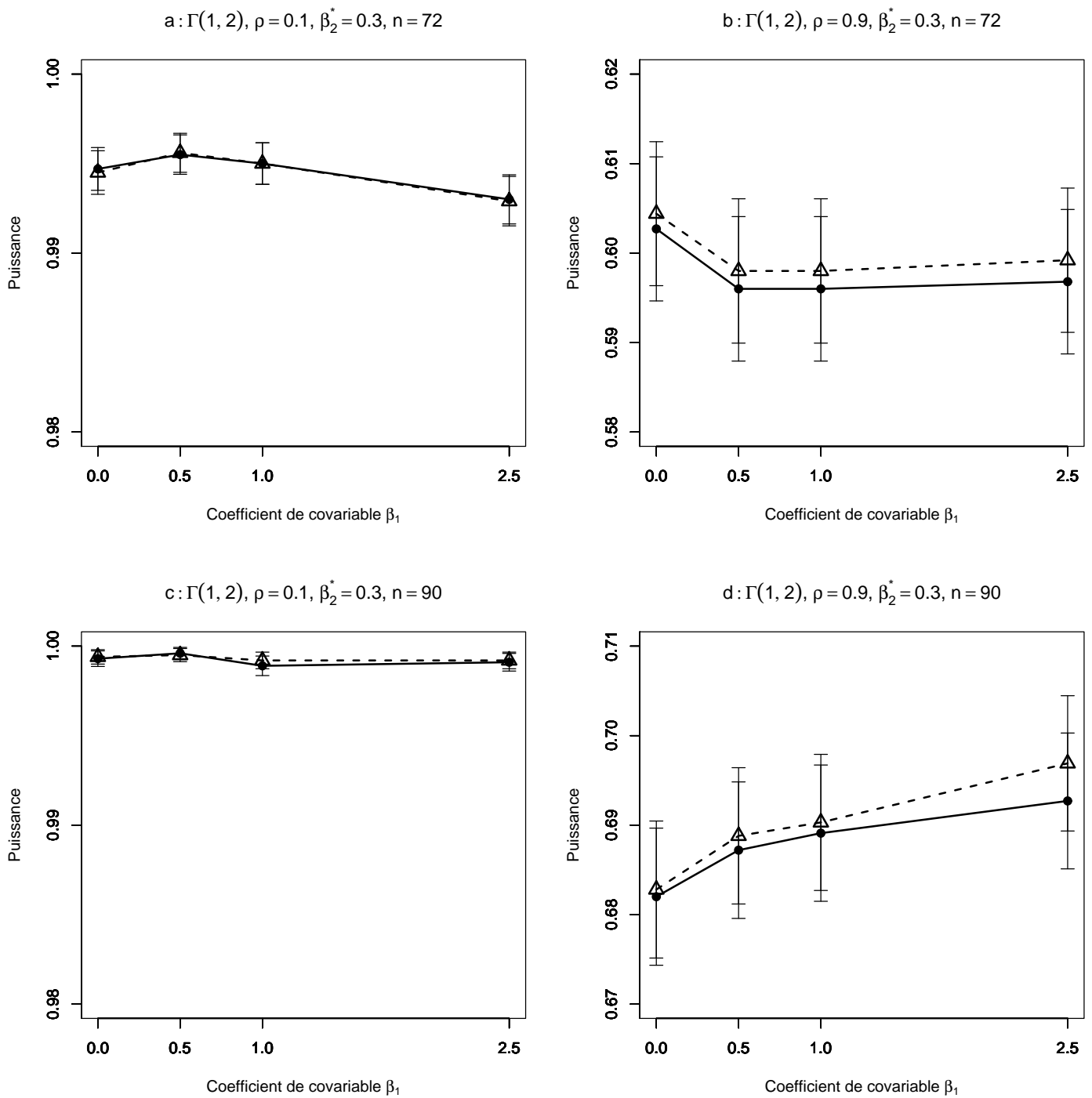


FIG. 4.55 – Puissance estimée du T-test de $H_0: \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

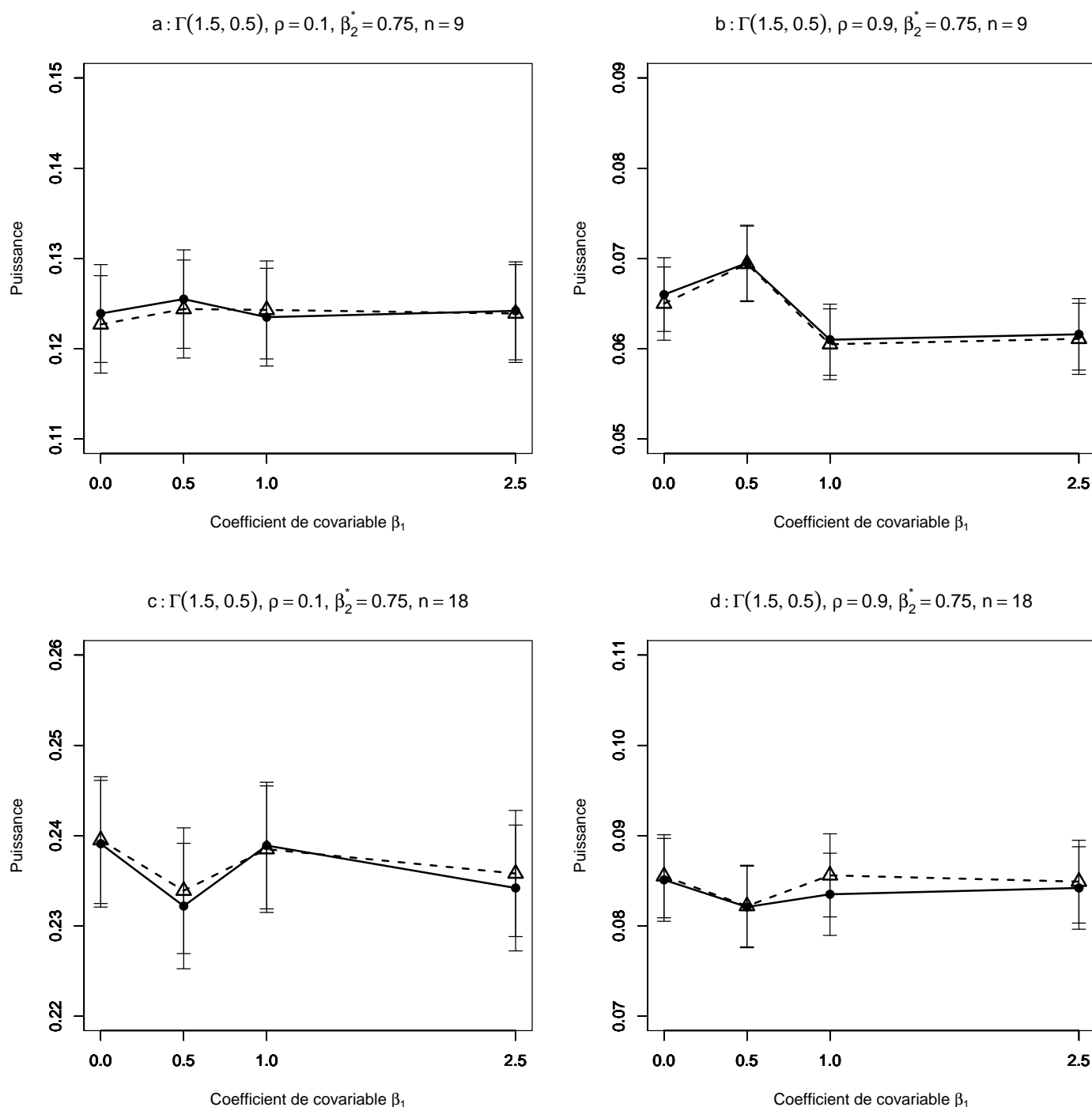


FIG. 4.56 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1.5,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

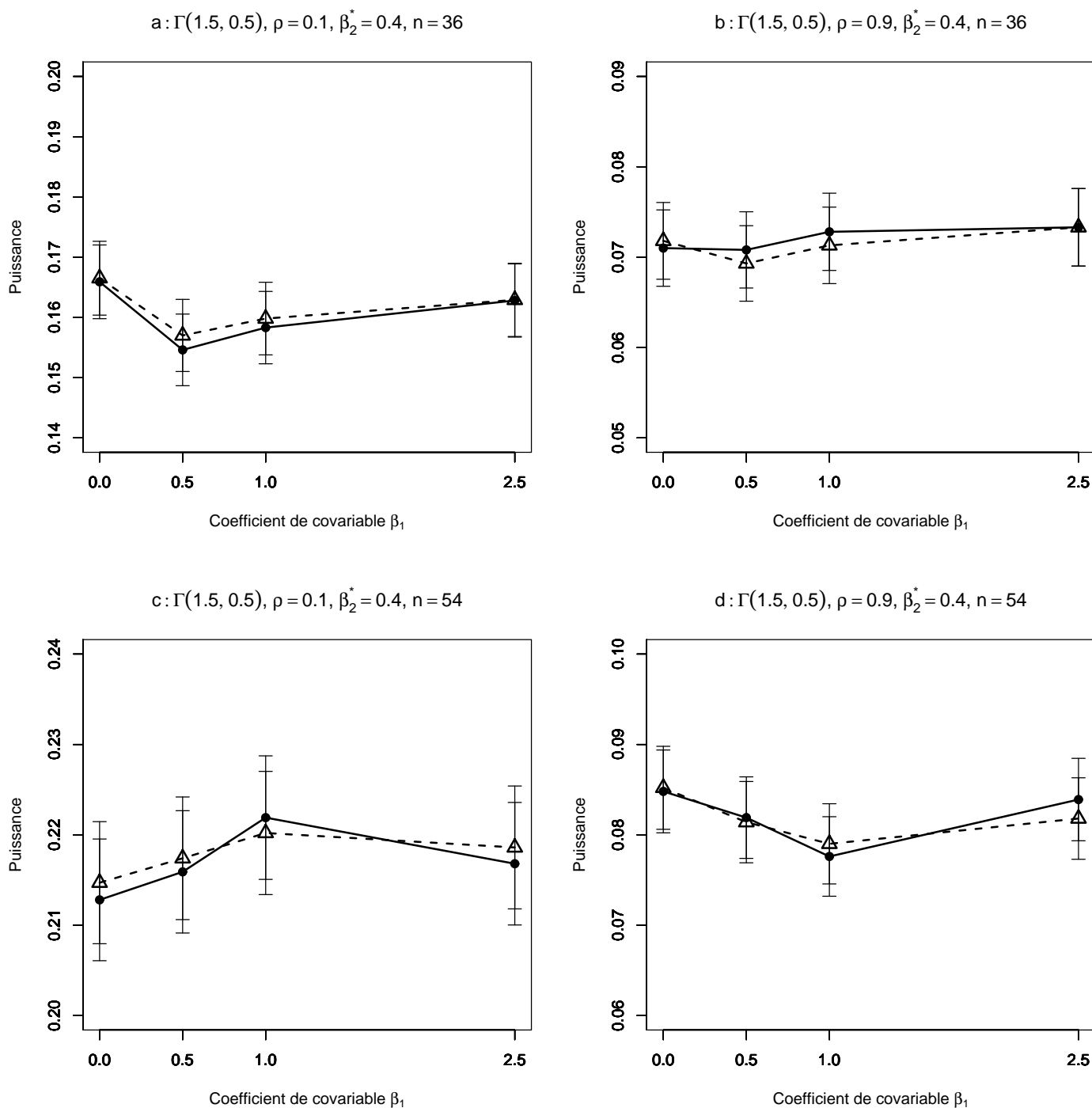


FIG. 4.57 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1.5,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

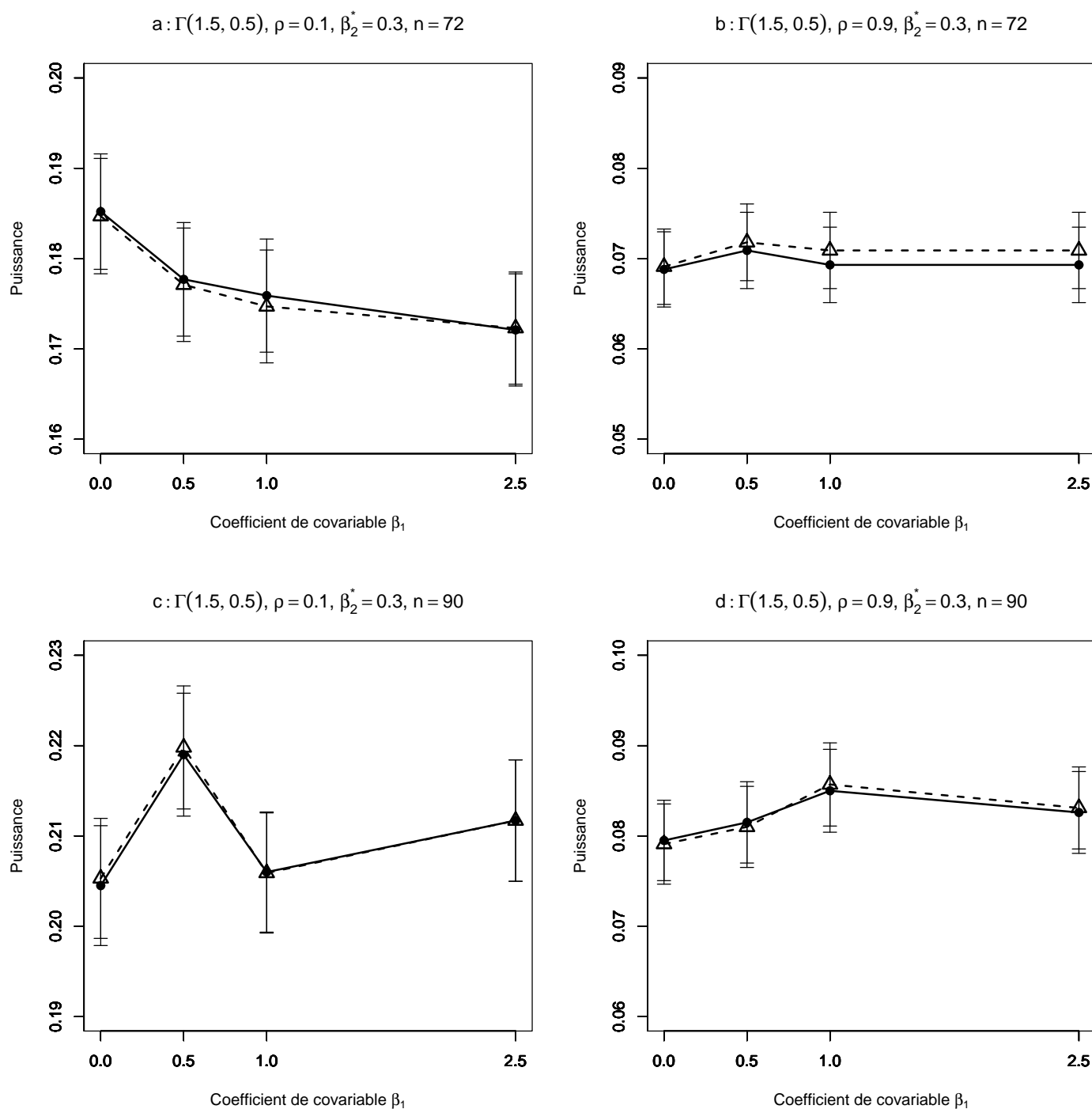


FIG. 4.58 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $--\triangle--$ et Freedman et Lane $—\bullet—$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1.5,0.5), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

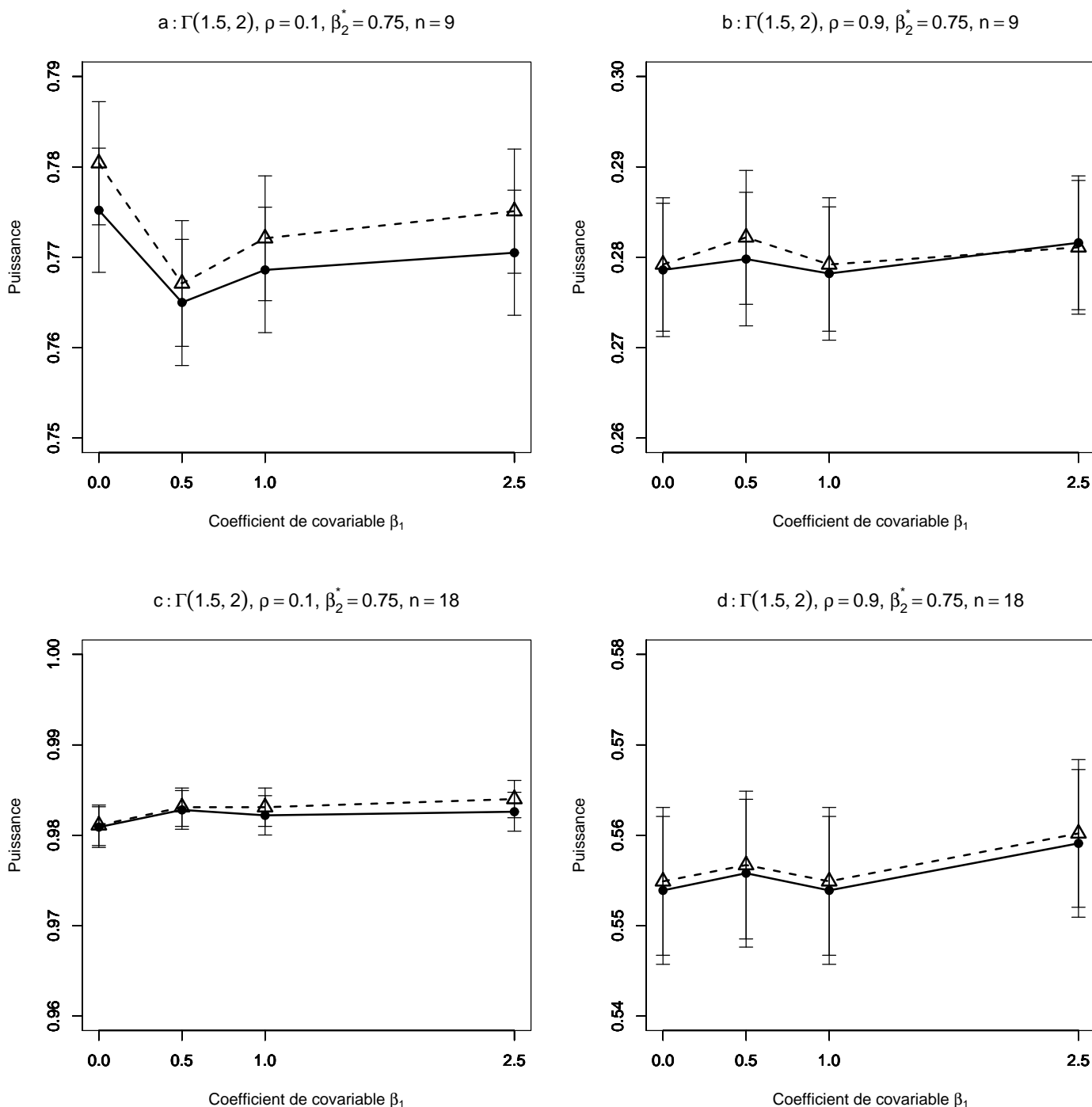


FIG. 4.59 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1.5,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

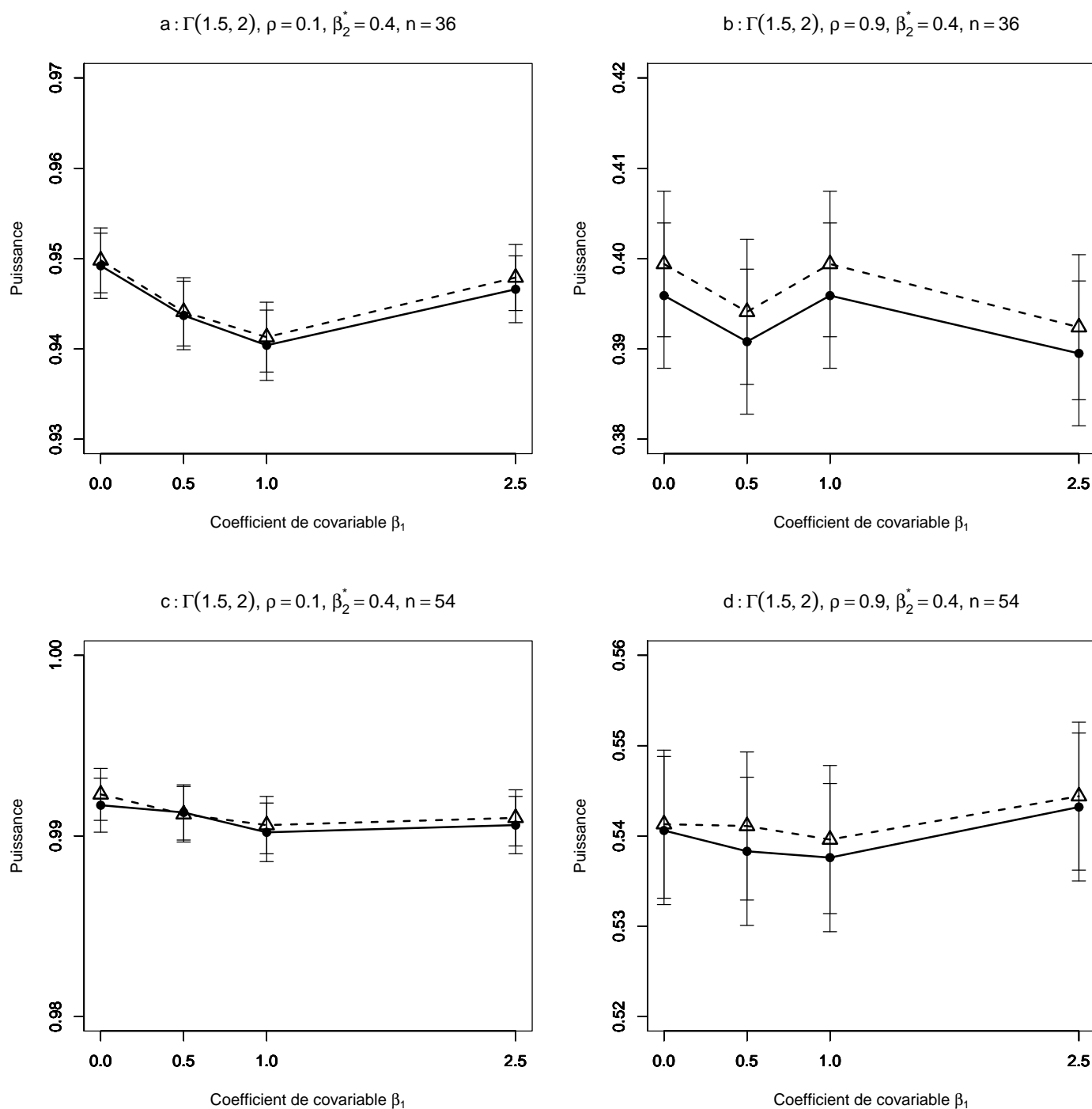


FIG. 4.60 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1.5,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

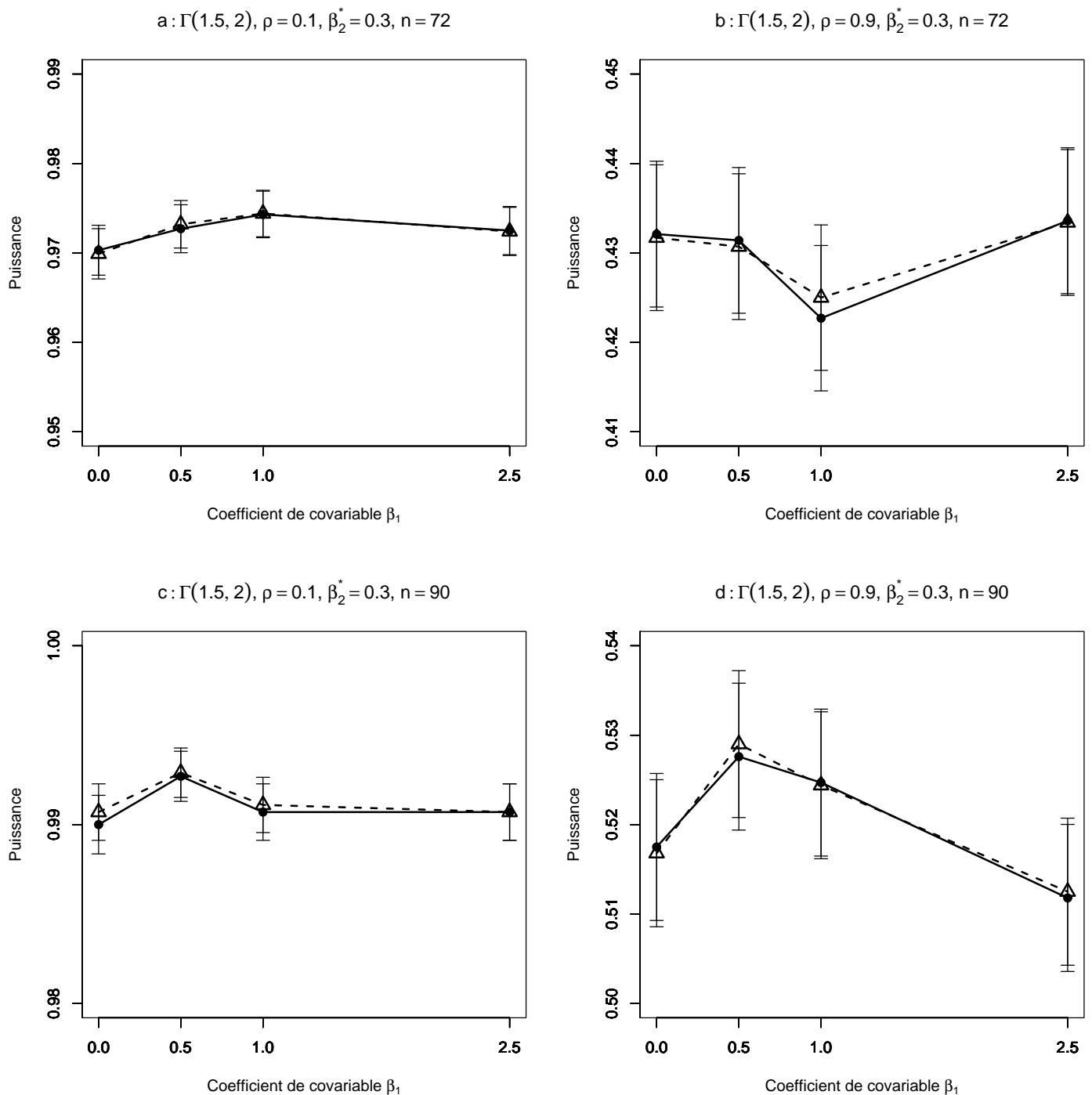


FIG. 4.61 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne— \triangle — et Freedman et Lane— \bullet —, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(1.5,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

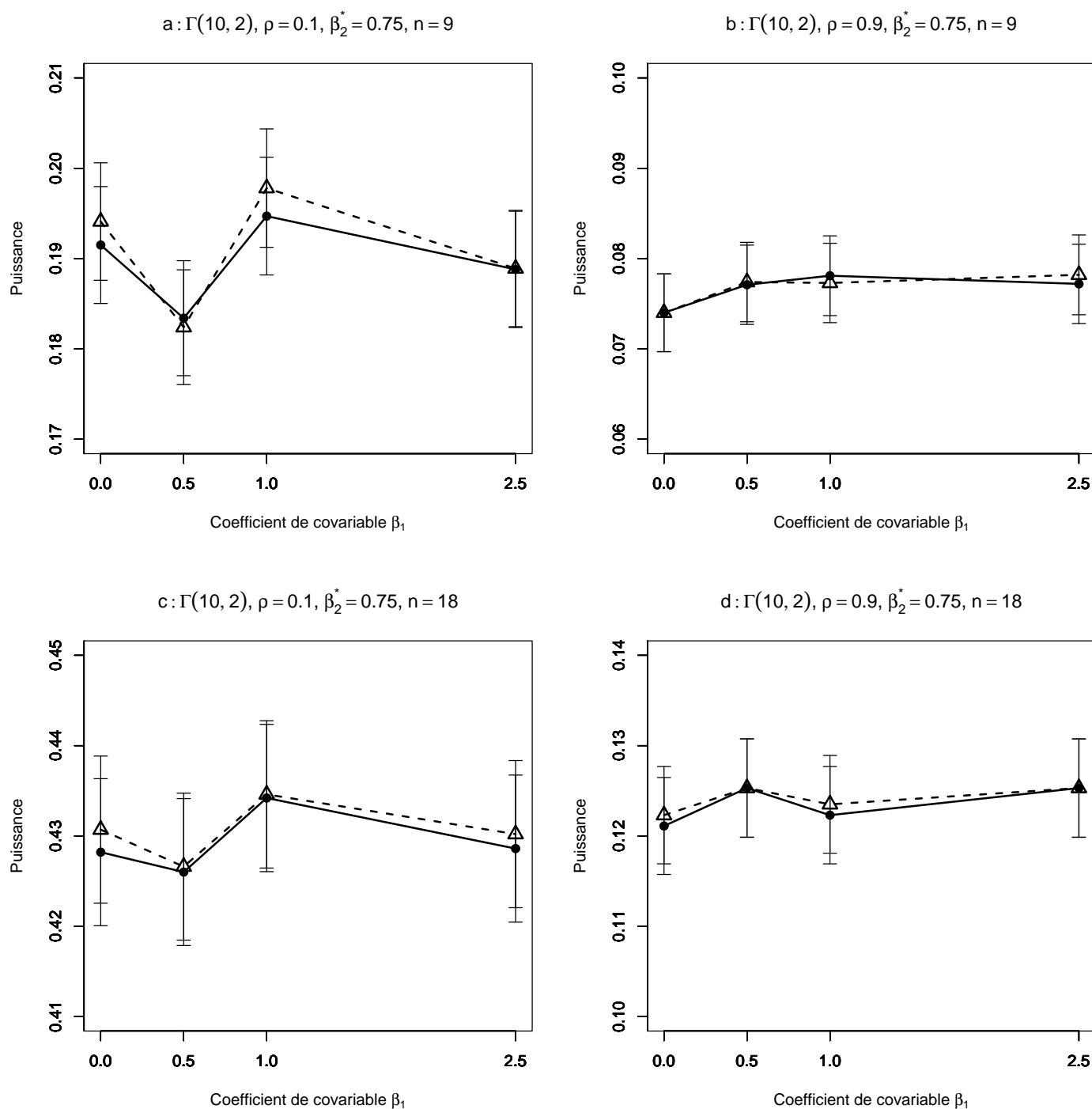


FIG. 4.62 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(10,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 9$; figure b : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 9$; figure c : $\rho = 0.1$, $\beta_2^* = 0.75$, $n = 18$, figure d : $\rho = 0.9$, $\beta_2^* = 0.75$, $n = 18$.

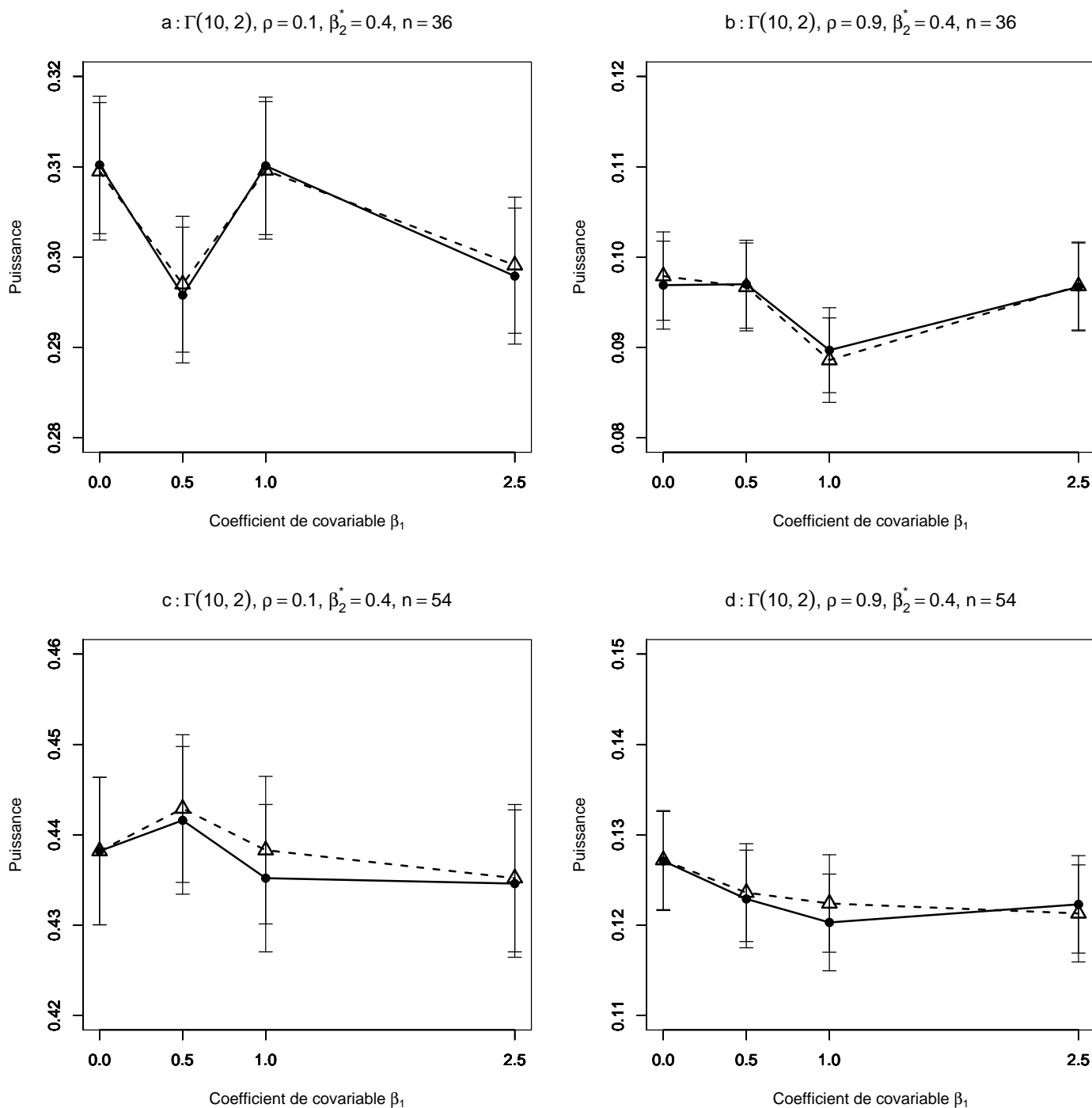


FIG. 4.63 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(10,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 36$; figure b : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 36$; figure c : $\rho = 0.1$, $\beta_2^* = 0.4$, $n = 54$, figure d : $\rho = 0.9$, $\beta_2^* = 0.4$, $n = 54$.

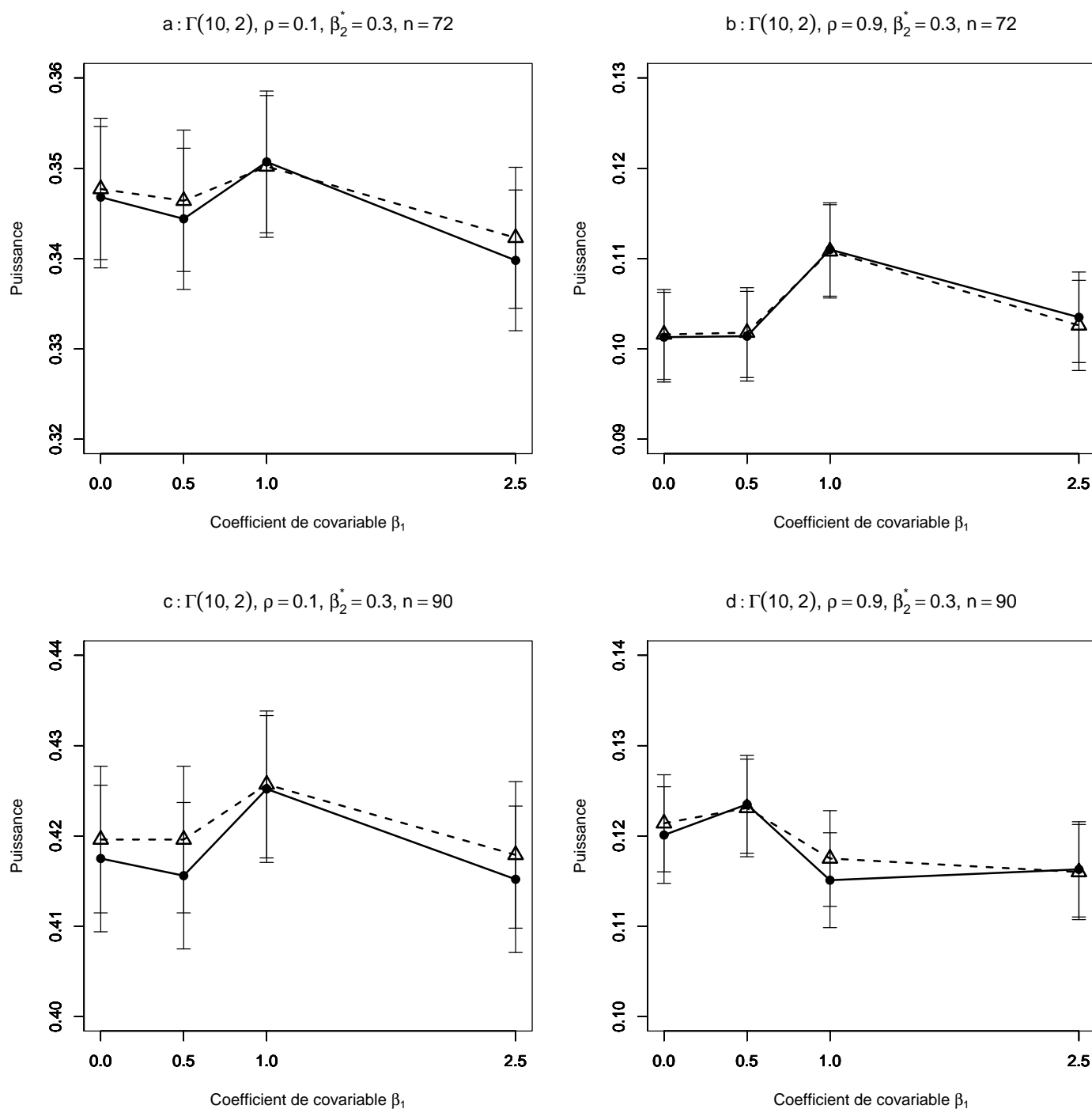


FIG. 4.64 – Puissance estimée du T-test de $H_0 : \beta_2 = \beta_2^*$ par les méthodes Gaussienne $-\triangle-$ et Freedman et Lane $-\bullet-$, en fonction de la taille du coefficient de covariable. 10000 n-échantillons du résidu ε ont été simulés suivant une loi Gamma(10,2), X^1 et X^2 sont toujours Gaussiennes $N(0,1)$. On a retenu $\beta_1=0, 0.5, 1, 2.5$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 72$; figure b : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 72$; figure c : $\rho = 0.1$, $\beta_2^* = 0.3$, $n = 90$, figure d : $\rho = 0.9$, $\beta_2^* = 0.3$, $n = 90$.

Annexe 3

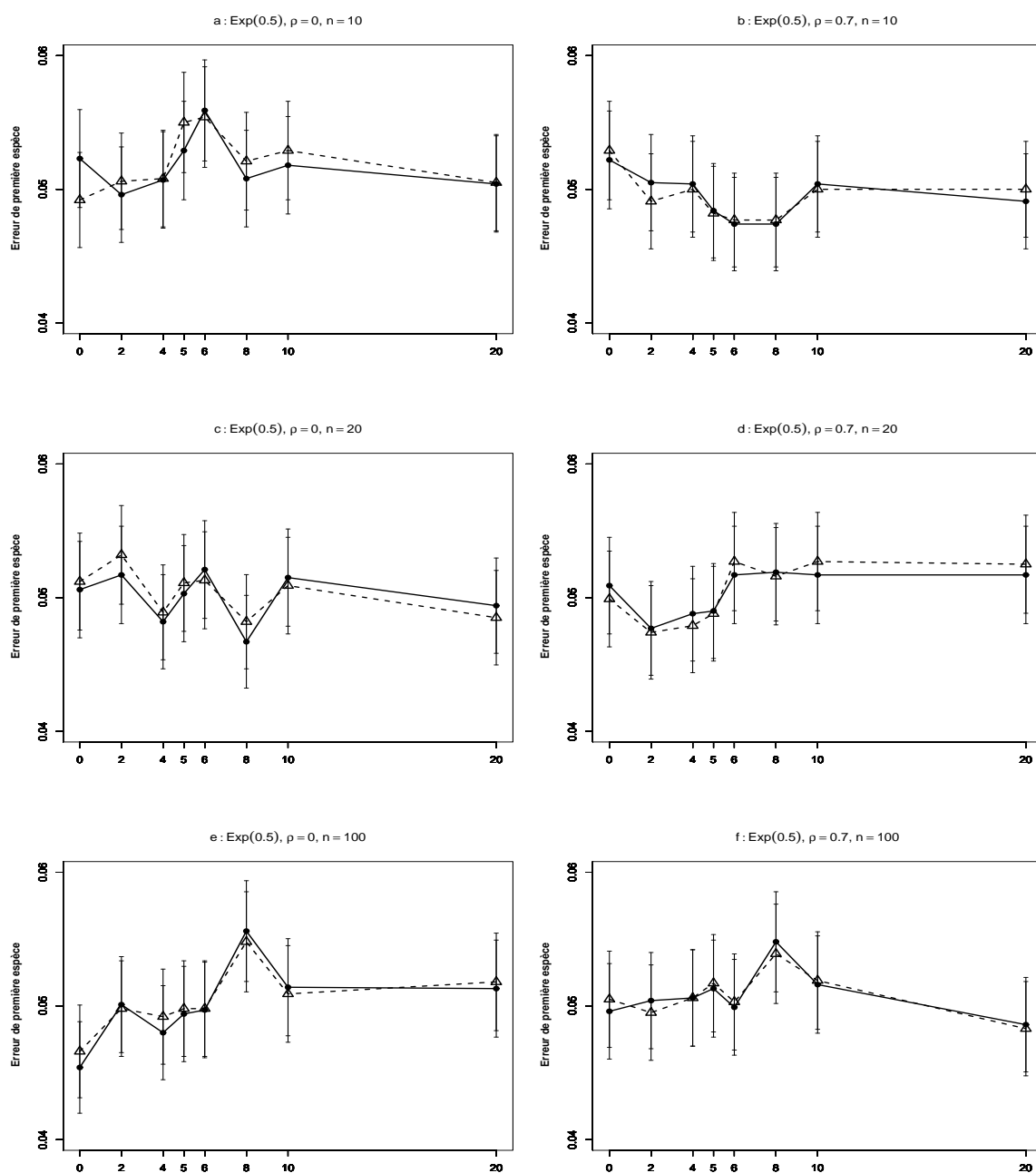


FIG. 4.65 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $-\triangle-$, et Freedman et Lane $-\bullet-$, en fonction du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Exp(0.5)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$ où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0, n = 10$; figure b : $\rho = 0.7, n = 10$; figure c : $\rho = 0, n = 20$; d : $\rho = 0.7, n = 20$; figure e : $\rho = 0, n = 100$; figure f : $\rho = 0.7, n = 100$.

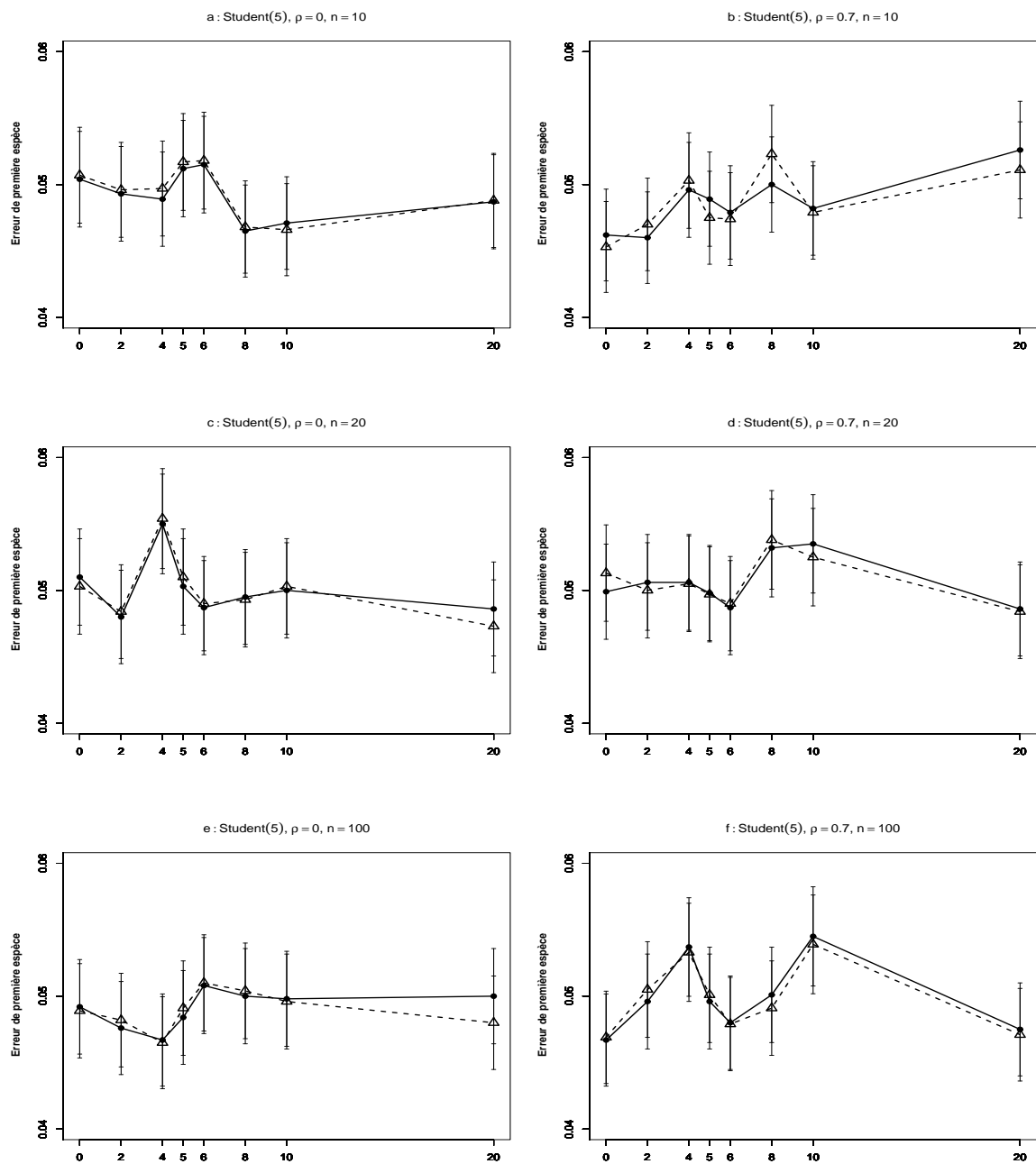


FIG. 4.66 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $--\triangle--$, et Freedman et Lane $-\bullet-$, en fonction du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de *Student*(5), X^1 et X^2 sont toujours Gaussiennes (0,1) où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

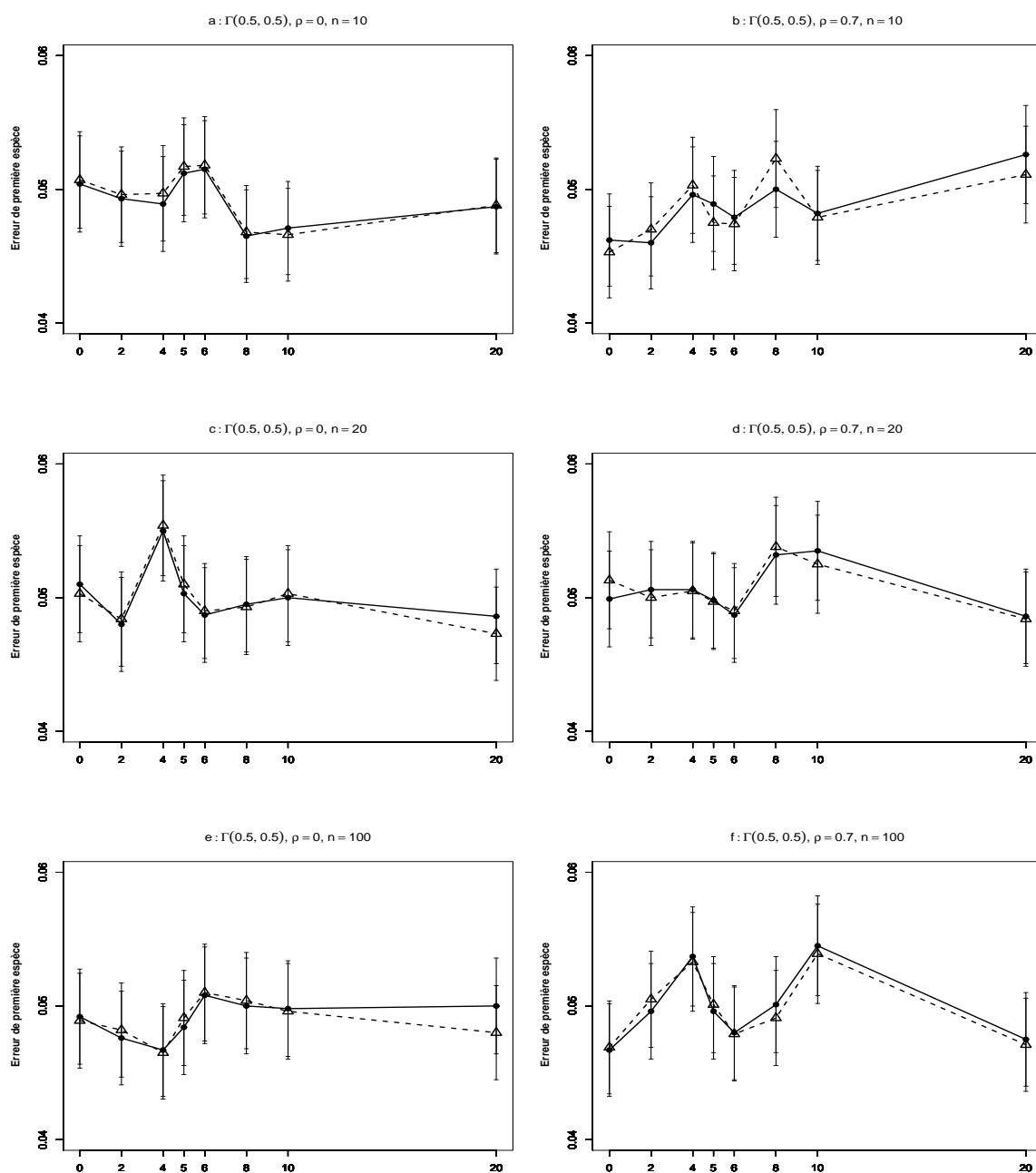


FIG. 4.67 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $\text{---}\triangle\text{---}$, et Freedman et Lane $\text{---}\bullet\text{---}$, en fonction du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $\Gamma(0.5, 0.5)$, X^1 et X^2 sont toujours Gaussiennes (0,1) où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0, n = 10$; figure b : $\rho = 0.7, n = 10$; figure c : $\rho = 0, n = 20$; d : $\rho = 0.7, n = 20$; figure e : $\rho = 0, n = 100$; figure f : $\rho = 0.7, n = 100$.

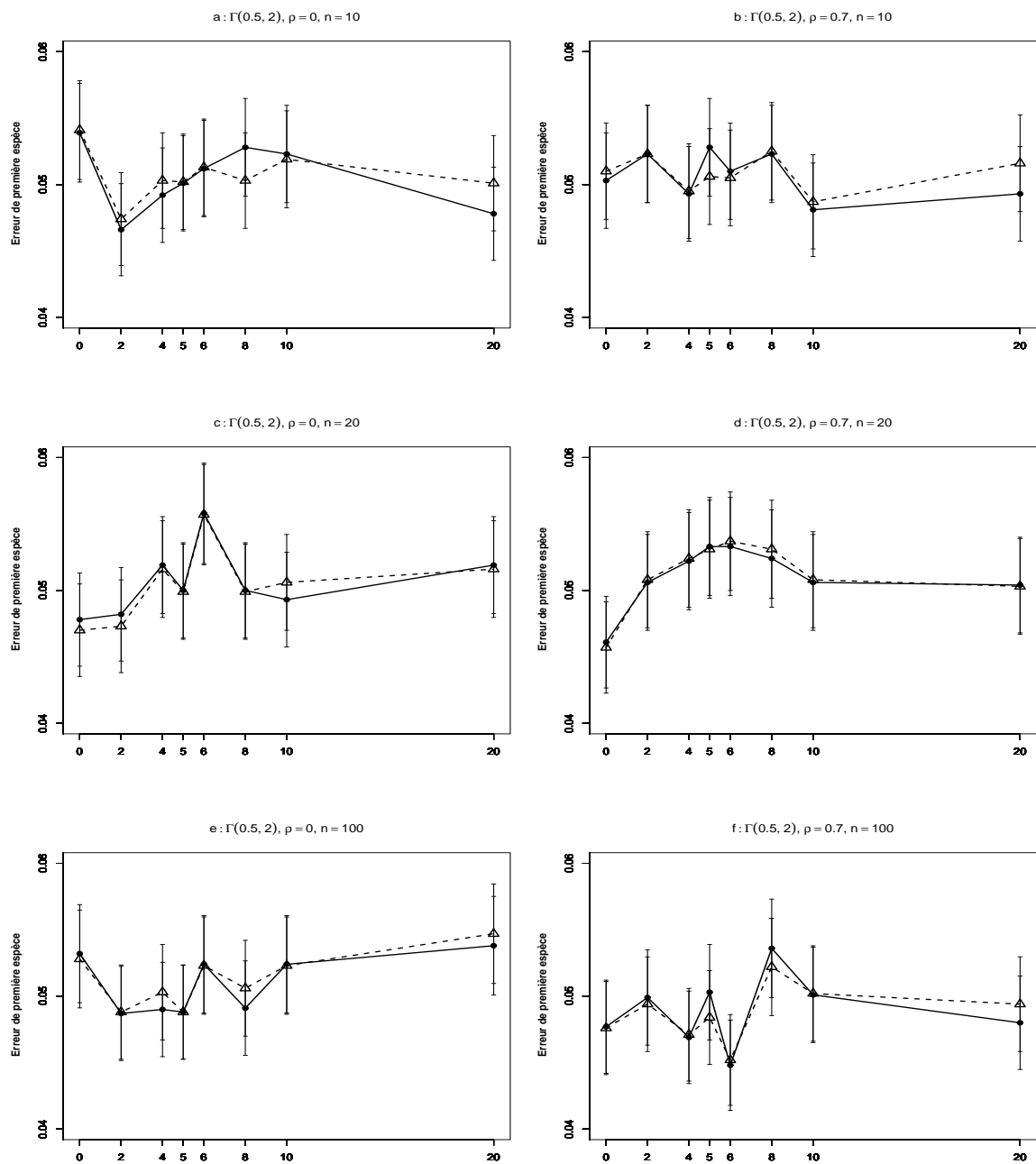


FIG. 4.68 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $--\triangle--$, et Freedman et Lane $-\bullet-$, en fonction du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $\Gamma(0.5, 2)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$ où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

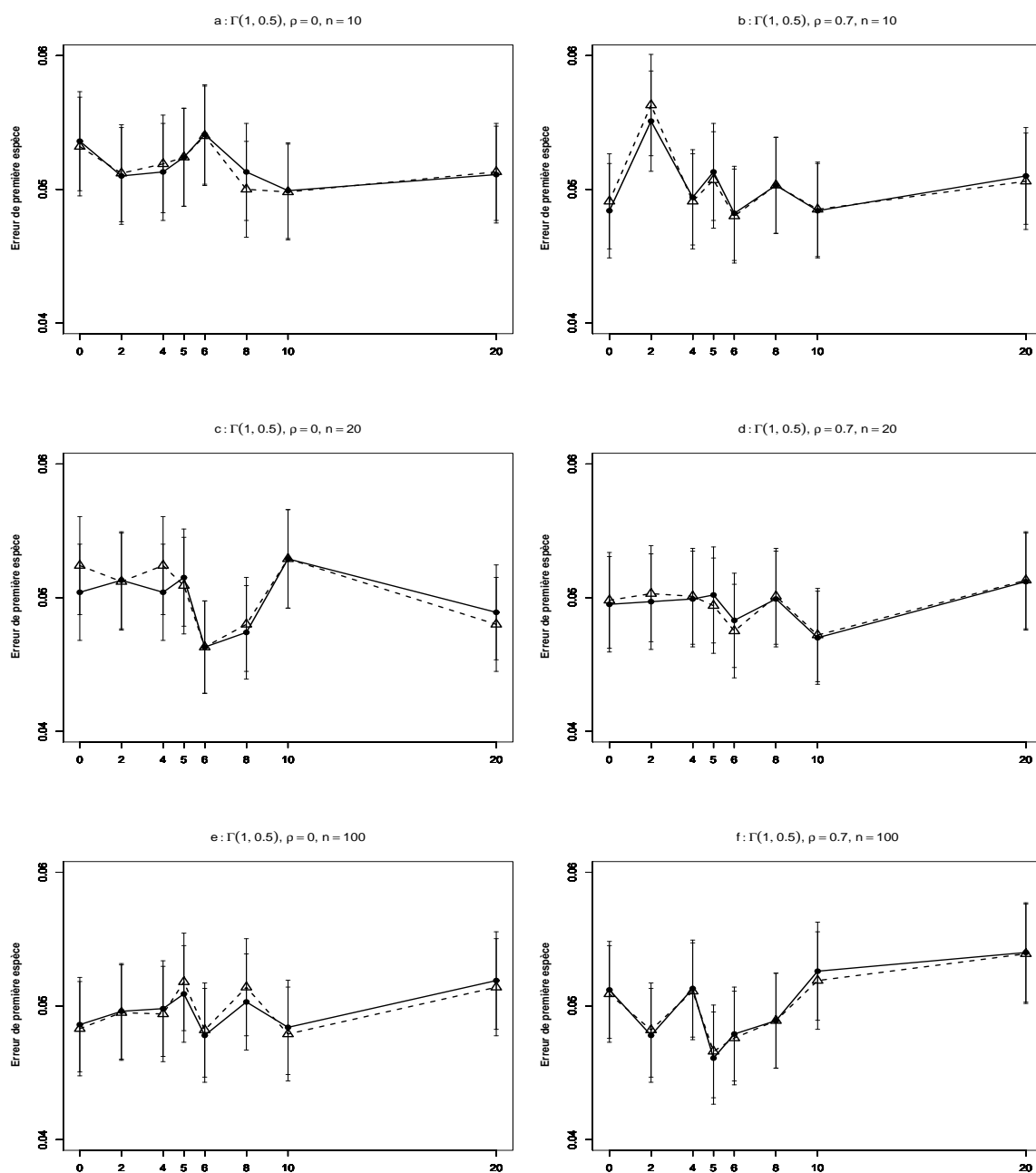


FIG. 4.69 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $\text{---}\triangle\text{---}$, et Freedman et Lane $\text{---}\bullet\text{---}$, en fonction du coefficient de covariable. 10000 n -échantillons du résidu ϵ ont été simulés suivant une loi de $\Gamma(1, 0.5)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$ où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

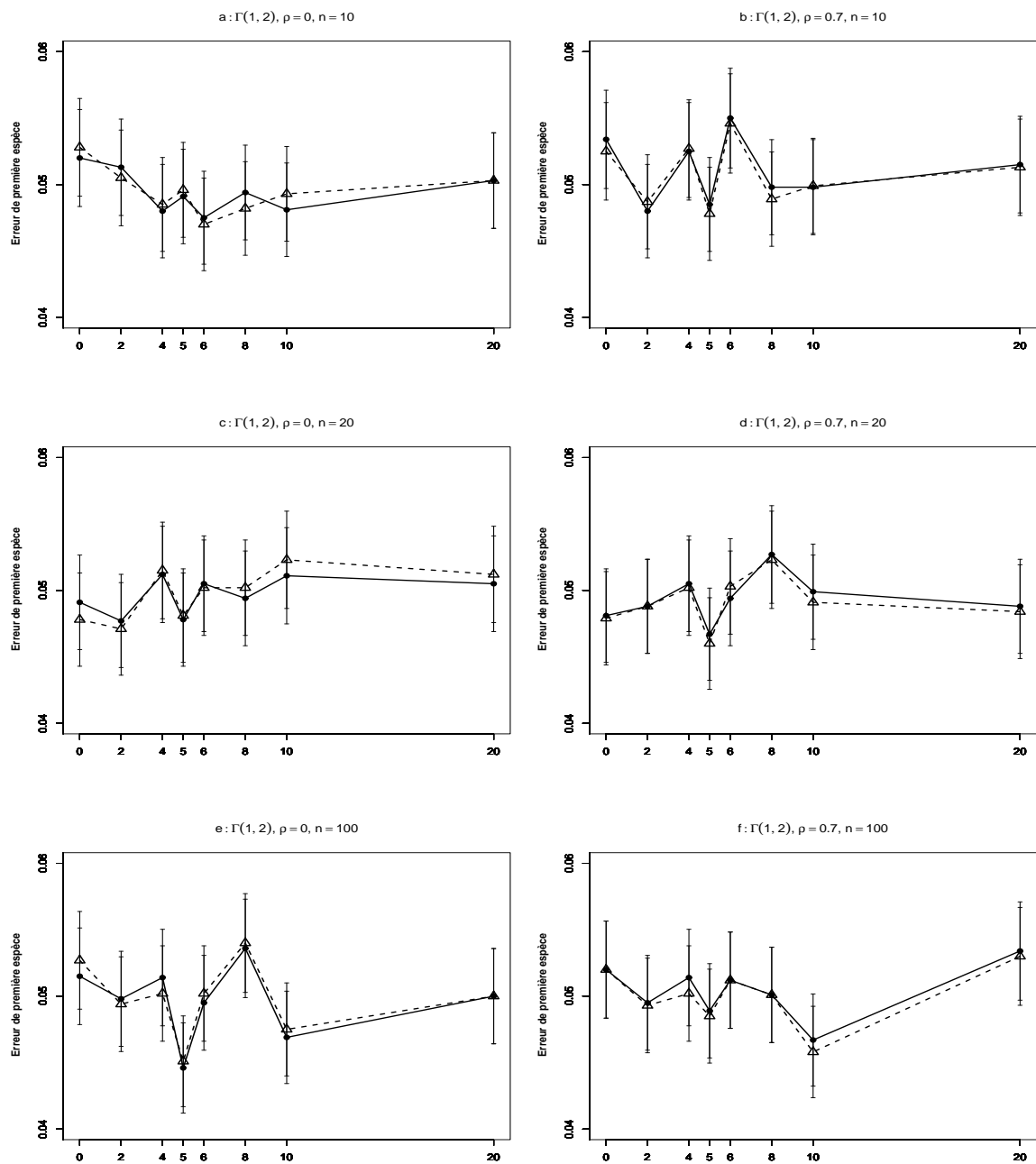


FIG. 4.70 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $--\triangle--$, et Freedman et Lane $-\bullet-$, en fonction du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $\Gamma(1, 2)$, X^1 et X^2 sont toujours Gaussiennes (0,1) où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0, n = 10$; figure b : $\rho = 0.7, n = 10$; figure c : $\rho = 0, n = 20$; d : $\rho = 0.7, n = 20$; figure e : $\rho = 0, n = 100$; figure f : $\rho = 0.7, n = 100$.

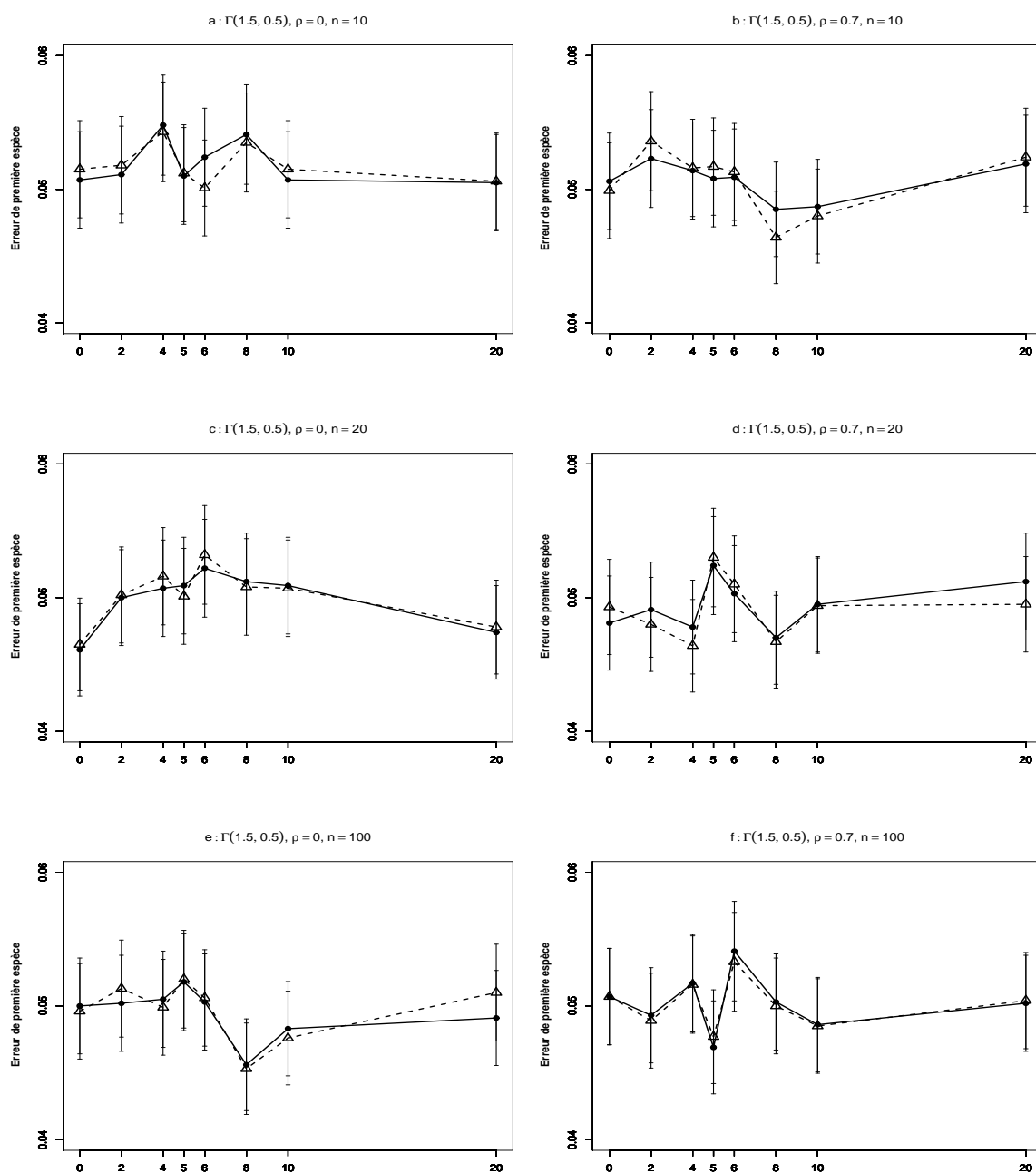


FIG. 4.71 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $-\triangle-$, et Freedman et Lane $-\bullet-$, en fonction du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $\Gamma(1.5, 0.5)$, X^1 et X^2 sont toujours Gaussiennes (0,1) où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0$, $n = 10$; figure b : $\rho = 0.7$, $n = 10$; figure c : $\rho = 0$, $n = 20$; d : $\rho = 0.7$, $n = 20$; figure e : $\rho = 0$, $n = 100$; figure f : $\rho = 0.7$, $n = 100$.

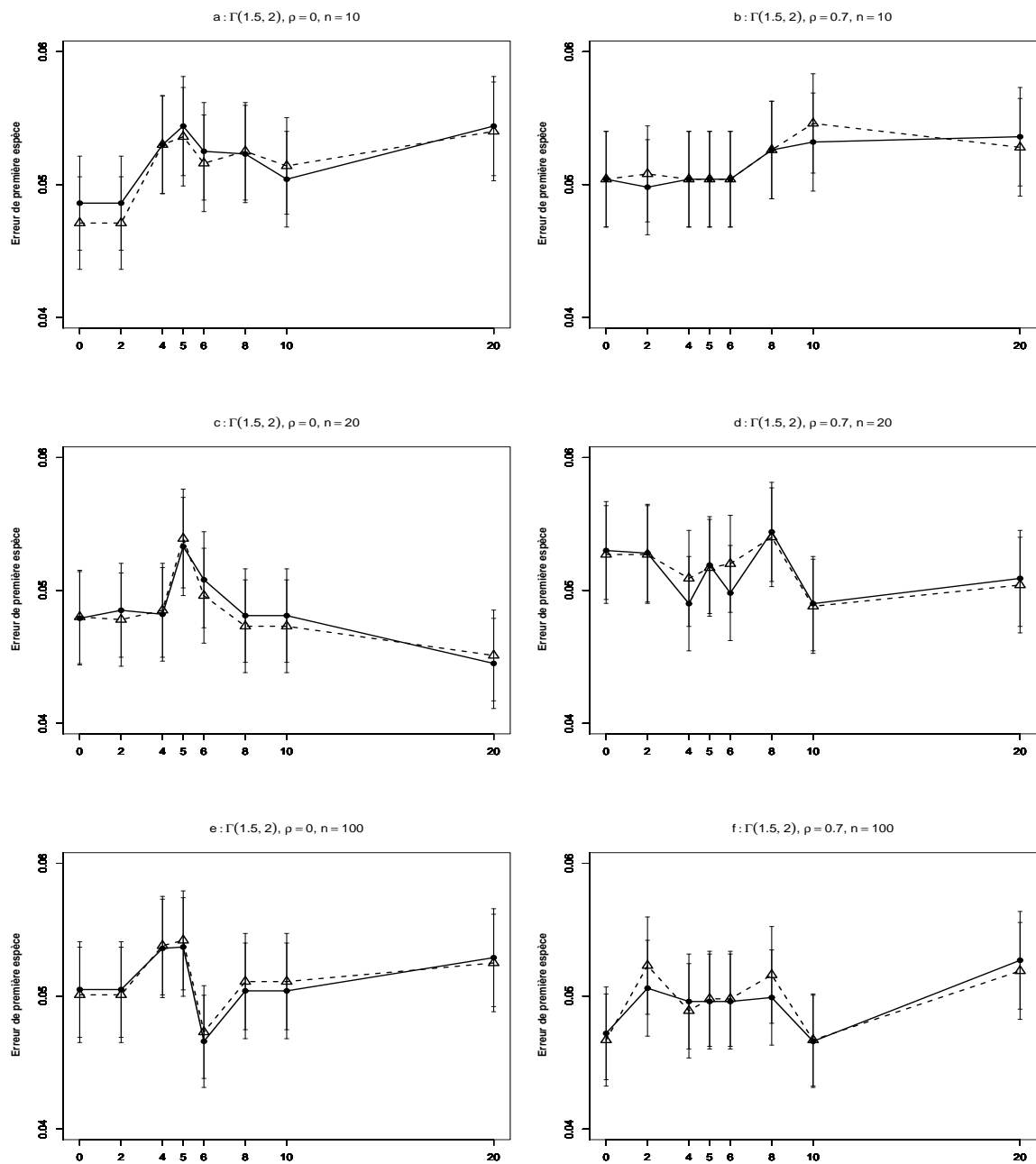


FIG. 4.72 – Erreur de première espèce estimée du T-test de $H_0 : \beta_2 = 0$ par les méthodes Gaussienne $--\triangle--$, et Freedman et Lane $—\bullet—$, en fonction du coefficient de covariable. 10000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $\Gamma(1.5, 2)$, X^1 et X^2 sont toujours Gaussiennes $(0,1)$ où X^1 a un point extrême. On a retenu $\beta_1 = 0, 2, 4, 5, 6, 8, 10, 20$. Les segments verticaux figurent l'intervalle de confiance à 95%. Figure a : $\rho = 0, n = 10$; figure b : $\rho = 0.7, n = 10$; figure c : $\rho = 0, n = 20$; d : $\rho = 0.7, n = 20$; figure e : $\rho = 0, n = 100$; figure f : $\rho = 0.7, n = 100$.

Annexe 4

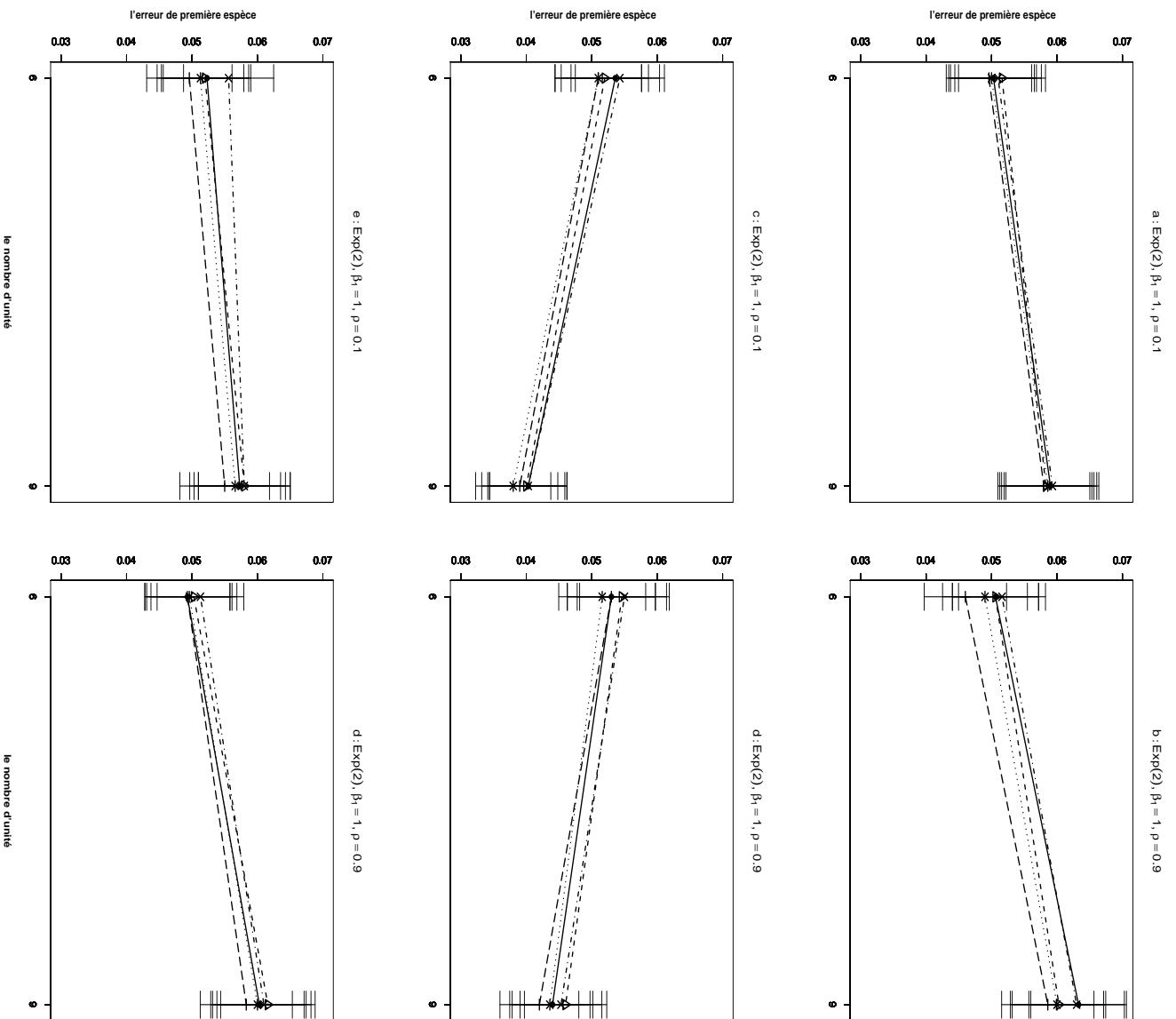


FIG. 4.73 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— \triangle —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Brack— $+$ — et Manly— \ast —. 30000 n-échantillons du résidu e ont été simulés suivant une loi de $\text{Exp}(2)$ et vérifient les propriétés d'une matrice de distance. La statistique de test est le T de Student pour toutes les figures. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures a , c et e . $\rho = 0.9$ pour les figures b , d et f . X^1 et X^2 sont des distances interpoints pour les figures a et b ; des carrés des distances pour les figures c et d ; des transformées logarithmiques des distances pour les figures e et f .

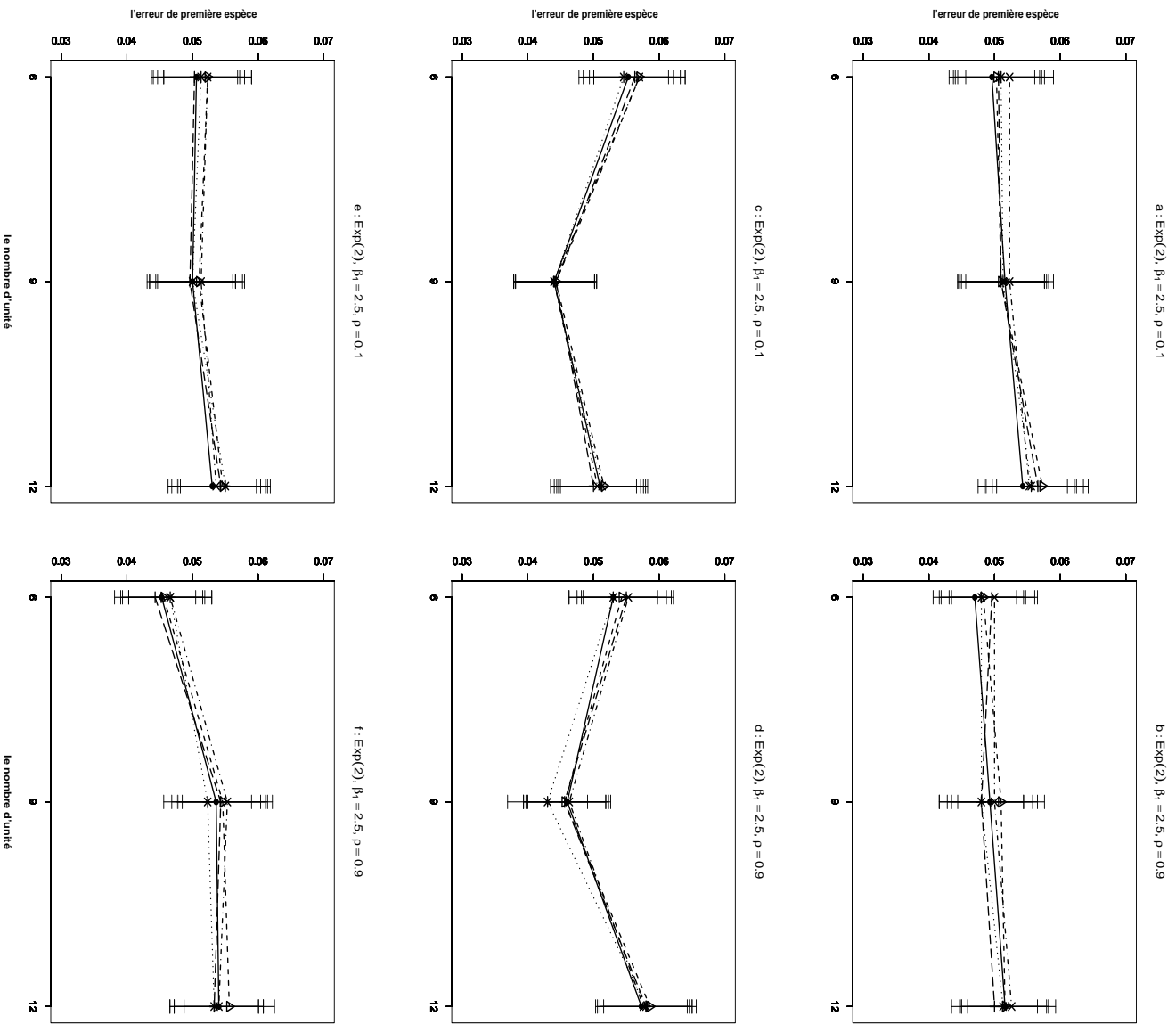


FIG. 4.74 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— \triangle —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Brack— $+$ — et Manly— \dots — $*$ — \dots . 30000 n-échantillons du résidu e ont été simulés suivant une loi de $\text{Exp}(2)$ et vérifient les propriétés d'une matrice de distance. La statistique de test est le T de Student pour toutes les figures. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures a , c et e . $\rho = 0.9$ pour les figures b , d et f . X^1 et X^2 sont des distances interpoints pour les figures a et b ; des carrés des distances pour les figures c et d ; des transformées logarithmiques des distances pour les figures e et f .

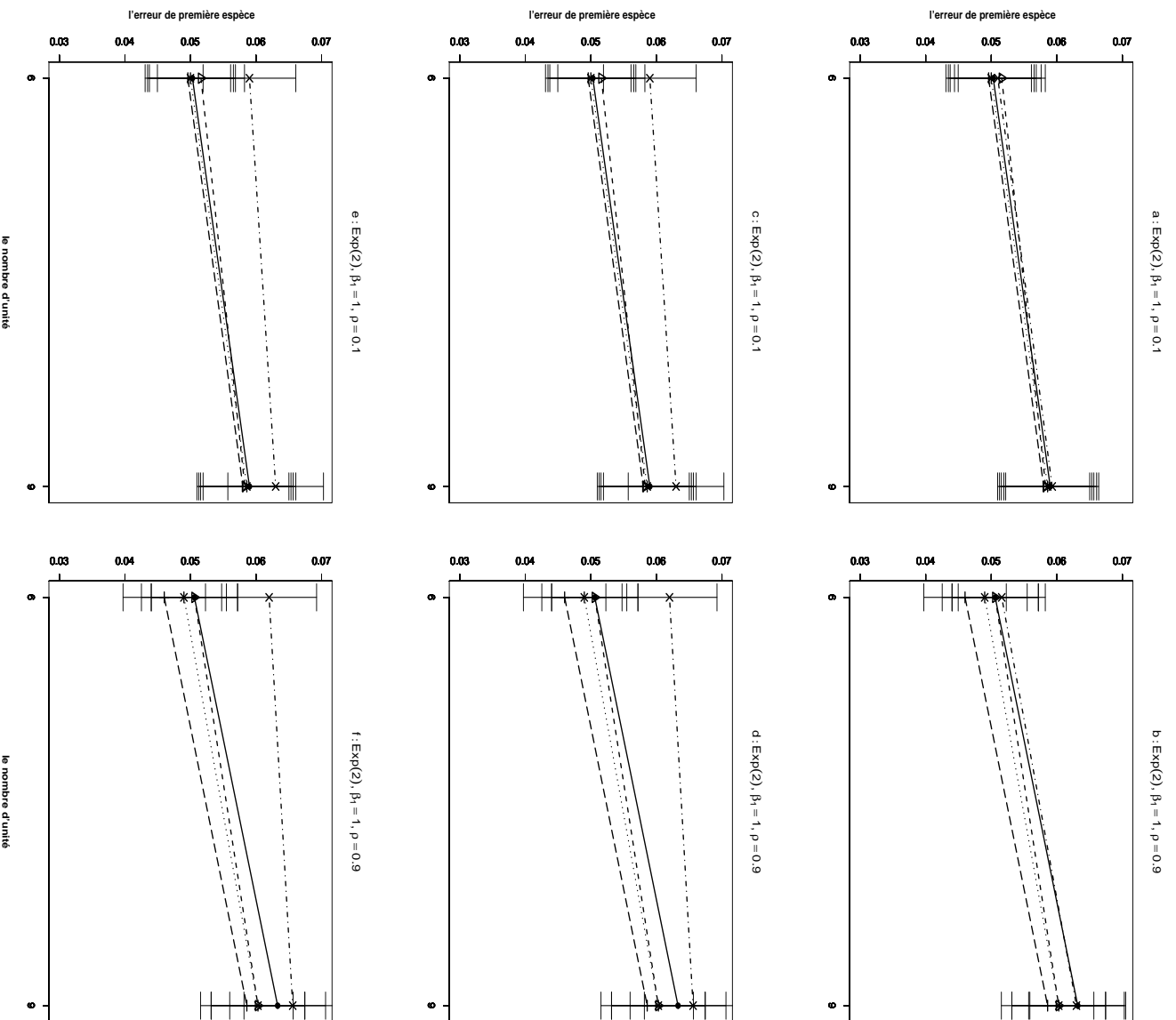


FIG. 4.75 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Brack— $+$ — et Manly— \dots —*. 30000 n-échantillons du résidu e ont été simulés suivant une loi de $Exp(2)$ et vérifient les propriétés d'une matrice de distance. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures a, c et e . $\rho = 0.9$ pour les figures b, d et f . X^1 et X^2 sont des distances interpolés pour toutes les figures. La statistique de test est : le T de Student pour les figures a et b ; le coefficient de corrélation partielle pour les figures c et d ; carrée du coefficient de corrélation partielle pour les figures e et f .

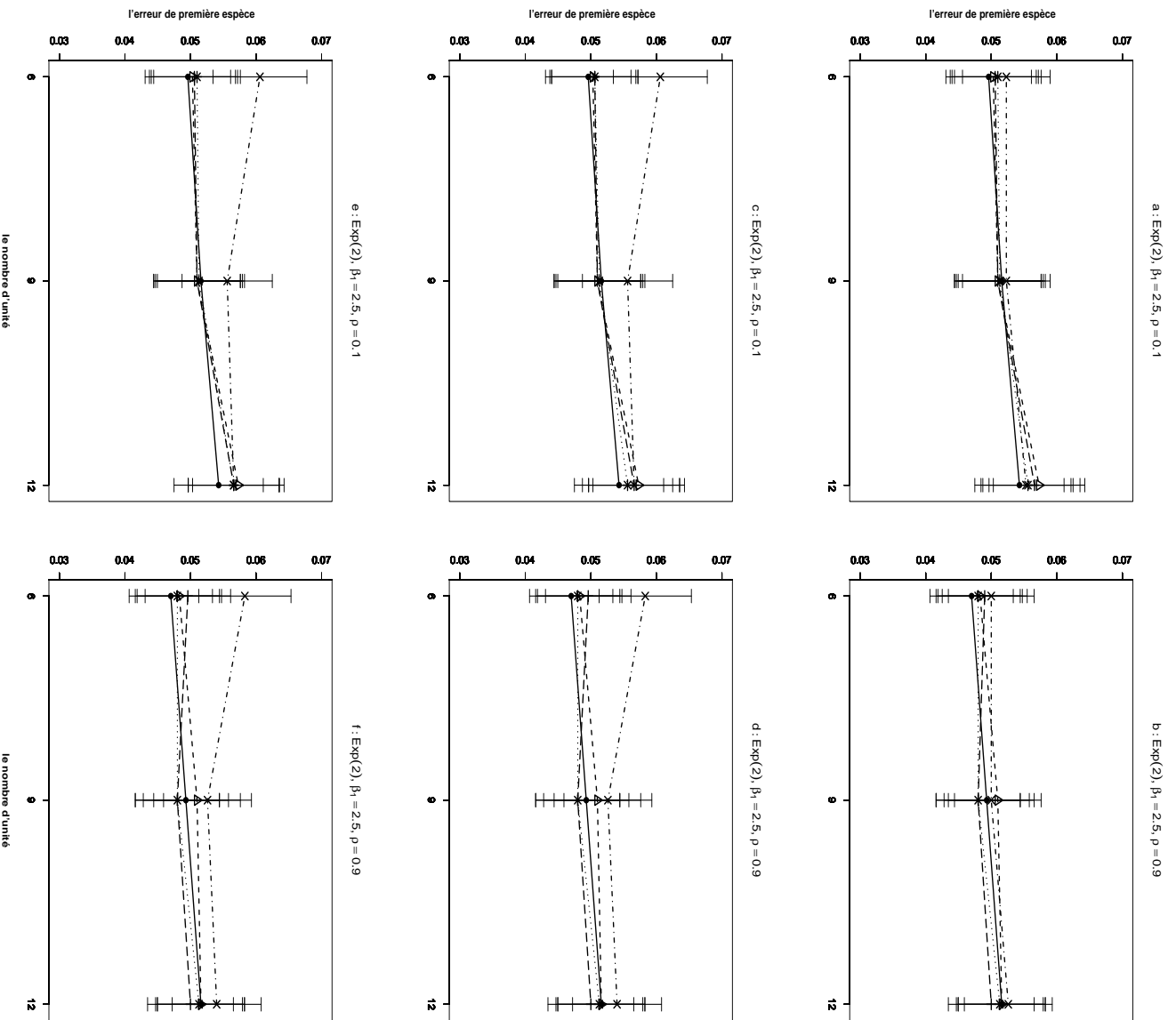
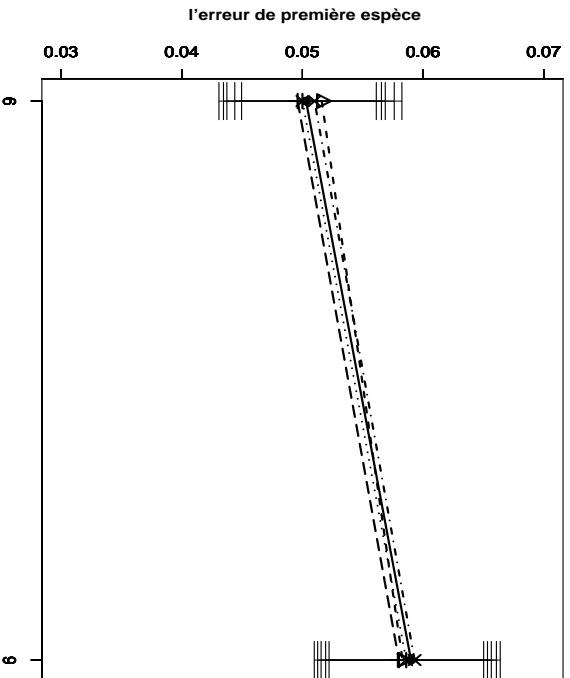
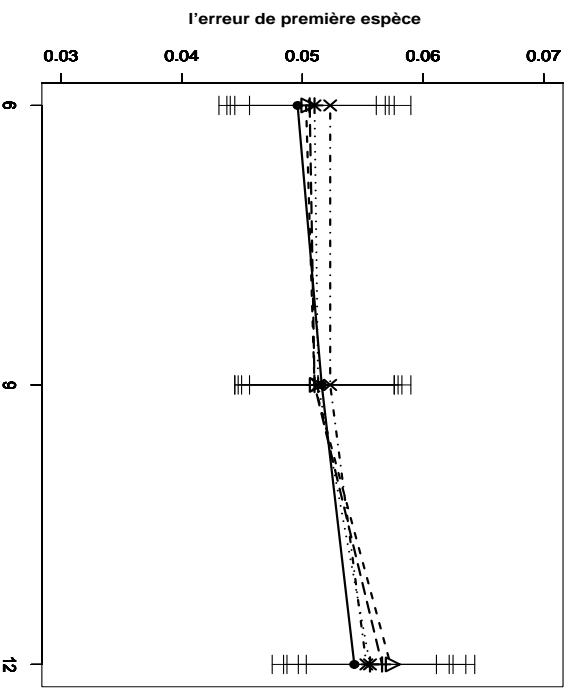


FIG. 4.76 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— \triangle —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Brack— $+$ — et Manly— \dots —*— \dots . 30000 n-échantillons du résidu e ont été simulés suivant une loi de $\text{Exp}(2)$ et vérifient les propriétés d'une matrice de distance. $\beta_1 = 1$ pour toutes les figures. $\rho = 0.1$ pour les figures a , c et e . $\rho = 0.9$ pour les figures b , d et f . X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est : le T de Student pour les figures a et b ; le coefficient de corrélation partielle pour les figures c et d ; carrée du coefficient de corrélation partielle pour les figures e et f .

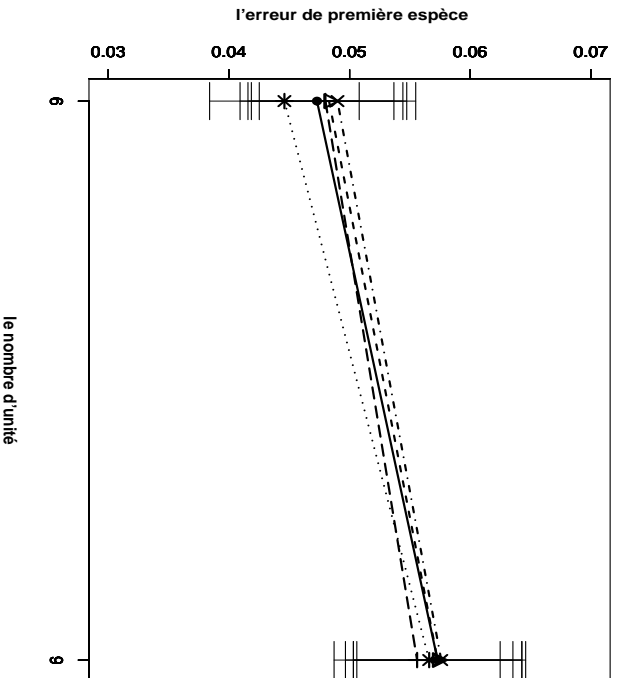
a : $\text{Exp}(2)$, $\rho = 0.1$, $\beta_1 = 1$



b : $\text{Exp}(2)$, $\rho = 0.1$, $\beta_1 = 2.5$



c : $\text{Exp}(2)$, $\rho = 0.1$, $\beta_1 = 1$



d : $\text{Exp}(2)$, $\rho = 0.1$, $\beta_1 = 2.5$

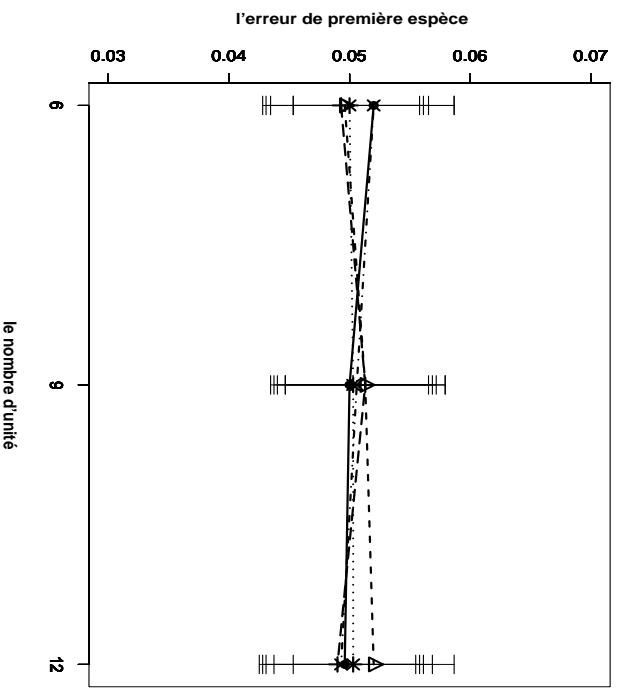


FIG. 4.77 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— $\cdot \cdot \times \cdot$ —; Ter Braek— $+$ — et Manly— $\dots \cdot \dots$. 30000 n-échantillons du résidu e ont été simulés suivant une loi de $\text{Exp}(2)$ et vérifient les propriétés d'une matrice de distance. X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est le T de Student pour toutes figures. $\rho_1 = 0.1$ pour toutes les figures. $\beta_1 = 1$ pour les figures a et c. $\beta_1 = 2.5$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

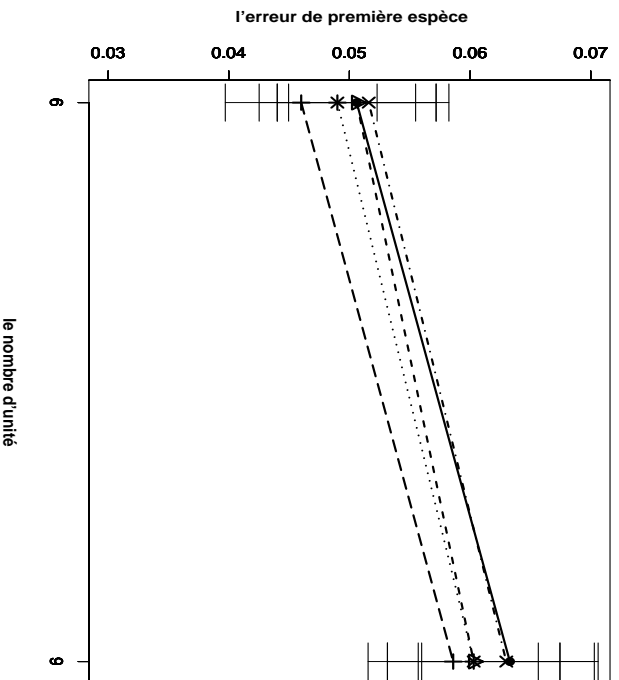
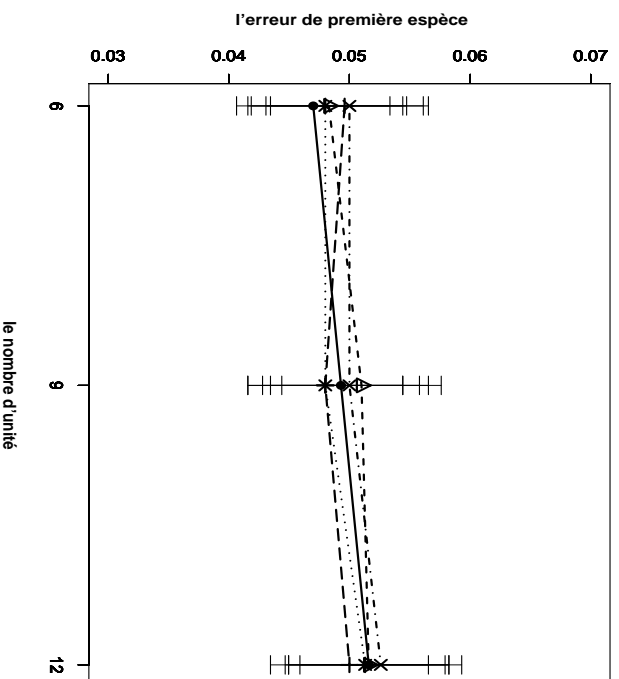
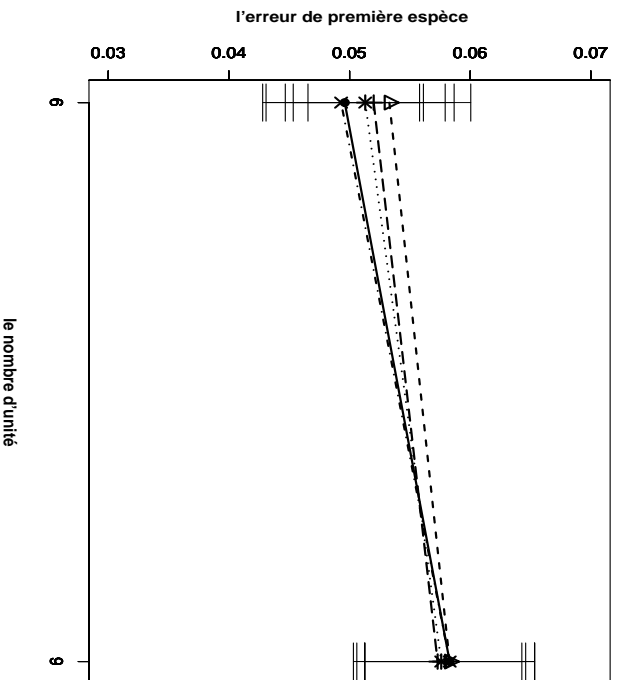
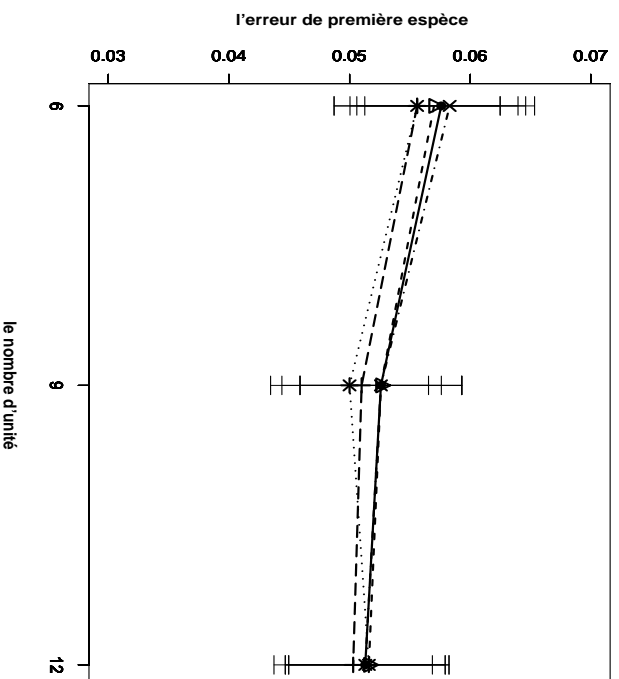
a : Exp(2), $\rho = 0.9$, $\beta_1 = 1$ b : Exp(2), $\rho = 0.9$, $\beta_1 = 2.5$ c : Exp(2), $\rho = 0.9$, $\beta_1 = 1$ d : Exp(2), $\rho = 0.9$, $\beta_1 = 2.5$ 

FIG. 4.78 – Erreur de première espèce estimée du test d'hypothèse $H_0 : \beta_2 = 0$ en fonction du nombre d'unités et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Braek— $+$ — et Manly— \dots — \ast — \dots . 30000 n-échantillons du résidu ϵ ont été simulés suivant une loi de $Exp(2)$ et vérifient les propriétés d'une matrice de distance. X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est le T de Student pour toutes figures. $\rho_1 = 0.9$ pour toutes les figures. $\beta_1 = 1$ pour les figures a et c. $\beta_1 = 2.5$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

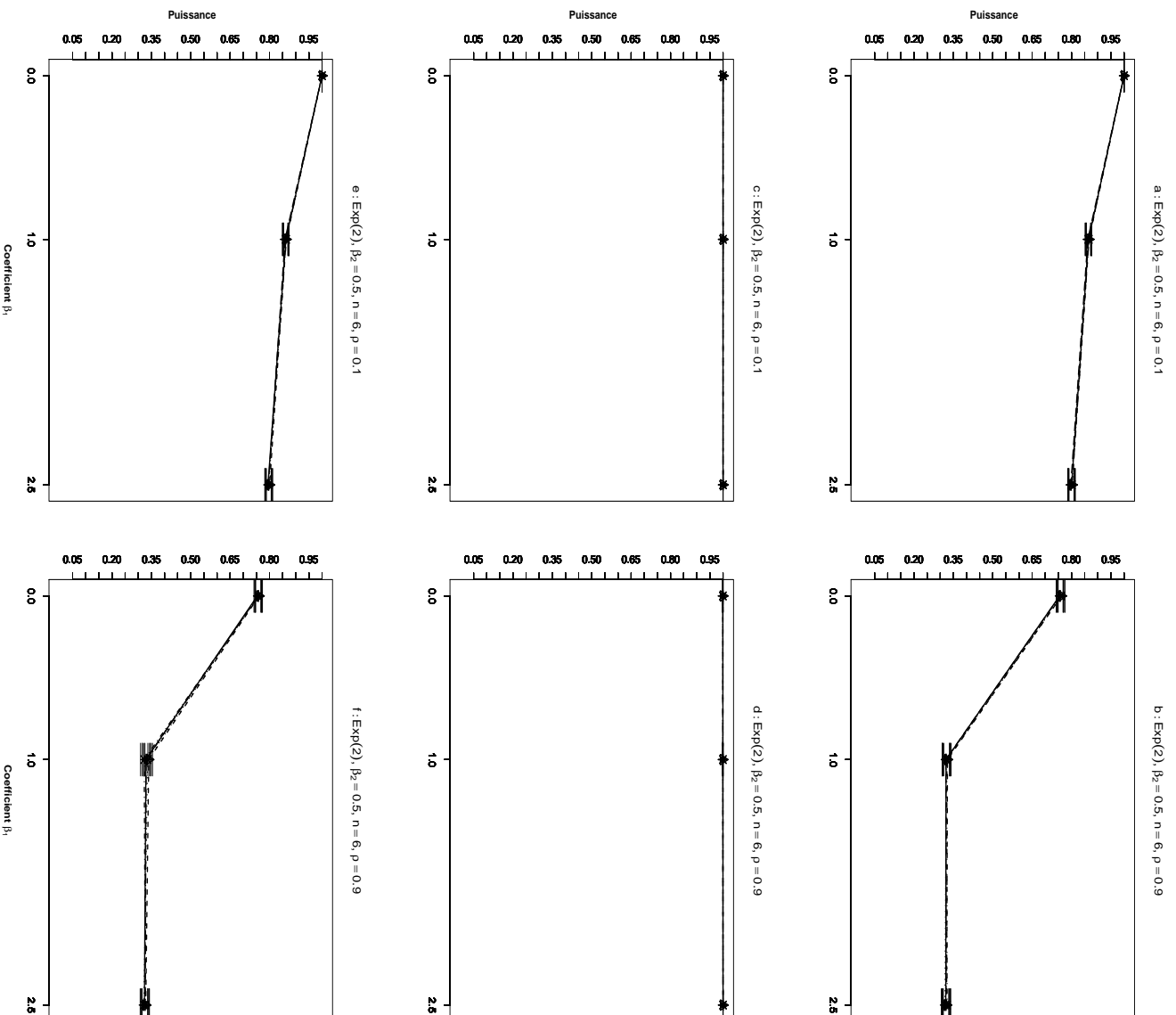


FIG. 4.79 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Braek— $+$ — et Manly— \ast —. 30000 n-échantillons du résidu e ont été simulés suivant une loi de $\text{Exp}(2)$ et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, $\beta_2 = 0.5$ et $n = 6$. $\rho = 0.1$ pour les figures a , c et e . $\rho = 0.9$ pour les figures b , d et f . X^1 et X^2 sont des distances interpoints pour les figures a et b ; des carrés des distances pour les figures c et d ; des transformées logarithmiques des distances pour les figures e et f .

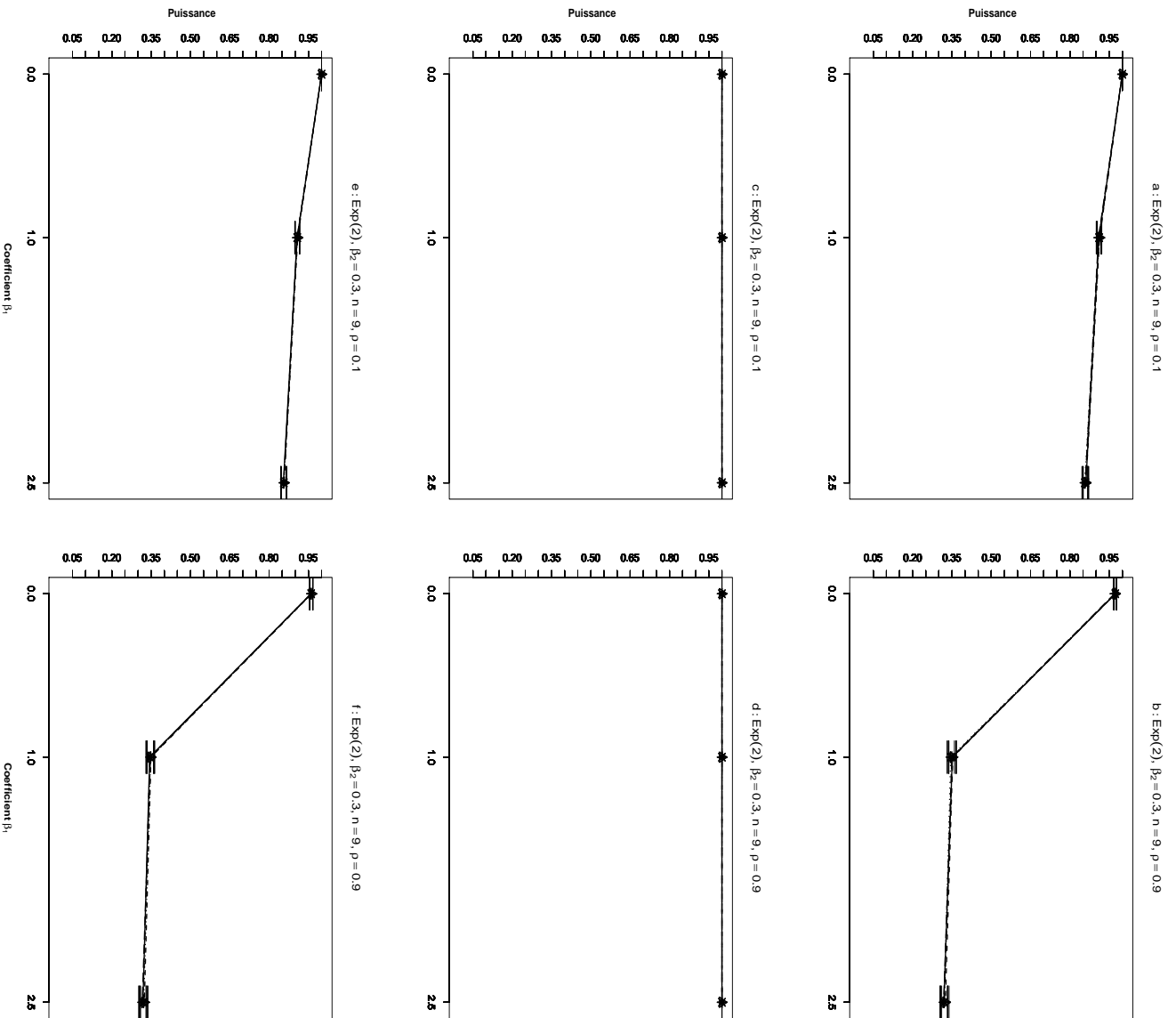


FIG. 4.80 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $--\Delta--$; Freedman et Lane $---\bullet---$; Kennedy $-\cdot\times-\cdot-$; Ter Brack $---+$ et Manly $\dots*\dots$. 30000 n-échantillons du résidu e ont été simulés suivant une loi de $Exp(2)$ et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, $\beta_2 = 0.3$ et $n = 9$. $\rho = 0.1$ pour les figures a , c et e . $\rho = 0.9$ pour les figures b , d et f . X^1 et X^2 sont des distances interpoints pour les figures a et b ; des carrés des distances pour les figures c et d ; des transformées logarithmiques des distances pour les figures e et f .

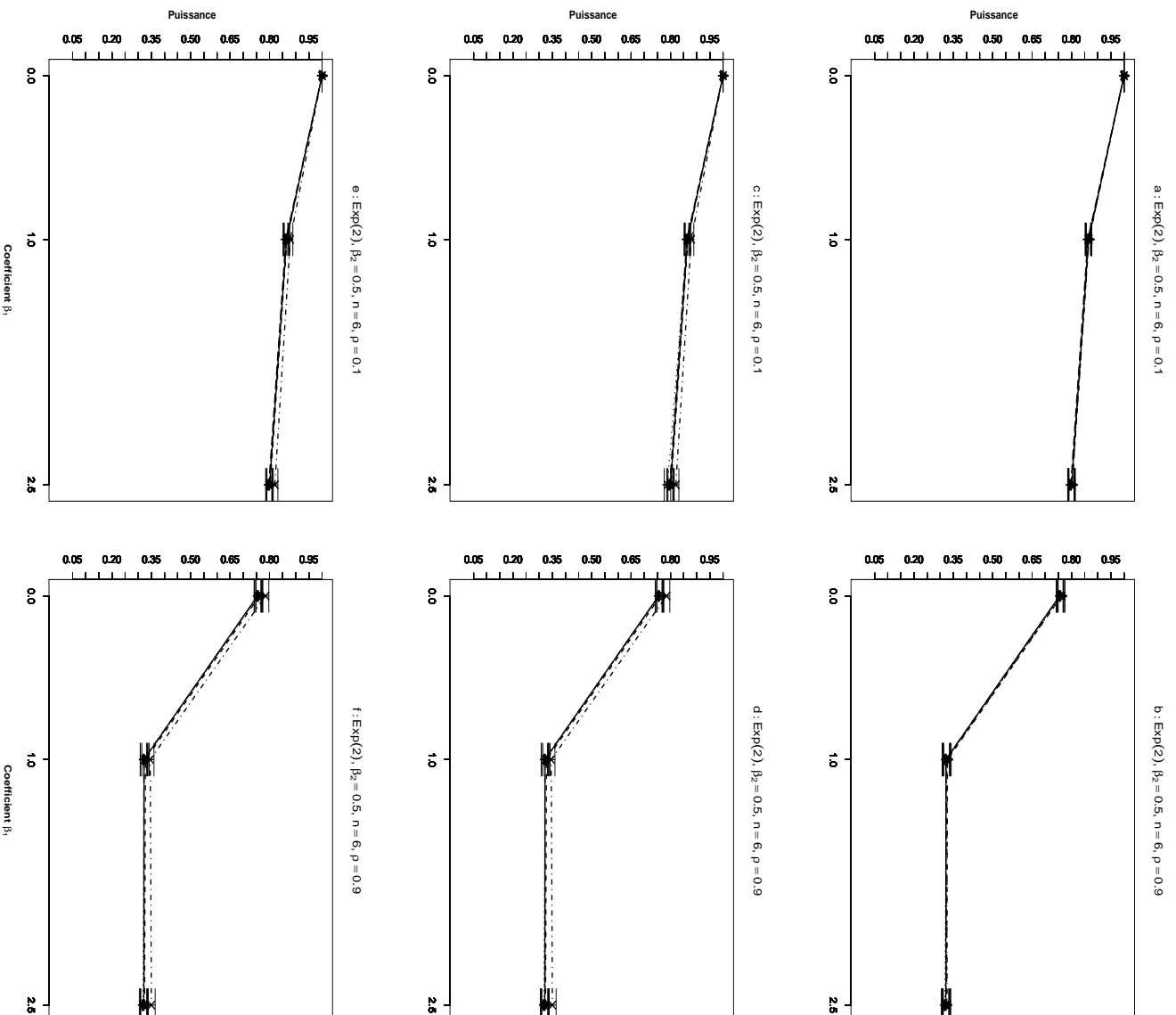


FIG. 4.81 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $\rightarrow \Delta$ Freedman et Lane $\rightarrow \bullet$; Kennedy $\rightarrow \times$; Ter Brack $\rightarrow +$ et Manly $\dots * \dots$. 30000 n-échantillons du résidu e ont été simulés suivant une loi de $Exp(2)$ et vérifient les propriétés d'une matrice de distance. Pour toute les figures $\beta_2 = 0.5$ et $n = 6$. $\rho = 0.1$ pour les figures a , c et e . $\rho = 0.9$ pour les figures b , d et f . X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est : T de Student pour les figures a et b ; le coefficient de corrélation partielle pour les figures c et d ; carrée du coefficient de corrélation partielle pour les figures e et f .

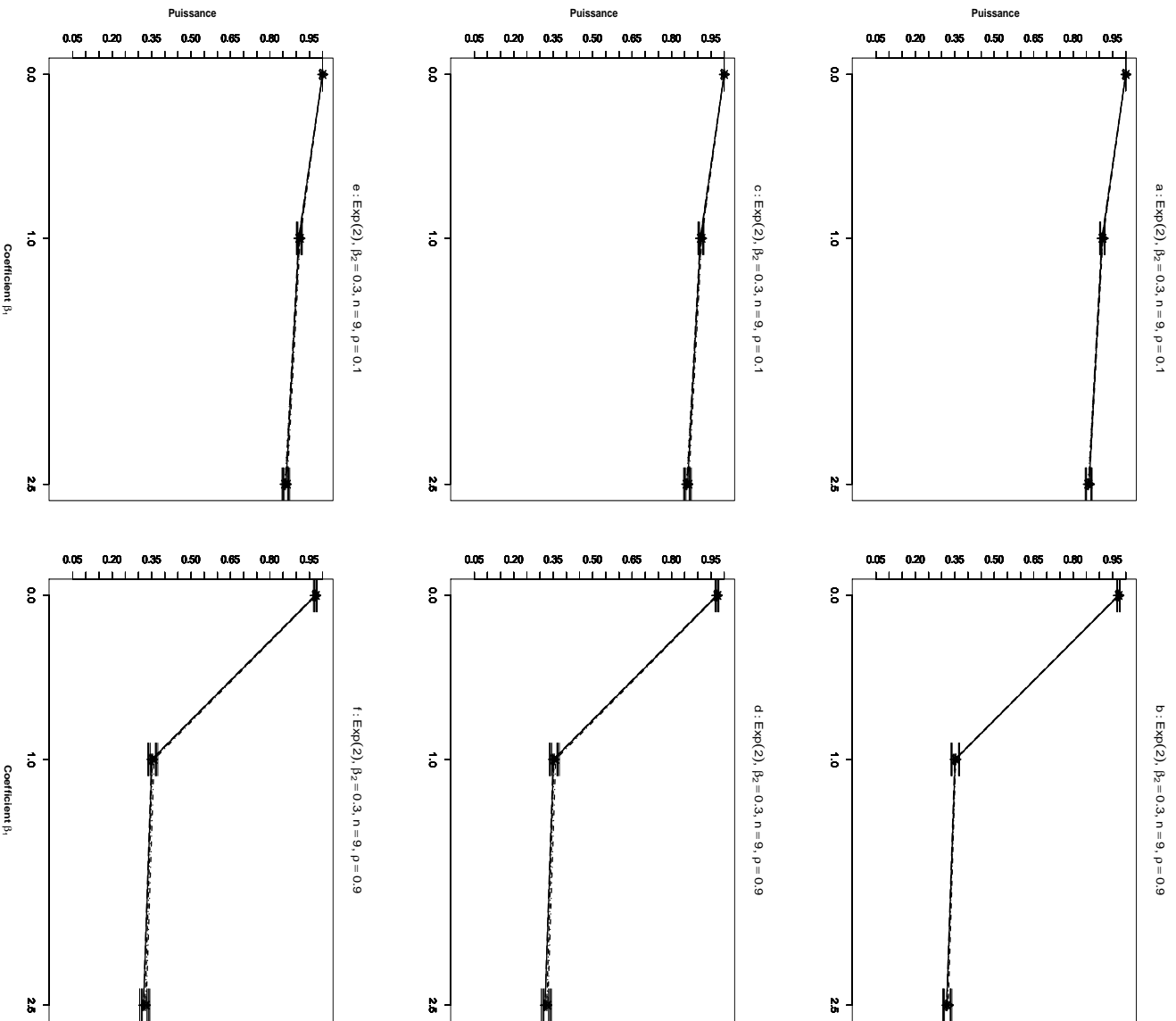


FIG. 4.82 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne— Δ —; Freedman et Lane— \bullet —; Kennedy— \times —; Ter Brack— $+$ — et Manly— \dots — \cdot 30000 n-échantillons du résidu e ont été simulés suivant une loi de $Exp(2)$ et vérifient les propriétés d'une matrice de distance. Pour toute les figures $\beta_2 = 0.3$ et $n = 9$. $\rho = 0.1$ pour les figures a , c et e . $\rho = 0.9$ pour les figures b , d et f . X^1 et X^2 sont des distances interpoints pour toutes les figures. La statistique de test est : T de Student pour les figures a et b ; le coefficient de corrélation partielle pour les figures c et d ; carrée du coefficient de corrélation partielle pour les figures e et f .

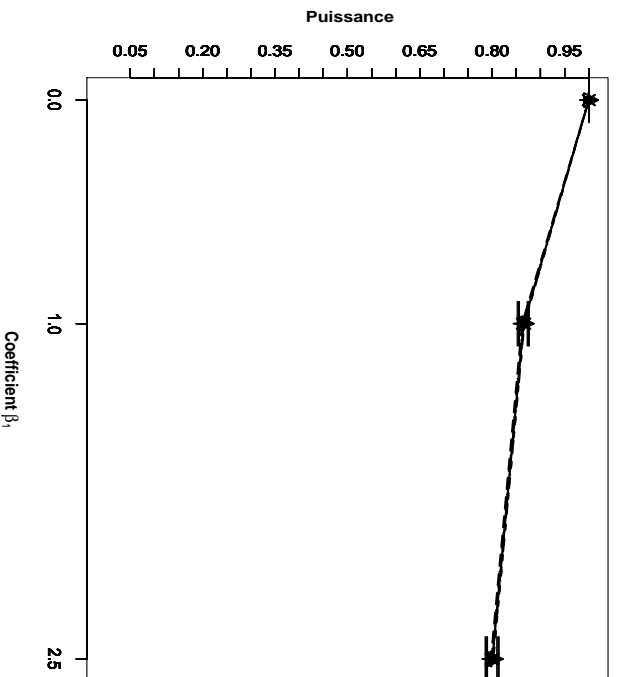
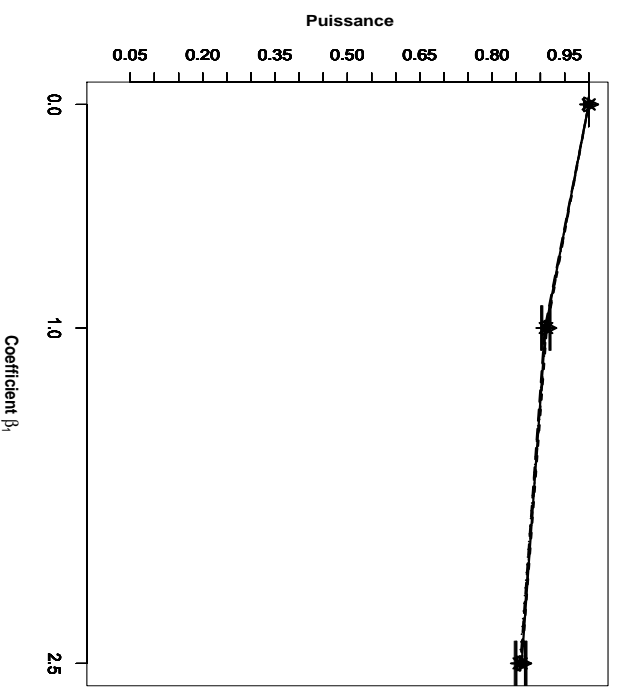
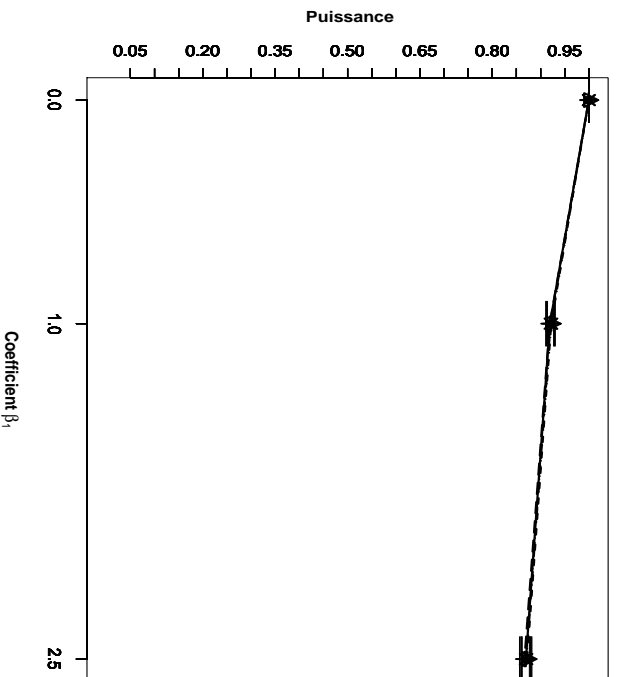
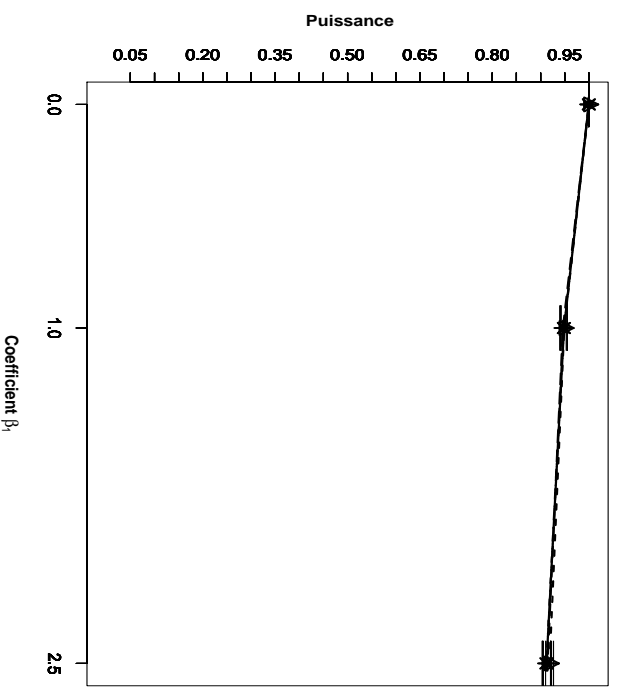
a : Exp(2), $\rho = 0.1$, $\beta_2 = 0.5$, $n = 6$ b : Exp(2), $\rho = 0.1$, $\beta_2 = 0.3$, $n = 9$ c : Exp(2), $\rho = 0.1$, $\beta_2 = 0.5$, $n = 6$ d : Exp(2), $\rho = 0.1$, $\beta_2 = 0.3$, $n = 9$ 

FIG. 4.83 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\Delta-$; Freedman et Lane $-\bullet-$; Kennedy $-\cdot \times \cdot-$; Ter Braek $-\cdot + \cdot-$ et Manly $\dots * \dots$. 30000 n-échantillons du résidu e ont été simulés suivant une loi de $Exp(2)$ et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, X^1 et X^2 sont des distances interpoints et $\rho_1 = 0.1$. $n = 6$, $\beta_1 = 0.5$ pour les figures a et c. $n = 9$, $\beta_1 = 0.3$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

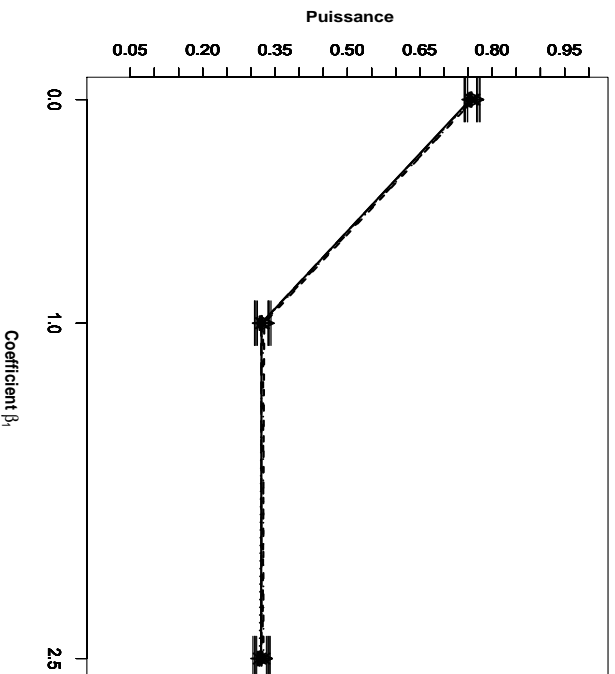
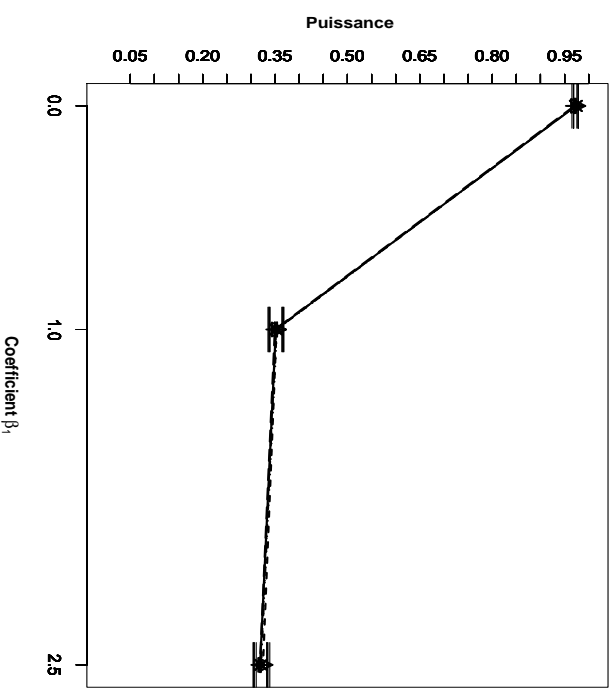
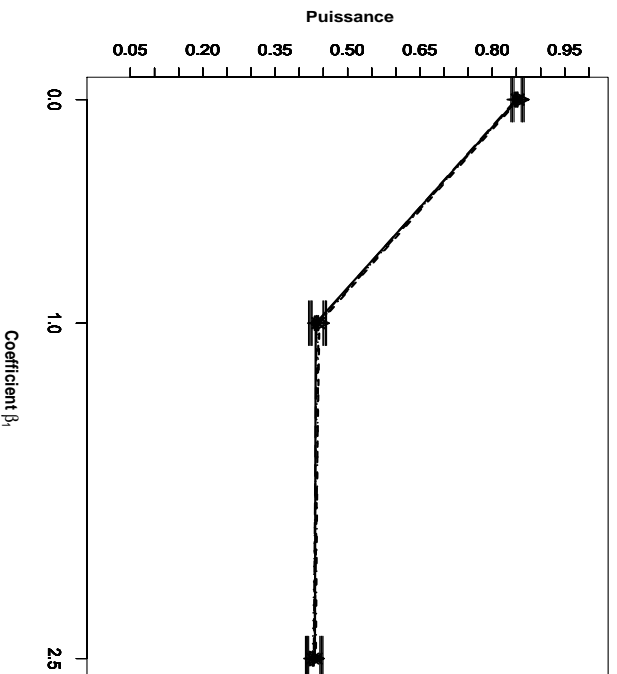
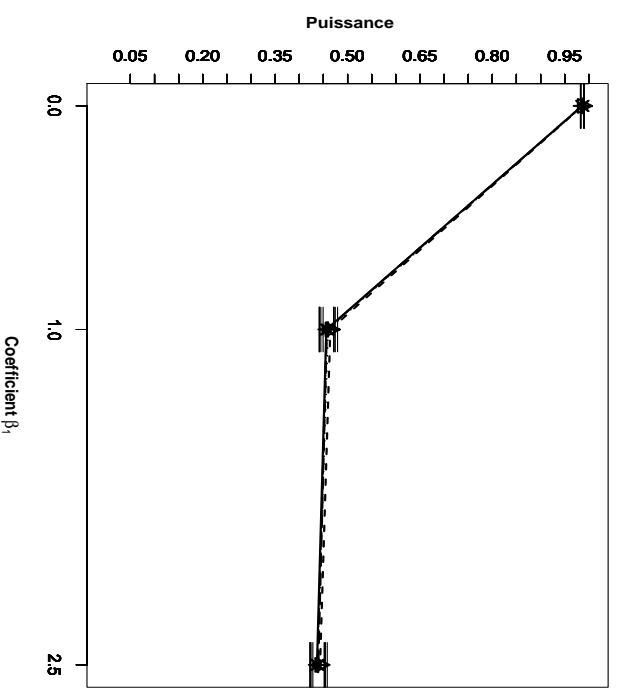
a : $\text{Exp}(2)$, $\rho = 0.9$, $\beta_2 = 0.5$, $n = 6$ b : $\text{Exp}(2)$, $\rho = 0.9$, $\beta_2 = 0.3$, $n = 9$ c : $\text{Exp}(2)$, $\rho = 0.9$, $\beta_2 = 0.5$, $n = 6$ d : $\text{Exp}(2)$, $\rho = 0.9$, $\beta_2 = 0.3$, $n = 9$ 

FIG. 4.84 – Puissance estimée du test d'hypothèse $H_0 : \beta_2 = \beta_2^*$ en fonction du coefficient de covariable β_1 et pour les méthodes Gaussienne $-\triangle-$; Freedman et Lane $-\bullet-$; Kennedy $-\times-$; Ter Braek $-\oplus-$ et Manly $-\ast-$. 30000 n-échantillons du résidu ϵ ont été simulées suivant une loi de $\text{Exp}(2)$ et vérifient les propriétés d'une matrice de distance. Pour toutes les figures la statistique de test est le T de Student, X^1 et X^2 sont des distances interpoints et $\rho_1 = 0.9$, $n = 6$, $\beta_1 = 0.5$ pour les figures a et c, $n = 9$, $\beta_1 = 0.3$ pour les figures b et d. Les tests présentés dans les figures a et b concernent des hypothèses alternatives bilatérales et dans les figures c et d des alternatives unilatérales.

Bibliographie

- [1] ANDERSON, M.J. et LEGENDRE, P. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *J. Statist. comput. Simulation.*, 62 :271–303, 1999.
- [2] ANTONIADIS, A., BERRUYER, J., et CARMONA, R. *Régression non linéaire et application*. Economica, 1992.
- [3] BARRA, J.R. *Notions fondamentales de statistique mathématique*. Paris, Dunod, 1971.
- [4] BERAN, R. Bootstrap method in statistics. *Jber. d. Dt. Math. verein.*, 86 :14–30, 1984.
- [5] BROWN, B.M. et MARITZ, L.S. Distribution-free methods in regression. *Austral. J. Statist.*, 24 :318–331, 1982.
- [6] COLLINS, M.F. A permutation test for planar regression. *Aust. J. Statist.*, 29 :303–308, 1987.
- [7] CZAPLEWSKI, R. L. et RIECH, R. M. Expected value and variance of moran’s bivariate spatial autocorrelation statistic for a permutation test. Research report, USDA Forest service, United states, April 1993.
- [8] D’AUBIGNY, G. *Introduction à l’économétrie (Cours de licence d’économie de Grenoble (2) 3ème année)*. UFR ESE, UPMF Grenoble 2, 2005.
- [9] DODGE, Y. et V., Rousson. *Analyse de régression appliquée*. Dunod, 2004.
- [10] DRAPER, D., HODGES, J.S., MALLOWS, C.L., et PREGIBON, D. Exchangeability and data analysis (with discussion). *Journal of the Royal Statistical Society*, A 156 :9–28, 1993.
- [11] DRAPER, N. R. et SMITH, H. *Applied Regression Analysis*. New York : John Wiley and Sons (Third Edition), 1998.
- [12] EDGINTON, E. S. *Randomization Tests*. New York : Marcel Dekker. (Third Edition), 1987.
- [13] EDGINTON, E. S. *Randomization Tests*. New York : Marcel Dekker. (Third Edition), 1995.
- [14] EFRON, B. Bootstrap methods : another look at the jackknife. *Annals of Statistics*, 21 :460–80, 1979.

- [15] EFRON, B. *The Jackknife, The Bootstrap and other Resampling plans*. SIMA, Philadelphia, 1982.
- [16] EFRON, B. et TIBSHIRANI, R. *An introduction to the Bootstrap*. Chapman and Hall, London., 1993.
- [17] FISHER, N.I et RICHARDS, J.D. On bootstrap hypothesis testing. *Austral. J. Classif.*, 32 :177–190, 1990.
- [18] FISHER, R.A. *Design of Experiments*. Edinburgh : Oliver and Boyd, 1935.
- [19] FREEDMAN, D. et LANE, D. A nonstochastic interpretation of reported significance levels. *J. Bus. Ecom. Statist.*, 1 :292–298, 1983.
- [20] GAIL, M.H Tan W.Y. et PIANTADOSI, S. tests for no treatment effect in randomized clinical trials. *Biometrika*, 75 :57–64, 1988.
- [21] GALAMBOS, J. Exchangeability. *Encyclopedia of Statistical Sciences*, NY :Wily 7 :573–577, 1986.
- [22] GOOD, P. *Permutation tests*. Springer, 2000.
- [23] H., Greene W. *Econometric Analysis*. Prentice Hall, 2004.
- [24] HALL, P. et TITTERINGTON, D.M. The effect of simulation order on level accuracy and power of monte carlo tests. *J. R. Statist. Soc. B*, 51 :459–467, 1989.
- [25] HALL, P. et WILSON, S. R. Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47, 757-762, 47 :757–762, 1991.
- [26] HAYASHI, F. *Econometrics*. Princeton University Press, princeton, N.J., 2000.
- [27] HINKLEY, D. V. Jackknifing in unbalanced situations. *Technometrics*, 19 :285–292, 1977.
- [28] HOGG, R.V. et SMITH, H. *Introduction to mathematical statistic*. New York. Macmillan Publishing Co. (Fourth Edition), 1978.
- [29] HOPE, A.C.A. A simplified monte carlo test procedure. *J. R. Statist. Soc. B*, 30 :582–598, 1968.
- [30] JOHNSTON, J. et DINARDO, J. *Econometric methods*. Mc GRAW-HILL International editional (Fourth edition), 1997.
- [31] KEMPTHORNE, O. *The desingn and Analysis of Exeriments*. New York : John Wiley and Sons., 1952.
- [32] KENDALL, M. J. et STUART, A. *The avdanced therory of statistics (volume 2 Inference and relationship)*. CHALES GRIFFIN and COMPANY LIMITED (Second edition), 1967.
- [33] KENNEDY, P.E. Randomization tests in econometrics. *J. Bus. Ecom. Statist.*, 13 :85–94, 1996.
- [34] KENNEDY, P.E. et CAD, B. S. Randomization tests for multiple regression. *Commun. Statist. Simulation Comput.*, 25 :923–936, 1996.

- [35] LEGENDRE, P. Comparison of permutation methods for the partial correlation and partial mantel tests. *J. Statist. comput. Simulation.*, 67 :37–73, 2000.
- [36] LEGENDRE, P. et LEGENDRE, L. *Numerical Ecology*. Amsterdam :Elsevier Science BV (Second English Edition), 1998.
- [37] LEHMAN, E.L. *Testing Statistical Hypotheses*. John Wiley and Sons, 1998.
- [38] LEHMAN, E.L. et CASELLA, G. *Theory of point estimation*. Springer, 1998.
- [39] MANEL, M. The detection of disease clustering and a generalized regression approach. *Cancer Res*, 27 :209–220, 1967.
- [40] MANLY, B.F.G. *Randomization, and Monte Carlo Methods in Biology*. London : Chapman and Hall. (First Edition), 1991.
- [41] MANLY, B.F.G. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. London : Chapman and Hall. (Second Edition), 1997.
- [42] MARRIOTT, F. H. C. Barnard's monte carlo tests : how many simulations? *Applied Statistics*, 28 :75–77, 1986.
- [43] MILLER, R. G. The jackknife-a review. *Biometrika*, 61 :1–15, 1974.
- [44] MILLER, R. G. Un unbalanced jackknife. *Ann. Statist.*, 2 :880–891, 1974.
- [45] MOOD, A. M., FRAYBILL, F. A., et C., Boes D. *Introduction to theory os statistics*. Mc Graw-Hill Kogakusha : Tokio, 1974.
- [46] NETER, J.W., WASSERMAN, W., et KUTNER, H.H. *Applied Linear Statistical Models*. Homewood, Illinois, Irwin, 1985.
- [47] OJA, H. On permutation tests in multiple regression and analysis of covariance problems. *Austral. J. Statist.*, 29 :91–100, 1987.
- [48] PITMAN, E.J.G. Significance tests which may be applied to samples from any population. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 4 :119–130, 1937a.
- [49] PITMAN, E.J.G. Significance tests which may be applied to samples from any population. ii. the correlation coefficient test. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 4 :225–232, 1937b.
- [50] PITMAN, E.J.G. Significance tests which may be applied to samples from any population. iii. the analysis of variance test. *Biometrika*, 29 :322–335, 1938.
- [51] QUENOUILLE, M. Approximate tests of correlation in time series. *J. Royal Statist. Soc.*, B(11) :18–44, 1958.
- [52] RAO, C.R. *Linear statistical inference and its applications*. J. Wiley and Sons, New York, (Second edition), 1973.
- [53] SMOUSE, P. E., LONG, J. C., et SOKAL, R. R. Multiple regression and correlation extensions of the mantel test of matrix correspondence. *Syst. Zool.*, 35 :627–632, 1986.
- [54] SRIVASTAVA, M. S. et KHATARI, C. G. *An introduction to multivariate statistics (volume 2 Inference and relationship)*. Nort Holland (Second edition), 1979.

- [55] STAPEL, D.M. et ter BRAAK, C.F.J. Randomization and bootstrap tests in factorial experiments : does analysis follow from design? Dans *International Biometrics Conference*, Münster, May 1994.
- [56] SUISSA, S. et SHUSTER, JJ. Are uniformly most powerful unbiased tests really best? *American Statistical*, 38 :204–206, 1984.
- [57] ter BRAAK, C.G.F. Ordination. In : *Data Analysis in Community and Landscape Ecology* (Eds., Jongman, R.H.G., ter Braak, C.J.F. and van Tongeren, O.F.R.), pages 91–169, 1987.
- [58] ter BRAAK, C.G.F. *Update Notes : CANOCO version 3.10*. Wageningen. The Netherlands : Agricultural Mathematics Group, 1990.
- [59] ter BRAAK, C.G.F. Permutation versus bootstrap significance tests in multiple regression and anova. In : *Bootstap and Related Techniques* (ed. K. H. Jöckel), Springer-Verlag, Berlin, pages 79–86, 1992.
- [60] TUKEY, J. W. Bias and confidence in not quite large samples (abstract). *Annals of Mathematical Statistics*, 29 :614–614, 1958.
- [61] WELCH, W.J. Construction of permutation tests. *J. Am. Statist. Ass.*, 85 :693–698, 1990.
- [62] WOLFE, D. A. On testing equality of related correlation coefficients. *Biometrika*, 63 :214–215, 1976.
- [63] WOLFE, D. A. A distribution-free test for related correlation coefficients. *Technometrics*, 19 :507–509, 1977.
- [64] WU, C. F. Jackknife, bootstrap and other resampling methods in regression analysis. *The annala of Statistics*, 14(4) :1261–1295, 1986.