



HAL
open science

Etude de la morphologie des indices d'association

Catherine Drouet d'Aubigny

► **To cite this version:**

Catherine Drouet d'Aubigny. Etude de la morphologie des indices d'association. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1980. Français. NNT: . tel-00292678

HAL Id: tel-00292678

<https://theses.hal.science/tel-00292678>

Submitted on 2 Jul 2008

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 à

l' Université Scientifique et Médicale de Grenoble

pour obtenir le grade de

DOCTEUR DE TROISIEME CYCLE

"Mathématiques Appliquées"

par

Catherine DROUET D'AUBIGNY



**ETUDE DE LA MORPHOLOGIE
DES INDICES D'ASSOCIATION**



Thèse soutenue le 26 septembre 1980 devant la Commission d'Examen :

Monsieur B. VAN CUTSEM Président

Messieurs F. BON

G. ROMIER

F. LONGEOT

Examineurs

UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE

Monsieur Gabriel CAU : Président

Monsieur Joseph KLEIN : Vice-Président

MEMBRES DU CORPS ENSEIGNANT DE L'U.S.M.G.

PROFESSEURS TITULAIRES

MM.	AMBLARD Pierre	Clinique de dermatologie
	ARNAUD Paul	Chimie
	ARVIEU Robert	I.S.N.
	AUBERT Guy	Physique
	AYANT Yves	Physique approfondie
Mme	BARBIER Marie-Jeanne	Electrochimie
MM.	BARBIER Jean-Claude	Physique expérimentale
	BARBIER Reynold	Géologie appliquée
	BARJON Robert	Physique nucléaire
	BARNOUD Fernand	Biosynthèse de la cellulose
	BARRA Jean-René	Statistiques
	BARRIE Joseph	Clinique chirurgicale A
	BEAUDOING André	Clinique de pédiatrie et puériculture
	BELORIZKY Elie	Physique
	BARNARD Alain	Mathématiques pures
Mme	BERTRANDIAS Françoise	Mathématiques pures
MM.	BERTRANDIAS Jean-Paul	Mathématiques pures
	BEZES Henri	Clinique chirurgicale et traumatologie
	BLAMBERT Maurice	Mathématiques pures
	BOLLIET Louis	Informatique (I.U.T. B)
	BONNET Jean-Louis	Clinique ophtalmologie
	BONNET-EYMARD Joseph	Clinique hépato-gastro-entérologie
Mme	BONNIER Marie-Jeanne	Chimie générale
MM.	BOUCHERLE André	Chimie et toxicologie
	BOUCHEZ Robert	Physique nucléaire
	BOUSSARD Jean-Claude	Mathématiques appliquées
	BOUTET DE MONVEL Louis	Mathématiques pures
	BRAVARD Yves	Géographie
	CABANEL Guy	Clinique rhumatologique et hydrologique
	CALAS François	Anatomie
	CARLIER Georges	Biologie végétale
	CARRAZ Gilbert	Biologie animale et pharmacodynamie

MM.	CAU Gabriel	Médecine légale et toxicologie
	CAUQUIS Georges	Chimie organique
	CHABAUTY Claude	Mathématiques pures
	CHARACHON Robert	Clinique ot-rhino-laryngologique
	CHATEAU Robert	Clinique de neurologie
	CHIBON Pierre	Biologie animale
	COEUR André	Pharmacie chimique et chimie analytique
	COUDERC Pierre	Anatomie pathologique
	DEBELMAS Jacques	Géologie générale
	DEGRANGE Charles	Zoologie
	DELORMAS Pierre	Pneumophtisiologie
	DEPORTES Charles	Chimie minérale
	DESRE Pierre	Métallurgie
	DODU Jacques	Mécanique appliquée (I.U.T. I)
	DOLIQUE Jean-Michel	Physique des plasmas
	DREYFUS Bernard	Thermodynamique
	DUCROS Pierre	Cristallographie
	FONTAINE Jean-Marc	Mathématiques pures
	GAGNAIRE Didier	Chimie physique
	GALVANI Octave	Mathématiques pures
	GASTINEL Noël	Analyse numérique
	GAVEND Michel	Pharmacologie
	GEINDRE Michel	Electroradiologie
	GERBER Robert	Mathématiques pures
	GERMAIN Jean-Pierre	Mécanique
	GIRAUD Pierre	Géologie
	JANIN Bernard	Géographie
	KAHANE André	Physique générale
	KLEIN Joseph	Mathématiques pures
	KOSZUL Jean-Louis	Mathématiques pures
	KRAVTCHENKO Julien	Mécanique
	LACAZE Albert	Thermodynamique
	LACHARME Jean	Biologie végétale
Mme	LAJZEROWICZ Janine	Physique
MM.	LAJZEROWICZ Joseph	Physique
	LATREILLE René	Chirurgie générale
	LATURAZE Jean	Biochimie pharmaceutique
	LAURENT Pierre	Mathématiques appliquées
	LEDRU Jean	Clinique médicale B
	LE ROY Philippe	Mécanique (I.U.T. I)

MM.	LLIBOUTRY Louis	Géophysique
	LOISEAUX Jean-Marie	Sciences nucléaires
	LONGEQUEUE Jean-Pierre	Physique nucléaire
	LOUP Jean	Géographie
Mlle	LUTZ Elisabeth	Mathématiques pures
MM.	MALINAS Yves	Clinique obstétricale
	MARTIN-NOEL Pierre	Clinique cardiologique
	MAYNARD Roger	Physique du solide
	MAZARE Yves	Clinique Médicale A
	MICHEL Robert	Minéralogie et pétrographie
	MICOUD Max	Clinique maladies infectieuses
	MOURIQUAND Claude	Histologie
	MOUSSA André	Chimie nucléaire
	NEGRE Robert	Mécanique
	NOZIERES Philippe	Spectrométrie physique
	OZENDA Paul	Botanique
	PAYAN Jean-Jacques	Mathématiques pures
	PEBAY-PEYROULA Jean-Claude	Physique
	PERRET Jean	Séméiologie médicale (neurologie)
	RASSAT André	Chimie systématique
	RENARD Michel	Thermodynamique
	REVOL Michel	Urologie
	RINALDI Renaud	Physique
	DE ROUGEMONT Jacques	Neuro-Chirurgie
	SARRAZIN Roger	Clinique chirurgicale B
	SEIGNEURIN Raymond	Microbiologie et hygiène
	SENGEL Philippe	Zoologie
	SIBILLE Robert	Construction mécanique (I.U.T. I)
	SOUTIF Michel	Physique générale
	TANCHE Maurice	Physiologie
	VAILLANT François	Zoologie
	VALENTIN Jacques	Physique nucléaire
Mme	VERAIN Alice	Pharmacie galénique
MM.	VERAIN André	Physique biophysique
	VEYRET Paul	Géographie
	VIGNAIS Pierre	Biochimie médicale

PROFESSEURS ASSOCIES

MM. CRABBE Pierre
SUNIER Jules

CERMO
Physique

PROFESSEURS SANS CHAIRE

Mlle	AGNIUS-DELORS Claudine	Physique pharmaceutique
	ALARY Josette	Chimie analytique
MM.	AMBROISE-THOMAS Pierre	Parasitologie
	ARMAND Gilbert	Géographie
	BENZAKEN Claude	Mathématiques appliquées
	BIAREZ Jean-Pierre	Mécanique
	BILLET Jean	Géographie
	BOUCHET Yves	Anatomie
	BRUGEL Lucien	Energétique (I.U.T. I)
	BUISSON René	Physique (I.U.T. I)
	BUTEL Jean	Orthopédie
	COHEN-ADDAD Jean-Pierre	Spectrométrie physique
	COLOMB Maurice	Biochimie médicale
	CONTE René	Physique (I.U.T. I)
	DELOBEL Claude	M.I.A.G.
	DEPASSEL Roger	Mécanique des fluides
	GAUTRON René	Chimie
	GIDON Paul	Géologie et minéralogie
	GLENAT René	Chimie organique
	GROULADE Joseph	Biochimie médicale
	HACQUES Gérard	Calcul numérique
	HOLLARD Daniel	Hématologie
	HUGONOT Robert	Hygiène et médecine préventive
	IDELMAN Simon	Physiologie animale
	JOLY Jean-René	Mathématiques pures
	JULLIEN Pierre	Mathématiques appliquées
Mme	KAHANE Josette	Physique
MM.	KRAKOWIACK Sacha	Mathématiques appliquées
	KUHN Gérard	Physique (I.U.T. I)
	LUU DUC Cuong	Chimie organique - pharmacie
	MICHOULIER Jean	Physique (I.U.T. I)
Mme	MINIER Colette	Physique (I.U.T. I)

MM.	PELMONT Jean	Biochimie
	PERRIAUX Jean-Jacques	Géologie et minéralogie
	PFISTER Jean-Claude	Physique du solide
Mlle	PIERY Yvette	Physiologie animale
MM.	RAYNAUD Hervé	M.I.A.G.
	REBECQ Jacques	Biologie (CUS)
	REYMOND Jean-Charles	Chirurgie générale
	RICHARD Lucien	Biologie végétale
Mme	RINAUDO Marguerite	Chimie macromoléculaire
MM.	SARROT-REYNAULD Jean	Géologie
	SIROT Louis	Chirurgie générale
Mme	SOUTIF Jeanne	Physique générale
MM.	STIEGLITZ Paul	Anesthésiologie
	VIALON Pierre	Géologie
	VAN CUTSEM Bernard	Mathématiques appliquées

MAITRES DE CONFERENCES ET MAITRES DE CONFERENCES AGREGES

MM.	ARMAND Yves	Chimie (I.U.T. I)
	BACHELOT Yvan	Endocrinologie
	BARGE Michel	Neuro-chirurgie
	BEGUIN Claude	Chimie organique
Mme	BERIEL Hélène	Pharmacodynamie
MM.	BOST Michel	Pédiatrie
	BOUCHARLAT Jacques	Psychiatrie adultes
Mme	BOUCHE Liane	Mathématiques (CUS)
MM.	BRODEAU François	Mathématiques (I.U.T. B) (Personne étrangère habilitée à être directeur de thèse)
	BERNARD Pierre	Gynécologie
	CHAMBAZ Edmond	Biochimie médicale
	CHAMPETIER Jean	Anatomie et organogénèse
	CHARDON Michel	Géographie
	CHERADAME Hervé	Chimie papetière
	CHIAVERINA Jean	Biologie appliquée (EFP)
	COLIN DE VERDIERE Yves	Mathématiques pures
	CONTAMIN Charles	Chirurgie thoracique et cardio-vasculaire
	CORDONNER Daniel	Néphrologie
	COULOMB Max	Radiologie
	CROUZET Guy	Radiologie

MM.	CYROT Michel	Physique du solide
	DENIS Bernard	Cardiologie
	DOUCE Roland	Physiologie végétale
	DUSSAUD René	Mathématiques (CUS)
Mme	ETERRADOSSI Jacqueline	Physiologie
MM.	FAURE Jacques	Médecine légale
	FAURE Gilbert	Urologie
	GAUTIER Robert	Chirurgie générale
	GIDON Maurice	Géologie
	GROS Yves	Physique (I.U.T. I)
	GUIGNIER Michel	Thérapeutique
	GUITTON Jacques	Chimie
	HICTER Pierre	Chimie
	JALBERT Pierre	Histologie
	JUNIEN-LAVILLAVROY Claude	O.R.L.
	KOLODIE Lucien	Hématologie
	LE NOC Pierre	Bactériologie-virologie
	MACHE Régis	Physiologie végétale
	MAGNIN Robert	Hygiène et médecine préventive
	MALLION Jean-Michel	Médecine du travail
	MARECHAL Jean	Mécanique (I.U.T. I)
	MARTIN-BOUYER Michel	Chimie (CUS)
	MASSOT Christian	Médecine interne
	NEMOZ Alain	Thermodynamique
	NOUGARET Marcel	Automatique (I.U.T. I)
	PARAMELLE Bernard	Pneumologie
	PECCOUD François	Analyse (I.U.T. B) (Personnalité étrangère habilitée à être directeur de thèse)
	PEFFEN René	Métallurgie (I.U.T. I)
	PERRIER Guy	Géophysique-glaciologie
	PHELIP Xavier	Rhumatologie
	RACHALL Michel	Médecine interne
	RACINET Claude	Gynécologie et obstétrique
	RAMBAUD Pierre	Pédiatrie
	RAPHAEL Bernard	Stomatologie
Mme	RENAUDET Jacqueline	Bactériologie (pharmacie)
MM.	ROBERT Jean-Bernard	Chimie-physique
	ROMIER Guy	Mathématiques (I.U.T. B) (Personnalité étrangère habilitée à être directeur de thèse)
	SAKAROVITCH Michel	Mathématiques appliquées

MM. SCHAEERER René	Cancérologie
Mme SEIGLE-MURANDI Françoise	Crytogamie
MM. STOEBSNER Pierre	Anatomie pathologie
STUTZ Pierre	Mécanique
VROUSOS Constantin	Radiologie

MAITRES DE CONFERENCES ASSOCIES

MM. DEVINE Roderick	Spectro Physique
KANEKO Akira	Mathématiques pures
JOHNSON Thomas	Mathématiques appliquées
RAY Tuhina	Physique

MAITRE DE CONFERENCES DELEGUE

M. : ROCHAT Jacques	Hygiène et hydrologie (pharmacie)
---------------------	-----------------------------------

Fait à Saint Martin d'Hères, novembre 1977

Je suis reconnaissante à Monsieur B. VAN CUTSEM, professeur à l'Université Scientifique et Médicale de Grenoble de l'honneur qu'il me fait en acceptant la présidence de ce jury.

Je remercie Monsieur F. BON, maître de recherche au C.N.R.S., d'avoir accepté de participer à ce jury. Les échanges que nous avons eus dans la préparation de l'école d'été m'ont permis d'enrichir ce travail.

Je remercie Monsieur F. LONGEOT, professeur à l'Université des Sciences Sociales de Grenoble de l'intérêt qu'il porte à la statistique et en particulier à ce travail en acceptant d'être membre du jury.

Je remercie Monsieur G. ROMIER, d'avoir accepté de diriger cette recherche et de m'avoir accordé sa confiance.

Mes remerciements vont aussi à Monsieur J. TRINQUART, de l'aide précieuse qu'il m'a apportée dans la réalisation des traitements Informatiques.

J'ai été très sensible à la qualité de frappe, malgré les courts délais, je tiens à en remercier les services de dactylographie.

Ce document doit beaucoup à la gentillesse de Monsieur ANGUILLE et au service de reprographie de l'I.R.M.A., qu'il en soient tous remerciés.

TABLE DES MATIERES

0 . Introduction

1. <u>Etude Morphologique des données</u>	0
1.1. La notion de système relationnel	0
1.2. Codage et échelles de mesure	2
1.3. Statistiques pertinentes	9
1.4. Mesurage et tableau de données	12
1.5. Mesurage et tri croisé	15
2. <u>Indépendance et Association</u>	18
2.1. Indépendance et probabilité	18
2.2. Indices de dépendance	24
2.2.1. Divergence entre lois de probabilité .	25
2.2.2. Exemples de mesures de la divergence	28
2.2.3. Indices de dépendance	30
3. <u>Indices de connexion</u>	34
3.1. Indices de connexion et logique P.R.E.	34
3.2. La notion d'Incertitude	36
3.2.1. Entropie et Incertitude	36
3.2.2. Dispersion et Incertitude	40
3.2.3. Entropie et dispersion	45
3.3. Notion de gain d'Incertitude et Indices de connexion	46

4. <u>Indices de CONCORDANCE</u>	51
4.1. Dualité et métrique de la covariance	52
4.2. Normalisation	54
4.3. Applications	56
4.3.1. Echelle nominale	57
4.3.1.1. Variables dichotomiques	57
4.3.1.2. Cas de variables nominales a plus de deux modalités	58
4.3.2. Echelle ordinale	59
4.3.2.1. Indices de degré 1	59
4.3.2.2. Indices de degré 2	60
4.3.3. Echelle hyperordinale	63
4.3.4. Echelle d'intervalle	64
4.4. Dualité et concordance	65
5. <u>Généralisation au cas vectoriel</u>	68
5.1. Liaison entre vecteurs aléatoires	68
5.1.1. Notions de dispersion d'un vecteur aléatoire	69
5.1.2. Indices de liaison multiple	72
5.2. Notion de liaison partielle	77
5.2.1. Liaison conditionnelle et liaison partielle	77
5.2.2. Décomposition de la liaison partielle	79
6. <u>Annexes -</u>	
Annexe 1	A1
Annexe 2	4 40
7. <u>Bibliographie</u>	

I N T R O D U C T I O N

Le travail que nous présentons est consacré à l'étude des principes généraux de construction d'indices mesurant l'existence et le degré de liaison entre deux éléments aléatoires (scalaires ou vectoriel).

Historiquement la recherche des corrélations est née de l'étude des covariations naturelles observées sur des échantillons d'individus, dans les sciences du comportement et biologiques, au début de ce siècle (GALTON, PEARSON, YULE). Cette démarche s'est essentiellement développée dans le cadre de l'analyse de données non expérimentales, de simples observations, issues d'enquêtes, contrairement aux techniques d'analyse de variance qui furent développées un peu plus tardivement dans le domaine agro-biologique (FISHER) sur des données expérimentales.

L'unité de ces techniques fondées sur le modèle linéaire explique la place prépondérante accordée au coefficient de corrélation linéaire et au rapport de corrélation dans l'indifférence quant au bien-fondé de leur utilisation. Cette situation privilégiée pose cependant deux questions, la première sur l'utilisation du modèle linéaire pour l'analyse de données de structure plus pauvre telles que les données ordinales, la deuxième sur la possibilité de développer des techniques mieux adaptées à la nature des données et conservant une flexibilité comparable à celle du modèle linéaire. Ces deux questions, à l'origine de ce travail, nous ont conduits à mener une réflexion préalable sur la structure des données et à étudier la morphologie des indices d'association existants.

Le chapitre 1 consacré à l'étude des systèmes relationnels introduits par SUPPES et ZINNES [73] nous permet d'étudier la structure des données. Toute variable est une manifestation chiffrée de propriétés des objets étudiés. De telles propriétés sont caractérisées par un ensemble de relations définies sur les objets et appelé système relationnel empirique auquel nous associons un système relationnel numérique.

La variable apparaît comme un homomorphisme liant ces deux systèmes relationnels c'est-à-dire un codage particulier d'une classe d'homomorphismes de même structure. Cette classe est appelée une échelle de mesure. Les principales échelles de mesure utilisées dans la suite de ce travail sont rappelées ainsi que les théorèmes d'existence dus à SUPPES

Cette présentation nous semble préférable à l'introduction classique des variables aléatoires puisqu'elle n'implique pas d'hypothèses sur la distribution en termes d'absolue continuité par exemple.

Nous définissons alors la notion de statistique pertinente, en termes d'invariance par rapport au groupe des recodages de la variable observée.

Ces notions nous permettent alors de montrer que à tout tableau de données, sont associés trois schémas de dualité (complémentaires) :

- le schéma "individus-catégories" qui permet de reconstruire la distribution des individus suivant les valeurs distinctes de la variable (tri à plat)
- le schéma "individus-variables" usuellement traité en analyse des données
- le schéma "catégories-variables" qui contient la même information que le précédent sous forme condensée.

L'analyse conjointe des deux schémas de dualité "individus-catégories" associés à deux variables permet de reconstituer le tri croisé de ces variables. On montre ainsi que l'étude conjointe de deux variables fournit trois informations complémentaires qui interviennent dans l'analyse de l'association entre ces deux variables :

- les marges du tri croisé permettent de comparer les distributions respectives des individus selon chaque variable
- le tableau de contingence fournit la distribution du couple
- les codages ou encore les échelles de mesure de chaque variable.

L'utilisation ou non de ces différents types d'informations nous permet de distinguer entre indices de liaison et indices de dépendance. Le chapitre 2 étudie la notion probabiliste de dépendance, dans le cadre de l'ensemble $L_{XY}^2(\Omega)$.

On montre tout d'abord la relation existant entre la notion de monotone dépendance et les bornes de FRECHET [32]

Nous définissons une relation d'ordre partielle sur les lois de probabilités bivariées liée à la notion de "quadrant dépendance" et montrons qu'elle implique une relation d'ordre partielle sur les covariances. Cette propriété est étendue aux recodages des variables initiales par des couples de fonctions concordantes et est à la base de la définition des indices de concordance présentés au chapitre 4.

Les notions introduites dans ce premier paragraphe nous permettent de montrer que l'on ne peut mesurer que l'écart à l'indépendance si on dispose comme seule information du tableau de contingence. Nous introduisons alors la notion de divergences entre lois de probabilité donnée par ALI et SILVEY [3] définie comme l'espérance d'une fonction convexe de la dérivée de Radon-Nikodym de l'une part rapport à l'autre. Cette définition générale nous permet de donner une présentation unifiée des différentes divergences classiquement utilisées.

Les chapitres suivants étudient les indices de liaison qui utilisent les codages. Nous reprenons la distinction de GINI [34] entre les indices mesurant l'écart à la situation de dépendance complète d'une variable Y en une variable X, appelés indice de connexion et les indices mesurant l'écart à la situation mutuelle dépendance complète, appelés indice de concordance.

La dépendance complète de la variable Y en la variable X est interprétée dans le chapitre 3 en termes de prédiction. La logique P.R.E. (Proportional Reduction of Error) due à COSTNER est alors proposée comme principe général de construction d'indices de connexion. Sa mise en oeuvre nécessite la définition de la notion d'incertitude qui généralise la variance comme mesure d'erreur de prédiction. Nous définissons donc l'incertitude comme une fonction concave de la loi de probabilité nous inspirant en cela des travaux de FERON [31] .

L'incertitude doit être une statistique pertinente. Lorsque la variable Y est d'échelle nominale, l'incertitude est mesurée en termes d'entropie, et apparaît comme optimum d'une fonction d'inexactitude.

Nous montrons dans le paragraphe suivant que cette approche peut être étendue à d'autres échelles de mesure ; en effet le concept de ψ -moyenne dû à BRONS et alii [14] est interprété dans ce cadre comme un indice de position centrale auquel nous associons une dispersion.

L'indice P.R.E. associé à une incertitude I s'écrit alors

$$\frac{I(Y) - E_X[I^X(Y)]}{I(Y)}$$

La logique P.R.E. permet donc de généraliser le problème de l'analyse de variance aux cadres d'hypothèses emboîtées, puisque la concavité des fonctions d'incertitude nous assure de la monotonie de l'incertitude par rapport à l'inclusion des sous-espaces. Il faut cependant remarquer que l'on a pas toujours une décomposition additive de l'incertitude.

Nous donnons divers exemples d'indices connus qui respectent cette logique.

La mutuelle dépendance complète entre les variables X et Y est interprétée dans le chapitre 4 en termes de concordance. Elle se traduit par l'existence simultanée de deux fonction f et y vérifiant presque sûrement

$$Y = f(X) \quad , \quad X = g(Y)$$

Les résultats exposés au chapitre 2 sur les fonctions concordantes nous permet de donner une présentation unifiée des indices de concordance associés à la géométrie de la covariance définie sur les codages dérivés construits à l'aide de fonctions concordantes.

Alors que les indices P.R.E. sont bornés par 0 et 1, par construction, l'utilisation de la covariance nécessite une normalisation permettant alors une interprétation de la valeur de l'indice dans une étude concrète. Nous distinguons deux types de normalisations fondées sur deux problématiques distinctes.

Le premier type proposé par KENDALL [46] est de nature probabiliste et utilise les propriétés des permutations. Sans hypothèse d'absence de liaison, l'appariement observé de n valeurs de X et de Y est considéré comme équiprobable parmi les $n!$ appariements possibles. Cette logique est utilisée par LERMAN [57] , LECALVE [54] qui construisent l'indice normalisé. :

$$\Phi \left[\frac{\text{cov} [f(X), g(Y)] - \mu}{\sigma} \right]$$

avec Φ la fonction de répartition de la loi normale $\mathcal{N}(0,1)$, et μ et σ respectivement la moyenne et la variance de la variable aléatoire covariance sous l'hypothèse d'absence de liaison.

Le second type de normalisation est de nature géométrique utilisant l'interprétation en termes de produit scalaire de la covariance. La concordance entre les variables X et Y est mesurée par le cosinus entre les codages dérivés.

Nous donnons dans ce paragraphe plusieurs exemples correspondant à des coefficients connus en dagageant pour chacun le principe de construction en termes de concordance.

La définition d'indices de concordance à l'aide de fonctions concordantes définies sur les paires (resp. les k -uples) conduit alors à étudier la dualité "paires (resp. k -uples) d'individus-variables".

Le chapitre 5 est consacré à la généralisation des notions de liaison au cas vectoriel. La variance du vecteur aléatoire X est alors un opérateur et on doit définir une fonctionnelle sur l'espace des opérateurs de rang déterminé afin de définir un indicateur scalaire de la variance. Nous proposons trois fonctionnelles particulières qui s'interprètent toutes trois dans le schéma de dualité associé au tableau X et définissant de plus un cosinus. Nous montrons qu'il n'y a pas d'extension unique de la concordance, bien que toutes s'expriment comme fonction des corrélations canoniques.

L'étude de la liaison entre plus de deux variables conduit à définir la notion de corrélation partielle que nous distinguons de la corrélation conditionnelle. La première est associée au schéma linéaire de la régression, alors que la deuxième correspond à la notion de conditionnement probabiliste. Si ces deux notions coïncident dans le cas gaussien, les exemples que nous donnons rappellent la non-équivalence de ces deux notions.

Nous montrons que la covariance partielle fait intervenir deux informations distinctes, la covariance partielle intra mesure la covariance des résidus à X_3 constant, la covariance partielle inter mesure la covariance des centres de gravité des sous-nuages résiduels définis par les modalités de X_3 . Cette décomposition est utilisée pour montrer à l'aide d'un exemple que la corrélation conditionnelle n'est pas la même suivant qu'on travaille sur les individus ou sur les paures.

L'étude des variables de types ordinales nous a conduits à mettre en évidence le rôle des ex aequos, par suite nous nous sommes limités parfois à une écriture adaptée à un ensemble de cardinal fini, afin d'attacher une mesure non nulle aux ex aequos. Par suite nous n'avons pas donné d'expressions de concordance théorique, et renvoyons à GOODMAN et KRUSKAL [37] pour la définition des indices théoriques.

Nous présentons en annexe des tableaux simulés permettant d'étudier le comportement de certains des indices présentés. Le cas des tableaux 2 x 2 est étudié plus particulièrement sur une simulation de 500 tableaux générés au hasard à l'aide d'une analyse en composantes principales.

1 - ETUDE MORPHOLOGIQUE DES DONNEES.

L'ensemble d'arbitraires qui préside à la construction d'un tableau de données impose au statisticien le choix de méthodes d'analyse, tout autant que les buts poursuivis. L'objet de ce chapitre est de caractériser les étapes suivies dans le processus de mesurage, qui conduit à des données chiffrées. Notre présentation des travaux de SUPPES et ZINNES [73] et TORGERSON [74] est limitée à l'exposé des concepts nécessaires à la suite de ce travail.

1.1. La notion de système relationnel.

Toute observation d'un ensemble d'objets empiriques, étudie leurs propriétés dont elle mesure les différentes manifestations. De telles propriétés n'existent que par rapport aux objets empiriques étudiés qui déterminent leur structure par des relations empiriques (c'est-à-dire les constatations subjectives de relation par l'observateur) qu'elles établissent entre eux. De façon symétrique, un objet empirique est caractérisé par un ensemble de propriétés. Ainsi un son peut-il être caractérisé par son intensité, sa hauteur, son timbre et la hauteur d'un son permet à l'observateur de constater la relation empirique selon laquelle un son est plus aigu qu'un autre.

Toute observation conduit donc à la définition de systèmes relationnels empiriques s.r.e traduisant les propriétés observées des objets empiriques étudiés.

Définition 1 .

On appelle système relationnel empirique s.r.e. le doublet

$$\Delta = \langle A, (R_j ; j \in J) \rangle$$

où Δ est un ensemble d'objets empiriques et $\{ R_j ; j \in J \}$ un ensemble de relations définis sur A.

Si R_j est une relation k_j -naire, nous dirons que $(k_j ; j \in J)$ est le type de Δ .

Exemples

Exemple 1 . A = ensemble des sons
 R₁ = la relation d'équivalence (notée usuellement I) par (a,b) ∈ I ssi a et b ont la même hauteur.
 R₂ = la relation d'ordre strict (notée usuellement P) définie par (a,b) ∈ P ssi a est strictement plus haut que b
 d'où le système empirique A = < A ; { I, P } >

Exemple 2. A = ensemble des C.S.P.
 R = la relation ternaire associée à l'opération de concatération o (a₁,a₂,a₃) ∈ R ssi a₁ o a₂ = a₃.

L'observation d'un s.r.e.se heurte à deux difficultés : d'une part le langage courant ne permet de distinguer qu'un faible nombre de manifestations d'une propriété, car limité par le nombre de mots, ainsi distingue-t-on un faible nombre de couleurs d'yeux ; d'autre part le langage courant utilise différentes mesures des manifestations d'une même propriété, ainsi mesure-t-on la distance entre deux points, en mètres en lancé de javelots en heures de marche en randonnée, etc...

Aussi est-on amené à définir la notion d'échelle de mesure qui permet de construire une mesure chiffrée d'une manifestation d'une propriété et d'étudier l'unicité d'un tel processus de construction. Une telle mesure chiffrée est construite pour des s.r.e. irréductibles, dont la définition fait l'objet du paragraphe suivant.

Définition 2.

Une relation d'équivalence sur A est une congruence sur le s.r.e.

$$A = \langle A, \{ R_j, j \in J \} \rangle \text{ de type } k_j, j \in J) \text{ si pour tout } j \text{ de } J \text{ on a}$$

$$\underline{a} = (a_1, \dots, a_{k_j}) \in A^{k_j}$$

$$\underline{a}' = (a'_1, \dots, a'_{k_j}) \in A^{k_j}$$

tels que a_i et a'_i sont équivalents pour tout i=1 ... k_j

alors R_j (a) = R_j (a')

Théorème 1 (PFANZAGL [65] page 21).

Pour tout s.r.e., il existe une unique relation de congruence la plus grossière.

La relation de congruence associe de façon canonique à chaque relation R_j la relation $\overline{R_j}$ définie sur l'espace quotient \overline{A} . On notera \overline{A} le s.r.e.

$$\overline{A} = \langle \overline{A}; (\overline{R_j}; j \in J) \rangle$$

Théorème 2 (PFANZAGL [65] page 22).

Le s.r.e. \overline{A} a pour congruence la plus grossière l'identité.

Remarque :

Un tel s.r.e. est dit irréductible.

Sauf indication contraire, nous supposons, dans la suite, sans perte de généralité, que tous les systèmes relationnels étudiés sont irréductibles.

1.2. Codages et Echelles de mesure.

L'analyse statistique d'un s.r.e. nécessite le recours à des observations chiffrées des propriétés des objets étudiés. Pour cela, on construit un système relationnel numérique s.r.n. respectant les relations du s.r.e. L'observation des manifestations est alors codée par un ensemble de valeurs. Ces notions sont étudiées dans ce paragraphe.

Définition 3.

Soient $\overline{A} = \langle A, (R_j; j \in J) \rangle$ et $\overline{B} = \langle B; (S_j; j \in J) \rangle$ deux systèmes relationnels de même type. L'application h définie sur \overline{A} à valeurs dans \overline{B} est un homomorphisme de \overline{A} dans \overline{B} ssi pour tout j , h est un homomorphisme de (A, R_j) dans (B, S_j) .

On note $\mathcal{H}(\overline{A}, \overline{B})$ l'ensemble des homomorphismes de \overline{A} dans \overline{B} .

Isomorphismes, automorphismes, endomorphismes de systèmes relationnels se définissent de façon analogue.

Exemple

$$\Delta A = \langle \mathbb{R}, (<, +) \rangle$$

$$\Delta B = \langle \mathbb{R}^+, (<, \times) \rangle$$

$$h = \mathbb{R} \rightarrow \mathbb{R}^+$$

$$x \rightarrow h(x) = \exp(x)$$

est un isomorphisme de ΔA dans ΔB .

Remarque

L'ensemble des automorphismes de ΔA , noté $\Gamma(\Delta A)$, forme un groupe pour la composition des applications.

Remarque

Soit h élément de $\mathcal{H}(\Delta A, \Delta B)$ où ΔA est un s.r.e. irréductible et ΔB un s.r.e. quelconque, alors h est injective.

Les concepts précédents introduits par SUPPES puis étudiés par PFANZAGL nous permettent de donner une définition algébrique aux notions de codages et de recodage employés en analyse des données. Nous utilisons alors les résultats de PFANZAGL pour définir la notion d'unicité d'un codage.

Définition 4.

On appelle codage r-dimensionnel d's.r.e. irréductible $\Delta A = \langle A, (R_j ; j \in J)$ dans un s.r.n. $\Delta B = \langle \mathbb{R}^r, (S_j ; j \in J) \rangle$ de même type tout élément h de $\mathcal{H}(\Delta A, \Delta B)$.

Pour tout élément a de A , $h(a)$ est appelé valeur de a , note de a , code de a ou encore score de a .

La propriété d'homomorphisme nous permet de tirer des conclusions sur les objets a à partir des relations entre les codes puisque

$$R_j(a_1, \dots, a_{k_j}) \Leftrightarrow S_j(h(a_1), \dots, h(a_{k_j}))$$

Par suite, plus on prend en compte de relations R_j dans la définition du codage et plus celui-ci est porteur d'information sur les objets empiriques.

Remarque

Si \bar{A} est le s.r.e. irréductible associé à A ; le diagramme suivant définit sans ambiguïté l'application h'

$$\begin{array}{ccc} \bar{A} & \xrightarrow{h} & B \\ \uparrow s & \nearrow h' = h \circ s & \\ A & & \end{array}$$

où h est un élément de $\mathcal{J}(A, B)$ et s la surjection canonique de A dans \bar{A} . h' est alors un homomorphisme que l'on notera encore h .

La contrainte imposée à h en termes d'homomorphismes, ne suffit pas en général, à définir de façon unique un codage, puisque tout élément de $\mathcal{J}(A, B)$ définit un codage admissible pour A . Cependant les propriétés internes de B nous permettent de définir une notion d'unicité d'un codage.

Définition 5.

Soient $A = \langle A, (R_j ; j \in J) \rangle$ et $A_0 = \langle A_0, (R_j^0 ; j \in J) \rangle$ deux s.r.e. tels que A_0 est inclus dans A et que R_j^0 est la restriction de R_j à A_0 , on appelle ensemble d'endomorphismes partiels de A l'ensemble

$$\Gamma(A, A_0) = \mathcal{J}(A_0, A)$$

Théorème 3 (PFANZAGL [65] page 26)

Soit A un système relationnel irréductible et B un système relationnel quelconque de même type que A . Soit h_0 un élément quelconque de $\mathcal{J}(A, B)$ alors

$$\mathcal{J}(A, B) = \{ g \circ h_0 ; g \in \Gamma(B, h_0(A)) \}$$

Lorsque B est un s.r.m., $h = g \circ h_0$ est un codage r -dimensionnel, obtenu par un recodage g qui respecte par construction la structure de A

$$\begin{array}{ccc} A & \xrightarrow{h_0} & h_0(A) \subset B \\ & \searrow h & \downarrow g \\ & & B \end{array}$$

Définition 6.

L'ensemble $\mathcal{H}(A, B)$ des codages numériques de A est appelé une échelle.

Ainsi il suffit pour décrire une échelle de connaître un de ses représentants h_0 ; alors toute transformation appartenant au groupe $\Gamma_B(h_0(A))$ définit un codage appartenant à l'échelle et donc équivalent à h_0 pour mesurer les propriétés de A . Les éléments de $\Gamma_B(h_0(A))$ sont souvent appelés des transformations admissibles de l'échelle.

Remarques

. La notion d'échelle proposée par SUPPES et PFANZAGL correspond à ce que nous appelons un codage. Notre changement de terminologie correspond à la nécessité de mettre en évidence l'importance de la notion de recodage. Il permet de plus de retrouver, dans le cadre du calcul des probabilités, le théorème de factorisation, nécessaire à la définition de l'espérance conditionnelle.

. L'exemple 2 montre que l'unicité d'une échelle n'existe que lorsque le s.r.n. B est fixé. Le choix de B reste arbitraire sous contrainte d'être une image homomorphe de A - par suite le choix d'une échelle de mesure est affaire de convenance.

L'étude des systèmes relationnels empiriques est en général limitée à des systèmes simples, correspondant à des procédés d'observations simples et limités. Les exemples d'échelles qui leur correspondent sont présentés ici selon les termes proposés par STEVENS [72]. Nous renvoyons le lecteur à SUPPES et ZINNES [73] pour la démonstration des théorèmes d'existence des échelles présentées.

Soit I une relation d'équivalence définie sur A et \bar{A} le s.r.e. associé au s.r.e. $A = \langle A, I \rangle$

Définition 7.

On appelle échelle nominale l'ensemble $\mathcal{H}(A, B)$ où
 $B = \langle \mathbb{R}, = \rangle$

SUPPES montre qu'une telle échelle existe ssi \bar{A} a au plus la puissance du continu. Elle donne alors comme information :

- la catégorie (classe de I-équivalence) à laquelle appartient l'objet a,
- le fait que a_1 et a_2 , éléments de A soient ou non équivalents pour I (c.a.d. le graphe de I).

Proposition 1.

$\mathcal{H}(\bar{A}, /B)$ est l'ensemble des fonctions injectives de A/I dans /R.

Soient $\bar{A} = \langle A, (I, P) \rangle$ un s.r.e. tel que I (resp. P) est une relation d'équivalence (resp d'ordre strict) définie sur A, alors le s.r.e. irréductible $\bar{A} = \langle A/I, \bar{P} \rangle$ est un ensemble ordonné.

Définition 2.

On appelle échelle ordinale l'ensemble $\mathcal{H}(\bar{A}, /B)$ où

$$/B = \langle /R, \langle \rangle \rangle$$

avec une relation d'ordre strict sur /R.

SUPPES et ZINNES (I [62] théorème 7 page 28) montre qu'une telle échelle existe ssi il existe un sous-ensemble de A/I au plus dénombrable, et dense dans A/I pour la topologie d'ordre.

Elle donne comme information :

- la catégorie (classe de I-équivalence) à laquelle appartient un objet,
- le graphe des relations d'ordre \bar{P} et P.

Proposition 2.

$\mathcal{M}(\bar{A}, /B)$ est l'ensemble des fonctions monotones de A/I dans /R.

Remarque

La définition 8 est restreinte à l'étude des relations des préordres sur A. La notion d'échelle ordinale s'étend sans difficulté à des structures plus pauvres. Ainsi définit-on de même les échelles quasi-ordinales associées à des relations d'indifférence I.

Exemples

- . Echelle de Moh de dureté des minéraux.
- . Echelle de BAPKAR de santé des bébés à la naissance.
- . Echelle de vitesse de sédimentation du sang.
- . Echelle de Beaufort de la force du vent.

Cependant la notion d'échelle ordinaire ne permet pas de prendre en compte des pratiques statistiques telle que le recours aux statistiques de rang, qui impose aux intervalles entre deux valeurs consécutives d'être de longueur égale. Cette notion contenue dans le concept de système empirique d'égale différence finie introduit par SUPPES nécessite l'introduction de deux nouvelles relations :

- la relation binaire R d'adjacence pour P définie par la proposition :
 $(a,b) \in R$ ssi $(a,b) \in P$ et $(a,c) \in P \Rightarrow I(b,c)$ ou $P(b,c) \forall a, b, c, \in A$
- la relation quaternaire D d'équiespacement D définie par les cinq propositions :
 - d_1 : transitivité $(a,b,c,d) \in D$ et $(c,d,e,f) \in D \Rightarrow (a,b,e,f) \in D$
 - d_2 : connexité forte $(a,b,c,d) \in D$ ou $(c,d,a,b) \in D$
 - d_3 : $(a,b,c,d) \in D \Rightarrow (a,c,b,d) \in D$
 - d_4 : antisymétrie $(a,b,c,d) \in D \Rightarrow (d,c,b,a) \in D$
 - d_5 : équiespacement $(a,b) \in R$ et $(c,d) \in R \Rightarrow (a,b,c,d) \in D$ et $(c,d,a,b) \in D$

Définition 9.

Le système relationnel $\mathcal{A} = \langle A, (I,P,R,D) \rangle$ est dit d'égale différence finie et $\mathcal{H}(\mathcal{A}, \mathcal{B})$ est une échelle hyperordinaire si $\mathcal{B} = \langle B, \langle \rangle$ où B est un sous-ensemble fini non vide de \mathcal{R}^T tel que l'écart entre deux valeurs adjacentes pour \langle est constante.

Exemple

Le codage par les rangs est un élément d'une échelle hyperordinaire pour lequel B est une section commençante de \mathcal{N} et \langle l'ordre naturel sur \mathcal{N} .

Proposition 3 (SUPPES [73] page 39)

$\Gamma_{/B}(h_o, \overline{/A})$ est le groupe des transformations linéaires sur $/B$.

Le caractère fini imposé à A assure l'existence du système relationnel $/B$ (SUPPES [], théorème 15, page 39). Une échelle hyperordinaire apporte pour information

- la catégorie à laquelle appartient l'objet a,
- le groupe des relations d'ordre \overline{P} et P,
- le positionnement des catégories définies par I en des points équivalents de la droite réelle.

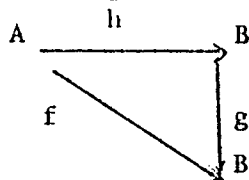
L'ensemble des systèmes relationnels précédents est en général étudié à l'aide de techniques statistiques qualifiées de non métriques, qualitatives ou encore non paramétriques. Les échelles de différence, de rapport ou d'intervalle étudiées ci-dessous sont souvent regroupées sous les termes de variables numériques ou quantitatives, marquant ainsi qu'elles sont définies par la seule donnée de $\Gamma_{/B}(h_o, \overline{/A})$ car elles correspondent à des s.r.e. pour lesquels les relations R_j sont des opérations.

Définition 10.

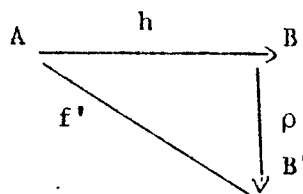
$\mathcal{H}(\overline{/A}, /B)$ est une échelle d'intervalle (resp. de différence, de rapport, absolue) ssi $\Gamma_{/B}(h_o, \overline{/A})$ est le groupe des transformations affines (resp. des translations, des homothéties, réduit à l'identité) sur $/R$.

La définition 10 utilise implicitement la structure de groupe additif de $(/R, +)$. Celui-ci est isomorphe au groupe $(/R^+, \times)$ qui permet une représentation équivalente de $/A$. Cette non unicité fonde par exemple le recours arbitraire aux modèles log-linéaires, logistiques et même gaussiens dans l'étude des tableaux de fréquence. On définit ainsi une nouvelle échelle par une transformation définie sur $/B$ et à valeurs dans un s.r.n. $/B'$ distinct de $/B$.

Recodage



Transformation des données



Cette notion de transformation des données est à distinguer de celle de codage dérivé introduite par SUPPES qui consiste à associer à \mathcal{A} le s.r.n. associé à l'une de ses puissances cartésiennes.

Définition 11.

Soit A' une puissance cartésienne de A et \mathcal{A}' le système relationnel induit par \mathcal{A} . On appelle codage dérivé tout codage de A' .

1.3. Statistiques pertinentes..

L'analyse des systèmes relationnels empiriques permet de délimiter le champ d'application des techniques statistiques usuelles, puisqu'elles sont fondées sur des codages des objets et nécessitent donc de connaître la structure des homomorphismes formant une échelle. Ainsi deux observations notant une différence de température l'un en degré Celsius, l'autre en degrés Fahrenheit, seront toujours d'accord sur le jugement J_1 et non sur le jugement J_2

$$J_1 : t(\omega_1) > t(\omega_2)$$

$$J_2 : t(\omega_1) + t(\omega_2) > t(\omega_3)$$

puisque les codages de la température en degrés Celsius t^c et en degré Fahrenheit t^f se correspondent par la relation

$$t^f = 1,8 t^c + 32$$

c'est-à-dire que la proposition J_2 définit une relation non pertinente pour une relation non pertinente pour une échelle d'intervalle.

Définition 12.

Soit \mathcal{A} un s.r.e. irréductible et \mathcal{B} un s.r.n. de même type que \mathcal{A} tels qu'il existe un élément h_0 de $\mathcal{H}(\overline{\mathcal{A}}, \mathcal{B})$. Une relation S k -naire sur \mathcal{B} est pertinente pour \mathcal{A} ssi pour tous h_1, h_2 éléments de $\mathcal{H}(\overline{\mathcal{A}}, \mathcal{B})$ et pour tout $a = (a_1, \dots, a_k)$ élément de A^k on a

$$((h_1(a_1), \dots, h_1(a_n)) \in S) \Leftrightarrow (h_2(a_1), \dots, h_2(a_k)) \in S)$$

c'est-à-dire $h_1^{-1}(S) = h_2^{-1}(S)$.

Tout codage h de $\mathcal{H}(\overline{\mathcal{A}}, \mathcal{B})$ définit sur A la relation $h^{-1}(S)$ associé à la relation S de \mathcal{B} qui est une interprétation de S dans A , on a le critère de pertinence :

Critère 1 :

Si h_0 est un codage fixé de $\mathcal{J}(\bar{A}, B)$. La relation S définie sur B est pertinente pour \bar{A} ssi

$$\mathcal{J}(\bar{A}, B) = \mathcal{J}(\langle A, ((R_E, 1 \in L), h^{-1}(S)) \rangle, \langle B, ((S_1, 1 \in L), S) \rangle)$$

La définition 12 nous permet de caractériser les relations pertinentes en termes de codage. En pratique il est cependant plus important de donner un critère en termes de recodage.

On pose donc $B_0 \subset B$ et $\Gamma_{/B}(B_0)$ et $\Gamma_{/B}$ les groupes d'endomorphismes de B_0 (resp B) dans B et g_0 un élément de $\Gamma_{/B}(B_0)$.

Définition 13.

La relation k -naire S définie sur B est g_0 -invariante ssi

$$S \cap B_0^k = g_0^{-1}(S)$$

Plus généralement, un ensemble de relation \mathcal{Y} définies sur B est $\Gamma_{/B}(B_0)$ -invariant ssi tout S de \mathcal{Y} est g -invariante pour tout g élément de $\Gamma_{/B}(B_0)$.

On en déduit immédiatement le critère.

Critère 2 (PFANZAGL [65] théorème 223 page 36)

Soit $\bar{B} = \langle B, (S_1, 1 \in L) \rangle$ un s.r.n. de même type que le s.r.e. \bar{A} .

Une relation S définie sur B est pertinente pour \bar{A} ssi elle est $\Gamma_{/B}(h_0(\bar{A}))$ - invariante pour un codage h_0 donné dans $\mathcal{J}(\bar{A}, \bar{B})$.

La notion de relation pertinente est utile à la définition d'une statistique pertinente.

Exemples :

. On considère sur \mathbb{R} la relation ternaire S :

$$(s_1, s_2, s_3) \in S \Leftrightarrow x_1 + x_2 > x_3$$

alors si $\mathcal{J}(\bar{A}, \bar{B})$ est une échelle d'intervalle, S n'est pas pertinente pour \bar{A} .

si $\mathcal{H}(\bar{A}, /B)$ est une échelle de rapport, S est pertinente pour $/A$.

. On considère sur $/R$ la relation 2k-naire S :

$(x_1, \dots, x_n, x'_1, \dots, x'_n) \in S \Leftrightarrow \frac{1}{n} \sum_{i=1}^n x_i > \frac{1}{n} \sum_{i=1}^n x'_i$
 alors si $\mathcal{H}(/A, /B)$ est une échelle d'intervalle, S est pertinente pour

. On considère sur $/R$ la relation binaire S souvent utilisée en statistique

$$y = \log x + \alpha \quad \text{où } \alpha \text{ ne dépend ni de } x \text{ ni de } y$$

Si h est d'échelle hyperordinaire et h' d'échelle de rapport alors

$$\begin{aligned} h(a) + n &= \text{Log } \beta h'(a) + \alpha + n - \log k \\ &= \text{Log } \beta h'(a) + \alpha' \end{aligned}$$

La relation S est donc pertinente - il n'en est pas de même si (par exemple) h' est d'échelle d'intervalle.

Définition 14.

On appelle statistique scalaire (resp vectorielle) toute fonction définie sur B^n à valeurs dans $/R$ (resp $/R^k$).

Définition 15.

Une statistique f définie sur B^n est pertinente pour $/A$ ssi la relation S 2n-naire définie sur B par

$(\underline{b}, \underline{b}') \in S$ ssi $f(\underline{b}) = f(\underline{b}')$ pour tous \underline{b} et \underline{b}' éléments de B^n est pertinente pour $/A$.

Proposition 5 (PFANZAGL [65] page 42)

Soit $/B$ un s.r.n. de même type que le s.r.e. $/A$ et f une statistique définie sur B. Alors les trois affirmations suivantes sont équivalentes :

- i) f est une statistique pertinente ;
- ii) si h_0 est un codage fixé de $\mathcal{H}(\bar{A}, /B)$, la relation n-aire sur B définie par $f^{-1} \circ f(\underline{b})$ est $\Gamma_{/B}(h_0(/A))$ invariante pour tout \underline{b} de B^n ;
- iii) la relation R 2n-naire sur $/A$ définie par $(\underline{a}, \underline{a}') \in R \Leftrightarrow foh(\underline{a}) = foh(\underline{a}') \quad \forall \underline{a}, \underline{a}' \in A^n$ ne dépend pas du codage h.

Exemples

- La moyenne arithmétique est pertinente pour des codages d'échelle d'intervalle puisque

$$\frac{1}{n} \sum_i b_i = \frac{1}{n} \sum_i b'_i \Leftrightarrow \frac{1}{n} \sum_i (\beta b_i + \alpha) = \frac{1}{n} \sum_i (\beta b'_i + \alpha)$$

mais n'est pas pertinente pour des codages d'échelle ordinale puisque

$$\frac{1}{n} \sum_i b_i = \frac{1}{n} \sum_i b'_i \neq \frac{1}{n} \sum_i f(b_i) = \frac{1}{n} \sum_i f(b'_i)$$

si f est monotone.

Dans ce dernier cas par contre la médiane est une statistique pertinente.

- Pour des codages d'échelle d'intervalle, la statistique

$$f : (b_1, \dots, b_n) \rightarrow (\bar{b} = \frac{1}{n} \sum_i b_i, S^2(\underline{b}) = \frac{1}{n} \sum_i (b_i - \bar{b})^2)$$

est pertinente mais le coefficient de variation $\frac{S(\underline{b})}{\bar{b}}$ ne l'est pas.

• 1.4. Mesurage et tableau de données.

Si le s.r.n. \mathcal{A} , de même type que le s.r.e. \mathcal{A} étudié est de dimension p , on notera le codage h élément de $\mathcal{H}(\overline{\mathcal{A}}, \mathcal{A})$

$$h(a) = (h^1(a), h^2(a), \dots, h^p(a))$$

où h^j est un codage unidimensionnel ou scalaire.

Le mesurage de h se traduit par un cube de données associé au multigraphe d'incidence

$$i : J \times A \times \mathcal{X} \rightarrow \{0, 1\}$$

$$(j, a, x) \rightarrow \begin{cases} 1 & \text{si } h^j(a) = x \\ 0 & \text{sinon} \end{cases}$$

où J représente l'ensemble des p codages h^j de A , appelé aussi variable, caractères observés et \mathcal{X} l'ensemble des valeurs prises par les h^j sur A , c'est-à-dire le sous-ensemble de \mathcal{R} :

$$\mathcal{X} = \bigcup_{j=1}^p \mathcal{X}^j \quad \text{où } \mathcal{X}^j = h^j(A)$$

Nous étudions dans la suite les relations algébriques existant entre J , A , x . Il convient cependant de remarquer que lorsque l'ensemble des objets

est regardé comme échantillon d'une population Ω , le procédé d'échantillonnage se traduit par la construction d'une probabilité sur Ω . De même les erreurs de mesure se traduisent par un processus d'échantillonnage sur les valeurs, enfin les problèmes de fidélité des variables, rencontrés en psychométrie (LORD et NOVICK)⁵⁸, et l'analyse factorielle classique ont reçu une interprétation rigoureuse en termes d'échantillonnage sur les variables (MULAÏK [62]) que l'on retrouve aussi en analyse de la variance à effets aléatoires. On est alors conduit à regarder h comme un élément aléatoire défini sur un espace probabilisé (Ω, A, P) .

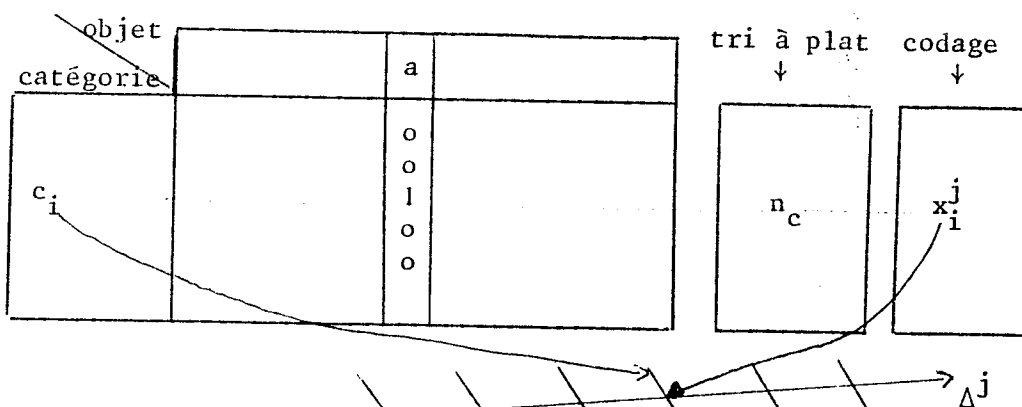
L'objet de ce travail n'était pas de nature confirmatoire, nous ne ferons qu'aborder les problèmes d'inférence, posés par l'étude des indicateurs de liaison, dans l'annexe 1 ; c'est pourquoi nous considérerons indifféremment que la population mesurée est la population entière ou est un échantillon.

Du point de vue algébrique, chaque élément j de J permet de mettre en dualité les objets éléments de A et les catégories éléments de \bar{A} , à l'aide de l'application

$$h^j = A \times X^j \rightarrow \{0, 1\}$$

$$(a, x) \rightarrow \begin{cases} 1 & \text{si } h^j(a) = x \\ 0 & \text{sinon} \end{cases}$$

que l'on traduit par le tableau des données H^j disjonctif complet associé à la relation de congruence I est correspondant à la strate j du cube de données (on note c la catégorie $(h^j)^{-1}(x)$).



Le tableau fournit trois informations, indépendamment de la structure algébrique de l'échelle étudiée :

- les lignes de H^j sont associées aux indicatrices des catégories, ou

encore modalités :

$$q_i : A \rightarrow \{0,1\}$$

$$a \rightarrow \begin{cases} 1 & \text{si } h^j(a) = x_i^j \text{ c'est-à-dire } a \in c_i \\ 0 & \text{sinon} \end{cases}$$

- le tri à plat, ou marge du tableau, fournit la distribution des objets de A dans les catégories ;
- le codage h^j fournit de plus la valeur x_i^j affectée à chacune des catégories c_i .

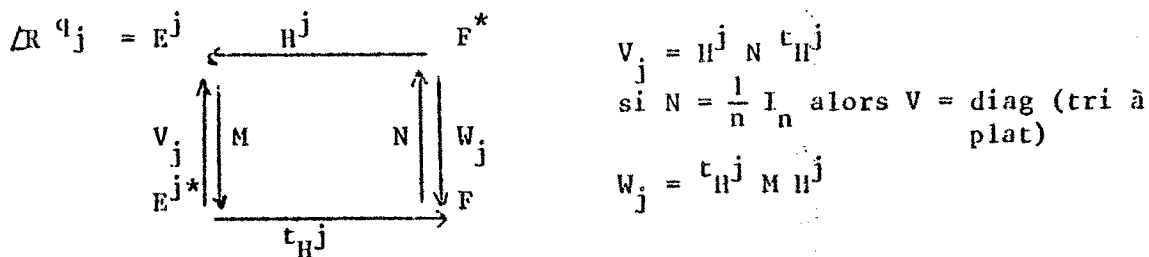
En faisant intervenir les q_j valeurs x_i^j prises par h^j sur \bar{A} on a :

$$h^j = \sum \{ x_i^j l_i ; i = 1 \dots q_j \}$$

Remarque

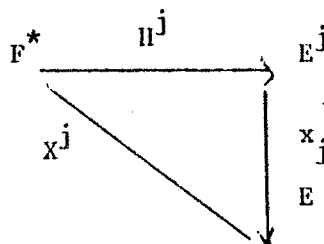
Le tri à plat obtenu à partir du tableau ci-dessus suppose implicitement l'équipondération des objets a. Il représente alors dans ce cas la distribution image par h^j dans x.

L'interprétation linéaire du tableau H^j , usuelle en analyse des données (CAILLES et PAGES [15], d'AUBIGNY [7]) est traduite par le schéma de dualité.



qui ne prend pas en compte l'ensemble des valeurs x^j . Celui-ci intervient dans la définition de la dualité "individus-variables". Le codage h^j est traduit par le vecteur des valeurs x^j associé à la forme linéaire x^j définie sur E^j . On définit ainsi l'application coordonnée (identifiée au codage h^j) par :

$$x^j = x^j_0 H^j$$



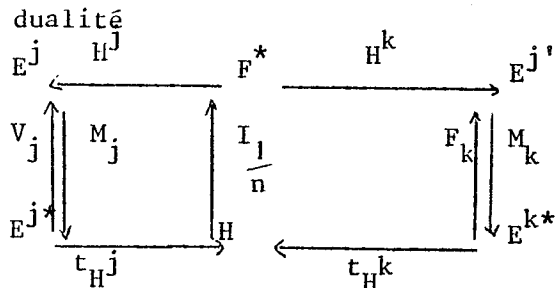
de vecteur associé dans les bases canoniques de E et F^* , le vecteur \underline{x}^j qui forme la $j^{\text{ème}}$ ligne du tableau de données X usuellement traité. X est l'interprétation linéaire du codage h .

Ainsi dans la dualité "individus-variables", on associe à chaque codage h^j une droite Δ^j dans E , et à chaque indicatrice $\mathbb{1}_i^j$, la fibre d'équation $h^j(.) = x_i^j$ lieu des points de coordonnées constante sur Δ^j , égale à x_i^j (D'AUBIGNY, BON, ROMIER [6]). Ceci donne une visualisation de l'échelle. Toute lecture utilisant la position relative des fibres ne peut être fondée que sur des relations pertinentes pour Δ .

1.5. Mesurage et tri croisé.

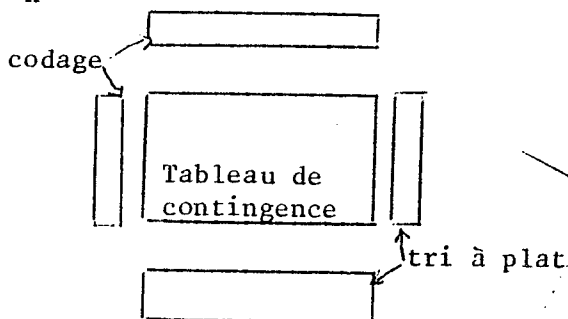
L'étude du tableau X nécessite de croiser les variables X^j deux à deux. Cette opération se traduit algébriquement de deux façons :

- la dualité "individus-catégories" nous conduit à traiter le problème classique de l'analyse des correspondances, traduit par le schéma de dualité



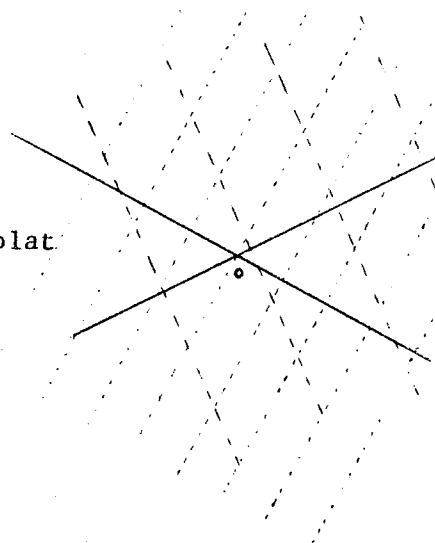
On fait donc l'analyse du tableau de contingence V_{jk} défini comme le tri croisé :

$$V_{jk} = \frac{1}{n} H^j H^k$$



de terme général

$$(V_{jk})_{ii'} = \frac{1}{n} \sum_{a \in A} \mathbb{1}_i^j(a) \mathbb{1}_{i'}^k(a)$$



- la dualité "individus-variables" se traduit dans le croisement des variables X^j et X^k par

$$\begin{aligned} \langle \underline{x}^j, \underline{x}^k \rangle_{\frac{1}{n}} &= \frac{1}{n} \sum_{a \in A} x^j(a) x^k(a) \\ &= \frac{1}{n} \sum_{a \in A} \left[\sum_i x_i^j \parallel_i^j(a) \right] \left[\sum_{i'} x_{i'}^k \parallel_{i'}^k(a) \right] \\ &= \sum_i \sum_j x_i^j x_{i'}^k (v_{jk})_{ii'} \end{aligned}$$

Cette expression montre que la statistique $\langle \underline{x}^j, \underline{x}^k \rangle_{\frac{1}{n}}$ n'est pertinente que pour des échelles absolues.

Remarque

Comme toute analyse canonique, l'analyse des correspondances du tableau V_{jk} construit les codages optimaux h_*^j et h_*^k des catégories définies par h^j et h^k , lus sur le premier axe factoriel, en ce sens qu'ils maximisent la valeur de $\langle \underline{x}^j, \underline{x}^k \rangle_{\frac{1}{n}} - h_*^j$ et h_*^k sont construits comme d'échelle d'intervalle.

Le tri croisé des variables X^j et X^k fournit trois informations qui interviennent dans l'étude de l'association entre ces deux variables :

- les marges du tableau de contingence permettent de comparer les distributions respectives des éléments de A dans les catégories définies par h^j et h^k (information associée à la dualité "individus-catégories") ;
- le tableau de contingence nous fournit la distribution conjointe du couple (X^j, X^k) que l'on compare au produit des marges dans l'analyse des correspondances, sans faire intervenir les valeurs des codages (information associée à la dualité "individus-catégories") ;
- l'étude du tableau de données X conduit à prendre en compte les codages h^j et h^k des lignes et colonnes du tableau de contingence, ce qui distingue l'analyse en composantes principales de l'analyse des correspondances (information associée à la dualité "individus-caractères").

La définition d'un indice d'association entre variables est confrontée à deux problèmes :

- doit-on raisonner à marges fixées ? ceci réfère à un modèle d'échantillonnage (multinomiale, hypergéométrique...). Les marges sont-elles connues ? Autant de questions qui interviennent dans les problèmes d'inférence ;

- doit-on prendre en compte les codages des marges ? si oui, on définit un indice de liaison - dans le cas contraire, on définit un indice de dépendance.

2 - INDEPENDANCE ET ASSOCIATION

Les axiomes du calcul des probabilités permettent de donner une définition non ambiguë de la notion d'indépendance de deux éléments aléatoires X et Y sous la forme :

$$P_{XY} = P_X \otimes P_Y$$

Nous montrons dans un premier paragraphe, qu'à l'opposé, la notion de dépendance parfaite reçoit une définition ambiguë puisqu'elle fait intervenir explicitement le codage des variables X et Y. C'est pourquoi les indices d'association étudiés dans le deuxième paragraphe, qui n'utilisent que l'information apportée par le tableau de contingence, ont pour point de mesurer l'écart entre le tableau observé et le tableau traduisant la situation d'indépendance.

L'utilisation du formalisme du chapitre précédent montre les liens entre ces indices et le choix d'un modèle d'analyse de la structure d'un tableau de contingence en termes de pertinence.

2.1. Indépendance et probabilités

Intuitivement, la variable Y est complètement dépendante de la variable X si Y est parfaitement prédictible par X. C'est ce que LANCASTER[44] formalise en termes probabilités en posant :

Définition 1 :

Soient X et Y deux variables aléatoires réelles définies sur l'espace probabilisé (Ω, A, p) , Y est dite complètement dépendante de X ssi il existe une application borélienne g vérifiant

$$p [Y = g(X)] = 1$$

X et Y sont dites mutuellement complètement dépendantes si chacune des variables est complètement dépendante de l'autre.

Une telle notion fonde l'emploi de la Sup-correlation

$$c'(X,Y) = \text{Sup} \{ \rho(f(X), g(Y)) : f \in L_X^2(\Omega), g \in L_Y^2(\Omega) \}$$

pour mesurer le degré d'association de deux variables de carré sommable. On a noté $L_X^2(\Omega)$ l'ensemble des fonctions boréliennes de X de carré sommable et ρ le coefficient de corrélation de BRAVAIS et PEARSON.

Ces notions justifient l'emploi de l'analyse canonique dans L_{XY}^2 et donnent un critère d'optimalité aux codages f et g obtenues par le premier axe en analyse des correspondances. Elles sont cependant mises en défaut par des arguments de convergence puisque KIMELDORF[48] construit une suite de couples de variables aléatoires mutuellement complètement indépendantes dont la limite en loi est un couple de variables indépendantes. Cet auteur introduit alors une notion de dépendance qui ne présente pas cette faiblesse :

Définition 2 :

La variable aléatoire Y est monotone dépendante de X s'il existe une fonction monotone g vérifiant :

$$P[Y = g(X)] = 1$$

les variables aléatoires X et Y sont mutuellement monotones dépendantes si chacune est monotone dépendante de l'autre.

Remarque :

Cette définition de la dépendance plus satisfaisante du point de vue probabiliste suppose des variables d'échelle ordinale. On notera cependant que les échelles nominales sont invariantes par permutation et donc la relation de monotone dépendance est pertinente pour de telles échelles.

La définition de KIMELDORF est justifiée par les travaux de FRECHET [32] qui montre que la fonction de répartition de la loi conjointe du couple (X,Y) vérifie :

$$H^-(x,y) \leq H(x,y) \leq H^+(x,y) \quad \forall (x,y), (x,y) \in \mathbb{R}^2$$

la borne supérieure (resp. inférieure) de FRECHET, notée H^+ (resp. H^-) est la fonction de répartition bivariée singulière qui affecte la probabilité

à l'ensemble $\{(x,y) : F(x) = G(y)\}$ (resp $\{(x,y) : F(x) + G(y) = 1\}$),
 F et G étant les fonctions de répartitions marginales de X et de Y. Alors :

$$H^+(x,y) = \text{Min} (F(x), G(y)) \quad \forall (x,y), (x,y) \in \mathbb{R}^2$$

$$H^-(x,y) = \text{Max} (0, F(x) + G(y) - 1) \quad \forall (x,y), (x,y) \in \mathbb{R}^2$$

La fonction de répartition

$$H_0(x,y) = F(x) G(y) \quad \forall (x,y), (x,y) \in \mathbb{R}^2$$

correspondant à l'indépendance du couple (X, Y) définit deux familles de loi bivariée, de marges F(x) et G(y) données,

$$\mathcal{F}^+ = \{H : H_0(x,y) \leq H(x,y) \leq H^+(x,y), \forall (x,y), (x,y) \in \mathbb{R}^2\}$$

$$\mathcal{F}^- = \{H : H^-(x,y) \leq H(x,y) \leq H_0(x,y), \forall (x,y), (x,y) \in \mathbb{R}^2\}$$

Définition 3 (LEHMANN [56]) :

La loi de fonction de répartition bivariée H est dite positive (resp. négative) quadrant dépendante (notée PQD (resp. NQD) ssi H est un élément de \mathcal{F}^+ (resp. \mathcal{F}^-).

Remarque :

La notion de quadrant dépendance ne peut s'appliquer aux variables d'échelle nominale, puisque la notion de fonction de répartition suppose définie une relation d'ordre partiel sur l'ensemble des valeurs et n'est pas invariante par permutation.

Cependant dans le cas d'un couple de variables dichotomiques, pour tout ordre choisi sur les modalités, la fonction de répartition correspondante appartient à \mathcal{F}^+ ou à \mathcal{F}^- selon le signe de $p_{11}p_{22} - p_{21}p_{12}$.

Exemples :

- La loi normale bivariée de coefficient de corrélation positive est P.Q.D.
- soit U et V deux variables indépendantes, alors la loi du couple (U, U + V) est PQD

- soient U et V deux variables indépendantes de loi absolument continue alors la loi du couple (X,Y) est PQD si
 - . $0 \leq \lambda < 1$, $X = U$ et $Y = \lambda U + \sqrt{1 - \lambda^2} V$
- toute loi de fonction de répartition
 - $H = (1 - \lambda) H_0 + \lambda H^+$ avec $0 \leq \lambda \leq 1$
 - est PQD

La relation qui lie la notion de monotone dépendance aux bornes de FRECHET est donnée par le théorème suivant dû à KIMELDORF [41] dans le cas d'une loi absolument continue et étendu au cas discret par AHMET et ALLII [2]

Théorème 1 :

Soit (X, Y) un couple de variables de fonction de répartition élément de \mathcal{F}^+ et absolument continue (resp. quelconque). Une condition nécessaire et suffisante pour qu'il existe une fonction g strictement croissante (resp. croissante) telle que $P(Y = g(X)) = 1$ est que H soit égale à H^+ .

Ce théorème montre que le concept de monotone dépendance traduit la situation la plus "éloignée" de la situation d'indépendance. On notera cependant que sa définition prend en compte le codage des variables puisque ce théorème montre que la borne supérieure de FRECHET est atteinte pour une distribution de \mathcal{F}^+ , lorsque X et Y appartiennent à la même échelle ordinale.

Entre les deux situations extrêmes que constituent dans \mathcal{F}^+ l'indépendance et la monotone dépendance, la quadrante dépendance nous permet d'ordonner les éléments de \mathcal{F}^+ :

Définition 4 :

Soient P_1 et P_2 deux lois de probabilité de fonctions de répartition H_1 et H_2 appartenant à \mathcal{F}^+ , P_2 est dite positivement quadrante dépendante que P_1 (noté $P_2 \succcurlyeq P_1$) ssi

$$\forall (x,y), (x,y) \in \mathbb{R}^2 \quad H_2(x,y) \geq H_1(x,y)$$

Remarque :

Lorsque $H_1(x,y)$ est égale au produit des marges, on retrouve la notion de positive quadrant dépendance.

Lorsqu'on se limite à l'étude de couples (X,Y) de variables aléatoires de carré sommable, la définition ci-dessus se traduit en termes géométriques dans L^2_{XY} , ceci fait l'objet des propositions suivantes.

Définition 5 :

Deux applications f et g de \mathbb{R}^n dans \mathbb{R} sont dites concordantes ssi lorsque $n-1$ quelconques des n coordonnées sont fixées, f et g sont des fonctions monotone de même sens de la coordonnée restante.

Lemme de Hoeffding

Soit H (resp. F et G) la fonction de répartition conjointe (resp. marginales) de la loi du couple (X,Y) de variables carré sommables, alors

$$\text{cov}(X,Y) = \int_{\mathbb{R}^2} (H(x,y) - F(x)G(y)) dx dy$$

si l'intégrale du deuxième membre existe.

Démonstration

$$\text{Notons } I(u,x) = \begin{cases} 1 & \text{si } u \leq x \\ 0 & \text{sinon} \end{cases}$$

soit (X_1, Y_1) et (X_2, Y_2) deux couples indépendants de même loi que (X, Y)

$$\begin{aligned} \text{cov}(X_1, Y_1) &= \frac{1}{2} E[(X_1 - X_2)(Y_1 - Y_2)] \\ &= \frac{1}{2} E\left[\int_{\mathbb{R}^2} (I(u, X_1) - I(u, X_2))(I(v, Y_1) - I(v, Y_2)) dv du\right] \end{aligned}$$

en utilisant FUBINI, on déduit le résultat.

Proposition 1

Soit P_1 et P_2 deux lois de probabilité telles que $P_2 \succ_{PQD} P_1$

Soit $\{(X_i, Y_i) ; i = 1..n\}$ un échantillon aléatoire de loi P_1 ou P_2 on pose $X = f(X_1, \dots, X_n)$ et $Y = g(Y_1, \dots, Y_n)$ où f et g sont deux fonctions concordantes alors si X et Y sont de carré intégrable

$$\text{cov}_2(X,Y) \geq \text{cov}_1(X,Y).$$

Démonstration (par récurrence)

Pour $n = 1$ le lemme de Hoeffding permet d'écrire

$$\text{cov}_2(X_1, Y_1) - \text{cov}_1(X_1, Y_1) = \int_{\mathbb{R}^2} (H_2(x, y) - H_1(x, y)) \, dx \, dy \geq$$

Le résultat s'en déduit par le changement de variables :

$$(X_1, Y_1) \rightarrow (X, Y)$$

par suite pour (x_i, y_i) fixé ($i = 2 \dots n$) le résultat est acquis puisque f et g sont concordantes :

$$\text{cov}_2 [f(X_1, x_2, \dots, x_n) g(Y_1, y_2, \dots, y_n)] - \text{cov}_1 [f(X_1, x_2, \dots, x_n) g(Y_1, y_2, \dots, y_n)] \geq 0$$

il en résulte que :

$$\begin{aligned} & E_2 [f(X_1, x_2, \dots, x_n) g(Y_1, y_2, \dots, y_n)] - E_1 [f(X_1, x_2, \dots, x_n) g(Y_1, y_2, \dots, y_n)] \\ & \geq E_2 [f(X_1, x_2, \dots, x_n)] E_2 [g(Y_1, y_2, \dots, y_n)] - E_1 [f(X_1, x_2, \dots, x_n)] \\ & \quad E_1 [g(Y_1, y_2, \dots, y_n)] \end{aligned}$$

Les fonctions $E_i [f(X_1, x_2, \dots, x_n)]$ et $E_i [g(Y_1, y_2, \dots, y_n)]$ ($i = 1, 2$) sont des fonctions concordantes sur \mathbb{R}^{n-1} par suite en prenant l'espérance de deux termes de l'inégalité, l'hypothèse de récurrence nous assure du résultat.

Corollaire 1

Sous les hypothèses de la proposition 1 :

$$i) (\text{cov}_2(X, Y) = \text{cov}_1(X, Y)) \Leftrightarrow (P_1 = P_2)$$

ii) $\text{cov}(X, Y) > 0$, l'égalité ayant lieu ssi X et Y sont indépendantes.

Remarque :

Le résultat ii) généralise aux lois quadrant dépendantes, un résultat connu dans le cas d'un vecteur gaussien bivarié.

Corollaire 2

Soit P_1 et P_2 deux lois de probabilité telles que $P_2 \underset{PQD}{\geq} P_1$
 soit $\{(X_i^k, Y_i^k) ; i = 1 \dots n, \}$ un échantillon aléatoire de la loi P_k ($k = 1, 2$)
 soit f et g deux fonctions concordantes sur \mathbb{R}^n .

On note Q_k la loi du couple $(f(X_1^k, X_2^k, \dots, X_n^k), g(Y_1^k, Y_2^k, \dots, Y_n^k))$ alors

$$Q_2 \underset{PQD}{\geq} Q_1$$

Démonstration

$$f_x^*(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } f(x_1, \dots, x_n) \leq x \\ 0 & \text{sinon} \end{cases}$$

$$g_y^*(y_1, \dots, y_n) = \begin{cases} 1 & \text{si } g(y_1, \dots, y_n) \leq y \\ 0 & \text{sinon} \end{cases}$$

sont deux fonctions concordantes sur \mathbb{R}^n pour tout x et tout y . Par suite les propriétés liant probabilité et espérance conditionnelle d'une indicatrice assure le résultat comme conséquence directe de la proposition 1.

Exemples

$$\text{- Notons } S(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases}$$

alors les coefficients τ de KENDALL, ρ_s de SPEARMAN, q de BLUMQUIST sont des fonctions croissantes de la loi de probabilité pour l'ordre $\underset{PQD}{\geq}$ puisqu'ils s'écrivent :

$$\tau = \text{cov}(S(X_1 - X_2), S(Y_1 - Y_2))$$

$$\rho_s = \text{cov}(S(X_1 - X_2), S(Y_1 - Y_3))$$

$$q = \text{cov}(S(X - m_e(X)), S(Y - m_e(Y)))$$

où m_e représente la médiane.

2.2. Indice d'indépendance :

Nous avons vu que la notion de l'association entre deux variables fondée sur la seule connaissance du tableau de contingence ne peut s'exprimer qu'en terme de non-indépendance. La construction d'une mesure d'association dans un tel cadre nécessite alors de mesurer l'écart entre la loi donnée P et la loi P_0 correspondant à la situation d'indépen-

dance. Les principales mesures proposées sont issues de la théorie de l'information et ont une structure commune que nous étudions.

2.2.1. Divergence entre lois de probabilité :

Soit P_1 et P_2 deux mesures de probabilité définies sur un espace probabilisable (E, Σ) dominées par une même mesure λ , on notera f_1 et des dérivées de RADON-NIKODYM des probabilités P_1 et P_2 par rapport à λ . Il est demandé à une mesure l'écart entre P_1 et P_2 de vérifier les quatre propriétés de base suivantes

la mesure doit être minimum (resp. maximum) lorsque P_1 et P_2 sont confondu (resp. singulières).

la mesure de l'écart ne peut augmenter par transformation des variables de loi P_1 et P_2 et doit être invariante par recodage, ce qui nous assure qu'elle ne dépend que du tableau de contingence, et diminue avec tout regroupement en classes (ce principe recouvre la notion d'exhaustivité étudiée par DEVILLE [26]).

dans un processus déchantillonnage, la mesure de l'écart entre les lois image augmentent avec la taille de l'échantillon.

enfin lorsque la mesure de l'écart entre deux lois est considérée comme un problème de discrimination, cette mesure doit être une fonction décroissante du taux de mal classés, encouru dans le problème de décision associé aux deux hypothèses

H_i : la loi de X est la loi p_i $i = 1, 2$

il résulte du théorème de BAYES que

$$p(H_i/x) = \frac{p(H_i) f_i(x)}{p(H_1) f_1(x) + p(H_2) f_2(x)} \quad i = 1, 2$$

on en déduit que $\frac{p(H_1/x)}{p(H_2/x)} = \frac{p(H_1)}{p(H_2)} \times \frac{f_1(x)}{f_2(x)}$

une telle relation montre que l'évidence en faveur de H_1 mesuré par le rapport des probabilités a posteriori augmente ou diminue par rapport à l'évidence en faveur de H_1 a priori selon que le rapport $\frac{f_1(x)}{f_2(x)}$ est supérieure ou inférieure à 1. L'étude de ce rapport de vraisemblance conduit ALI et SILVEY [3] à poser la définition.

Définition 6 :

Soit c une fonction réelle convexe continue définie sur \mathbb{R}^+ , on appelle divergence entre les lois P_1 et P_2 la quantité

$$d(P_1, P_2) = E_2 \left[c \left(\frac{f_1(X)}{f_2(X)} \right) \right]$$

où E_2 est l'espérance mathématique par rapport à la loi P_2

Remarque :

Si on choisit la fonction c telle que $c(1) = 0$, alors une divergence nulle implique que P_1 est égale à P_2 .

Remarque :

Cette définition suppose que P_1 est absolument continue par rapport à P_2 . ALI et SILVEY généralisent cette définition au cas où P_1 a une composante singulière par rapport à P_2 .

La divergence $d(P_1, P_2)$ vérifie les quatre propriétés citées ci-dessus. Cela fait l'objet des propositions suivantes :

Proposition 2 :

Soit X une variable aléatoire à valeurs dans (E, Σ) , de loi P_i , et T une fonction mesurable définie sur (E, Σ) de loi $P'_i = (i = 1, 2)$ alors

$$d(P'_1, P'_2) \leq d(P_1, P_2)$$

l'égalité ayant lieu si T est exhaustive

Démonstration :

Les probabilités $P'_i = P_i T^{-1}$ ($i = 1, 2$) sont absolument continues par rapport à la mesure $\lambda' = \lambda T^{-1}$. Si les P_i sont absolument continues par rapport à λ ; leurs dérivées de Radon-Nikodym $g_i = \frac{dP'_i}{d\lambda'}$ vérifient alors

$$\frac{g_1}{g_2} \circ T = E_2^T \left[\frac{f_1}{f_2}(X) \right] \quad P_2 \text{ p.s.}$$

où $E_2^T(\cdot)$ est l'espérance conditionnelle en T pour la loi P_2 par suite il résulte de l'inégalité de Jensen :

$$\begin{aligned}
 d(P_1, P_2) &= E_2 \left[c\left(\frac{f_1}{f_2}(X)\right) \right] = E_2 \left[E_2^T \left(c\left(\frac{f_1}{f_2}(X)\right) \right) \right] \\
 &\geq E_2 \left[c\left(\frac{g_1}{g_2} T(X)\right) \right] = d(P'_1, P'_2)
 \end{aligned}$$

si T est exhaustive alors il résulte du théorème de factorisation que

$$\frac{g_1}{g_2} \circ T = \frac{f_1}{f_2} \quad \lambda \text{ p.s.}$$

Remarque

ALI et SILVEY montrent que cette propriété est encore vérifiée dans le cas où P_1 n'est pas absolument continue par rapport à P_2 .

Corollaire 3

$d(P_1, P_2)$ est minimum si P_1 est égale à P_2

Démonstration

Si $P_1 = P_2$ alors $\frac{f_1}{f_2} = 1$ d'où $d(P_1, P_2) = c(1)$

Soit T l'application mesurable constante alors P'_1 est égale à P'_2 d'après la proposition 2 $d(P_1, P_2) \geq d(P'_1, P'_2) = c(1)$ pour tout P_1 et P_2

Corollaire 4

Soit $P_i^{(n)}$ la puissance tensorielle $n^{\text{ième}}$ de la loi P_i ($i = 1, 2$) alors $d(P_1^{(n)}, P_2^{(n)})$ est une fonction croissante de n .

La démonstration est immédiate en considérant l'application qui à l'élément (x_1, x_2, \dots, x_n) associe (x_1, x_2, \dots, m) où $m < n$

Remarque

Soit f une fonction réelle, définie sur \mathbb{R} , croissante, alors si d est une mesure de divergence, $f(d)$ vérifie les propriétés d'une divergence. On appellera donc encore $f(d(P_1, P_2))$ une divergence entre P_1 et P_2 .

2.2.2. Exemples de mesure de la divergence

La notion de divergence entre lois de probabilité a été introduite par KULLBACK, à partir de l'information de SHANNON

Définition 7

On appelle information discriminante de KULLBACK et LEIBLER de P_1 contre P_2 l'expression

$$I(P_1, P_2) = \int_E \log \left[\frac{f_1(x)}{f_2(x)} \right] f_1(x) d\lambda(x)$$

$I(P_1, P_2)$ est la divergence associée à la fonction

$$c(t) = t \log t$$

on remarque que $I(P_1, P_2)$ n'est pas une distance.

Nous étendons alors la définition précédente à des formes plus générales de la mesure de l'information.

Définition 8

On appelle information discriminante d'ordre α , d'espace β de P_1 contre P_2 l'expression

$$I_{\alpha\beta}(P_1, P_2) = \frac{-1}{1-2^{1-\beta}} \left[1 - \left(\int_E \left(\frac{f_1(x)}{f_2(x)} \right)^\alpha f_2(x) d\lambda(x) \right)^{\frac{\beta-1}{\alpha-1}} \right]$$

Proposition 3

$I_{\alpha\beta}(P_1, P_2)$ est une divergence si α et β vérifient les conditions $\alpha > 1$ ou $\alpha < 1$ et $\beta < 1$

La démonstration résulte de la remarque ci-dessus

- lorsque α et β tendent vers 1 on retrouve l'information discriminante de KULLBACK et LEIBLER

- lorsque $\alpha = \beta = 2$

$$I_{22}(P_1, P_2) = 2 \left[\int_E \frac{f_1^2(x)}{f_2(x)} d\lambda(x) - 1 \right]$$

$I_{22}(P_1, P_2)$ est la divergence associée à la fonction
 $c(t) = (t - 1)^2$

qui permet de retrouver la distance du χ^2 centrée sur la loi P_2 .
 L'importance accordée à cette information discriminante en analyse des
 données est justifiée par le fait que l'on retrouve une distance hilbertien
 ne.

- lorsque $\alpha = \beta = \frac{1}{2}$

$$I_{1/2 \ 1/2}(P_1, P_2) = \frac{1}{\sqrt{2} - 1} \left[1 - \int_E \sqrt{f_1(x) f_2(x)} d\lambda(x) \right]$$

$I_{1/2 \ 1/2}(P_1, P_2)$ est la divergence associée à la fonction
 $c(t) = (\sqrt{t} - 1)^2$

On retrouve la distance de MATUSITA [53] qui, elle aussi, est hilbertienne
 associée au produit scalaire

$$\langle P_1, P_2 \rangle = \int_E \sqrt{f_1(x) f_2(x)} d\lambda(x)$$

appelé "affinité de BHATTACHARYA"

La méthode d'analyse sphérique proposée par DOMENGES et VOLLE [27] utilise
 les propriétés géométriques de cette divergence tout comme l'analyse des
 correspondances utilise celles de la distance χ^2 .

Soit A (resp. \bar{A}) la région d'acceptation de la loi P_1 (resp. P_2) dans le
 problème de discrimination de ces deux lois, l'erreur totale de discrimi-
 nation

$$\varepsilon(A) = P_1(\bar{A}) + P_2(A)$$

est minimum pour

$$A^* = \{x / \frac{f_1(x)}{f_2(x)} > 1\}$$

on vérifie alors :

$$\begin{aligned} \langle P_1, P_2 \rangle &= \int_A \sqrt{\frac{f_1(x)}{f_2(x)}} f_2(x) d\lambda(x) + \int_{\bar{A}} \sqrt{\frac{f_2(x)}{f_1(x)}} f_1(x) d\lambda(x) \\ &\geq P_2(A^*) + P_1(\bar{A}^*) \end{aligned}$$

par suite la distance de MATUSITA est une fonction monotone décroissante du taux optimal de mal classés.

Ces exemples montrent que le concept d'information discriminante permet de construire des mesures de divergence. On peut cependant construire des mesures de divergence respectant la définition 6 bien que n'étant pas une information discriminante $I_{\alpha\beta}$. Tel est le cas de la "distance en variation" proposée par KOLMOGOROFF et étudiée par ADHIKARI et JOSHI [1]

$$d(P_1, P_2) = \frac{1}{2} \int | f_1(x) - f_2(x) | dx$$

cette divergence est associée à la fonction

$$c(t) = | t - 1 |$$

elle est reliée à l'erreur $\epsilon(A^*)$ par la relation

$$1 - d(P_1, P_2) = \epsilon(A^*)$$

Notons de plus la divergence proposée par RAO sous le nom de "distance de l'erreur minimum commune" et étudiée par ADHIKARI et JOSHI ne respecte pas la définition 6 tout en vérifiant la proposition 2. Son expression est :

$$d(P_1, P_2) = 1 - \alpha, \quad \alpha = \text{Inf} \{ P_1(\bar{A}) : A \text{ tel que } P_1(\bar{A}) = P_2(A) \}$$

2.2.3. Indices de dépendance

Nous avons vu que les indices de dépendance mesurent l'écart à la situation d'indépendance, leurs constructions reposent alors sur le concept de divergence appliquée aux deux distributions P_{XY} et $P_X \otimes P_Y$ données par la table de contingence.

Cette pratique se heurte à l'absence de normalisation des indices ainsi construits.

Deux types de normalisation sont possibles : la première s'applique lorsque l'ensemble E d'objets étudiés a été construit comme échantillon. Dans ce cas, on calcule la quantité suivante comprise entre 0 et 1 :

$$P(\chi^2 \geq d(P_{XY}, P_X \otimes P_Y))$$

puisque toutes les divergences respectant la définition 8 ont pour loi asymptotique la loi $\chi^2_{(r-1)(c-1)}$ (BERKSON [11]). Le plus souvent une telle interprétation n'est pas admissible ou bien encore les conditions d'approximation de la loi de l'indice par un χ^2 ne sont pas remplies, on doit donc normaliser l'indice utilisé suivant une argumentation non probabiliste qui dépend de son expression. Ce problème est illustré ici par les exemples suivants choisis parce qu'ils sont calculés dans S.P.S.S. [79] et figurent donc dans les tableaux simulés qui illustrent ce travail.

Exemple 1 : la statistique du ϕ^2

$$\phi^2 = \frac{\sum_{i=1}^r \sum_{j=1}^c p_{ij}^2}{\sum_{i=1}^r p_{i.} \sum_{j=1}^c p_{.j}} - 1$$

CRAMER [18] montre que ϕ^2 varie entre 0 et $q-1$ où q est le minimum du nombre de lignes r et du nombre de colonnes c du tableau de contingence étudié. Il propose l'indice :

$$v = \left(\frac{\phi^2}{q-1} \right)^{1/2}$$

une telle normalisation semble plus fondée que celle proposée par PEARSON, fondée sur l'étude du cas particulier de variables gaussiennes. Il propose le "coefficient de contingence"

$$c = \left(\frac{\phi^2}{1 + \phi^2} \right)^{1/2}$$

qui n'attend jamais la borne supérieure 1.

Une troisième normalisation, proposée par TCHUPROV sera justifiée par la suite :

$$T = \left(\frac{\phi^2}{\sqrt{(r-1)(c-1)}} \right)^{1/2}$$

Exemple 2 : l'information minimum discriminante de KULLBACK et LEIBLER

$$MDI = - \sum_{i=1}^r \sum_{j=1}^c p_{ij} \log \frac{p_{ij}}{p_{i.} p_{.j}}$$

$$= \sum_{i=1}^r p_i \cdot \log p_i + \sum_{j=1}^c p_j \log p_j - \sum_{i=1}^r \sum_{j=1}^c p_{ij} \log p_{ij}$$

Le coefficient d'incertitude normalisé MDI de façon à mesurer la réduction d'entropie H (au sens de SHANNON) qui est gagnée par la connaissance du tri croisé par rapport à la seule connaissance des tris à plat

$$U = 1 - \frac{\sum_{i=1}^r \sum_{j=1}^c p_{ij} \log p_{ij}}{\sum_{i=1}^r p_i \log p_i + \sum_{j=1}^c p_j \log p_j} = \frac{\text{MDI}}{H(p_X) + H(p_Y)}$$

BELL [] propose deux autres normalisations

$$U_1 = \frac{\text{MDI}}{\text{Min}(H(p_X), H(p_Y))}$$

$$U_2 = \frac{\text{MDI}}{\text{Max}(H(p_X), H(p_Y))}$$

U_1 (resp. U_2) égale 1 lorsqu'il existe une inclusion (resp. égalité) entre les partitions induites par X et Y sur E

Exemple 3 : distance de MATUSITA

$$d_A = \left(\sum_{i=1}^r \sum_{j=1}^c (\sqrt{p_{ij}} - \sqrt{p_i \cdot p_j})^2 \right)^{1/2}$$

est un indice naturellement normalisé à $\sqrt{2}$ près, puisque toute loi de probabilité somme à 1.

Le comportement des indices calculés par S.P.S.S. est présenté dans les simulations données en annexe. La comparaison des tableaux 33 à 42 est particulièrement centrée sur l'effet de regroupements en classes, et des normalisations.

L'étude des tables de contingence obtenues par croisement de deux variables nominales au moins, a fait l'objet d'une très vaste littérature, axée sur la recherche de modèles assez complexes pour ajuster convenablement la loi observée P_1 , mais assez simples pour permettre l'interprétation.

Les méthodes proposées, sont en général axés sur la recherche d'interactions dans des modèles construits par analogie avec l'analyse de la variance.

Les fréquences observées dans les cases d'un tableau de contingence sont interprétés comme les valeurs d'une variance réponse d'échelle de rapport, qu'il convient d'expliquer à l'aide des interactions et effets principaux de facteurs qualitatifs dont le croisement définit la table de contingence. Le recours à un modèle linéaire (CARLIER [16], DAUDIN [24]) log-linéaire (HABERMAN [40], BISHOP et alli [12] , COLOMBIER [18]) ou autre est par une large part arbitraire dans la mesure où chacun définit un système relationnel numérique associé au système de relationnel empirique. Le choix d'un modèle est alors fonction de la simplicité de sa résolution mathématique liée à la pertinence de la divergence choisie pour juger de la plus ou moins grande adéquation du modèle, traduite par une "formule de décomposition de la divergence" analogue à la décomposition de la variance. Ainsi le modèle log-linéaire est associé au MDI et modèle linéaire au ϕ^2 . L'analyse de la divergence apparaît donc comme un équivalent non paramétrique de l'analyse de la variance. GARNER et MAC GILL [33] montre que la variance et la divergence ne mesure pas exactement les mêmes caractéristiques puisque la divergence dépend du nombre de catégories observées alors que la variance dépend aussi des codages de ces catégories, par suite certains termes d'interactions significatifs pour un des modèles peuvent ne pas l'être pour l'autre.

3. INDICES DE CONNEXION

Nous avons montré dans le chapitre précédent que la complète dépendance de la variable Y en la variable X se traduisait par le fait que Y apparaissait comme une fonction de X . Les indices d'association mesurant l'écart à la situation de dépendance complète sont appelés indices de connexion selon une terminologie due à GINI [34] lorsqu'on ne recherche pas une mesure de la mutuelle complète dépendance.

Définition 1.

Soient X et Y deux variables aléatoires définies sur une même population, Y est dite connexe à X s'il existe une fonction f telle que $Y = f(X)$

Cette définition accorde un rôle dissymétrique aux variables X et Y traduit dans le langage courant par des dénominations distinctes attribuées à X et à Y . Suivant le domaine d'application X sera appelée facteur, variable indépendante, explicative ou encore prédictive alors que Y sera appelée variable réponse, dépendante, expliquée ou prédite. Le problème de prédiction de Y par X ainsi posé peut se formuler de la façon suivante : connaissant la valeur affectée à un objet donné par le codage X , peut-on prédire la valeur de cet objet pour le codage Y ? La réponse à cette question nécessite de définir deux choses :

- la nature de la relation liant Y à X , traduite par la donnée d'un modèle structural de liaison, de nature algébrique,
- un procédé de prédiction définissant des règles de recodage adapté au modèle.

aussi ces deux points interviennent-ils dans la définition d'un indice de connexion.

3.1. Indice de connexion et logique P.R.E.

En 1954, GOODMAN et KRUSKAL [37], dans un souci d'interprétabilité des valeurs de la mesure de liaison ont développé une procédure de construction d'indices de connexion à partir de règles de prédiction, pour l'étude des tableaux de contingence, le succès de prédiction étant alors mesuré en termes de probabilité d'erreur. Cette logique P.R.E. (Proportional Reduction of Error) a été formalisée par COSTNER qui étend la mesure de l'erreur de prédiction à des variables d'échelle quelconque.

La procédure P.R.E. de COSTNER repose sur la donnée de :

1. une règle (C) de prédiction liée au codage Y lorsque X est connu
2. une règle (I) de prédiction liée au codage Y lorsque X est inconnu
3. une définition de l'erreur de prédiction dans le cas (C) et (I) et une définition de sa mesure.

l'indice P.R.E. correspondant est alors donné par le rapport.

$$\frac{\text{erreur selon (I)} - \text{erreur selon (C)}}{\text{erreur selon (I)}}$$

qui mesure donc la Proportion de Réduction de l'Erreur de prédiction de Y due à la connaissance de X.

Exemple 1.:

Le coefficient λ de GUTTMANN présenté par GOODMAN et KRUSKAL [37] comme un indice associé à une prédiction optimale.

X et Y sont deux variables d'échelles nominales.

la procédure P.R.E. consiste alors en :

1. on prédit pour Y le mode de la distribution conditionnelle
2. on prédit pour Y le mode de la distribution marginale
3. l'erreur est dans les deux cas la probabilité hors-mode l'indice P.R.E. correspondant s'écrit :

$$\lambda_{Y/X} = \frac{(1 - \text{Max}_j p.j) - (1 - \sum_{i=1}^c p.i \text{Max}_j \frac{p.j}{p.i})}{1 - \text{Max}_j P.j}$$

Exemple 2.:

Le rapport de corrélation η^2

Soient X et Y deux variables d'échelle d'intervalle supposées de variance finie.

La procédure P.R.E. consiste en :

1. on prédit pour Y la valeur de l'espérance conditionnelle notée $E^X(Y)$ au point $X = x$
2. on prédit pour Y son espérance
3. l'erreur selon (C) est égale a :

$$E [(Y - E^X(Y))^2]$$

l'indice P.R.E. correspondant s'écrit :

$$\eta^2_{Y/X} = \frac{E[(Y-E(Y))^2] - E\{(Y-E^X(Y))^2\}}{E[(Y-E(Y))^2]}$$

l'interprétation P.R.E. du rapport de corrélation est fondée sur une double interprétation de la variance. D'une part, celle-ci mesure la dispersion de la loi de Y en termes de moment centré d'ordre 2 ; elle apparaît donc comme une incertitude sur Y due au caractère aléatoire de cette variable. D'autre part, elle mesure l'erreur quadratique moyenne commise en remplaçant la valeur prise par Y sur tout objet par la valeur moyenne - cette dernière interprétation privilégie le rôle des codages.

Nous étudions dans la suite la notion d'incertitude suivant les voies de FERON [27] pour en déduire une famille d'indices de connexion.

3.2. La notion d'incertitude

La définition d'un indice P.R.E. que l'on ait toujours une réduction positive de l'erreur. Dans les deux exemples précédents, cette propriété résulte de concavité de la mesure d'erreur. On pose donc la

Définition 2.

On appelle incertitude de X notée $I(P_X)$, toute fonctionnelle positive, concave, continue, définie sur le simplexe des lois de probabilités sur E.

3.2.1. Entropie et incertitude

La partition induite sur Ω par une variable d'échelle nominale nous permet d'utiliser les concepts de la théorie de l'information. L'entropie de SHANNON définie par :

$$H(P) = - \sum_i^c p_i \log p_i$$

vérifie la propriété

$$H(P) = \text{Min} \left\{ - \sum_i p_i \log q_i ; Q \in S_c \right\}$$

où S_c est le simplexe des lois de probabilités :

$$S_c = \{ P = (p_1, p_2, \dots, p_c) ; \sum p_i = 1, p_i \geq 0 \}$$

cette propriété d'optimalité est généralisée, par SHARMA et GUPTA [70] dans les termes présentés ci-dessous.

Définition 3.

Soit ψ une fonction réelle, continue, strictement monotone, définie sur \mathbb{R} et W une fonction poids telle que :

$$\sum_i W(p_i) = 1 \quad \forall (p_1, \dots, p_c) \in S_c$$

on appelle inexactitude la fonction réelle $F(P, Q)$ définie sur $S \times S$ par

$$F(P, Q) = \psi^{-1} \left[\sum_{i=1}^c w(p_i) \psi(F(p_i, q_i)) \right]$$

Remarque

La définition de $F(P, Q)$ généralise la notion de moyenne pondérée.

Définition 4.

On appelle entropie de la loi P , la valeur de la fonction $F(P, Q)$ en son minimum (resp. maximum) lorsque Q décrit S_c , si F est convexe (resp. concave)

Remarques

Lorsque ψ et W sont l'identité, on retrouve l'entropie de SHANNON pour :

$$F(p_i, q_i) = -\log q_i$$

SHARMA et GUPTA [61] montre que l'on retrouve les entropies :

$$I(P, \alpha, \beta) = (1 - 2^{1-\beta})^{-1} \left[1 - \left(\sum_i p_i^\alpha \right)^{\frac{\beta-1}{\alpha-1}} \right] \quad \alpha \neq 1, \beta \neq 1$$

ainsi que les cas limite ($\alpha \rightarrow 1$, $\beta \rightarrow 1$, $\alpha \rightarrow \infty$) comme optima associés à des fonctions d'inexactitude non additive.

Exemples 3.

Nous donnons quelques exemples que nous reprendrons par la suite.

(nous conservons abusivement la notation $I(P, \alpha, \beta)$ pour les cas limite).

3.1. $I(P, 1, 1) = -\sum p_i \log p_i$ (entropie de SHANNON)

3.2. $I(P, 2, 2) = 2 \left[1 - \sum p_i^2 \right]$ (entropie quadratique)

cette entropie est égale à l'erreur de prédiction proportionnelle associée au coefficient τ de GOODMAN et KRUSKAL [37]

3.3. $I(P, \frac{1}{2}, \frac{1}{2}) = (1 - \sqrt{2})^{-1} \left[\sum \sqrt{p_i} - 1 \right]$ (associée à l'affinité de

BHATTACHARYA)

3.4. $I(P,0,0) = m-1$ si m valeurs p_i sont strictement positives (associé au ϕ^2)

3.5. $I(P,\infty,2) = 1 - \text{Max}_i p_i$

Cette entropie mesure le pourcentage de mal classés quand on décide d'affecter tout individu à la classe modale (cf. exemple 1).

Propriétés 1.

- P1 : $I(P,\alpha,\beta)$ est une fonction continue et symétrique en p_1, p_2, \dots, p_c
 P2 : $I(P,\alpha,\beta)$ est une fonction soit concave, soit convexe de P
 P3 : $0 \leq I(P,\alpha,\beta) \leq I(\pi,\alpha,\beta) \quad \forall \alpha, \beta$, où π est la distribution uniforme de S_c
 $I(P,\alpha,\beta) = 0$ ssi il existe i_0 tel que $P = S_{i_0}$ (distribution de Dirac)
 P4 : $I(PT^{-1},\alpha,\beta) \leq I(P,\alpha,\beta)$ pour toute application mesurable T avec égalité si T est exhaustive.

Démonstration

P1 est évidente. La continuité assure la convergence des estimateurs de maximum de vraisemblance. La démonstration de P2 est donnée par VAN der PYL [75]. Les propriétés 4 et 5 sont des conséquences immédiates des propriétés analogues démontrées sur les divergences dans la proposition 2 du chapitre 2.

Remarque

Dans le cas $c = 2$, EFRON [28] retrouve certains des exemples donnés à l'aide de la fonction d'inexactitude.

$$F(P,Q) = \sum_i p_i f(q_i)$$

où f est une fonction continue, décroissante, et telle que :

$$F(P,P) = \text{Inf}\{ F(P,Q) ; Q \in S_2 \} (*)$$

l'extension de ce procédé à $c \geq 3$ ne vérifie la relation (*) que si $f(x) = -\log x$ contrairement à ce que sous-entend cet auteur.

Proposition 1

Soit S_{i_0} la distribution de Dirac de S_c au point i_0 et la distribution uniforme de S_c . Alors tout élément p de S_c vérifie :

$$d(P_0, \pi) \geq d(P, \pi)$$

pour toute mesure de divergence vérifiant la définition 2.6

Démonstration :

$$d(P, \pi) = \frac{1}{c} \sum_{i=1}^c C(p_i, c) \leq \frac{1}{c} \sum_i p_i C(c) \leq \frac{C(c)}{c}$$

$$d(P_0, \pi) = \frac{1}{c} \sum_{i=1}^i C(S_0(i), c) = \frac{C(c)}{c} + \frac{(c-1)}{c} C(0)$$

d'où le résultat.

La loi observée se situe donc toujours entre deux situations extrêmes correspondant respectivement à une variable constante, et donc parfaitement prévisible, et à une variable équadistribuée, et donc parfaitement imprévisible.

Proposition 2

Soit X une variable de la loi P_X appartenant à S, alors la quantité

$$D(P_X) = d(P_0, \pi) - d(P_X, \pi)$$

est une incertitude.

Cette proposition nous permet de donner une interprétation simple des indices présentés ci-dessus.

- 3.1. $(C(t) = t \log t) \implies (D(P_X) = I(P_X, 1, 1))$
 3.2. $(C(t) = (t-1)^2) \implies (D(P_X) = \frac{c}{2} I(P_X, 2, 2))$
 3.3. $(C(t) = (\sqrt{t} - 1)^2) \implies (D(P_X)) = \sqrt{q}(1-\sqrt{2}) I(P_X, \frac{1}{2}, \frac{1}{2})$
 3.4. $C(t) = \begin{cases} 1 & \text{si } t = 0 \\ 0 & \text{si } t \neq 0 \end{cases} \implies (D(P_X) = q I(P_X, 0, 0))$

Remarque :

A notre connaissance, il n'existe pas de divergence permettant de retrouver l'exemple 3.5. Elle reçoit cependant une interprétation en termes de codages au paragraphe suivant.

L'incertitude de variables d'échelles nominales a été définie dans ce paragraphe en termes d'entropie. Cette approche peut être étendue à des variables appartenant à d'autres échelles. Ainsi la généralisation de la notion d'entropie, proposée par RENYI [63] permet de mesurer l'incertitude dans le cas de variables d'échelle d'intervalle.

Exemple 4.

Si X est une variable gaussienne de loi $\mathcal{N}(m, \sigma)$ alors

$$I(P_X, 1, 1) = 1 + \log(\sqrt{2\pi} \sigma)$$

$$I(P_X, 2, 2) = 1 - \frac{1}{2\sqrt{\pi} \sigma}$$

$$I(P_X, \frac{1}{2}, \frac{1}{2}) = K\sigma - 1 \quad \text{avec } K > 0$$

On remarque que dans ces trois cas, l'entropie est une fonction croissante de la variance, et traduit donc bien une notion d'incertitude.

Cependant l'hypothèse gaussienne, suppose que les variables sont d'échelle d'intervalle et pas plus que la variance, ces trois incertitudes ne sont des statistiques pertinentes.

3.2.2. Dispersion et incertitude

L'interprétation de la variance en termes de prédiction met en relief le rôle particulier joué par la notion de moyenne pour mesurer la dispersion d'un codage. Nous montrons que la notion de moyenne généralisée introduite par BRONS et alii [2] permet de construire une large classe d'incertitude s'interprétant en termes de dispersion d'un codage unidimensionnel.

Soit $\phi(\cdot, \cdot)$ une fonction définie sur $\mathbb{R} \times \mathbb{R}$ à valeurs dans $\overline{\mathbb{R}}$ telle que

- $\phi(x, \theta)$ est croissante en θ pour tout x
- $\phi(x, \theta) \geq 0$ pour $\theta > x$ et $\phi(x, \theta) \leq 0$ pour $\theta < x$
- $\phi(x, \theta)$ est une fonction borélienne en x pour tout θ .

On note qu'alors ϕ est une fonction borélienne du couple (x, θ) et que la fonction de θ

$$\gamma(\theta) = \int_X \phi(x, \theta) P(dx)$$

vérifie (lorsqu'elle est définie)

$$(\theta_1 < \theta_2) \implies (\gamma(\theta_1) \leq \gamma(\theta_2))$$

Définition 5

Le nombre réel μ est une ϕ moyenne de P_X si :

$$\gamma(\theta) \leq 0 \quad \text{pour } \theta < \mu \quad \text{et } \gamma(\theta) \geq 0 \quad \text{pour } \theta > \mu$$

on note $M(P_X, \phi)$ l'ensemble des ϕ moyennes de P_X .

Exemples 5

5.1. si $\phi(x, \theta) = \theta - x$ alors $\mu = E X$

$$5.2. \text{ si } \phi(x, \theta) = \begin{cases} \frac{1}{p} & \text{pour } \theta > x \\ 0 & \text{pour } \theta = x \\ \frac{-1}{1-p} & \text{pour } \theta < x \end{cases} \text{ pour } 0 < p < 1 \text{ alors}$$

μ est un p - fractile de P_X .

5.3. si $\phi(x, \theta) = \text{syn}(\theta - x)$ alors, μ est une médiane de P_X .

Proposition 3 BRONset alli [14])

- si $\alpha > 0$ alors $M(P_X, \phi) = M(\alpha P_X, \phi)$

- $M(P_X, \phi)$ est un intervalle fermé

- si $\gamma(\mu) = 0$ alors $\mu \in M(P_X, \phi)$ et réciproquement, si θ est un point intérieur de $M(P_X, \phi)$ alors $\gamma(\theta) = 0$

- $M(P_X, \phi)$ est vide ssi ou bien $\gamma(\theta)$ est définie, non nulle et de signe constant pour tout θ , ou bien $\gamma(\theta)$ n'est pas définie pour deux valeurs au moins de θ .

- si pour tout θ et θ' vérifiant $\theta < \theta'$, on a :

$$P_X[\{\phi(X, \theta) < \phi(X, \theta')\}] > 0$$

alors $M(P_X, \phi)$ contient au plus un point.

- $\phi(X, \theta)$ est constante P_X pp en θ dans l'intérieur de $M(P_X, \phi)$.

Corollaire

Si P_X est finie sur un intervalle de support I et si ϕ est bornée sur $I \times I$ alors $M(P_X, \phi) \neq \emptyset$.

Définition 6 :

Soit X une variable à valeurs dans l'ensemble $\{x_1, \dots, x_n\}$ inclus dans \mathbb{R} de loi P_X vérifiant

$$P_X(B) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_B(x_k)$$

alors μ est une ϕ -moyenne des valeurs x_1, x_2, \dots, x_n ssi

μ est une ϕ -moyenne de P_X .

Il résulte de cette définition que si ϕ est bornée, alors tout ensemble de valeurs x_1, x_2, \dots, x_n inclus dans \mathbb{R} admet au moins une ϕ -moyenne et d'après la proposition elle est unique si $\phi(x, \theta)$ est une fonction strictement croissante de θ pour tout x . Lorsque (x_1, x_2, \dots, x_n) sont issus d'un processus d'échantillonnage, on parlera de ϕ -moyenne empirique.

Le théorème suivant généralise la propriété d'optimalité de $\mu = E[X]$

$$E[(X-\mu)^2] \leq E[(X-\theta)^2] \quad \text{pour tout } \theta$$

au cas de fonctions $\phi(x, \theta)$ plus générales.

Théorème 1. (BRONS et alii [12])

Soit λ une mesure borélienne σ finie, si la fonction ϕ vérifie de plus $\phi(x, x) = 0$ pour tout x , si μ appartient à $M(P_X, \phi)$ et si $\gamma(\mu)$ est définie alors la fonction Φ définie par

$$\Phi(x, \theta) = \begin{cases} \int_{]x, \theta[\cup]\theta, x[} |\phi(x, u)| \lambda(du) & \text{si } \theta > \mu \text{ ou si } \theta = \mu \text{ et } \gamma(\mu) \geq 0 \\ \int_{]x, \theta] \cup]\theta, x]} |\phi(x, u)| \lambda(du) & \text{si } \theta < \mu \text{ ou si } \theta = \mu \text{ et } \gamma(\mu) < 0 \end{cases}$$

alors

$$\int \Phi(x, \mu) P_X(dx) \leq \int \Phi(x, \theta) P_X(dx) \quad \text{pour tout } \theta$$

l'égalité est réalisée ssi on a

- soit $\int \Phi(x, \mu) P_X(dx) = +\infty$

- soit $\gamma(u) = 0$ λ p.p. sur $] \theta, \mu [\cup] \mu, \theta [$ (resp. $] \theta, \mu [\cup] \mu, \theta [$)

si $\gamma(\mu) \geq 0$ (resp. < 0)

Remarques :

- dans le cas où λ est non atomique, la définition de $\Phi(x, \theta)$ devient

$$\Phi(x, \theta) = \begin{cases} \int_{]x, \theta[\cup]\theta, x[} |\phi(x, u)| \lambda(du) & \text{pour tout } \theta \end{cases}$$

nous nous limiterons à ce cas dans la suite.

Définition 7.

On appelle dispersion de P_X , le nombre réel positif

$$D_\phi(X) = \int \Phi(x, \mu) P_X(dx)$$

où μ est une ϕ -moyenne.

Remarque

La définition de la notion de dispersion est bien cohérente puisque d'après le théorème 1 $D_\phi(X)$ est indépendante du représentant choisi dans $M(P_X, \phi)$.

Proposition 4

La dispersion $D_\phi(X)$ est une incertitude.

Démonstration :

Soit P_1, P_2, P trois lois de probabilités de ϕ -moyenne μ_1, μ_2, μ vérifiant

$$P = \alpha P_1 + (1-\alpha) P_2 \quad \text{avec } 0 \leq \alpha \leq 1.$$

$$\begin{aligned} D_\phi(X) &= \int \Phi(x, \mu) P(dx) = \alpha \int \Phi(x, \mu) P_1(dx) + (1-\alpha) \int \Phi(x, \mu) P_2(dx) \\ &\geq \alpha \int \Phi(x, \mu_1) P_1(dx) + (1-\alpha) \int \Phi(x, \mu_2) P_2 dx \end{aligned}$$

d'après le théorème 1, ce qui assure le résultat.

Remarque :

Cette propriété de concavité indispensable à la définition d'indice d'association, nous a conduit à restreindre la définition d'une dispersion au cas où la mesure sur les paramètres est non atomique. Dans le cas contraire on montre par exemple que la dispersion associée à la moyenne arithmétique peut ne pas être concave lorsque l'on choisit $\lambda = \delta_{\mu}$.

Proposition 5

- $D_\phi(X) \geq 0$
- $(D_\phi(X) = 0) \iff (X \in M(P_X, \phi))$

Les propriétés de forte consistence de la ϕ -moyenne empirique, étudiées par BRONS et alii[12] sont rappelées en annexe.

Exemples 6

6.1. si λ est la mesure de Lebesgue,

$$\text{si } \phi(x, \theta) = 2(\theta - x)$$

alors $\mu = E[X]$ et $D(X) = V(X)$, où $V(X)$ est la variance de X .

6.2. si λ est la mesure de Lebesgue

$$\text{si } \phi(x, \theta) = k |x - \theta|^{k-1} \text{sgn}(x - \theta)$$

alors μ est le moment typique de Fréchet d'ordre k et la dispersion associée est égale à $E[|X - \mu|^k]$ appelé écart moyen d'ordre k .

6.3. si λ est la mesure de Lebesgue

$$\text{si } \phi(x, \theta) = \text{sgn}(x - \theta) \quad (k=1)$$

alors μ est la médiane et $D(X)$ est l'écart absolu moyen.

6.4. si λ est la mesure de Lebesgue

$$\text{si } \rho(x, \theta) = \log \theta - \log x \quad \text{pour } \theta \text{ et } x > 0$$

alors μ est la moyenne géométrique, d'expression $\mu = \exp\{E[\log X]\}$

$$D(X) = E[X] - \mu = E[\exp\{\log X\}] - \exp\{E[\log X]\}$$

6.5. si λ est la mesure de Lebesgue

$$\text{si } \rho(x, \theta) = \frac{1}{2} \left(\frac{1}{x} - \frac{1}{\theta} \right) \quad \text{alors } \mu \text{ est la moyenne harmonique}$$

$$\mu = (E[X^{-1}])^{-1} \quad \text{et}$$

$$D(X) = \frac{1}{\mu^2} - \frac{1}{E(X^2)} = (E[X^{-1}])^2 - (E[X^2])^{-1}$$

Remarque :

Les moments typiques d'ordre k ont été étudiés sous le nom de "r-means" par HUBER [44] dans le cadre de la statistique robuste. De plus dans le cas discret, les cas limites $k = 1$ et $k = \infty$ correspondent respectivement au mode et au mid-range de la distribution.

Ils ne respectent cependant pas exactement notre définition d'une ϕ -moyenne puisque la fonction Φ n'est pas dérivable au sens des fonctions.

- si $k = 1$ on pose

$$\phi(x, \theta) = |\text{sgn}(x - \theta)|$$

et donc

$$E[\phi(X, \theta)] = 1 - P(X = \theta)$$

$$D(X) = \min_{\theta} [1 - P(X = \theta)] = 1 - \max_i P(X = x_i) = I(P, \infty, 2)$$

alors la " ϕ -moyenne" est égale au mode et la dispersion à la fréquence hors-mode.

- si $k = \infty$ on pose :

$$\Phi(x, \theta) = \max_i |x_i - \theta|$$

$$E[\Phi(X, \theta)] = \max_i |x_i - \theta|$$

$$\min_{\theta} E[\Phi(X, \theta)] = \min_{\theta} \max_i |x_i - \theta|$$

la solution explicite de ce problème est donnée par le mid-range

$$\mu = \frac{1}{2} (\min_i X_i + \max_i x_i)$$

$$D(X) = \frac{1}{2} (\max_i x_i - \min_i x_i)$$

$D(X)$ est donc égale à l'étendue au coefficient $\frac{1}{2}$ près.

3.2.3. Entropie et dispersion

La notion de dispersion s'interprète directement en termes d'entropie, dans le cas d'une variable dichotomique. Celle-ci définit en effet une partition de l'ensemble des objets en deux classes C et \bar{C} de probabilités respectives p et $1-p$. De façon équivalente, elle définit les variables indicatrices $\mathbb{1}_C$ et $\mathbb{1}_{\bar{C}}$ qui se correspondent par dédoublement.

Le paramètre p s'interprète donc comme une probabilité sur l'ensemble des objets mais aussi comme moyenne d'un codage.

Soit $\Phi(x, \theta)$ une fonction réelle définie sur $[0,1] \times [0,1]$ vérifiant

$$\Phi(0, \theta) = \Phi(1, 1-\theta)$$

alors

$$E[\Phi(X, \theta)] = p \Phi(1, \theta) + (1-p) \Phi(1, 1-\theta)$$

il suffit alors de poser

$$F(p_i, q_i) = \Phi(1, q_i)$$

pour obtenir

$$E[\Phi(X, \theta)] = F(P, Q) \quad \text{où}$$

$$P = \{p, 1-p\} \text{ et } Q = \{\theta, 1-\theta\}$$

ainsi par exemple l'entropie quadratique $I(P, 2, 2)$ est égale à la variance de l'indicatrice $\mathbb{1}_C$:

$$I(P,2,2) = \text{Var}(l_C) = p(1-p)$$

A l'opposé, il n'existe pas de codage numérique canonique d'une variable polychotomique permettant d'associer directement dispersion et entropie. Les q modalités de la variable fournissent trois informations :

- la partition induite sur l'ensemble des objets Λ
- la relation d'équivalence R sur $A \times A$ associée à la partition
- le vecteur aléatoire des q indicatrices des q classes de R -équivalence.

Chacune d'elles permet de définir des incertitudes qui pour certains choix de codage s'interprètent à la fois en termes de dispersion et d'entropie.

Exemple 7

- codage de la relation d'équivalence

$$Z(k,1) = \begin{cases} \frac{1}{p_i} - 1 & \text{si } X(k) = X(1) = i \\ -1 & \text{sinon} \end{cases}$$

$$\text{alors } P(Z = \frac{1}{p_i} - 1) = p_i^2.$$

$$D(X) = \text{Var}(Z) = C - 1 = I(P_X, 0, 0)$$

nous montrons de plus que cette incertitude s'exprime dans le cadre multivarié défini par le vecteur des q indicatrices comme la trace de l'opérateur de covariances.

3.3. Notion de gain d'incertitude et indice de connexion

Soit X et Y deux variables aléatoires définies sur un même ensemble d'objets à valeurs dans \mathbb{R} . On note P_Y^X une version de la loi conditionnelle de Y sachant X , et S le simplexe des lois de probabilités.

Définition 8

Soit I une incertitude définie sur S , on appelle gain d'incertitude la quantité

$$\Delta I(Y/X) = I(P_Y) - E_X [I(P_Y^X)]$$

L'étude de variables de carré sommable montre que si on n'a pas toujours la relation.

$$\text{Var}(Y/X = x) \leq \text{Var}(Y) \quad \text{pour tout } x$$

par contre

$$E_X [\text{Var}(Y)] \leq \text{Var}(Y)$$

plus généralement l'inégalité de JENSEN permet de démontrer.

Proposition 6

Soit f une fonction réelle positive concave et continue définie sur S.

Alors

$$f(P_Y) \geq E_X[f(P_Y^X)]$$

Corollaire

$$\Delta I(Y/X) \geq 0$$

Ce corollaire montre que le gain d'incertitude mesure la réduction de l'erreur de prédiction de Y apportée par la connaissance de X. Bien que cette prédiction se fonde sur un modèle d'association, on vérifie la proposition.

Proposition 7

Si X et Y sont indépendantes, alors $\Delta I(Y/X) = 0$. La réciproque est vraie si I est une fonction strictement concave.

Démonstration

X et Y sont indépendantes ssi

$$P_Y^X = P_Y P_X \text{ p.p.}$$

Par suite, l'inégalité de JENSEN étant stricte, du fait de la stricte concavité de I, le résultat est immédiat.

Définition 9

On appelle indice de connexion de Y à X la quantité

$$C(Y/X) = \frac{\Delta I(Y/X)}{I(P_Y)}$$

Il résulte des propositions précédentes que :

- $0 \leq C(Y/X) \leq 1$
- $C(Y/X) = 1$ si le modèle de prédiction est vérifié
- $C(Y/X) = 0$ traduit l'absence d'association. Celle-ci équivaut à l'indépendance lorsque l'incertitude choisie est strictement concave.

Exemples 8

$$8.1. \quad I(P_Y) = \text{Var}(Y)$$

dans ce cas $C(Y/X) = \eta^2(Y/X)$

$$(C(Y/X) = 1) \iff (\exists g : Y = g(X))$$

$$(C(Y/X) = 0) \iff E_Y^X \equiv c \quad P_X \text{ p.p.}$$

$$\eta^2(Y/X) = 1 - \frac{E_X(\text{Var}(Y))}{\text{Var}(Y)}$$

$$8.2. \quad I(P_Y) = E[|Y - \text{med}(Y)|]$$

on obtient ainsi un indice de connexion de Y à X utile en régression L^1 pour laquelle :

$$(C(Y/X) = 1) \iff (\exists g : Y = g(X))$$

où $g(x)$ est une médiane de Y conditionnelle à $X = x$, puisque le modèle de prédiction consiste à coder chaque objet par la médiane $\text{med}(Y)$.

$$(C(Y/X) = 0) \iff (\text{Med}(Y) \equiv c \quad P_X \text{ p.p.})$$

$$8.3. \quad I(P_Y) = 1 - \text{Max}_j p_{.j}$$

dans ce cas, $C(Y/X)$ est égal au coefficient $\lambda_{Y/X}$ de GUTTMAN étudié dans l'exemple 1 de ce chapitre.

$C(Y/X) = 0$ ssi la loi de Y conditionnelle à $X = x_i$ est une Dirac dépendant de x :

$C(Y/X) = 0$ ssi les distributions conditionnelles de Y ont même mode que la distribution marginale

$$8.4. \quad I(P_Y) = 1 - \sum_j p_{.j}$$

dans ce cas, $C(Y/X)$ est égal au coefficient $\tau_{Y/X}$ de GOODMAN et KRUSKAL [] pour lequel le modèle de prédiction est le suivant :

- si $X = x_i$ on prédit $Y = y_j$ avec la probabilité p_{ij}^i

- si X est inconnu on prédit $Y = y_j$ avec la probabilité $p_{.j}$

$$C(Y/X) = \tau_{Y/X} = \frac{\sum_i \sum_j \frac{1}{p_{i.}} (p_{ij} - p_{i.} p_{.j})}{1 - \sum_j p_{.j}}$$

$C(Y/X) = 1$ ssi la loi de Y conditionnellement à $X = x_i$ est une Dirac dépendant de x_i

$C(Y/X) = 0$ ssi X et Y sont indépendantes.

$$8.5. \quad I(P_Y) = - \sum_j p_{\cdot j} \log p_{\cdot j}$$

dans ce cas $C(Y/X)$ est égal au coefficient d'"incertitude"

$$U_{Y/X} = \frac{\sum_i p_{i\cdot} \log p_{i\cdot} + \sum_j p_{\cdot j} \log p_{\cdot j} - \sum_j \sum_i p_{ij} \log p_{ij}}{\sum_j p_{\cdot j} \log p_{\cdot j}}$$

$$C(Y/X) = 1 \quad \underline{\text{ssi}} \quad Y \text{ est une fonction de } X$$

$$C(Y/X) = 0 \quad \underline{\text{ssi}} \quad X \text{ et } Y \text{ sont indépendantes.}$$

Les indices de connexion présentés ci-dessus font intervenir dans leurs modèles de prédiction, les codages des objets. Une telle démarche ne permet pas de prendre en compte des règles de prédiction portant sur des variables d'échelles ordinales. Dans ce cas, le formalisme mis en place s'avère fécond sur les codages dérivés définis sur les paires d'objets. On définit ainsi le coefficient proposé par KIM [47] obtenu en construisant les codages dérivés Z (resp. T) défini par la signature du préordre induit sur l'ensemble des objets par la variable X (resp Y)

$$Z(k,l) = \text{sgn} (X_1 - X_k)$$

$$T(k,l) = \text{sgn} (Y_1 - Y_k)$$

On montre dans la suite que :

$$\text{Var}(T) = 1 - \sum_j p_{\cdot j}^2$$

$$\text{Var} (T/Z = 1) = \text{Var} (T/Z = -1) = (P(C) + P(D)) (1 - \sum_i p_{i\cdot}^2)^{-1}$$

$$P(C) = \sum_i \sum_j \sum_{i' > i} \sum_{j' > j} p_{ij} p_{i'j'} \quad P(D) = \sum_i \sum_j \sum_{i' < i} \sum_{j' > j} p_{ij} p_{i'j'}$$

$$\text{Var} (T/Z = 0) = \sum_i p_{i\cdot}^2 - \sum_i \sum_j p_{ij}^2 (\sum_i p_{i\cdot}^2)^{-1}$$

d'où

$$E_Z(\text{Var}(T)) = P(C) + P(D) + \sum_i p_{i\cdot} - \sum_i \sum_j p_{ij}$$

le coefficient de connexion est alors :

$$C(Y/X) = \eta_{T/Z}^2$$

L'utilisation des codages dérivés Z et T définis sur les couples d'objets, prend en compte l'information ordinale portée par la variable X. Il n'en est pas de même de la statistique de WALLIS [77] qui utilise pour codage dérivé de Y la statistique de rang T et s'exprime comme

$$\eta^2 (Y/X) = \eta^2_{T/X} = \frac{\chi_r^2}{n(r-1)}$$

où χ_r^2 est la statistique de Friedman utilisée en analyse de variance non paramétrique qui s'écrit :

$$\chi_r^2 = \frac{12}{nr(r+1)} \sum_{i=1}^r r_i^2 - 3n(p+1)$$

$$\text{où } r_i = \sum_{j=1}^c t_j^i n_{ij}$$

Ces deux indices permettent de montrer l'existence d'une connexion monotone de Y à X mais pas de la décrire. Ce problème de régression est pris en compte par la définition de l'indice de corrélation monotone proposé par d'AUBIGNY [8].

$$v^2(Y/X) = 1 - \frac{E_X^{\sim X} [\text{Var}(Y)]}{\text{Var}(Y)}$$

qui vaut 1 ssi Y est une fonction monotone de X qui réalise la borne de FRECHET et 0 si les variables sont indépendantes. Cependant cet indice nécessite une extension de la définition de l'incertitude au cas du conditionnement par rapport à un σ -treillis. Cette démarche équivaut à définir l'incertitude comme un minimum de la fonction $E[\Phi(X, \theta)]$ sous la contrainte de monotonie du recodage m de la variable X, on impose donc à la ϕ -moyenne de Y d'appartenir au cône des fonctions monotones de X. L'utilisation des indices d'incertitude sans contrainte fonde le recours au modèle linéaire et au coefficient de détermination (ρ^2) comme indice de connexion de Y à X associé à la variance lorsqu'on contraint l'espérance conditionnelle de Y à être une fonction affine de X (il est connu que les deux notions coïncident dans le cas gaussien). La logique P.R.E. permet ainsi d'étendre la logique du modèle linéaire au cadre d'hypothèses emboîtées quelconques ; cette possibilité est exploitée par exemple dans l'étude des tables de contingence suivant le modèle log-linéaire en terme d'entropie de SHANNON (BISHOP et alii [12]), suivant le modèle linéaire sur les indicatrices (MARGOLIN et LIGHT [60]).

4 - MESURE DE CONCORDANCE

Les indices de connexion étudiés dans le chapitre précédent mesurent l'écart à la situation de complète dépendance, traduite par l'existence d'une fonction liant Y à X. Cette transformation des données permet alors d'associer le système relationnel numérique Δ_{B_Y} image du système relationnel empirique Δ par Y en système relationnel numérique Δ_{B_X} . Les différents indices proposés se différencient suivant la structure d'homomorphisme imposée à cette transformation par une hypothèse d'échelle de mesure de Y. Lorsqu'on étudie la situation de mutuelle complète dépendance, traduite par l'existence d'un isomorphisme liant Δ_{B_X} et Δ_{B_Y} , on utilise alors des indices de concordance, selon la terminologie de GINI [34].

Définition 1.

X et Y, deux variables aléatoires définies sur une même population sont dites concordantes ssi il existe deux fonctions f et g telles que

$$Y = f(X) \text{ et } X = g(Y)$$

Remarque

Le fondement intuitif de la notion de concordance de deux variables est que les deux échelles définissent une même relation sur les objets, et mesurent donc un même facteur sous-jacent, ou les mêmes propriétés. L'isomorphisme liant Δ_{B_X} et Δ_{B_Y} nous autorise à parler abusivement de recodages pour désigner les transformations f et g.

L'étude de l'existence des fonctions f et g trouve un cadre naturel dans l'espace des variables de carrés sommables, puisque dans cet espace la complète dépendance mutuelle impose au coefficient de sup. corrélation

$$\rho^*(X, Y) = \text{Sup} \{ \rho(\phi(X), \psi(Y)) ; \phi \in L_X^2, \psi \in L_Y^2 \}$$

où L_X^2 (resp L_Y^2) est l'ensemble des fonctions de X (resp. Y) et ρ le coefficient de corrélation de Bravais et Pearson, d'être égal à 1.

Cette étude peut alors être menée en termes géométriques puisque le produit scalaire défini sur $L^2(\Omega, a, p)$ coïncide dans l'orthogonale de la droite des constantes avec la métrique de la covariance. On notera ce produit scalaire

$$\langle X^\circ, Y^\circ \rangle = \text{COV} (X, Y)$$

Lorsque l'on résonne à marges fixées, le terme de Hoeffding présenté au chapitre 2 nous permet de construire une relation d'ordre partiel sur les couples d'éléments de $L^2(\Omega, A, P)$ puisque

$$\left(\underset{\text{P.Q.D.}}{P_{X_1 Y_1}} \geq P_{X_2 Y_2} \right) \Rightarrow (\langle X_1, Y_1 \rangle \geq \langle X_2, Y_2 \rangle)$$

De plus la proposition 1 chapitre 2 nous assure que cette relation d'ordre est conservée par tout couple de transformations concordantes des variables. Par suite la construction d'indices de concordance repose sur l'utilisation de telles transformations qui devront de plus être pertinentes, ainsi que des outils permettant une interprétation des codages dérivés dans la métrique de la covariance.

Cette étude doit cependant être complétée par une méthode de normalisation des indices ainsi construits afin de faciliter l'interprétation de leurs valeurs dans une application donnée.

4.1. Dualité et métrique de la covariance.

L'expression de la covariance des variables Z et T, définies sur une population finie Ω , dépend de la dualité utilisée.

Lorsqu'on étudie le croisement des variables Z et T, indépendamment des valeurs affectées aux catégories définies par celles-ci, la dualité "Individus-catégories" permet uniquement de reconstituer le tableau de contingence.

$$V = (V_{ij})_{i=1 \dots r} \quad r_{ij} = 1 \dots i$$

$$V_{ij} = \frac{1}{n} \sum_{\omega} \mathbb{1}_{i(\omega)} \mathbb{1}_{j(\omega)}$$

et ne permet donc pas de mesurer la covariance de Z et T. La nécessité de recourir aux codages peut être traitée de deux façons. Soit on utilise ceux fournis par les coordonnés des catégories sur le premier axe d'une analyse des correspondances, qui réalisent la sup-corrélation de Z et T (LANCASTER [52]), soit on utilise la dualité "Individus-variables" afin de mesurer une covariance à l'aide des codages observés.

Notons \bar{z} et \bar{t} les moyennes observées respectives de Z et T. On calcule :

$$\begin{aligned} \text{COV}(Z, T) &= \frac{1}{n} \sum_{\omega} (z_{\omega} - \bar{z})(t_{\omega} - \bar{t}) \\ &= \langle \underline{z}^{\circ}, \underline{t}^{\circ} \rangle_{D_1/n} \end{aligned} \quad (1)$$

ou plus généralement $\langle \underline{z}^{\circ}, \underline{t}^{\circ} \rangle_{D_P}$ lorsque la loi P définie sur Ω est une pondération pas seulement uniforme. Cette expression montre que l'on s'intéresse aux relations entre codages Z et T , observées sur les individus. Elle doit être comparée à l'expression suivante, dont l'équivalence à (1) est démontrée au chapitre I

$$\text{COV}(Z, T) = \sum_{i=1}^r \sum_{j=1}^c h_i^1 h_j^2 v_{ij} \quad (2)$$

qui utilise la dualité "catégories-variables" associée aux codages \underline{h}^1 et \underline{h}^2 des valeurs prises par Z° et T° , de façon classique V s'interprète comme matrice, dans les bases canoniques, de E_2^* et E_1^* , d'une explication linéaire. De même les codages \underline{h}^1 et \underline{h}^2 définissent le tenseur

$$O = (O_{ij})_{i=1, \dots, r} \quad F_{ij} = 1, \dots, c$$

de terme général

$$O_{ij} = h_i^1 h_j^2$$

élément du dual de $\mathcal{L}(E_1, E_2^*)$. Cette dualité permet d'interpréter la covariance :

$$\langle O, V \rangle = \text{Trace}({}^t O V) = \text{Tr}({}^t O V) \quad (3)$$

Dans cette optique, on croise sur l'ensemble des cases du tri croisé, la variable "loi conjointe de (Z, T) " avec la variable "codage conjoint de (Z, T) ". On donne ainsi une interprétation géométrique aux indices proposés par HILDEBRAND et al. [42].

La correspondance entre les formules (1), (2) et (3) permet d'étudier la concordance, en termes de covariance dans l'un de ces trois espaces. La structure vectorielle de ces espaces, ainsi que la structure euclidienne induite par la métrique de la covariance, conduisent à identifier deux variables définissant un même sous-espace vectoriel comme deux codages d'une même échelle d'intervalle. Cette contrainte

de représentation nécessite de recourir à un recodage des données à l'aide de fonctions concordantes pour étudier la mutuelle dépendance de niveau d'échelle quelconque.

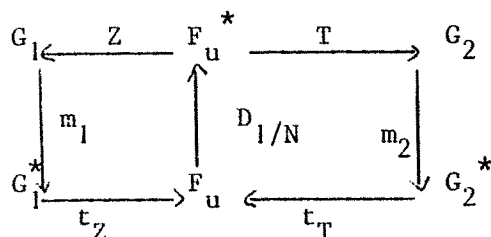
Définition 2.

On appelle indice de concordance de X et Y, d'ordre k toute fonction monotone de COV (Z,T), où

$$Z = f(X_1, X_2, \dots, X_k), \quad T = g(Y_1, Y_2, \dots, Y_k)$$

et f et g sont des fonctions concordantes définies sur le k-échantillon de (X,Y).

On mène alors une analyse linéaire de la concordance entre les tableaux Z et T d'une part en étudiant la dualité "unités statistiques-variables" associée aux tableaux Z et T des données recodées, à laquelle correspond le schéma



$N = \text{cardinal}(U)$

$U = A^k$, ensemble des unités statistiques

d'autre part en étudiant la dualité, "unités statistiques-catégories" associée aux tableaux H_1 et H_2 des indications des catégories de Z et T.

Les hypothèses sur le système relationnel empirique s'expriment en termes de relations k-aires sur les objets qui se traduisent par le choix de fonctions concordantes d'ordre k. Ainsi l'étude d'une relation binaire nécessite d'utiliser des fonctions concordantes définies sur les couples.

4.2. Normalisation

Les résultats du chapitre 2 font de la covariance un indicateur monotone décroissant de l'écart à la situation de dépendance complète. Cependant, si elle s'annule dans le cas extrême de l'indépendance,

l'interprétation de la valeur prise dans une étude concrète reste difficile, puisqu'il n'existe pas de valeur de référence caractérisant la situation de dépendance complète.

Deux types de normalisation sont utilisés, fondés sur deux problématiques distinctes.

Le premier type, proposé par KENDALL [46] pour les coefficients de corrélation ordinaux, généralisé par DANIELS [23] est de nature probabiliste, fondé sur les propriétés de permutations.

On note \tilde{H}_1 (resp \tilde{H}_2) le tableau disjonctif complet bloc diagonal associé à la variable Z (resp. T). Le tableau H_1 (resp. H_2) se déduit de \tilde{H}_1 (resp. \tilde{H}_2) par une permutation σ_1 (resp. σ_2) des colonnes, de matrice de permutation σ_1 (resp. σ_2)

$$H_i = \tilde{H}_i \sigma_i, \quad i = 1, 2$$

le tableau de contingence V s'écrit alors

$$V = \frac{1}{N} H_1 {}^t H_2 = \frac{1}{N} \tilde{H}_1 \sigma_1 {}^t \sigma_2 {}^t \tilde{H}_2$$

et la covariance entre Z et T s'écrit

$$\begin{aligned} \text{COV}(Z, T) &= \text{COV}_{\sigma}(Z, T) = \text{Tr}({}^t \text{OV}) = \frac{1}{N} \text{Tr}({}^t \text{O} \tilde{H}_1 \sigma {}^t H_2) \\ &= \frac{1}{N} \text{Tr}({}^t \tilde{H}_2 {}^t \text{O} \tilde{H}_1 \sigma) \\ &= \frac{1}{N} \sum_u z_u {}^t \sigma(u) \end{aligned}$$

Les distributions marginales de Z et T étant données, le tableau de contingence est entièrement caractérisé par la permutation $\sigma = \sigma_1 {}^t \sigma_2$. KENDALL [46] propose de traduire l'hypothèse de non-relation entre Z et T par l'hypothèse d'équiprobabilité sur l'ensemble des permutations définies sur les N unités statistiques. La normalisation consiste alors à calculer l'indice de concordance en terme de vraisemblance de la covariance sous l'hypothèse de non-relation.

L'indice $\text{COV}_{\sigma}(Z, T)$ définit une U-statistique, dont on peut calculer la moyenne et la variance permettant d'obtenir la loi asymptotique sous l'hypothèse de non-relation en utilisant le théorème de WALD-WOLFOWITZ [76]. HUBERT [45], LERMAN [57], LECALVE [54] donnent le calcul des deux premiers moments lorsque Z et T sont des fonctions d'ordre 1 ou 2.

L'indice de concordance s'écrit :

$$C(X,Y) = F \left[\frac{\text{COV}(Z,T) - E [\text{COV}_\sigma(Z,T)]}{\sqrt{\text{Var} [\text{COV}_\sigma(Z,T)]}} \right]$$

où F est la fonction de répartition de $\text{COV}_\sigma(Z,T)$ sous l'hypothèse de non-relation.

Remarques :

- . Un tel indice calculé sous l'hypothèse de non-relation est donc construit pour mesurer l'écart à la situation d'indépendance. Son interprétation comme indice de concordance est donc ambiguë.
- . Un tel indice vérifie bien notre définition, du fait de la monotonie de F. Le second type de normalisation, de nature géométrique, utilise la structure de produit scalaire de $\text{COV}(Z,T)$. On définit alors pour mesurer la concordance entre X et Y, une classe de cosinus dont l'interprétation varie avec les différentes transformations Z et T

$$C(X,Y) = \frac{\text{COV}(Z,T)}{\sqrt{\text{COV}(Z,Z) \text{COV}(T,T)}}$$

Cet indice prend la valeur 1 (resp. -1) lorsque Z et T décrivent une même droite dans F, c'est-à-dire si X et Y sont les codages d'une même échelle.

Proposition (LE CALVE [54])

L'indice centré réduit coïncide avec le cosinus lorsque les fonctions Z et T sont de degré 1.

4.3. Applications.

Le choix des transformations F et G dépend naturellement de l'hypothèse d'échelle à laquelle appartient chaque variable et donc du système relationnel correspondant. La covariance mesurant une proximité entre les codages dérivés, cette proximité doit pouvoir s'interpréter comme proximité entre les codages ou les relations empiriques. Les indices

présentés dans la suite sont ordonnés suivant le niveau d'échelle des variables.

4.3.1. Echelle nominale

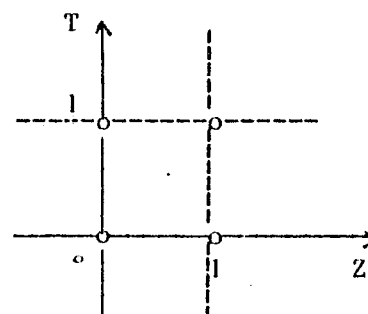
4.3.1.1. Variables dichotomiques

Nous avons montré le statut particulier des variables dichotomiques. En effet la relation d'équivalence qu'elles définissent, est totalement décrite par la donnée de l'indicatrice de l'une de ses modalités. Les deux dualités "individus-catégories", "individus-variables" coïncident alors et le choix des fonctions concordantes devient immédiat :

- l'ensemble des unités statistiques est l'ensemble des objets ;
- $Z(u)$ = (resp. $T(u)$) est l'indicatrice d'une des deux modalités.

Le tableau de contingence s'écrit

Z \ T	0	1	:	:
0	p ₀₀	p ₀₁	:	p _{0.}
1	p ₁₀	p ₁₁	:	p _{1.}
	p _{.0}	p _{.1}	:	1



nuage des individus

On lui associe la représentation des individus (telle que l'épaisseur des points est proportionnelle à la fréquence des cases).

On obtient l'indice de degré 1 (différence des lois conditionnelles) :

$$\phi = \frac{p_{11} - p_{1.}p_{.1}}{\sqrt{p_{1.}p_{.1}p_{.0}p_{0.}}} = \frac{p_{11}p_{00} - p_{10}p_{01}}{\sqrt{p_{.1}p_{1.}p_{.0}p_{0.}}} = \frac{p_{11}}{p_{.1}} - \frac{p_{10}}{p_{.0}}$$

la mutuelle complète dépendance de X ou Y se traduit par la construction des deux fonctions de recodages, linéaires :

$$Z = \beta_{T/Z} T \text{ où } \beta_{T/Z} = \frac{\text{COV}(Z, T)}{\text{COV}(T, T)} = \frac{p_{11} - p_{1.}p_{.1}}{p_{.0}p_{.1}}$$

on retrouve alors

$$\phi^2 = \beta_{T/Z} \beta_{Z/T}$$

cet indice vaut + 1 ou -1 ssi Z est une fonction linéaire de T, et 0 ssi Z et T sont indépendantes.

4.3.1.2. Cas des variables nominales à plus de deux modalités

Les variables X et Y définissent les relations d'équivalence R et sur l'ensemble A dont les classes d'équivalence sont les catégories. On décrit alors cette relation binaire à l'aide des codages définis sur les paires.

Exemple 1

LERMAN propose un codage des paires à l'aide de l'indicatrice de la relation \mathcal{R} :

$$\mathcal{R} \rightarrow Z(a_1, a_2) = \begin{cases} 1 & \text{si } X(a_1) = X(a_2) \\ 0 & \text{sinon} \end{cases} \quad \mathcal{R}$$

alors l'indice proposé est obtenu en contrant et en réduisant le produit scalaire

$$\langle Z, T \rangle_{D_{1/N}} = \sum_i \sum_j p_{ij}^2$$

qui mesure la proximité entre les deux relations, en mesurant la probabilité de leurs intersections.

Exemple 2

L'indice précédent est un indice différent de la statistique de TCHUPROW

$$\phi^2 = \frac{\sum_i \sum_j \frac{p_{ij}^2}{p_i \cdot p_j} - 1}{\sqrt{r-1} \sqrt{c-1}} = \frac{\phi^2}{\sqrt{r-1} \sqrt{c-1}}$$

associée au codage des paires :

$$Z(a_1, a_1) = \begin{cases} \frac{1}{n_i} & \text{si } X(a_1) = X(a_2) \text{ et appartiennent à la classe } C_i \\ 0 & \text{sinon} \end{cases} \quad \mathcal{R}$$

alors

$$t^2 = \frac{\text{COV}(Z^\circ, T^\circ)}{\sqrt{\text{COV}(Z^\circ, Z^\circ) \cdot \text{COV}(T^\circ, T^\circ)}}$$

on vérifie aisément que la distance entre les deux variables Z et T construite à l'aide de ce produit scalaire s'écrit (CAILLEZ et PAGES [13]):

$$d^2(\delta_i, \delta_j) = \sum_{i=1}^r \sum_{j=1}^c n_{ij} \frac{p(C_i \Delta C_j)}{p(C_i)p(C_j)}$$

où $C_i \Delta C_j$ est la différence symétrique entre les classes C_i et C_j et mesure donc une proximité entre les deux relations.

4.3.2. Echelles ordinales

4.3.2.1. Indice de degré 1

Lorsqu'on étudie la concordance entre deux variables d'échelle ordinale on définit un indice d'ordre 1, qui calcule le degré de quadrant dépendance de (X,Y) en comparant les codages des objets à une valeur de position centrale. On choisit généralement la médiane qui fut proposé par BLUMQUIST [13], dans la construction de l'indice

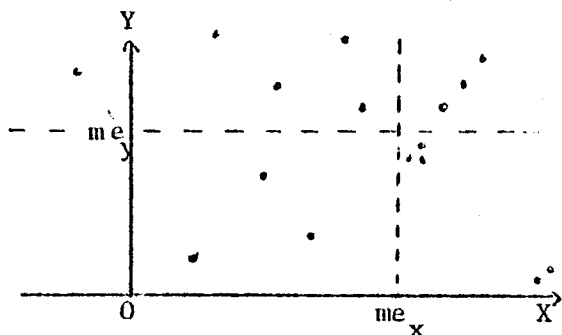
$$q = P [\text{sg}(X - \text{me}_x) (Y - \text{me}_y) > 0] - P [\text{sg}(X - \text{me}_x) (Y - \text{me}_y) < 0]$$

Soit Z et T les codages :

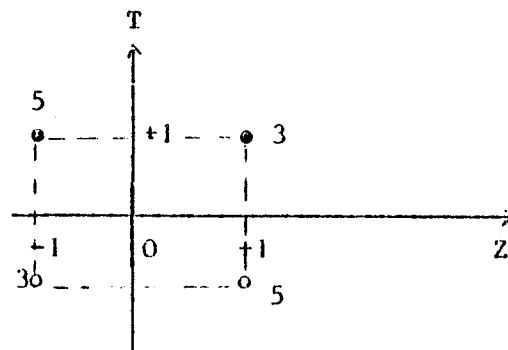
$$Z = \text{sg}(X - \text{me}_x) \quad , \quad T = \text{sg}(Y - \text{me}_y)$$

alors

$$q = \frac{\text{COV}(Z, T)}{\sqrt{\text{COV}(Z, Z) \cdot \text{COV}(T, T)}}$$



Dualité "individu-variables"



Dualité "unités statistiques-variables"

4.3.2.2. Indice de degré 2

Dire que les variables X et Y sont concordantes, c'est dire qu'elles appartiennent à la même échelle et donc qu'elles se déduisent l'une de l'autre par une transformation monotone. Cela se traduit sur les relations de préordre induites, \mathcal{R} et \mathcal{Y} induites sur la relation

$$((a_1, a_2) \in \mathcal{R}) \Leftrightarrow ((a_1, a_2) \in \mathcal{Y})$$

On utilise alors le codage dérivé des paires à l'aide des fonctions concordantes

$$f(x_1, x_2) = \text{sg}(x_1 - x_2), g(y_1, y_2) = \text{sg}(y_1 - y_2)$$

les codages dérivés $Z = f(X(a_1), X(a_2)), T = g(Y(a_1), Y(a_2))$ permettent d'associer deux tableaux disjonctifs complets de la forme :

		-----	(a ₁ , a ₂)	-----	← paires
	+ 1	-----	1		
	0		0		↔ (Z(a ₁ , a ₂) = + 1)
codage	-1		0		

dont le croisement permet de comptabiliser dans le tableau de contingence sous les symboles respectifs C, D, $\frac{E}{Z\bar{T}}$, $\frac{E}{ZT}$, les nombres de

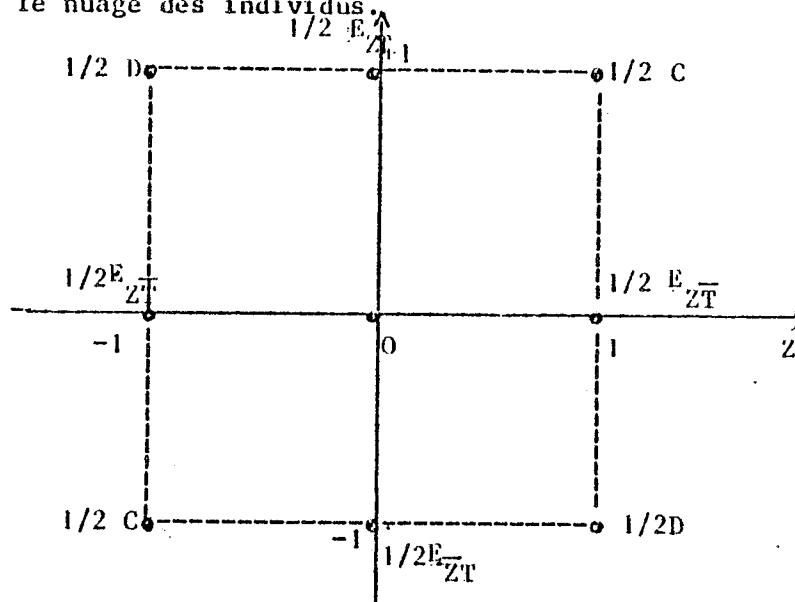
T \ Z	-1	:	0	:	1	:	marge	
-1	$\frac{1}{2} C$:	$\frac{1}{2} \frac{E}{Z\bar{T}}$:	$\frac{1}{2} D$:		$\frac{1}{2} C = \sum_i \sum_j \sum_{i' > i} \sum_{j' > j} p_{ij} p_{i'j'}$
0	$\frac{1}{2} \frac{E}{Z\bar{T}}$:	E_{ZT}	:	$\frac{1}{2} \frac{E}{Z\bar{T}}$:	E_2	$\frac{1}{2} D = \sum_i \sum_j \sum_{i' > i} \sum_{j' < j} p_{ij} p_{i'j'}$
1	$\frac{1}{2} D$:	$\frac{1}{2} \frac{E}{Z\bar{T}}$:	$\frac{1}{2} C$:		$\frac{1}{2} \frac{E}{Z\bar{T}} = \sum_i \sum_j \sum_{j' > j} p_{ij} p_{ij'}$
marge		:	E_T	:		:		$E_{ZT} = \sum_i \sum_j p_{ij}^2$

paires concordantes, discordantes et exaequo, selon la terminologie de KENDALL [46].

Remarque

Le calcul des formules suppose que les incidences i et j sont dans le même ordre que les catégories.

Dans la dualité "individus-variables" associée à ces codages, correspond le nuage des individus.



Lorsque les relations \mathcal{R} et \mathcal{S} sont des relations d'ordre, les lignes $Z = 0$ et $T = 0$ sont vides, on construit alors à partir de la covariance un indice qui pour les deux normalisations coïncident, c'est l'indice de KENDALL [46]

$$\tau_a = \frac{C - D}{n^2}$$

τ_a est égal à 1 (resp. -1) ssi les deux ordres sont identiques (resp. inverses).

Lorsqu'il y a des exaequos, les indices se différencient suivant le traitement qu'ils leur accordent. On définit alors une série d'indices de concordance de préordres :

L'indice γ de GOODMAN et KRUSKAL [37] est construit en ignorant les effectifs correspondants aux lignes et aux colonnes. On travaille donc conditionnellement à l'ensemble des non-exaequos

$$\gamma = \frac{C - D}{C + D}$$

Ceci correspond à ne retenir que les quatre sommets du carré dans l'analyse du nuage des individus. On obtient alors les coefficients de régression

$$\beta_{Z/T} = \beta_{T/Z} = \gamma$$

Dans le cas particulier de variables dichotomiques on retrouve le coefficient Q de YLE (cf. BISHOP et al. [12]) qui traite les variables dichotomiques comme étant ordonnées. L'étude des valeurs de γ est poursuivie dans l'annexe 1.

Pour tenir compte des exaequos, KENDALL [46] propose une modification de son indice τ_a en construisant

$$\tau_b = \frac{C - D}{\sqrt{(1 - E_Z)(1 - E_T)}} = \frac{\text{COV}(Z, T)}{\sqrt{\text{COV}(Z, Z) \cdot \text{COV}(T, T)}}$$

τ_B ne peut atteindre les bornes 1 ou -1 sur des tableaux rectangulaires puisque le modèle même de concordance est déjà violé, une telle situation traduisant l'existence de deux systèmes relationnels, définis sur A.

KENDALL propose une normalisation à 1

$$\tau_c = \frac{(C - D)m}{n^2(m - 1)}, \quad m = \min(r, c), \quad \frac{r - 1}{r} = \text{Max Var}(z)$$

Les indices de régression

$$d_{Z/T} = \frac{C - D}{1 - E_T}, \quad d_{T/Z} = \frac{C - D}{1 - E_Z}$$

sont appelés les coefficients d de SOMERS [62].

On remarque encore que $\tau_B^2 = d_{Z/T} \cdot d_{T/Z}$

Remarques :

- . L'indice de LERMAN [57], calculé dans le cas de préordres ne coïncide pas avec l'indice τ_b .
- . Tous ces indices s'expriment en fonction de la différence C - D valeur du produit scalaire associée à la distance de la différence symétrique, qui définit une géométrie euclidienne dans l'espace de relations.

4.3.3. Echelle hyperordinaire

L'un des indices les plus utilisés en pratique, du fait de la simplicité des calculs est le coefficient de Spearman

$$r_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i - S_i)^2$$

R_i (resp. S_i) est le rang de a_i parmi les $X(a)$ (resp. $Y(a)$). Cette expression montre que le coefficient est l'indice d'association calculé à l'aide des fonctions concordantes :

$$Z(a_i) = R_i(X_1, X_2, \dots, X_n) - \frac{n+1}{2}$$

$$T(a_i) = S_i(X_1, X_2, \dots, X_n) - \frac{n+1}{2}$$

pour la normalisation du cosinus.

On veillera cependant à ce que l'affectation de rangs à des objets définit une relation ternaire sur A . Le coefficient de Spearman suppose une information ordinaire plus riche que celle traitée par les indices ordinaux précédents. Ceci justifie le rappel de la proposition 3, chapitre I. L'expression correcte de cet indice est donc donnée en termes de fonctions concordantes d'ordre 3 :

$$f(x_1, x_2, x_3) = \text{sg}(x_1 - x_2)$$

$$f(y_1, y_2, y_3) = \text{sg}(y_1 - y_3)$$

La classe des indices de concordance hyperordinaire est très vaste puisque les fonctions scores utilisées en statistiques non paramétriques pour définir des statistiques linéaires de rang (HAJEK et SIDAK [41]) sont des fonctions concordantes ; ainsi toute fonction score ρ définit un indice de concordance hyperordinaire

$$C_\rho(X, Y) = E_\Phi \left[\frac{\sum_{i=1}^h \rho(R_i) \rho(S_i)}{\sum_{i=1}^h \rho^2 \left(i - \frac{h+1}{2} \right)} \right]$$

où h définit le degré de l'indice et Φ est la distribution de probabilité définissant la fonction score ρ .

Remarques (POON [57])

- si $h = 2$ et $\rho = \text{sg}(x)$, on retrouve le τ de KENDALL ;
- si Φ est la loi $(0,1)$ et $h = 3$, on retrouve le coefficient de corrélation associé au test de Fisher-Yates ;
- si $h = 1$ et $\rho = \text{sg} [x - \frac{1}{2}(n + 1)]$, on retrouve le q de BLUMQUIST.

3.3.4. Les échelles d'intervalle

Lorsque les variables X et Y sont d'échelles d'intervalle on retrouve les concepts usuels de L^2 appliqués à une population finie. On définit alors le coefficient de corrélation linéaire comme coefficient de concordance d'ordre 1 associé aux fonctions concordantes

$$f(x) = x - \bar{x}$$

$$g(y) = y - \bar{y}$$

La grande sensibilité de cet indice aux valeurs extrêmes, conduit à retenir d'autres fonctions concordantes.

Exemple 1

$$f(x) = g(x) = |x|^\alpha \text{sg } x ; \quad \alpha \geq 0$$

$$C_\alpha(X,Y) = \text{COV} \left[f\left(\frac{X - a}{D(X)^{1/2}}\right), g\left(\frac{Y - b}{D(Y)^{1/2}}\right) \right]$$

définit un indice de concordance associé aux valeurs des caractéristiques de position (resp. de dispersion) a et b (resp. D_x et D_y). Plus α tend vers 0, et moins C_α dépend des valeurs extrêmes, le cas limite $\alpha = 0$ redonne le q de BLUMQUIST.

Exemple 2

Le même souci de construire des indices robustes aux valeurs extrêmes conduit à la définition du coefficient w insensibilisé de corrélation

$$C_{w,k}(X,Y) = \text{COV} \left[f\left(\frac{X - a}{D(X)^{1/2}}\right), g\left(\frac{Y - b}{D(Y)^{1/2}}\right) \right]$$

où k est un réel ≥ 0 fixé et

$$\begin{aligned}
 & k \text{ si } x > k \\
 f(x) = g(x) = x & \text{ si } x \in [-k, +k] \\
 & -k \text{ si } x < -k
 \end{aligned}$$

Les simulations faites par POON [66] montrent que l'influence des valeurs sur $C_{w,k}(X,Y)$ diminue comme de juste avec les valeurs de k . L'idée intuitive conduisant au choix de la fonction.

$$f(x) = g(x) = x \mathbb{1}_{[-k,+k]}(x)$$

conduisant à un coefficient de corrélation "trimmed" ne définit pas un indice de concordance parce que f et g ne sont pas monotones.

4.4. Dualité et concordance.

Les indices présentés dans le paragraphe précédent, en termes de cosinus apparaissent comme des coefficients de corrélation linéaire sur les codages dérivés Z et T . Ils mettent ainsi en dualité l'espace des variables avec l'espace des objets. Cette propriété est fondée sur l'expression de la covariance donnée en 4.1. L'étude de relations k -naires définissant un système relationnel empirique /A nécessite une écriture plus adaptée de la covariance. A titre d'exemple, nous étudions le cas binaire et ternaire.

$$\begin{aligned}
 \text{COV}(X,Y) &= \sum_a q_a (x_a - \bar{x})(y_a - \bar{y}) ; \sum_a q_a = 1 \\
 &= \frac{1}{2} \sum_{a_1} \sum_{a_2} q_{a_1} q_{a_2} (x_{a_1} - x_{a_2})(y_{a_1} - y_{a_2})
 \end{aligned} \tag{4}$$

de même vérifie-t-on

$$\text{COV}(X,Y) = \sum_{a_1} \sum_{a_2} \sum_{a_3} q_{a_1} q_{a_2} q_{a_3} (x_{a_1} - x_{a_2})(y_{a_1} - y_{a_2}) \tag{5}$$

De telles expressions permettent de mettre en dualité les paires d'objets (resp. les triplets d'objets) et les codages dérivés.

Soit A la matrice d'incidence

$${}^t A = I_n \otimes {}^t J_{\underline{n}} - {}^t J_{\underline{n}} \otimes J_n$$

et les codages dérivés définis sur les paires par

$$Z(a_1, a_2) = x_{a_1} - x_{a_2}, \quad T(a_1, a_2) = t_{a_1} - t_{a_2}$$

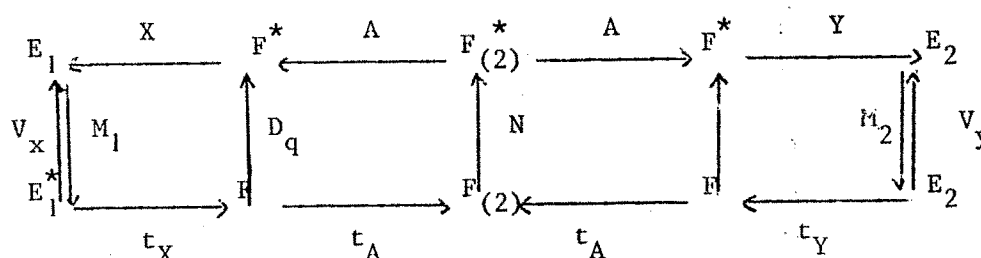
alors

$$Z = AX \text{ et } T = AY$$

Plus généralement si X (resp. Y) est un tableau de données $p \times n$ (resp. $q \times n$) correspondant à l'observation de p (resp. q) variables X^1, \dots, X^p (resp. $Y^1 \dots Y^q$) alors le tableau Z (resp. T) des codages dérivés $Z^1 \dots Z^p$ (resp. $T^1 \dots T^q$) s'écrit

$$Z = AX \text{ (resp. } T = AY)$$

la transformation linéaire A permet de construire le nouveau schéma de dualité.



si on choisit pour métrique N

$$N = D_q \otimes D_q$$

il vient que

$$\langle \underline{z}^1, \underline{z}^2 \rangle_N = 2 \langle \underline{x}^1 - \bar{x}^1 \underline{J}_n, \underline{x}^2 - \bar{x}^2 \underline{J}_n \rangle_{D_q} = 2 \text{COV}(X^1, X^2)$$

par suite la géométrie définie sur l'espace des variables centrées par la dualité "objets-variables" est identique à celle définie par la dualité "paires d'objets-variables", de plus la matrice A assure que le sous-espace de $F^{(2)}$ engendré par Z est identique au sous-espace de F engendré par X. La correspondance qui existe entre les deux schémas de dualité associé à X et à Z est dû au fait que les fonctions concordantes utilisées dans le calcul de $\text{COV}(X^i, X^L)$ vérifient

$$f_{ii'} = g_{ii'} = \rho$$

A l'opposé dans le calcul du coefficient de Spearman on a $f_{ii'} \neq g_{ii'}$, et $f_{ii'} = f_{i''i}$ ce qui interdit, semble-t-il, de construire un schéma de dualité liant les triplets d'individus aux variables recodées associé à l'indice de Spearman ; on utilise donc le recodage monotone par les rangs pour mettre en dualité individus et variables.

Les propriétés de la différence symétrique, comme métrique euclidienne dans l'espace des relations, assure la pertinence du schéma du dualité

associée au tableau Z liant les paires d'objets aux variables. Elles justifient l'analyse en composantes principales sur matrice de corrélation de KENDALL et donnent un moyen d'interpréter les résultats en repérant non pas les objets mais les paires d'objets contribuant à l'inertie.

Dans l'étude des variables nominales, la fonction de concordance définissant les codages des paires

$$Z(a_1, a_2) = \begin{cases} \frac{1}{n_i} & \text{si } X(a_1) = X(a_2) = C_i \\ 0 & \text{sinon} \end{cases}$$

justifient de même la pratique de l'analyse en composantes principales sur matrice de concordance de Tchuprow (χ^2) encore appelée M-analyse optimale (cf. CAZES et al. [17]).

Toutefois dans les deux cas, la fonction de concordance n'étant pas linéaire, on ne peut représenter les individus.

5 GENERALISATION AU CAS VECTORIEL

Les méthodes usuelles d'analyse linéaire des données relèvent de deux grandes classes de problèmes distinguant les analyses des liaisons internes à un paquet de variables, telles que l'analyse en composantes principales, et les analyses des interrelations entre éléments de deux paquets de variables. L'étude de ce deuxième problème fait l'objet de ce chapitre. On notera cependant que la mesure de la qualité d'une représentation en analyse en composantes principales relève du deuxième problème. De tels problèmes sont rencontrés en pratique dans l'étude du modèle linéaire multivarié, les problèmes de régression et d'analyse en composantes principales de paquets de variables qualitatives par exemple.

Les problèmes de dépendance étudiés dans le chapitre 2, reçoivent un traitement identique, en termes d'information discriminante dans le cas multivarié, lorsque l'on travaille de façon globale. Les problèmes de description de modèle de dépendance partielle étudiée par exemple par DAUDIN [24], BISHOP et al. [12], HABERMAN, COLOMBIER [18] seront évoqués dans le cadre de la mesure d'une association partielle.

Dans un premier paragraphe nous distinguons les problèmes de mesures de connexion et de concordance, dans l'étude d'une association dans le cadre multivarié. La description de vecteurs aléatoires de dimension supérieure ou égale à trois pose le problème de mesures d'une association partielle ; ce point fait l'objet du deuxième paragraphe, ainsi que sa généralisation aux cas multivariés, nécessaire, par exemple, en régression pas à pas sur variables qualitatives.

5.1. Liaison entre deux vecteurs aléatoires.

La définition d'indicateurs de liaison dans le cadre multivarié se heurte à la difficulté de généraliser à ce cadre les notions de ϕ -moyenne, d'incertitude et de covariance. Nous rappelons dans la suite les principaux résultats démontrés à ce sujet dans l'ensemble des vecteurs de L^2 et on étudie ensuite les géométries associées aux opérateurs trace et déterminant qui leur correspondent dans le cas de population finie.

5.1.1. Notions de dispersion d'un vecteur aléatoire

Soit $X = (X^1, X^2, \dots, X^p)$ un vecteur aléatoire défini sur (Ω, \mathcal{A}, P) .

Soit f^* une forme linéaire sur l'espace des individus E , elle définit la variable aléatoire :

$$\langle X, f^* \rangle : \omega \rightarrow \langle X(\omega), f^* \rangle$$

Définition 1.

Le vecteur de E noté $E[X]$ et défini par

$$(\forall f^* \in E^*) \quad \langle E(X), f^* \rangle = E(\langle X, f^* \rangle)$$

existe si $\langle X, f^* \rangle$ est P -intégrable pour tout f^* .

On l'appelle espérance de X ou moyenne de X .

Remarque :

Si $\{e_1, e_2, \dots, e_p\}$ est une base de E , les coordonnées de $E(X)$ dans cette base s'écrivent :

$$E(X)_i = E(\langle X, e_i^* \rangle) = E[X^i]$$

où e_i^* (resp. X^i) est le i ème élément de la base duale de (resp. la i ème coordonnée de X dans la base $\{e_1, \dots, e_p\}$).

Définition 2.

Si pour tout f^* élément de E^* , $\langle X, f^* \rangle$ est de variance finie, la fonction de f^* définie par

$$\text{Var}_X(f^*) = E[(\langle X, f^* \rangle - \langle E(X), f^* \rangle)^2]$$

est appelée variance de X .

Soit \underline{a}^* élément de E^*

$$\underline{a}^* = \sum_{i=1}^p a_i e_i^*$$

$$\text{Var}_X(\underline{a}^*) = E\left[\left(\sum_{i=1}^p a_i (X^i - E(X^i))\right)^2\right]$$

$$= \sum_{i=1}^p \sum_{j=1}^p a_i a_j V_X(i, j)$$

définit la forme quadratique sur E^* de matrice dans la base $(\underline{e}_1^*, \dots, \underline{e}_p^*)$

$$V_X = (V_X(i,j))_{i=1 \dots p, j=1 \dots p}$$

$$V_X(i,j) = \text{COV}(X^i, X^j)$$

Soit M un élément de $\mathcal{L}(E, E^*)$, alors la dualité de $\mathcal{L}(E^*, E^*)$ et $\mathcal{L}(E^*, E)$ nous permet de définir le produit scalaire

$$\langle M, V_X \rangle = \text{Tr}(V_X M)$$

Lorsqu'on dispose d'un tableau de données X mettant en dualité l'ensemble des p variables X^1, \dots, X^p est un ensemble fini d'objets (Ω, A, P) l'analyse classique du schéma de dualité qui suppose les matrices des applications écrites dans les bases canoniques.

$$\begin{array}{ccc}
 \Delta^{\mathbb{R}^p} = E & \xrightarrow{X} & F^* \\
 \begin{array}{c} \uparrow \\ V_X \\ \downarrow \end{array} & & \begin{array}{c} \downarrow \\ W \\ \uparrow \end{array} \\
 E^* & & F = \Delta^{\mathbb{R}^n} \\
 & \xrightarrow{t_X} & \\
 & & \begin{array}{c} \uparrow \\ D_p \\ \downarrow \end{array}
 \end{array}
 \quad
 \begin{array}{l}
 V_X = X D_p^t X \\
 W = t_X M X
 \end{array}$$

permet de montrer que

$$\text{Tr}(V_X M) = \frac{1}{2} \sum_{i,-j} p_i p_j \| \underline{x}_i - \underline{x}_j \|^2_M = \text{Tr}(W D_p)$$

Cette inertie du nuage des individus dans la métrique M s'interprète comme la somme des variances des X^i lorsque M est la matrice identité, et les variables X^i centrées.

On peut donc retenir cette quantité comme indicateur de dispersion qui dans le cas $p = 1$ redonne la variance.

Remarque :

$V_X M$ (resp. $W D_p$) est la matrice du projecteur sur l'espace engendré E_X (resp. F_X) par les vecteurs colonnes (resp. lignes) du tableau X .

L'ensemble autodual des automorphismes de E (resp. F) permet comme

précédemment de définir le produit scalaire entre projecteurs :

$$(V_{X_1} M, V_{X_2} M) = T_r(V_{X_1} M V_{X_2} M)$$

$$(\text{resp } ((W_1 D_p, W_2 D_p)) = T_r(W_1 D_p W_2 D_p))$$

Ce produit scalaire doit être distingué du précédent qui était associé à la norme euclidienne de X

$$T_r(V_X M) = E [\| X - E(X) \|_M^2]$$

Le produit scalaire ((.,.)) est associé au coefficient R.V. d'ESCOUFIER [29].

L'utilisation de la trace pour mesurer une dispersion est fondée sur l'inertie du nuage des individus autour de leur barycentre, cette généralisation exprimée dans l'espace des variables présente (parfois!) l'inconvénient de ne pas tenir compte des corrélations entre ces variables. On peut remarquer que la variance s'écrit aussi dans le cas unidimensionnel

$$\begin{aligned} \sigma_y^2 &= \frac{1}{2} E [(Y_1 - Y_2)^2] \\ &= \frac{1}{2} E [| \begin{matrix} 1 & 1 \\ Y_1 & Y_2 \end{matrix} |] \end{aligned}$$

où Y_1 et Y_2 sont deux variables de même loi que Y et indépendantes. et $|A|$ est le déterminant de la matrice A.

Cette propriété de U-statistique de la variance s'étend dans le cas multivarié à la quantité

$$\sigma_X^2 = \frac{1}{(p+1)!} E [| \begin{matrix} 1 \dots 1 \\ X_1 \dots X_{p+1} \end{matrix} |] = \frac{1}{p!} E [| \begin{matrix} 1 \dots 1 & 1 \\ X_1 & X_p & E[X] \end{matrix} |]$$

où X_1, \dots, X_{p+1} sont des vecteurs aléatoires de même loi que X dans l'espace des individus, $|V_X|$ est égal proportionnel à la somme des carrés des volumes de tous les parallélotopes construits sur p vecteurs \underline{x}_i choisis par les n (les variables étant supposées centrées) dans l'espace des variables, $|V_X|$ est proportionnel au carré du volume du parallélotope de côtés principaux les \underline{x}^i .

Remarque :

Du point de vue algébrique, le déterminant définit un cosinus lorsqu'on l'interprète dans l'algèbre extérieure d'ordre p définie sur F (cf. GREUB [35]) : si on note (X) le déterminant de $W D_p$ et (X, Y) le déterminant du projecteur sur $F_X \cap F_Y$ alors

$$\cos \theta = \frac{(X, Y)}{(X)(Y)}$$

Cette définition suppose que X et Y soit de même ordre.

Cette définition tient compte de l'orientation des sous-espaces F_X et F_Y et vérifie

- $\cos \theta = 1$ ssi $F_X = F_Y$ en tant qu'espaces orientés ;
- $\cos \theta = -1$ ssi $F_X = F_Y$ et les orientations sont opposées ;
- $\cos \theta = 0$ ssi $F_X \cap F_Y \neq 0$

La norme de Banach de l'opérateur de variance covariance permet de définir une autre généralisation de la dispersion. Dans l'espace des individus, cette norme est égale à l'inertie expliquée sur le premier axe principal.

5.1.2. Indices de liaison multiple

L'étude des indices de connexion d'un vecteur aléatoire $Y = (Y^1, \dots, Y^q)$ au vecteur aléatoire X s'est essentiellement développée dans le cadre des variables d'échelle d'intervalle de variance finie. Nous suivons ici cette voie afin de déterminer les arbitraires conduisant aux nombreux indices proposés.

Le cas le plus élémentaire correspond à $p = q = 1$. En ce cas le modèle de prédiction de Y consiste à coder l'individu ω par la valeur $\hat{y}(\omega)$, où \hat{y} est la variable fonction linéaire de X obtenue dans une régression au sens des moindres carrés.

$$\hat{y} = \frac{\text{COV}(X, Y)}{\text{Var}(X)} X$$

les variables étant supposées centrées.

L'indice P.R.E. associé à la variance s'écrit :

$$r^2 = 1 - \frac{\text{Var}(y - \hat{y})}{\text{Var}(y)} = \frac{\text{Var}(\hat{y})}{\text{Var}(y)}$$

On retrouve le carré du coefficient de corrélation linéaire du fait de l'additivité de la variance.

On vérifie alors :

$$r^2 = T_r(I - D) \quad , \quad D = V_y^{-1} V(y - \hat{y})$$

où I, matrice unité, D, $V_y^{-1} V(y - \hat{y})$ sont des matrices scalaires.

On vérifie tout aussi aisément

$$r^2 = |I - D| = |V_{\hat{y}}| |V_y|^{-1}$$

Le cas intermédiaire $q = 1$ $p > 1$ conduit à prédire pour l'individu ω la valeur $\hat{y}(\omega)$ obtenue en construisant la projection au sens des moindres carrés de la variable y sur le sous-espace linéaire engendré par X.

L'indice P.R.E. s'écrit alors

$$R^2 = R_{Y.X^1, \dots, X^r}^2 = 1 - \frac{V_{y - \hat{y}}}{V_y} = \frac{V_{YX} V_X^{-1} V_{XY}}{V_Y}$$

où $V_{YX} = Y D_r^t X$.

On retrouve le coefficient de corrélation multiple que l'on peut écrire de façons équivalentes.

Proposition 1 :

Les expressions suivantes du coefficient R^2 sont équivalentes :

$$\rho_1^2 = \text{Sup} \left\{ \frac{\| \underline{u} \|_M^2}{\| \underline{u} \|^2} ; \underline{u} \in E_2^* \right\} \quad , \quad M = V_y^{-1} V_{yx} V_x^{-1} V_{xy}$$

$$\gamma_1 = \frac{|V_{yx} V_x^{-1} V_{xy}|}{|V_y|} = |V_{yx} V_x^{-1} V_{xy} V_y^{-1}| = |V_{\hat{y}} V_y^{-1}|$$

$$\gamma_2 = 1 - \frac{|v_y - \hat{y}|}{|v_y|} = 1 - |v_y^{-1} v_y - \hat{y}|$$

$$\gamma_3 = \frac{T_r(v_y^{-1} v_{\hat{y}})}{T_r(v_y - \hat{y} v_y)}$$

$$\gamma_4 = T_r(v_y^{-1} v_{\hat{y}})$$

ρ_1^2 est donc la norme de Banach de l'opérateur $v_y^{-1} v_{\hat{y}}$.

γ_1 et γ_2 utilisent la décomposition de la variance.

γ_3 s'exprime en fonction du rapport des normes des opérateurs de variance covariances lorsque la métrique sur l'espace des individus est la métrique de Mahalanobis définie sur les résidus.

Dans le cas $q > 1$ et $p > 1$, on prédit pour ω le vecteur $\hat{y}(\omega)$ obtenu dans la régression linéaire multivariée de Y en X .

Les généralisations de l'expression précédente se différencient soit par le choix de la norme, soit par le fait qu'on norme $v_{\hat{y}}$ ou $v_y - \hat{y}$. Nous donnons leurs expressions dans une base quelconque ainsi que dans la base des composantes canoniques. Le coefficient ρ_1^2 (HOTELLING 1936) est égal à la plus grande corrélation canonique.

Les expressions utilisant le déterminant construites à partir de $|v_{\hat{y}}|$

$$\gamma_1 = \frac{|v_{yx} \ v_x^{-1} \ v_{xy} \ v_y^{-1}|}{\prod_{i=1}^q \rho_i^2} \quad (\text{HOTELLING 1936})$$

$$\gamma_4 = \sqrt[q]{\gamma_1} \quad (\text{HOOPER 1959}), \text{ moyenne géométrique des } \rho_i^2.$$

Les expressions utilisant le déterminant, construites à partir de $|v_y - \hat{y}|$

$$\gamma^2 = 1 - |v_y - \hat{y} v_y^{-1}| = 1 - \prod_{i=1}^q (1 - \rho_i^2) \quad (\text{HOTELLING, 1936 [43]}).$$

Le test d'analyse de variance proposé par WILKS utilise $1 - \gamma_2$

$$\gamma_5 = 1 - \sqrt[q]{1 - \gamma_2} = 1 - \sqrt[q]{M(1 - \rho_1^2)} \quad (\text{CRAMER et NICE WANDER 1979 [21]})$$

$$I(1,2) = -\frac{1}{2} \log \frac{\begin{vmatrix} v_x & v_{xy} \\ v_{yx} & v_y \end{vmatrix}}{\begin{vmatrix} v_x & 0 \\ 0 & v_y \end{vmatrix}} = -\frac{1}{2} \log \prod_{i=1}^q (1 - \rho_i^2) \quad (\text{KULLBACK 1958 [50]})$$

Cet indice est construit à partir de l'information discriminante, étudiée dans le chapitre 2, calculée dans le cas de vertène gaussien. On vérifie que

$$\begin{vmatrix} V_x & V_{xy} \\ V_{yx} & V_y \end{vmatrix} = |V_x| |V_y - \hat{y}|$$

Cette partition en deux classes d'indices est due à la non additivité du déterminant. C'est pourquoi la généralisation de la statistique F, due à BARTLET 1948 [9]

$$\gamma_7 = |V_{\hat{y}}| |V_y - \hat{y}|^{-1} = \prod_i \frac{\rho_i^2}{1 - \rho_i^2}$$

n'est pas réductible aux indices précédents.

Les expressions utilisant la trace se différencient parce que

$$\text{Tr}(AB) \neq \text{Tr}(A) \text{Tr}(B)$$

$$\gamma_3 = \frac{\text{Tr}(V_q V_y^{-1} (y - \hat{y}))}{\text{Tr}(V_y V_y^{-1} (y - \hat{y}))} = 1 - \frac{p}{\sum_{i=1}^p \frac{1}{1 - \rho_i^2}} \quad (\text{COXHEAD, 1974 [20]})$$

γ_3 est la moyenne harmonique des $(1 - \rho_i^2)$

$$\gamma_6^r = \frac{1}{q} \text{Tr}(V_y^{-1} V_{\hat{y}}) = \frac{1}{q} \sum \rho_i^2$$

γ_6 est la moyenne arithmétique des ρ_i^2 utilisée dans le test de PILLAI de même LAWLEY 1948 [53] utilise la statistique de test

$$\gamma_8 = \text{Tr}(V_{\hat{y}} V_y^{-1} (y - \hat{y})) = \sum \frac{\rho_i}{1 - \rho_i^2}$$

qui est équivalente à la statistique proposée par KULLBACK.

$$J(1,2) = \gamma_8 - q$$

le coefficient RV d'ESCOUFIER 1973 [29]

$$\frac{\text{Tr}(V_{\hat{y}} V_y^{-1})}{\sqrt{pq}} = \frac{1}{\sqrt{pq}} \sum \rho_i^2$$

est le cosinus entre les projecteurs W_1^D , W_2^D .

Proposition 2 (RAMER et NICEWANDER [21])

$$\gamma_1 \leq \rho_q^2 \leq \gamma_4 \leq \gamma_6 \leq \gamma_5 \leq \gamma_3 \leq \rho_1^2 \leq \gamma_2$$

où ρ_q^2 est la plus petite corrélation canonique.

L'ensemble de ces indices est lié à l'analyse canonique bien que traduisant une notion de connexion. Ceci résulte d'une logique linéaire de prédiction. On peut cependant définir dans ce cadre des indices plus directement liés à une logique de régression tel que l'indice de redondance

$$G_1^2 = \frac{1}{q} \sum_{i=1}^q \|W_i D_p \underline{y}^i\|_M^2$$

qui mesure l'inertie du nuage projeté.

La logique linéaire sous-jacente à la construction de ces indices limite leur utilisation aux variables d'échelle d'intervalle - puisque les opérations de projection sur les sous-espaces canoniques ne définissent pas en général des relations pertinentes - pour un système relationnel plus pauvre. Les variables d'échelle nominale peuvent cependant être analysées en termes de connexion puisque l'ensemble des indices associés à une entropie sont calculables dans le cas multivarié à partir de la partition produit, définie par chaque paquet de variables. Une telle pratique est cependant limitée aux traitements d'un grand nombre d'objets de façon à assurer un effectif minimal dans chaque case.

De la même façon le calcul d'une incertitude dans chaque classe de la partition induite par les catégories du vecteur X, permet de généraliser, avec la même contrainte sur les effectifs, le calcul des indices de connexion associé à une dispersion. On étend ainsi la logique de l'analyse de la variance. On peut ainsi vérifier la pertinence d'hypothèses d'appartenance d'une ϕ -moyenne à un sous-espace vectoriel donné de E. Un cas particulier de cette approche a été développé par NELDER et WEDERBURN [63] dans un modèle linéaire sous hypothèses de structure exponentielle, de plus l'utilisation de la régression L^1 à des fins de robustesse est une technique récente qui relève de ce principe ARMSTRONG et FROME [5].

La définition d'indices de concordance, rencontrés dans l'analyse de données cubiques nécessitent la donnée d'un produit scalaire entre deux tableaux. De tels exemples ont été donnés plus haut.

5.2. Notions de liaison partielle.

L'étude de l'association de deux éléments aléatoires choisis parmi p se rencontrent dans de nombreuses méthodes d'analyse de vecteurs aléatoires telles que l'analyse discriminante pas à pas, la sélection de variables en régression, la description de tables de contingence multidimensionnelles. La distinction entre étude de la dépendance et étude de la liaison est là encore nécessaire. Le problème est compliqué par le fait qu'il n'est pas indifférent de travailler sur les objets ou sur les paires d'objets.

Nous rappelons tout d'abord la distinction entre association partielle et association conditionnelle et montrons alors l'insuffisance des méthodes linéaires usuelles pour traduire la notion de liaison partielle.

5.2.1. Liaison conditionnelle et liaison partielle

Soit X un vecteur aléatoire p -dimensionnel, décomposé en $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, de matrice de variance-covariance :

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

où Σ_{22} est supposée inversible. Posons alors

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

Définition 3.

On appelle coefficient de corrélation conditionnelle des composantes X_1^i et X_1^j de X_1 sachant X_2 , la variable aléatoire $\rho_{ij/2}$ définie par

$$\rho_{ij/2} = \frac{\sigma_{ij/2}}{\sqrt{\sigma_{ii/2} \sigma_{jj/2}}}$$

où $\sigma_{ij/2}$ est le terme (ij) de la matrice de variance covariance conditionnelle de X_1 sachant X_2 .

Dans le cas d'un vecteur gaussien la matrice de variance-covariance

conditionnelle est constante et égale à $\Sigma_{11.2}$ (cf. GRAYBILL [38]). Le coefficient de corrélation conditionnelle s'interprète géométriquement comme le cosinus de l'angle formé par les facteurs résidus obtenus dans la régression linéaire de X_1^i (resp. X_1^j) sur X_2 . Cette interprétation géométrique simple conduit à définir un indice en termes de moindres carrés sans référence à l'hypothèse gaussienne, obtenu par analogie au cas gaussien.

Définition 4.

On appelle coefficient de corrélation partielle de X_1^i avec X_1^j , à X_2 contrôlé, la quantité $\rho_{ij.2}$ définie à partir de la matrice $\Sigma_{11.2}$ par

$$\rho_{ij.2} = \frac{\sigma_{ij.2}}{\sqrt{\sigma_{ii.2}\sigma_{jj.2}}}$$

L'équivalence entre ces deux notions est limitée aux distributions de probabilités pour lesquelles $\rho_{ij/2}$ est une variable p.s. constante, associée à une fonction de régression linéaire. De nombreux contre-exemples ont été proposés afin de montrer l'originalité de chaque notion. Nous rappelons ici ceux donnés par GOKHALE [37].

C.E. 1 :

Soit $X = (X^1, X^2, X^3)$, $X_1 = (X^1, X^2)$

On suppose que $X^1 = X^2 X^3$ et que X^2 et X^3 sont centrées et indépendantes ;

on obtient :

$$E(X^1/X^3 = x_3) = x_3 \quad E[X^2/X^3 = x_3]$$

$$\rho_{12/3}(x_3) = \text{sg}(x_3)$$

d'autre part

$$\rho_{12} = \rho_{13} = \rho_{23} = \rho_{12.3} = 0$$

Remarque :

La pratique usuelle consistant à résumer l'information portée par la variable $\rho_{ij/2}$ par son espérance, ne suffit pas à identifier les deux notions, puisque on peut rendre $E[\rho_{ij/2}]$ aussi proche de ± 1 que l'on veut par le choix d'une distribution fortement dissymétrique.

C.E. 2 :

Soit $X = (X^1, X^2, X^3)$ et $X_1 = (X^1, X^2)$

On suppose que

$$X_1^1 = z(X^3) + U$$

$$X_1^2 = z(X^3) + V$$

où $z(\cdot)$ est une fonction non linéaire et X^3, U, V sont des variables indépendantes, de variance finie. Alors

$$\rho_{12/3}(x_3) = 0$$

cependant

$$\rho_{12.3} = (V(Z) V(X^3) - \text{COV}^2(Z, X^3)) / D$$

peut être rendu arbitrairement proche de 1 en faisant tendre vers 0 les variances de U et V puisque

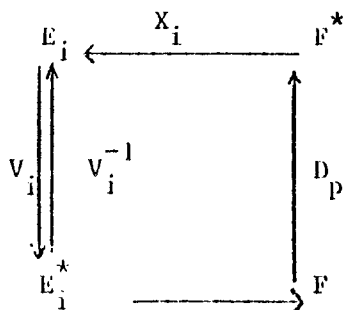
$$D^2 = [V(X^3)(V(Z) + V(U)) - \text{COV}^2(Z, X^3)][V(X^3)(V(Z) + V(U)) - \text{COV}^2(Z, X^3)]$$

Lorsque X^1, X^2, X^3 sont observés sur une population finie, ces deux notions distinctes se traduisent par des manipulations distinctes des tableaux de tri croisé de X_1 et X_2 . La mesure de la corrélation conditionnelle conduit en effet à construire une famille de tableaux de tri croisé conditionnel à X^3 , indexée par les valeurs de la variable filtre X^3 , cette opération est fort distincte des transformations nécessaires à la mesure de la corrélation partielle ; dans ce cas en effet le tableau de contingence reste inchangé, et les codages des lignes et colonnes deviennent ceux des résidus de la régression linéaire de X^1 et X^2 sur X^3 .

5.2.2. Décomposition de la corrélation partielle

Nous avons vu dans le paragraphe précédent que la matrice de variance covariance conditionnelle est constante dans le cas gaussien ; en l'absence d'une telle hypothèse, on doit distinguer les éléments de la partition de l'ensemble des individus A induite par le vecteur qui conditionne. Celui-ci joue donc un double rôle dans l'analyse comme indicateur qualitatif de cette partition et comme vecteur explicatif dans la régression. Ces deux aspects sont étudiés dans ce paragraphe.

L'analyse linéaire du tableau $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ est associée aux schémas de dualité



$$\begin{aligned}
 V_i &= X_i D_p {}^t X_i \\
 V_{ij} &= X_i D_p {}^t X_j \\
 D_p &= \text{Diag}(p_a), \quad p_a > 0, \quad \sum_a p_a = 1
 \end{aligned}$$

X_j définir sur A la partition $\{A_j, j = 1, \dots, J\}$ de poids

$$q_j = \sum_{a \in A_j} p_a, \quad j = 1, \dots, J$$

et dans E_i ($i = 1, 2$) la partition en nuages d'individus

$$\mathcal{N}_i^j = \left\{ \left(\underline{x}_{ia}, \frac{p_a}{q_j} \right), a \in A_j, j = 1, \dots, J, i = 1, 2 \right\}$$

de barycentres respectifs

$$\underline{g}_i^j = \sum_{a \in A_j} \frac{p_a}{q_j} \underline{x}_{ia}$$

et de forme quadratique d'inertie

$$V_i^j = \sum_{a \in A_j} p_a \underline{x}_{ia} {}^t \underline{x}_{ia} - q_j \underline{g}_i^j {}^t \underline{g}_i^j$$

$\frac{1}{q_j} V_i^j$ est alors la matrice de variance associée au nuage \mathcal{N}_i^j .

La matrice de variance intra-classe V_i^w est alors définie par

$$V_i^w = \sum V_i^j$$

On note D_q la matrice diagonale des poids q_j et G_i le tableau

$$G_i = (\underline{g}_i^1, \dots, \underline{g}_i^J)$$

la forme quadratique d'inertie des centres de gravité est associée à la matrice de variance inter-classe :

$$V_i^b = G_i D_q {}^t G_i$$

On retrouve la formule de la décomposition de la variance

$$V_i = V_i^b + V_i^w$$

De la même façon, on obtient la décomposition de la matrice des covariances croisées :

$$V_{12} = V_{12}^b + V_{12}^w$$

Définition 5.

On appelle matrice de variance covariance partielle de X^1 avec X^2 à X^3 contrôlé, la matrice

$$V_{12.3} = V_{12} - V_{13} V_3^{-1} V_{32}$$

dont les éléments sont les covariances partielles à X_3 contrôlé.

Proposition 3

Les éléments de la décomposition de $V_{12.3}$ ont pour expression

$$V_{12.3}^w = V_{12}^w$$

$$V_{12.3}^b = V_{12}^b - V_{13} V_3^{-1} V_{32}$$

Remarque :

Le codage du vecteur X_3 n'intervient que dans l'expression de l'élément inter-classe.

Démonstration :

La régression de X_i ($i = 1, 2$) en X_3 permet de construire les tableaux de résidus (resp. barycentres).

$$Y_i = X_i - V_{i3} V_3^{-1} X_3 \quad (\text{resp. } \hat{g}_i^j)$$

et les matrices de variance-covariance associées :

$$V_{ik.3}^j = \sum_{a \in A_j} p_a y_{ia}^t y_{ka}^t - q_j \hat{g}_i^j t_{\hat{g}_k^j}, \quad i, k = 1, 2$$

$$\begin{aligned} &= V_{ik}^j - q_j \sum_{a \in A_j} (p_a \frac{x_{ia}^t}{g_3^j} V_3^{-1} - V_{i3} V_3^{-1} \frac{1}{g_3^j} t_{x_{ka}}^t) + \dots \\ &\dots + q_j V_{i3} V_3^{-1} \frac{1}{g_3^j} t_{g_3^j}^j V_3^{-1} V_{3k} - q_j \hat{g}_i^i t_{\hat{g}_k^k}^j \end{aligned}$$

en exprimant le dernier terme de la somme en fonction des \hat{g}_i^j on obtient le résultat après sommation.

$$V_{ik.3}^b = \hat{G}_i D_q \hat{G}_k = \sum_{j \in J} q_j \hat{g}_i^j t_{\hat{g}_k^j}$$

$$= \sum_{j \in J} q_j \hat{g}_i^j t_{\hat{g}_k^j} - \sum_j q_j \hat{g}_i^j t_{g_3^j}^j V_3^{-1} V_{3k} - \sum_j q_j V_{i3} V_3^{-1} V_3^j t_{g_3^j}^j + \dots$$

$$+ \sum_j q_j v_{i3} v_3^{-1} \underline{g}_3^j t_{\underline{g}_3^j} v_3^{-1} v_{3k}$$

$$\sum_j q_j \underline{g}_i^j t_{\underline{g}_3^j} = \sum_j q_j \sum_a \frac{1}{q_j} p_a x_{ia}^j t_{x_{3a}^j} = v_{i3} \quad (i = 1, 2, 3)$$

d'où

$$v_{ik.3}^b = v_{ik} - v_{i3} v_3^{-1} v_{3k} \quad \square$$

Dans le cas où on étudie la corrélation partielle entre deux variables X^1 et X^2 à X^3 contrôlée, la matrice des covariances partielles s'écrit :

$$v_{ik.3} = v_{ik}^w + v_{ik}^b - v_{i3} v_3^{-1} v_{3k}$$

$$v_{12.3}^w = \sum_j q_j \sigma_{12/3}(j)$$

$$v_{12}^b = \text{COV}_3(G_1, G_2)$$

$$v_{i3} v_3^{-1} v_{32} = \frac{\text{COV}(X_i, X_3) \text{COV}(X_3, X_2)}{\text{Var}(X_3)}$$

on en déduit les coefficients de corrélation partielle.

$$\text{totale} : \rho_{12.3} = \frac{v_{12.3}}{\sqrt{v_{11.3} v_{22.3}}}$$

$$\text{Intra} : \rho_{12.3}^w = \frac{v_{12}^w}{\sqrt{v_1^w v_2^w}} = \rho_{12}^w$$

$$\text{Inter} : \rho_{12.3}^b = \frac{v_{12.3}^b}{\sqrt{v_{11.3}^b v_{22.3}^b}}$$

auxquels sont associés les coefficients de régression partielle.

$$\beta_{12.3} = \frac{v_{12}^w + v_{12.3}^b}{v_2^w + v_{22.3}^b}$$

$$= \frac{v_2^w}{v_2^w + v_{22.3}^b} \beta_{12.3}^w + \frac{v_{22.3}^b}{v_2^w + v_{22.3}^b} \beta_{12.3}^b$$

$$\beta_{12}^w = \frac{v_{12}^w}{v_2^w}, \quad \beta_{12.3}^b = \frac{v_{12.3}^b}{v_{22.3}^b}$$

Le coefficient de régression partielle s'écrit donc dans le cas général comme une moyenne pondérée d'un terme qui s'interprète comme une "moyenne" des pentes dans chaque nuage et d'un terme qui s'interprète comme la pente de la régression des barycentres des résidus.

Trois cas particuliers nous intéressent. Tout d'abord lorsque toutes les valeurs de X^3 observées sont distinctes (cas de l'absolue continuité):

$$\beta_{12.3} = \beta_{12.3}^b$$

Lorsque la variable contrôlée est dichotomique :

$$\beta_{12.3} = \beta_{12}^w$$

Dans ce cas la covariance conditionnelle coïncide avec la covariance partielle. Lorsque les trois variables sont trichotomiques et de distribution symétrique, on obtient le même résultat. Une telle situation se rencontre en particulier dans le croisement des codages dérivés de trois variables ordinales.

Le calcul de la corrélation partielle peut être faite dans le modèle linéaire associé à la dualité "individus-variables", ou bien encore à celui qui est associé à la dualité "paires-variables".

Nous reprenons ici l'exemple donné par DAUDIN [24] dans le but de montrer la distinction entre mesure d'une corrélation partielle et mesure d'une liaison conditionnelle.

Les calculs que nous présentons montrent que l'étude de la répartition des individus définit une notion d'interaction distincte de celle qui est mesurée par des comparaisons par paires. Plus que le recours à la géométrie du modèle linéaire, c'est cette distinction qui fait de tout indice de corrélation partielle utilisant le codage des paires un indicateur inadaptable à l'étude de l'indépendance conditionnelle.

Exemple :

Tris croisés de X_1, X_2 filtré par X_3

$X_3 = 0$

X_2	0	1
X_1		
0	3	1
1	0	2

$X_3 = 1$

X_2	0	1
X_1		
0	0	2
1	3	1

Tables marginales associées

X_2	0	1
X_1		
0	3	3
1	3	3

X_3	0	1
X_1		
0	4	2
1	2	4

X_3	0	1
X_2		
0	3	3
1	3	3

$$\rho_{12}^2 = 0$$

$$\rho_{13}^2 = \frac{1}{3}$$

$$\rho_{23}^2 = 0$$

Remarque :

Dans le cas du tableau 2 x 2 :

$$\rho^2 = \rho = \tau_b$$

Soit Z_i le codage dérivé de X_i défini sur les paires par la fonction concordante $sg(x)$, $i = 1, 2, 3$.

On obtient les tableaux conditionnels suivants :

$Z_3 = 1$

Z_2	-1	0	1	
Z_1				
-1	3	10	3	16
0	6	4	6	16
1	0	4	0	4
	9	18	9	36

$Z_3 = 0$

Z_2	-1	0	1	
Z_1				
-1	6	4	6	16
0	6	28	6	40
1	6	4	6	16
	18	36	18	72

Le tableau $Z_3 = -1$ s'obtient à partir du tableau $Z = 1$ par symétrie par rapport à la ligne $Z_1 = 0$.

$$\tau_{12/Z_3} = 1 = \tau_{12/Z_3} = -1 = \tau_{12/Z_3} = -1 = 0$$

La corrélation conditionnelle sur les paires est donc constante et nulle par suite la corrélation intra-classe est nulle.

Puisque le codage est trichotomique, la corrélation partielle, égale à la corrélation intraclasse, est donc nulle.

Lorsque l'on raisonne non plus à partir des paires mais sur le codage par X_3 des individus, on obtient les tris croisés :

$X_3 = 0$

$X' \backslash Y'$	-1	0	1	
-1	6	2	0	8
0	3	14	3	20
1	0	2	6	8
	9	20	9	36

$X' \backslash Y'$	-1	0	1	
-1	0	2	6	8
0	3	14	3	20
1	6	2	0	8
	9	20	9	

X_3

$$\tau_b = \frac{\sqrt{2}}{2}$$

$$\tau_b = -\frac{\sqrt{2}}{2}$$

Ces calculs montrent que la critique faite au coefficient de Tchuprow partiel fondée sur la remarque que

$$\phi_{12.3} = 0$$

alors que on a par exemple

$$\frac{p(X_1 = X_2 = X_3 = 0)}{p(X_3 = 0)} = \frac{1}{2} \neq \frac{p(X_1 = X_3 = 0)}{p(X_3 = 0)} \cdot \frac{p(X_2 = X_3 = 0)}{p(X_3 = 0)} = \frac{2}{3} \times \frac{1}{2} =$$

s'applique à tout autre indice fondé sur un codage. Le paradoxe réside en effet dans la volonté d'étudier la structure de dépendance à l'aide d'indices mesurant la liaison en termes de concordance entre les codages

La décomposition proposée de la corrélation partielle, si elle permet d'éclairer la nature des problèmes rencontrés dans l'utilisation du modèle linéaire, ne prétend donc pas résoudre le problème dû à l'emboîtement des espaces

$$L_{X_1}^2 \oplus L_{X_3} \subset L_{X_1 X_3}^2$$

La connaissance du projecteur sur $L_{X_1 X_3}^2$ nécessaire au calcul de la sup corrélation, permet seule d'étudier les structures de dépendance.

Lorsque l'on veut étudier la concordance partielle entre codages multidimensionnels, l'étude de la concordance entre résidus issus d'une régression multivariée, utilise la proposition 3 et les résultats du paragraphe 5.1 pour construire des indices de corrélation partielle.

A N N E X E 1

Cette annexe présente un ensemble de tris croisés construits artificiellement en collaboration avec Messieurs F. BON et G. d'AUBIGNY dans le cadre de l'école d'Informatique et Mathématiques en Sciences Sociales du C.N.R.S. . Leur traitement, à l'aide de S.P.S.S. illustre les différents concepts d'association présentés dans ce travail.

Les chapitre précédents nous permettent de classer les indices d'association calculés par ce logiciel selon le tableau ci-dessous.

ECHELLE	ASSOCIATION		
	LIAISON		DEPENDANCE
	CONCORDANCE	CONNEXION	
NOMINALE	u symétrisé (*) λ symétrisé (*)	$\lambda_{Y/X}$ de GUTTMAN $u_{Y/X}$ incertitude	χ^2 V de CRAMER C de contingence de PEARSON
ORDINALE	τ_b et τ_c de KENDALL γ de GOODMAN et KRUSKAL	$d_{Y/X}$ de SOMMERS	
D'INTERVALLE	r de BRAVAIS et PEARSON	$\eta^2_{Y/X}$ rapport de corrélation	

Les indices marqués de (*) sont une forme symétrisée des indices de connexion correspondants, ajustés pour mesurer une liaison mutuelle. Ils s'expriment comme une combinaison connexe de ceux-ci

$$\lambda = \frac{I(P_X, \infty, 2)}{I(P_X, \infty, 2) + I(P_Y, \infty, 2)} \lambda_{Y/X} + \frac{I(P_Y, \infty, 2)}{I(P_X, \infty, 2) + I(P_Y, \infty, 2)} \lambda_{X/Y}$$

$$U = \frac{I(P_X, 1, 1)}{I(P_X, 1, 1) + I(P_Y, 1, 1)} U_{Y/X} + \frac{I(P_Y, 1, 1)}{I(P_X, 1, 1) + I(P_Y, 1, 1)} U_{X/Y}$$

ils réalisent donc un compromis entre les deux prédictions dissymétriques possibles pondérées par la précision relative de chacune des prédictions. Cette normalisation ne semble pas présenter de justification interprétable du point de vue statistique, contrairement aux indices de concordance qui s'expriment en termes de moyenne géométrique des coefficients de régression (r^2 , r^2 , ϕ^2 de Tchuprov, qui n'est pas calculé par S.P.S.S.).

L'étude des tableaux ci-joints utilise comme référence le numéro de la page figurant sur le listage S.P.S.S. .

Les tableaux 63, 64, 65 montrent une situation d'absence de liaison fonctionnelle associée à un fort écart de l'indépendance. Ceci traduit le fait que les tris croisés présentés contiennent trois informations distinctes portant sur les marges, les codages et le tableau de contingence. Ces trois éléments interagissent cependant : ainsi la comparaison des tableaux 3 et 4 montre que la liaison fonctionnelle mesurée par η^2 peut être augmentée par un jeu sur les effectifs.

Les tableaux suivants permettent d'étudier l'influence des codages : les tableaux 3 et 19 qui se correspondent par la permutation des lignes 1 et 3 et des colonnes 2 et 3 traduisent un recodage bijectif laissant invariant les indices de dépendance, et les indices de liaison entre échelles nominales - alors que la permutation permettant de passer du tableau 23 au tableau 24, change sensiblement les indices ordinaux.

Les tableaux 4, 16, 72 d'une part, 29, 33 d'autre part, qui se correspondent par une transformation monotone croissante de codage des lignes laissant invariants les indices d'indépendances, et les indices de liaison entre échelles nominales et ordinales.

Les tableaux 16 et 17 d'une part et 29, 31 de l'autre qui se correspondent par recodage linéaire laissent invariants tous les indices .

Le tableau 4 et ses dérivés montrent en plus une relation de caractère dissymétrique traduit par l'écart entre les valeurs des deux indices de connexion de même type.

Les tableaux 14 et 27 se correspondent par une permutation sur les cases qui ne s'écrit pas comme un produit de permutation des lignes et des colonnes. Tous les indices bougent, en particulier l'indice λ est nul dans le tableau 27 malgré des indices de dépendance non nuls.

On présente dans les tableaux 33 à 42 une série de regroupement des lignes et, on colonnes, montrant l'influence d'une transformation non exhaustive : on retrouve que le χ^2 diminue. Les tableaux 23, 24, 25 et 26 montrent des situations de liaison symétrique (tableaux carrés) ou dissymétriques (tableaux rectangulaires) . Les tableaux 54, 55 et 78 montrent trois situations où l'indice de liaison monotone γ , ne tenant pas compte des exaequos, est égal à 1 alors que les autres indices de liaison monotone ne sont pas maximum.

Les tableaux 109 et 111 sont obtenus à partir du tableau 107 en conditionnant par la variable sexe, afin de montrer la difficulté de l'interprétation de l'indice γ partiel.

FILE EXECUTIVE (CREATION DATE = 12/10/79) EXERCISE SUR LES LIATIONS
 DEVENIR CHOSE STABILITE DE
 BY AGE3A

		AGE3A			ROW
COUNT	1	2	3	TOTAL	
1	130	7	33	170	
	76.5	4.1	19.4	28.6	
	65.7	2.4	24.7		
	21.6	1.2	5.5		
2	54	42	62	158	
	34.2	26.6	39.2	26.6	
	27.3	14.7	55.9		
	9.1	7.1	10.4		
3	14	237	16	267	
	5.2	88.8	6.0	44.9	
	7.1	82.9	14.4		
	2.4	39.8	2.7		
COLUMN	198	286	111	595	
TOTAL	33.3	48.1	18.7	100.0	

PAW CHI SQUARE = 392.77930 WITH 4 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.57451
 CONTINGENCY COEFFICIENT = 0.63059
 LAMBDA (ASYMMETRIC) = 0.49390 WITH REVENUE DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.47881
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.33982 WITH REVENUE DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.34598
 KENDALL'S TAU B = 0.29555. SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.28145. SIGNIFICANCE = 0.0000
 GAMMA = 0.37165
 SOMERS'S D (ASYMMETRIC) = 0.30097 WITH REVENUE DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.29551
 FTA = 0.75441 WITH REVENUE DEPENDENT.
 PEARSON'S P = 0.32040 SIGNIFICANCE = 0.0000

= 0.46278 WITH AGE3A DEPENDENT.
 = 0.35237 WITH AGE3A DEPENDENT.
 = 0.29023 WITH AGE3A DEPENDENT.

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

12/10/79 PAGE 4

FILE EX011A1 (CREATION DATE = 12/10/79) EXERCISE SUPPLIES LIAISONS
DEVELOPMENT C.P. SYSTEM AT ION OF BY AGE3R
PAGE 1 OF 1

AGE3R	COUNT	ROW PCT	COL PCT	TOT PCT	ROW TOTAL
1	150	7	63	220	220
2	34	3.2	28.6	37.0	37.0
3	39.1	2.4	56.8	14.6	14.6
TOTAL	198	1.2	19.6	57	57
	14	25.8	16	28.8	28.8
	4.9	89.6	5.6	48.4	48.4
	7.1	90.2	14.4	2.7	2.7
	2.4	43.4	2.7	111	111
COLUMN TOTAL	198	246	18.7	595	100.0

ROW CHI SQUARE = 410.71704 WITH 4 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.58749
 CONTINGENCY COEFFICIENT = 0.63905
 LAMBDA (ASYMMETRIC) = 0.59609 WITH PEVFRU2 DEPENDENT. = 0.50485 WITH AGE3R DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.55632
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.40768 WITH PEVFRU2 DEPENDENT. = 0.39531 WITH AGE3R DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.40140
 KENDALL'S TAU B = 0.26713. SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.24652. SIGNIFICANCE = 0.0000
 GAMMA = 0.32710
 SOMERS'S D (ASYMMETRIC) = 0.26373 WITH PEVFRU2 DEPENDENT. = 0.27058 WITH AGE3R DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.26711
 ETA = 0.40707 WITH PEVFRU2 DEPENDENT. = 0.27090 WITH AGE3R DEPENDENT.
 PEARSON'S R = 0.25701 SIGNIFICANCE = 0.0000

FILE FXP01A1 (CREATION DATE = 12/10/79) EXERCICE SUR LES LIANSONS
 FCHFL ATTITUDE A L'EGARD DE LA VIOLENCE C R O S T A B U L A T I O N O F
 BY AGE71

FCHFL	AGE71	1	2	3	4	5	6	7	ROW TOTAL
1	COUNT	55	3	3	1	1	0	1	64
	PCH PCT	85.9	4.7	4.7	1.6	1.6	0.0	1.6	10.8
	COL PCT	78.6	3.7	2.9	1.7	1.2	0.0	0.7	
	TOT PCT	9.2	0.5	0.5	0.2	0.2	0.0	0.2	
2	COUNT	10	6	9	2	0	1	1	91
	PCH PCT	11.0	74.7	9.9	2.2	0.0	1.1	1.1	15.3
	COL PCT	14.3	42.9	11.8	3.4	0.0	1.8	0.7	
	TOT PCT	1.7	11.4	1.5	0.3	0.0	0.2	0.2	
3	COUNT	2	4	10	6	1	0	1	99
	PCH PCT	2.0	9.1	10.8	6.1	1.0	0.0	1.0	16.6
	COL PCT	2.9	11.0	14.4	10.3	1.2	0.0	0.7	
	TOT PCT	0.3	1.5	1.4	1.0	0.2	0.0	0.2	
4	COUNT	2	1	8	40	4	0	2	57
	PCH PCT	3.5	1.8	14.0	70.2	7.0	0.0	3.5	9.6
	COL PCT	2.9	1.2	7.8	69.0	4.7	0.0	1.4	
	TOT PCT	0.3	0.2	1.3	6.7	0.7	0.0	0.3	
5	COUNT	1	0	1	7	71	6	1	87
	PCH PCT	1.1	0.0	1.1	8.0	81.6	6.9	1.1	14.6
	COL PCT	1.4	0.0	1.0	12.1	83.5	10.9	0.7	
	TOT PCT	0.2	0.0	0.2	1.2	11.9	1.0	0.2	
6	COUNT	0	1	0	1	6	43	5	56
	PCH PCT	0.0	1.8	0.0	1.8	10.7	76.8	8.9	9.4
	COL PCT	0.0	1.2	0.0	1.7	7.1	78.2	3.5	
	TOT PCT	0.0	0.2	0.0	0.2	1.0	7.2	0.8	
7	COUNT	0	0	1	1	2	5	132	141
	PCH PCT	0.0	0.0	0.7	0.7	1.4	3.5	93.6	23.7
	COL PCT	0.0	0.0	1.0	1.7	2.4	9.1	92.3	
	TOT PCT	0.0	0.0	0.2	0.2	0.3	0.8	22.2	
	COLUMN TOTAL	70	42	102	58	85	55	143	595
	TOTAL	11.8	13.8	17.1	9.7	14.3	9.2	24.0	100.0

PAW CHI SQUARE = 2173.83252 WITH 36 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.79033
 CONTINGENCY COEFFICIENT = 0.88606
 LAMBDA (ASYMMETRIC) = 0.76652 WITH ECHEL DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.76600
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.64600 WITH ECHEL DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.64599
 KENDALL'S TAU B = 0.80946. SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.87360. SIGNIFICANCE = 0.0000
 GAMMA = 0.93422
 SOMERS'S D (ASYMMETRIC) = 0.88957 WITH ECHEL DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.88926 WITH ECHEL DEPENDENT.
 = 0.64598 WITH AGE71 DEPENDENT.
 = 0.76549 WITH AGE71 DEPENDENT.

FILE FACIAT (CREATION DATE = 12/10/75) EXERCISE SUR LES LIAISONS
 MFCHEI FICHEL DE PERMISSIVITE C R U S S T A R U L A T I O N O F
 MFCHEI MFCHEI BY AGE7J
 PAGE 1 OF 1

AGE7J	1	2	3	4	5	6	7	ROW TOTAL
COUNT	27	1	2	1	1	3	66	101
ROW FCI	26.7	1.0	2.0	1.0	1.0	3.0	65.3	17.0
COL DCI	34.6	1.2	2.0	1.7	1.2	5.5	46.2	
TOT PCI	4.5	0.2	0.3	0.2	0.2	0.5	11.1	
1	5	35	4	2	3	22	3	74
2	6.8	47.3	5.4	2.7	4.1	29.7	4.1	12.4
3	7.1	42.7	3.9	3.4	3.5	40.0	2.1	
4	0.8	5.9	0.7	0.3	0.5	3.7	0.5	
5	2	4	41	7	36	3	1	94
6	2.1	4.3	43.6	7.4	38.3	3.2	1.1	15.8
7	2.9	4.9	40.2	12.1	42.4	5.5	0.7	
8	0.3	0.7	6.9	1.2	6.1	0.5	0.2	
9	2	1	8	40	4	0	2	57
10	3.5	1.8	14.0	70.2	7.0	0.0	3.5	9.6
11	2.9	1.2	7.4	69.0	4.7	0.0	1.4	
12	0.3	0.2	1.3	6.7	0.7	0.0	0.3	
13	1	5	40	6	36	3	1	92
14	1.1	5.4	43.5	6.5	39.1	3.3	1.1	15.5
15	1.4	6.1	34.2	10.3	42.4	5.5	0.7	
16	0.2	0.8	6.7	1.0	6.1	0.5	0.2	
17	5	34	5	1	3	22	3	73
18	6.8	46.6	6.8	1.4	4.1	30.1	4.1	12.3
19	7.1	41.5	4.9	1.7	3.5	40.0	2.1	
20	0.8	5.7	0.8	0.2	0.5	3.7	0.5	
21	2	2	2	1	2	2	67	104
22	26.9	1.9	1.9	1.0	1.9	1.9	64.4	17.5
23	40.0	2.4	2.0	1.7	2.4	3.6	46.9	
24	4.7	0.2	0.3	0.2	0.3	0.3	11.3	
COLUMN TOTAL	75	13.0	17.1	58	85	55	143	595
	11.8	1.9	2.5	9.7	14.3	9.2	24.0	100.0

RAY CHI SQUARE = 995.75635 WITH 13 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.52413
 CONTINGENCY COEFFICIENT = 0.74119
 LAMBDA (ASYMMETRIC) = 0.13605 WITH 13 DEGREES OF FREEDOM. SIGNIFICANCE = 0.4737
 LAMBDA (SYMMETRIC) = 0.36585
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.37035 WITH 13 DEGREES OF FREEDOM. SIGNIFICANCE = 0.4737
 KENDALL'S TAU B = -0.00708. SIGNIFICANCE = 0.4737
 KENDALL'S TAU C = -0.00206. SIGNIFICANCE = 0.4737
 GAMMA = -0.00234
 SOMERS'S D (ASYMMETRIC) = -0.00209 WITH 13 DEGREES OF FREEDOM. SIGNIFICANCE = 0.4737
 SOMERS'S D (SYMMETRIC) = -0.00208
 ETA = 0.01471 WITH 13 DEGREES OF FREEDOM. SIGNIFICANCE = 0.4737
 PFADSPONS P = -0.00239 SIGNIFICANCE = 0.4737

DEPENDENT.

DEPENDENT.

DEPENDENT.

DEPENDENT.

DEPENDENT.

DEPENDENT.

DEPENDENT.

DEPENDENT.

FILE EXOLIAI (CREATION DATE = 12.10.79) EXERCICE SUR LES LIAISONS
 MECHEL FICHELE DE PERMISSIBILITE T A H U L A T I O N O F
 CONTROLLING FOR.. BY AGE7J *****
 SFXF *****

***** VALUE.. 1 ***** PAGE 1 OF 1

MECHEL	AGE7J	1	2	3	4	5	6	7	ROW TOTAL
1	14	0	1	1	1	0	2	33	51
	27.5	0.0	2.0	2.0	2.0	0.0	3.9	64.7	17.1
	39.9	0.0	2.0	2.0	2.0	0.0	7.1	45.2	
	4.7	0.0	0.3	0.3	0.3	0.0	0.7	11.1	
2	18	2	2	3	1	1	11	2	37
	5.4	48.6	5.4	2.7	2.7	2.7	2.7	5.4	12.4
	5.6	45.0	4.0	3.3	3.3	3.3	2.7	2.7	
	0.7	6.0	0.7	0.3	0.3	0.3	0.7	0.7	
3	1	2	20	4	1	1	1	1	47
	2.1	4.3	42.6	8.5	38.3	7.1	2.1	2.1	15.8
	2.8	5.0	40.0	13.3	43.9	3.0	1.4	1.4	
	0.3	0.7	6.7	1.3	6.0	0.3	0.3	0.3	
4	1	0	4	2	1	0	0	1	28
	3.6	0.0	14.3	7.4	7.1	0.0	0.0	0.0	9.4
	2.8	0.0	8.0	66.7	4.9	0.0	0.0	0.0	
	0.3	0.0	1.3	6.7	9.7	0.0	0.0	0.0	
5	1	2	20	3	1	2	2	0	46
	2.2	4.3	43.5	6.5	4.3	4.3	0.0	0.0	15.4
	2.8	5.0	40.0	10.0	4.3	7.1	0.0	0.0	
	0.3	0.7	6.7	1.0	0.0	0.7	0.0	0.0	
6	3	17	5	1	1	1	11	2	37
	8.1	45.9	5.0	2.7	2.7	2.7	29.7	5.4	12.4
	8.3	42.5	0.0	2.1	2.4	39.3	2.7	2.7	
	1.0	5.7	0.7	0.3	0.3	3.7	0.7	0.7	
7	14	1	1	0	1	1	1	34	52
	26.9	1.9	1.9	0.0	1.9	1.9	1.9	65.4	17.4
	39.9	2.5	2.0	0.0	2.4	3.6	46.6	46.6	
	4.7	0.3	0.3	0.0	0.3	0.3	11.4	11.4	
COLUMN TOTAL	36	40	50	30	41	28	73	298	298
	12.1	13.4	16.8	10.1	13.8	9.4	24.5	100.0	100.0

PAV CHI SQUARE = 502.56421 WITH 36 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.53017
 CONTINGENCY COEFFICIENT = 0.79231
 LAMBDA (ASYMMETRIC) = 0.33740 WITH MECHEL DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.36516
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.38084 WITH MECHEL DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.34398
 KENDALL'S TAU B = -0.00165. SIGNIFICANCE = 0.4852
 KENDALL'S TAU C = -0.00163. SIGNIFICANCE = 0.4852
 CONDITIONAL GAMMA = -0.00185
 = 0.39556 WITH AGE7J DEPENDENT.
 = 0.38717 WITH AGE7J DEPENDENT.

AGE7J

AGE7J	1	2	3	4	5	6	7	ROW TOTAL
1	13	1	1	0	1	1	33	50
	26.0	2.0	2.0	0.0	2.0	2.0	66.0	16.6
	38.2	2.4	1.9	0.0	2.3	3.7	47.1	
	4.4	0.3	0.3	0.0	0.3	0.3	11.1	
2	3	17	2	1	2	11	1	37
	8.1	45.9	5.4	2.7	5.4	29.7	2.7	12.5
	8.8	40.5	3.8	3.6	4.5	40.7	1.4	
	1.0	5.7	0.7	0.3	0.7	3.7	0.3	
3	1	2	21	3	18	2	0	47
	2.1	4.3	44.7	6.4	38.3	4.3	0.0	15.8
	2.9	4.8	40.4	10.7	40.9	7.4	0.0	
	0.3	0.7	7.1	1.0	6.1	0.7	0.0	
4	1	1	4	20	2	0	1	29
	3.4	3.4	13.8	69.0	6.9	0.0	3.4	9.8
	2.9	2.4	7.7	71.4	4.5	0.0	1.4	
	0.3	0.3	1.3	6.7	0.7	0.0	0.3	
5	0	3	20	3	18	1	1	46
	0.0	6.5	43.5	6.5	39.1	2.2	2.2	15.5
	0.0	7.1	38.5	10.7	40.9	3.7	1.4	
	0.0	1.0	6.7	1.0	6.1	0.3	0.3	
6	2	17	3	0	2	11	1	36
	5.6	47.2	8.3	0.0	5.6	30.6	2.8	12.1
	5.9	40.5	5.8	0.0	4.5	40.7	1.4	
	0.7	5.7	1.0	0.0	0.7	3.7	0.3	
7	14	1	1	1	1	1	33	52
	26.9	1.9	1.9	1.9	1.9	1.9	63.5	17.5
	41.2	2.4	1.9	3.6	2.3	3.7	47.1	
	4.7	0.3	0.3	0.3	0.3	0.3	11.1	
COLUMN TOTAL	34	42	52	28	44	27	70	297
	11.4	14.1	17.5	9.4	14.8	9.1	23.6	100.0

PAW CHI SQUARE = 496.15674 WITH 36 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 GAMMA'S Y = 0.52766
 COEFFICIENCY COEFFICIENT = 0.79092
 LAMBDA (ASYMMETRIC) = 0.33469 WITH NCFEL DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.36653
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.37336 WITH NCFEL DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.37669
 KENDALL'S TAU B = -0.00241. SIGNIFICANCE = 0.4785
 KENDALL'S TAU C = -0.00238. SIGNIFICANCE = 0.4785
 CONDITIONAL GAMMA = -0.00270
 SOMERS'S D (ASYMMETRIC) = -0.00242 WITH NCFEL DEPENDENT.
 SOMERS'S D (SYMMETRIC) = -0.00241
 ETA = 0.00685 WITH NCFEL DEPENDENT.
 PEARSON'S R = -0.00209 SIGNIFICANCE = 0.4657
 SUMMARY GAMMAS FOR CROSS-tabULATION OF NCFEL BY AGE7J = 0.32375 WITH AGE7J DEPENDENT.
 = -0.00239 WITH AGE7J DEPENDENT.
 = 0.38008 WITH AGE7J DEPENDENT.
 ZERO-ORDER GAMMA = -0.00234
 FIRST-ORDER CAPITAL GAMMA = -0.00228

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

12/10/79 PAGE 14

FILE EXO1IAI (CREATION DATE = 12/10/79) EXERCICE SUR LES LIAISONS
 * * * * * C R O S T A B U L A T I O N O F * * * * *
 PFVFIJJA * * * * * BY AGF3A * * * * * PAGE 1 OF 1

AGE3A	1.1	2.1	3.1	ROW TOTAL
COUNT				
POI PCI	130	7	33	170
COL PCI	76.5	4.1	19.4	28.6
TOT PCI	65.7	2.4	29.7	
PFVENIJA	21.8	1.2	5.5	
5000.	54	42	62	158
	34.2	26.0	39.2	26.6
	27.3	14.7	55.9	
	9.1	7.1	10.4	
10000.	14	237	16	267
	5.2	88.8	6.0	44.9
	7.1	82.9	14.4	
	2.4	39.8	2.7	
COLUMN TOTAL	198	286	111	595
	33.3	48.1	18.7	100.0

CHI SQUARE = 392.77930 WITH 4 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.57451
 CONTINGENCY COEFFICIENT = 0.63059
 LAMBDA (ASYMMETRIC) = 0.49390 WITH REVENUA DEPENDENT.
 LAMBDA (ASYMMETRIC) = 0.47881
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.33982 WITH REVENUA DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.34598
 KENDALL'S TAU B = 0.29555 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.28145 SIGNIFICANCE = 0.0000
 GAMMA = 0.37165
 SOMERS'S D (ASYMMETRIC) = 0.30097 WITH REVENUA DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.29551
 ETA = 0.76318 WITH REVENUA DEPENDENT.
 PEARSON'S R = 0.26015 SIGNIFICANCE = 0.0000
 = 0.38117 WITH AGE3A DEPENDENT.
 = 0.29023 WITH AGE3A DEPENDENT.
 = 0.35237 WITH AGE3A DEPENDENT.
 = 0.46278 WITH AGE3A DEPENDENT.

FILE EXCELTAI (CREATION DATE = 12/10/79) FAEMCICE SUR LES LIAISONS
 REVENUE C P U S T A P U L A T I O N O F
 BY AGE3B

REVENUE	AGE3B	1.1	2.1	3.1	ROW TOTAL
3000.		150	7	63	270
		68.2	3.2	28.6	37.0
		75.8	2.4	56.8	
		25.2	1.2	10.6	
5000.		34	21	32	87
		39.1	24.1	36.8	14.6
		17.2	7.3	28.8	
		5.7	3.5	5.4	
10000.		14	258	16	288
		4.9	49.6	5.6	48.4
		7.1	90.2	14.4	
		2.4	43.4	2.7	
COLUMN TOTAL		198	286	111	595
TOTAL		33.3	48.1	18.7	100.0

CHI SQUARE = 410.71704 WITH 4 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CAMPBELL'S V = 0.58749
 CONTINGENCY COEFFICIENT = 0.63905
 LAMBDA (ASYMMETRIC) = 0.59609 WITH REVENUE DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.55032
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.40768 WITH REVENUE DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.40140
 KENDALL'S TAU B = 0.26713 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.24662 SIGNIFICANCE = 0.0000
 GAMMA = 0.32710
 SOMERS'S D (ASYMMETRIC) = 0.26373 WITH REVENUE DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.26711
 FTA = 0.41954 WITH REVENUE DEPENDENT.
 PEARSON'S R = 0.23982 SIGNIFICANCE = 0.0000
 = 0.27090 WITH AGE3B DEPENDENT.
 = 0.27058 WITH AGE3B DEPENDENT.
 = 0.39531 WITH AGE3B DEPENDENT.

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

12/10/79 PAGE 17

FILE FXT01JAI (CREATION DATE = 12/10/79) EXERCICE SUR LES LIAISONS
 * * * * * C R O S S T A B U L A T I O N O F * * * * *
 * * * * * R E V E N R * * * * * B Y * * * * * A G E 3 B * * * * *

REVENR	AGE3B				ROW TOTAL
	1.1	2.1	3.1		
7000.	150	7	63	220	
	68.2	3.2	28.6	37.0	
	75.8	2.4	56.8		
	25.2	1.2	10.6		
11000.	34	21	32	87	
	39.1	24.1	36.8	14.6	
	17.2	7.3	28.8		
	5.7	3.5	5.4		
21000.	14	258	16	288	
	4.9	89.6	5.6	48.4	
	7.1	90.2	14.4		
	2.4	43.4	2.7		
COLUMN TOTAL	198	286	111	595	
	33.3	48.1	18.7	100.0	

CHI SQUARE = 410.71704 WITH 4 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.58749
 CONTINGENCY COEFFICIENT = 0.63905
 LAMBDA (ASYMMETRIC) = 0.59609 WITH REVENR DEPENDENT. = 0.50485 WITH AGE3B DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.55032
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.40768 WITH REVENR DEPENDENT. = 0.39531 WITH AGE3B DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.40140
 KENDALL'S TAU B = 0.26713 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.24562 SIGNIFICANCE = 0.0000
 GAMMA = 0.32710
 SOMERS'S D (ASYMMETRIC) = 0.26373 WITH REVENR DEPENDENT. = 0.27058 WITH AGE3B DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.26711
 ETA = 0.81958 WITH REVENR DEPENDENT. = 0.27090 WITH AGE3B DEPENDENT.
 PFARSON'S R = 0.23982 SIGNIFICANCE = 0.0000

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

12/10/79 PAGE 19

FILE EX01141 (CREATION DATE = 12/10/79) EXERCISE SUP LES LIATIONS
 REVENUS C F C S T A R U L A T I O N O F * * * * *
 BY AGE3C * * * * * PAGE 1 OF 1

REVENUS	AGE3C	COUNT	ROW PCT	COL PCT	TOT PCT	ROW TOTAL
1.	1.	14	16	2.1	3.1	267
	2.	5.2	6.0			44.9
	3.	7.1	14.4			
2.	1.	2.4	2.7			
	2.	54	62			158
	3.	34.2	39.2			26.6
3.	1.	27.3	55.5			26.6
	2.	9.1	10.4			
	3.	130	33			170
TOTAL		198	111			595
TOTAL		33.3	18.7			100.0

CHI SQUARE = 392.77930 WITH 4 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.57451
 CONTINGENCY COEFFICIENT = 0.63059
 LAMBDA (ASYMMETRIC) = 0.49390 WITH REVENUS DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.47881
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.33982 WITH REVENUS DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.37598
 KENDALL'S TAU B = -0.69925 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = -0.66588 SIGNIFICANCE = 0.0000
 GAMMA = -0.87928
 SOMERS'S D (ASYMMETRIC) = -0.71208 WITH REVENUS DEPENDENT.
 SOMERS'S D (SYMMETRIC) = -0.69914
 FTA = 0.75441 WITH REVENUS DEPENDENT.
 PEARSON'S R = -0.72464 SIGNIFICANCE = 0.0000

= 0.46278 WITH AGE3C DEPENDENT.
 = 0.35237 WITH AGE3C DEPENDENT.

= -0.68666 WITH AGE3C DEPENDENT.
 = 0.74746 WITH AGE3C DEPENDENT.

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

12/10/79 PAGE 23

FILE EX01A1 (CREATION DATE = 12/10/79) EXERCICE SUR LES LIAISONS
 * * * * * C P O S T A R U L A T I O N O F * * * * *
 * * * * * R E V E N U * * * * * P Y A G E 3 0 * * * * *

PAGE 1 OF 1

AGE30

	1.1	2.1	3.1	ROW TOTAL
COUNT				
ROW PCT				
COL PCT				
TOT PCT	1.1	2.1	3.1	
REVENUE				
3000.	176	0	0	170
	100.0	0.0	0.0	26.6
	100.0	0.0	0.0	
	28.6	0.0	0.0	
5000.	0	0	158	158
	0.0	0.0	100.0	25.6
	0.0	0.0	100.0	
	0.0	0.0	25.6	
10000.	0	267	0	267
	0.0	100.0	0.0	44.9
	0.0	100.0	0.0	
	0.0	44.9	0.0	
COLUMN	170	267	158	595
TOTAL	26.6	44.9	26.6	100.0

CHI SQUARE = 1185.99878 WITH 4 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 1.00000
 CONTINGENCY COEFFICIENT = 0.81650
 LAMBDA (ASYMMETRIC) = 1.00000 WITH REVENUE DEPENDENT.
 LAMBDA (SYMMETRIC) = 1.00000
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 1.00000 WITH REVENUE DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 1.00000
 PEARSON'S TAU H = 0.26271 SIGNIFICANCE = 0.0000
 PEARSON'S TAU C = 0.25476 SIGNIFICANCE = 0.0000
 GAMMA = 0.26271
 SOMER'S D (ASYMMETRIC) = 0.26271 WITH REVENUE DEPENDENT.
 SOMER'S D (SYMMETRIC) = 0.26271
 ETA = 1.00000 WITH REVENUE DEPENDENT.
 PEARSON'S R = 0.26364 SIGNIFICANCE = 0.0000

DEPENDENT.

DEPENDENT.

DEPENDENT.

DEPENDENT.

FILE FAXIAL (CREATION DATE = 12/10/79) EXERCICE SUR LES LIAISONS

REVENUE C H O S T A B U I A T I O N O F * * * * * PAGE 1 OF 1

AGE3E

REVENUE	1.1	2.1	3.1	ROW TOTAL
COUNT	170	0	0	170
POW FCT	100.0	0.0	0.0	28.6
COL FCT	100.0	0.0	0.0	
TOT FCT	28.6	0.0	0.0	

	0	158	0	158
	0.0	100.0	0.0	26.6
	0.0	100.0	0.0	
	0.0	26.6	0.0	

	0	0	267	267
	0.0	0.0	100.0	44.9
	0.0	0.0	100.0	
	0.0	0.0	44.9	

COLUMN TOTAL	170	158	267	595
TOTAL	28.6	26.6	44.9	100.0

CHI SQUARE = 1189.99478 WITH 4 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 1.00000
 CONTINGENCY COEFFICIENT = 0.81650
 LAMBDA (ASYMMETRIC) = 1.00000 WITH REVENUE DEPENDENT.
 LAMBDA (SYMMETRIC) = 1.00000
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 1.00000 WITH REVENUE DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 1.00000
 KENDALL'S TAU B = 1.00000 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.96973 SIGNIFICANCE = 0.0000
 GAMMA = 1.00000
 SOMER'S D (ASYMMETRIC) = 1.00000 WITH REVENUE DEPENDENT.
 SOMER'S D (SYMMETRIC) = 1.00000
 ETA = 1.00000 WITH REVENUE DEPENDENT.
 PFAPSON'S P = 0.97711 SIGNIFICANCE = 0.0000

DEPENDENT.

= 1.00000 WITH AGE3E

DEPENDENT.

DEPENDENT.

= 1.00000 WITH AGE3E

DEPENDENT.

DEPENDENT.

= 1.00000 WITH AGE3E

DEPENDENT.

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

12/10/79 PAGE 25

FILE FATAL (CREATION DATE = 12/10/79) EXERCICE SUP LES LIAISONS
REVENUE C F U S T A H U L A T I O N O F
BY AGF4F

PAGE 1 OF 1

AGE4F

REVENUE	1.1	2.1	3.1	4.1	ROW TOTAL
COUNT	170	0	0	0	170
POW PCT	100.0	0.0	0.0	0.0	28.6
COL PCT	100.0	0.0	0.0	0.0	
TOT PCT	28.6	0.0	0.0	0.0	

5000.	0	0	1.6	0	158
POW PCT	0.0	0.0	100.0	0.0	26.6
COL PCT	0.0	0.0	100.0	0.0	
TOT PCT	0.0	0.0	26.6	0.0	

10000.	0	237	0	30	267
POW PCT	0.0	48.8	0.0	11.2	44.9
COL PCT	0.0	100.0	0.0	100.0	
TOT PCT	0.0	35.8	0.0	5.0	

COLUMN TOTAL	170	237	1.8	30	595
	28.6	39.4	26.6	5.0	100.0

CHI SQUARE = 1189.99854 WITH 6 DEGRES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 1.00000
 CONTINGENCY COEFFICIENT = 0.81650
 LAMBDA (ASYMMETRIC) = 1.00000 WITH REVENUE DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.95627
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 1.00000 WITH REVENUE DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.9134
 KENDALL'S TAU B = 0.33530 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.33509 SIGNIFICANCE = 0.0000
 GAMMA = 0.34556
 SOMER'S D (ASYMMETRIC) = 0.32534 WITH REVENUE DEPENDENT.
 SOMER'S D (SYMMETRIC) = 0.33514
 ETA = 1.00000 WITH REVENUE DEPENDENT.
 PEARSON'S R = 0.35207 SIGNIFICANCE = 0.0000

= 0.91620 WITH AGE4F DEPENDENT.
 = 0.87151 WITH AGE4F DEPENDENT.
 = 0.34556 WITH AGE4F DEPENDENT.
 = 0.87187 WITH AGE4F DEPENDENT.

FILE EXCELTAJ (CREATION DATE = 12/10/79) EXERCICE SUR LES LIAISONS
 **
 REVENUG **
 BY AGE4G **
 PAGE 1 OF 1

	AGE4G				ROW TOTAL
	1.1	2.1	3.1	4.1	
COUNT					
ROW PCT					
COL PCT					
TOT PCT					
REVENUG					
3000.	170	0	0	0	170
	100.0	0.0	0.0	0.0	28.6
	100.0	0.0	0.0	0.0	
	28.6	0.0	0.0	0.0	
5000.	0	158	0	0	158
	0.0	100.0	0.0	0.0	26.6
	0.0	100.0	0.0	0.0	
	0.0	26.6	0.0	0.0	
10000.	0	0	237	30	267
	0.0	0.0	88.8	11.2	44.9
	0.0	0.0	100.0	100.0	
	0.0	0.0	39.8	5.0	
COLUMN TOTAL	170	158	237	30	595
TOTAL	28.6	26.6	39.8	5.0	100.0

CHI SQUARE = 1189.99854 WITH 6 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CREMER'S V = 1.00000
 CONTINGENCY COEFFICIENT = 0.61650
 LAMBDA (ASYMMETRIC) = 1.00000 WITH REVENUG DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.95627
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 1.00000 WITH REVENUG DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.93134
 KENDALL'S TAU B = 0.97031 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.96473 SIGNIFICANCE = 0.0000
 GAMMA = 1.00000
 SOMERS'S D (ASYMMETRIC) = 0.94150 WITH REVENUG DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.56987
 FTA = 1.00000 WITH REVENUG DEPENDENT.
 PEARSON'S R = 0.95609 SIGNIFICANCE = 0.0000
 = 0.91620 WITH AGE4G DEPENDENT.
 = 0.87151 WITH AGE4G DEPENDENT.

FILE EX01A1 (CREATION DATE = 12/10/79) EXECUTICE SUR LES LIAISONS
 APPREUJ BY TEMPC7 TEMPERATURE EN DEGRE CELSIUS
 PAGE 1 OF 1

APPREUJ	10.1	12.1	14.1	16.1	18.1	20.1	22.1	ROW TOTAL
1.	55	3	3	1	1	0	1	64
	85.9	4.7	4.7	1.6	1.6	0.0	1.6	10.8
	78.6	3.7	2.4	1.7	1.2	0.0	0.7	
	9.2	0.5	0.5	0.2	0.2	0.0	0.2	
2.	10	64	4	2	0	1	1	91
	11.0	74.7	4.9	2.2	0.0	1.1	1.1	15.3
	14.3	82.4	8.8	3.4	0.0	1.8	0.7	
	1.7	11.4	1.5	0.3	0.0	0.2	0.2	
3.	2	5	40	6	1	0	1	99
	2.0	9.1	80.7	6.1	1.0	0.0	1.0	16.6
	2.9	11.0	78.4	10.3	1.2	0.0	0.7	
	0.3	1.5	13.4	1.0	0.2	0.0	0.2	
4.	2	1	8	40	4	0	2	57
	3.5	1.8	14.0	70.2	7.0	0.0	3.5	9.6
	2.9	1.2	7.8	69.0	4.7	0.0	1.4	
	0.3	0.2	1.3	6.7	0.7	0.0	0.3	
5.	1	0	1	7	71	6	1	87
	1.1	0.0	1.1	8.0	81.6	6.9	1.1	14.6
	1.4	0.0	1.0	12.1	83.5	10.9	0.7	
	0.2	0.0	0.2	1.2	11.9	1.0	0.2	
6.	0	1	0	1	6	43	5	56
	0.0	1.8	0.0	1.8	10.7	76.8	8.9	9.4
	0.0	1.2	0.0	1.7	7.1	78.2	3.5	
	0.0	0.2	0.0	0.2	1.0	7.2	0.8	
7.	0	0	1	1	2	5	132	141
	0.0	0.0	0.7	0.7	1.4	3.5	93.6	23.7
	0.0	0.0	1.0	1.7	2.4	9.1	92.3	
	0.0	0.0	0.2	0.2	0.3	0.8	27.2	
COLUMN TOTAL	70	82	102	58	85	55	143	595
	11.6	13.8	17.1	9.7	14.3	9.2	24.0	100.0

CHI SQUARE = 2173.83252 WITH 36 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.78033
 CONTINGENCY COEFFICIENT = 0.64606
 LAMBDA (ASYMMETRIC) = 0.76652 WITH AMPUP DOWN DEPENDENT. = 0.76549 WITH TEMPC7 DOWN DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.76600
 UNCEP TAINTY COEFFICIENT (ASYMMETRIC) = 0.04600 WITH AMPUP DOWN DEPENDENT.
 UNCEP TAINTY COEFFICIENT (SYMMETRIC) = 0.64599
 KENDALL'S TAU B = 0.64946 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.67340 SIGNIFICANCE = 0.0000
 GAMMA = 0.93822
 SOMER'S D (ASYMMETRIC) = 0.68957 WITH AMPUP DOWN DEPENDENT. = 0.68934 WITH TEMPC7 DOWN DEPENDENT.

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

FILE FXP0161 (CREATION DATE = 12/10/79) EXERCICE SUR LES LIANSONS
 * * * * * C R O S S T A B U L A T I O N O F * * * * *
 * * * * * A M P L I T U D E A U T E M P C 7 * * * * * T E M P E R A T U R E E N D E G R E C F L S I U S * * * * *
 * * * * * P A G E 1 O F 1

AMPLITUD	TEMPC7	10.1	12.1	14.1	15.1	18.1	20.1	22.1	ROW TOTAL
1.		55	3	3	1	1	0	1	64
		85.9	4.7	4.7	1.6	1.6	0.0	1.6	10.8
		78.6	3.7	2.9	1.7	1.2	0.0	0.7	
		4.2	0.5	0.5	0.2	0.2	0.0	0.2	
2.		10	68	9	2	0	1	1	91
		11.0	74.7	9.9	2.2	0.0	1.1	1.1	15.3
		14.3	82.9	4.8	3.4	0.0	1.8	0.7	
		1.7	11.4	1.5	0.3	0.0	0.2	0.2	
3.		2	9	40	6	1	0	1	99
		2.0	9.1	80.8	6.1	1.0	0.0	1.0	16.6
		2.9	11.0	78.4	10.3	1.2	0.0	0.7	
		0.3	1.5	13.4	1.0	0.2	0.0	0.2	
5.		2	1	8	40	4	0	2	57
		3.5	1.8	14.0	70.2	7.0	0.0	3.5	9.6
		2.9	1.2	7.8	49.0	4.7	0.0	1.4	
		0.3	0.2	1.3	6.7	0.7	0.0	0.3	
7.		1	0	1	7	71	6	1	87
		1.1	0.0	1.1	8.0	81.6	6.9	1.1	14.6
		1.4	0.0	1.0	12.1	83.5	10.9	0.7	
		0.2	0.0	0.2	3.2	11.9	1.0	0.2	
10.		0	1	0	1	6	43	5	56
		0.0	1.8	0.0	1.6	10.7	76.8	8.9	9.4
		0.0	1.2	0.0	1.7	7.1	78.2	3.5	
		0.0	0.2	0.0	0.2	1.0	7.2	0.8	
15.		0	0	1	1	2	5	132	141
		0.0	0.0	0.7	0.7	1.4	3.5	93.6	23.7
		0.0	0.0	1.0	1.7	2.4	9.1	92.3	
		0.0	0.0	0.2	0.2	0.3	0.8	22.2	
COLUMN TOTAL		70	82	162	58	85	55	143	595
		11.8	13.8	17.1	9.7	14.3	9.2	24.0	100.0

CHI SQUARE = 2173.83252 WITH 36 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.78033
 CONTINGENCY COEFFICIENT = 0.684006
 LAMBDA (ASYMPTOTIC) = 0.76652 WITH AMPLITUDE/DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.76600
 UNCERTAINTY COEFFICIENT (ASYMPTOTIC) = 0.64600 WITH AMPLITUDE/DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.64599
 KENDALL'S TAU B = 0.60966 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.67340 SIGNIFICANCE = 0.0000
 GAMMA = 0.69322
 SORBERS D (ASYMPTOTIC) = 0.88497 WITH AMPLITUDE/DEPENDENT.
 = 0.88934 WITH TEMPC7 DEPENDENT.
 = 0.64598 WITH TEMPC7 DEPENDENT.
 = 0.76549 WITH TEMPC7 DEPENDENT.

FILE FORMAL (CREATION DATE = 12/10/79) EXERCICE SUR LES LIAISONS
 APPREH4 APPREH4 (C H O S T A H U L A T I O N O F
 APPREH4 APPREH4 AU TRAVAIL BY TEMPC7 TEMPERATURE EN DEGRE CELSIUS
 PAGE 1 OF 1

APPREH4	COURT	ROW PCT	COL PCT	TOT PCT	10.1	12.1	14.1	16.1	18.1	20.1	22.1	ROW TOTAL
2.00	67	31.5	42	9	2	1	1	1	1	1	3	254
4.50	95.7	47.6	40.2	15.5	2.4	1.8	1.8	0.4	0.3	0.2	2.1	42.7
6.00	11.3	13.4	15.5	1.5	0.3	0.2	0.2	0.2	0.3	0.2	0.5	144
7.00	3	0.7	6.3	47	75	6	6	4.2	10.9	2.1	2.1	24.2
	4.3	1.2	9.4	41.0	88.2	10.9	10.9	10.9	10.9	10.9	10.9	56
	0.5	0.2	1.5	7.9	12.6	1.0	1.0	1.0	1.0	1.0	0.5	9.4
	0	1	0	1	6	43	43	43	43	43	5	141
	0.0	1.8	0.0	1.8	10.7	76.8	76.8	76.8	76.8	76.8	8.9	23.7
	0.0	1.2	0.0	1.2	7.1	78.2	78.2	78.2	78.2	78.2	3.5	56
	0.0	0.2	0.0	0.2	1.0	7.2	7.2	7.2	7.2	7.2	0.8	9.4
	0	0	1	1	2	5	5	5	5	5	132	141
	0.0	0.0	0.7	0.7	1.4	3.5	3.5	3.5	3.5	3.5	93.6	23.7
	0.0	0.0	1.0	1.0	2.4	9.1	9.1	9.1	9.1	9.1	92.3	141
	0.0	0.0	0.2	0.2	0.3	0.8	0.8	0.8	0.8	0.8	22.2	141
COLUMN TOTAL	70	82	162	58	85	55	143	143	143	143	143	595
	11.8	13.8	17.1	9.7	14.3	9.2	24.0	24.0	24.0	24.0	24.0	100.0

CHI SQUARE = 1253.90820 WITH 14 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.83813
 CONTINGENCY COEFFICIENT = 0.82352
 LAMBDA (ASYMMETRIC) = 0.82698 WITH APPREH4 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.60656
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.70286 WITH APPREH4 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.56417
 KENDALL'S TAU B = 0.83543 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.85139 SIGNIFICANCE = 0.0000
 GAMMA = 0.85605
 SOMERS'S D (ASYMMETRIC) = 0.75875 WITH APPREH4 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.83157
 FTA = 0.93104 WITH APPREH4 DEPENDENT.
 FTAUSS = 0.90390 SIGNIFICANCE = 0.0000
 = 0.90401 WITH TEMPC7 DEPENDENT.
 = 0.44027 WITH TEMPC7 DEPENDENT.
 = 0.47120 WITH TEMPC7 DEPENDENT.

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

FILE EXH1AT (CORRELATION DATE = 12/10/79) EXERCICE SUR LES LIAISONS
 ** * * * * * ** C R O S T A B U L A T I O N O F * * * * * ** * * * * * ** * * * * * **
 ** * * * * * ** A R R E T U R A U I N A V A I L ** * * * * * ** * * * * * ** * * * * * ** * * * * * **
 ** * * * * * ** T E M P C 4 1 T E M P E R A T U R E E N D E G R E C E L S I U S ** * * * * * ** * * * * * ** * * * * * **
 ** * * * * * ** P A G E 1 O F 1

ARRIURE7	TEMPC41	12.1	17.1	20.1	22.1	ROW TOTAL
1.	61	2	0	1	1	64
	95.3	3.1	0.0	1.6	1.6	10.8
	24.0	1.4	0.0	0.7	0.7	
	10.3	0.3	0.0	0.2	0.2	
2.	87	2	1	1	1	91
	95.6	2.2	1.1	1.1	1.1	15.3
	34.3	1.4	1.8	0.7	0.7	
	14.6	0.3	0.2	0.2	0.2	
3.	91	7	0	1	1	99
	41.9	7.1	0.0	1.0	1.0	16.6
	35.4	4.9	0.0	0.7	0.7	
	15.3	1.2	0.0	0.2	0.2	
6.	11	44	0	2	2	57
	14.3	77.2	0.0	3.5	3.5	9.6
	4.3	30.8	0.0	1.4	1.4	
	1.8	7.4	0.0	0.3	0.3	
7.	2	76	6	1	1	87
	2.3	49.7	6.9	1.1	1.1	14.6
	0.3	54.5	10.9	0.7	0.7	
	0.3	13.1	1.0	0.2	0.2	
10.	1	7	43	5	5	56
	1.8	12.5	76.8	8.9	8.9	9.4
	0.4	4.9	78.2	3.5	3.5	
	0.2	1.2	7.2	0.6	0.6	
15.	1	3	5	132	141	141
	0.7	2.1	3.5	93.6	93.6	23.7
	0.4	2.1	9.1	92.3	92.3	
	0.2	0.5	0.6	22.2	22.2	
TOTAL	254	143	95	143	143	595
	42.7	24.0	9.2	24.0	24.0	100.0

CHI SQUARE = 1256.04956 WITH 14 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CONTINGENCY COEFFICIENT = 0.43685
 LAMBDA (ASYMMETRIC) = 0.42375
 LAMBDA (SYMMETRIC) = 0.41066
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.47263 WITH ARRURE7 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.51616
 KENDALL'S TAU B = 0.83357 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.84330 SIGNIFICANCE = 0.0000
 GAMMA = 0.95241
 SOMER'S D (ASYMMETRIC) = 0.91826 WITH ARRURE7 DEPENDENT.
 SOMER'S D (SYMMETRIC) = 0.75670 WITH TEMP41 DEPENDENT.
 LAMBDA (ASYMMETRIC) = 0.82698 WITH TEMP41 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.70583 WITH TEMP41 DEPENDENT.

FILE EXCELTAJ (CREATION DATE = 12/10/79) EXERCICE SUR LES LIAISONS
 ARDEUR4 ARDEUR AU TRAVAIL C R O S S T A B U L A T I O N O F * * * * *
 BY TEMPC41 TEMPERATURE EN DEGRE CELSIUS * * * * *
 PAGE 1 OF 1

		TEMPC41				ROW
COUNT I		12.1	17.1	20.1	22.1	TOTAL
ARDEUR4	ROW PCT I	239	11	1	3	254
	COL PCT I	94.1	4.3	0.4	1.2	42.7
	TOT PCT I	94.1	7.7	1.8	2.1	
		40.2	1.8	0.2	0.5	
		13	172	6	3	144
		9.0	84.7	4.2	2.1	24.2
		5.1	85.3	10.9	2.1	
		2.2	20.5	1.0	0.5	
		1	7	43	5	56
		1.8	12.5	76.8	8.9	9.4
		0.4	4.9	78.2	3.5	
		0.2	1.2	7.2	0.8	
		1	3	5	132	141
		0.7	2.1	3.5	93.6	23.7
		0.4	2.1	9.1	92.3	
		0.2	0.5	0.8	22.2	
	COLUMN TOTAL	254	143	15	143	595
		42.7	24.0	9.2	24.0	100.0

CHI SQUARE = 1248.85962 WITH 9 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.83645
 CONTINGENCY COEFFICIENT = 0.82299
 LAMBDA (ASYMMETRIC) = 0.82698 WITH ARDEUR4 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.82698
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.68915 WITH ARDEUR4 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.68958
 KENDALL'S TAU B = 0.90572 SIGNIFICANCE = 0.0000
 FORDAL'S TAU C = 0.83802 SIGNIFICANCE = 0.0000
 GAMMA = 0.97145
 SOMERS'S D (ASYMMETRIC) = 0.90604 WITH ARDEUR4 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.90572
 ETA = 0.92930 WITH ARDEUR4 DEPENDENT.
 PEARSON'S R = 0.92924 SIGNIFICANCE = 0.0000
 = 0.82698 WITH TEMPC41 DEPENDENT.
 = 0.69000 WITH TEMPC41 DEPENDENT.

FILE EXONIAL (CREATION DATE = 12/10/79) EXERCICE SUP LES LIAISONS
 ARDEUR4 ARDEUR AU TRAVAIL C H O S T A H U L A T I O N O F * * * * *
 TEMPERATURE EN DEGRE CELSIUS * * * * *
 TEMPC42 * * * * *
 PAGE 1 OF 1

TEMPC42

COUNT	ROW PCT	COL PCT	TOT PCT	10.1	12.1	15.1	20.1	ROW TOTAL
2.00	67	80	1.1	6	254			
	26.4	31.5	39.8	2.4	42.7			
	95.7	97.6	63.1	2.1				
	11.3	13.4	17.0	1.0				
4.50	3	1	5.6	84	144			
	2.1	0.7	38.9	58.3	24.2			
	4.3	1.2	35.0	29.7				
	0.5	0.2	9.4	14.1				
6.00	0	1	1	54	56			
	0.0	1.8	1.8	96.4	9.4			
	0.0	1.2	0.6	19.1				
	0.0	0.2	0.2	9.1				
7.00	0	0	2	139	141			
	0.0	0.0	1.4	98.6	23.7			
	0.0	0.0	1.3	49.1				
	0.0	0.0	0.3	23.4				
COLUMN TOTAL	70	82	14.0	283	595			
TOTAL	11.9	13.8	26.9	47.6	100.0			

CHI SQUARE = 470.16772 WITH 9 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CONTINGENCY COEFFICIENT = 0.51322
 LAMBDA (ASYMMETRIC) = 0.66438
 LAMBDA (ASYMMETRIC) = 0.39003 WITH ARDEUR4 DEPENDENT.
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.34916
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.40329
 KENDALL'S TAU B = 0.72763 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.66096 SIGNIFICANCE = 0.0000
 GAMMA = 0.96420
 SOMERS'S D (ASYMMETRIC) = 0.74140 WITH ARDEUR4 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.72750
 ETA = 0.84714 WITH ARDEUR4 DEPENDENT.
 PEARSON'S F = 0.83368 WITH TEMPC42 DEPENDENT.
 PEARSON'S F = 0.83368 WITH TEMPC42 DEPENDENT.
 PEARSON'S F = 0.83368 WITH TEMPC42 DEPENDENT.
 PEARSON'S F = 0.83368 WITH TEMPC42 DEPENDENT.

EXERCICE SUR LES LIAISONS
 CHRONOLOGIQUE AU TRAVAIL
 TEMPERATURE EN DEGRE CELSIUS
 BY TEMPC3

ARDEUR3	TEMPC3	ROW TOTAL
2.00	12.1	21.1
4.50	17.1	254
6.50	21.1	42.7
8.50	25.1	144
10.50	29.1	24.2
12.50	33.1	197
14.50	37.1	33.1
16.50	41.1	198
18.50	45.1	143
20.50	49.1	24.0
TOTAL		595
		100.0

CHI SQUARE = 0.9270950 WITH 4 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.15613
 CONTINGENCY COEFFICIENT = 0.77463
 LAMBDA (ASYMPTOTIC) = 0.85630 WITH ARDEUR3 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.85630
 UNCERTAINTY COEFFICIENT (ASYMPTOTIC) = 0.09752 WITH ARDEUR3 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.09770
 PEARSON'S TAU B = 0.90512 SIGNIFICANCE = 0.0000
 PEARSON'S TAU C = 0.88170 SIGNIFICANCE = 0.0000
 GAMMA = 0.97920
 SOMERS'S D (ASYMPTOTIC) = 0.90533 WITH ARDEUR3 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.90512
 ETA = 0.92363 WITH ARDEUR3 DEPENDENT.
 PEARSON'S R = 0.92362 SIGNIFICANCE = 0.0000

TEMPC3 = 0.69787 WITH TEMPC3 DEPENDENT.
 TEMPC3 = 0.90491 WITH TEMPC3 DEPENDENT.
 TEMPC3 = 0.92407 WITH TEMPC3 DEPENDENT.

ROW	T1	250.I	750.I	1250.I	1750.I	2250.I	ROW TOTAL
5.	TOT PCT I	10 I	0 I	0 I	0 I	0 I	10
	I	100.0 I	0.0 I	0.0 I	0.0 I	0.0 I	20.0
	I	45.5 I	0.0 I	0.0 I	0.0 I	0.0 I	
	I	20.0 I	0.0 I	0.0 I	0.0 I	0.0 I	
15.	TOT PCT I	10 I	0 I	0 I	0 I	0 I	10
	I	100.0 I	0.0 I	0.0 I	0.0 I	0.0 I	20.0
	I	45.5 I	0.0 I	0.0 I	0.0 I	0.0 I	
	I	20.0 I	0.0 I	0.0 I	0.0 I	0.0 I	
25.	TOT PCT I	2 I	8 I	0 I	0 I	0 I	10
	I	20.0 I	80.0 I	0.0 I	0.0 I	0.0 I	20.0
	I	9.1 I	88.9 I	0.0 I	0.0 I	0.0 I	
	I	4.0 I	16.0 I	0.0 I	0.0 I	0.0 I	
35.	TOT PCT I	0 I	1 I	7 I	2 I	0 I	10
	I	0.0 I	10.0 I	70.0 I	20.0 I	0.0 I	20.0
	I	0.0 I	11.1 I	100.0 I	33.3 I	0.0 I	
	I	0.0 I	2.0 I	14.0 I	4.0 I	0.0 I	
45.	TOT PCT I	0 I	0 I	0 I	4 I	6 I	10
	I	0.0 I	0.0 I	0.0 I	40.0 I	60.0 I	20.0
	I	0.0 I	0.0 I	0.0 I	66.7 I	100.0 I	
	I	0.0 I	0.0 I	0.0 I	8.0 I	12.0 I	
	COLUMN TOTAL	22	9	7	6	6	50
	TOTAL	44.0	18.0	14.0	12.0	12.0	100.0

CHI SQUARE = 114.14141 WITH 16 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.75545
 CONTINGENCY COEFFICIENT = 0.83390
 LAMBDA (ASYMMETRIC) = 0.62500 WITH T1 DEPENDENT.
 LAMBDA (ASYMMETRIC) = 0.64706 DEPENDENT. = 0.67857 WITH T2 DEPENDENT.
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.65797 WITH T1 DEPENDENT.
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.69136 DEPENDENT. = 0.72831 WITH T2 DEPENDENT.
 KENDALL'S TAU B = 0.88621 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.84400 SIGNIFICANCE = 0.0000
 GAMMA = 1.00000
 SOMERS'S D (ASYMMETRIC) = 0.93054 WITH T1 DEPENDENT. = 0.84400 WITH T2 DEPENDENT.
 SOMERS'S D (ASYMMETRIC) = 0.88516 DEPENDENT. = 0.96576 WITH T2 DEPENDENT.
 ETA = 0.94174 WITH T1 DEPENDENT. = 0.96576 WITH T2 DEPENDENT.

FILE EX01FAT (IMMEDIATION DATE = 02/26/80) EXERCICE SUR LES LIAISONS

T1 NUE APRES PAR SEMAINE C H O N S T A B I L I T E D E C H I F F R E A F F A I R E
BY T1

COUNT	ROW PCT	COL PCT	TOT PCT	50.I	100.I	250.I	401.I	650.I	1250.I	1600.I	2050.I	ROW TOTAL
5.	I	I	I	9	1	0	0	0	0	0	0	10
	I	I	I	90.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0
	I	I	I	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
	I	I	I	18.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	
15.	I	I	I	0	0	9	1	0	0	0	0	10
	I	I	I	0.0	0.0	90.0	10.0	0.0	0.0	0.0	0.0	20.0
	I	I	I	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	
	I	I	I	0.0	0.0	18.0	2.0	0.0	0.0	0.0	0.0	
25.	I	I	I	0	0	0	0	10	0	0	0	10
	I	I	I	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	20.0
	I	I	I	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	
	I	I	I	0.0	0.0	0.0	0.0	20.0	0.0	0.0	0.0	
35.	I	I	I	0	0	0	0	0	0	0	0	10
	I	I	I	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0
	I	I	I	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	
	I	I	I	0.0	0.0	0.0	0.0	20.0	0.0	0.0	0.0	
45.	I	I	I	0	0	0	0	0	9	1	0	10
	I	I	I	0.0	0.0	0.0	0.0	0.0	90.0	10.0	0.0	20.0
	I	I	I	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	
	I	I	I	0.0	0.0	0.0	0.0	0.0	18.0	2.0	0.0	
COLUMN	9	1	9	18.0	2.0	10	9	10	18.0	2.0	10	50
TOTAL	18.0	2.0	18.0	2.0	20.0	18.0	2.0	20.0	100.0	20.0	100.0	

CHI SQUARE = 200.0000 WITH 28 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
CRAMER'S V = 1.00000

CONTINGENCY COEFFICIENT = 0.89443
LAMBDA (ASYMMETRIC) = 1.00000 WITH T1 DEPENDENT.

LAMBDA (SYMMETRIC) = 0.96250
UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 1.00000 WITH T1 DEPENDENT. = 0.92500 WITH T3 DEPENDENT.

UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.94287
KENDALL'S TAU B = 0.98677 SIGNIFICANCE = 0.0000
KENDALL'S TAU C = 1.00000 SIGNIFICANCE = 0.0000
GAMMA = 1.00000

SOMERS'S D (ASYMMETRIC) = 0.97371 WITH T1 DEPENDENT. = 1.00000 WITH T3 DEPENDENT.
SOMERS'S D (SYMMETRIC) = 0.98668
ETA = 1.00000 WITH T1 DEPENDENT. = 0.99749 WITH T3 DEPENDENT.

DEPENDENT.

DEPENDENT. = 0.89191 WITH T3

DEPENDENT.

DEPENDENT.

***** C R O S S T A B U L A T I O N O F *****
 V A 0 1 B Y V A 0 2 *****
 ***** PAGE 1 OF 1 *****

		V A 0 2			
COUNT	I	1	2	3	I
ROW					
PCT					
TOT					

1	0	100	0	0	100
	0.0	100.0	0.0	0.0	16.7
	0.0	33.3	0.0	0.0	
	0.0	16.7	0.0	0.0	

2	100	0	200	0	300
	33.3	0.0	66.7	0.0	50.0
	100.0	0.0	100.0	0.0	
	16.7	0.0	33.3	0.0	

3	0	200	0	0	200
	0.0	100.0	0.0	0.0	33.3
	0.0	66.7	0.0	0.0	
	0.0	33.3	0.0	0.0	

COLUMN	100	300	200		600
TOTAL	16.7	50.0	33.3		100.0

RAW CHI SQUARE = 600.0000 WITH 4 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.70711
 CONTINGENCY COEFFICIENT = 0.70711
 LAMBDA (ASYMMETRIC) = 0.66667 WITH VA01 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.66667
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.68533 WITH VA01 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.68533
 KENDALL'S TAU B = -0.09091. SIGNIFICANCE = 0.0072
 KENDALL'S TAU C = -0.08333. SIGNIFICANCE = 0.0072
 GAMMA = -0.11111
 SOMERS'S D (ASYMMETRIC) = -0.09091 WITH VA01 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = -0.09091
 ETA = 0.24254 WITH VA01 DEPENDENT. = 0.24254 WITH VA02 DEPENDENT.
 PEARSON'S R = -0.05882. SIGNIFICANCE = 0.0751

FILE ECOLETE (CREATION DATE = 08/29/80)

***** CROSS TABULATION OF *****
VA03 BY VA04 *****
***** PAGE 1 OF 1 *****

		VA04			ROW TOTAL
	COUNT	1	2	3	
1	0	50	0	0	50
	0.0	100.0	0.0	0.0	9.3
	0.0	16.7	0.0	0.0	
	0.0	8.3	0.0	0.0	
2	100	0	200	0	300
	33.3	0.0	66.7	0.0	50.0
	100.0	0.0	100.0	0.0	
	16.7	0.0	33.3	0.0	
3	0	250	0	0	250
	0.0	100.0	0.0	0.0	41.7
	0.0	83.3	0.0	0.0	
	0.0	41.7	0.0	0.0	
COLUMN TOTAL	100	300	200	600	
	16.7	50.0	33.3	100.0	

RAW CHI SQUARE = 600.00001 WITH 4 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.70711
 CONTINGENCY COEFFICIENT = 0.70711
 LAMBDA (ASYMMETRIC) = 0.83333 WITH VA03 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.75000
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.75471 WITH VA03 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.71835
 KENDALL'S TAU B = -0.18835. SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = -0.16667. SIGNIFICANCE = 0.0000
 GAMMA = -0.22222
 SOMERS'S D (ASYMMETRIC) = -0.18182 WITH VA03 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = -0.18824
 ETA = 0.53452 WITH VA03 DEPENDENT. = 0.24254 WITH VA04 DEPENDENT.
 PEARSON'S R = -0.12964 SIGNIFICANCE = 0.0007
 = 0.66667 WITH VA04 DEPENDENT.
 = 0.68533 WITH VA04 DEPENDENT.

 VA05 ***** CROSS TABULATION OF *****
 BY VA06 *****
 ***** PAGE 1 OF 1 *****

		VA06			
COUNT	I	1	2	3	I
ROW PCT I		0	100	0	100
COL PCT I		0.0	100.0	0.0	16.7
TOT PCT I		0.0	50.0	0.0	
		0.0	16.7	0.0	

2		200	0	200	400
		50.0	0.0	50.0	66.7
		100.0	0.0	100.0	
		33.3	0.0	33.3	

3		0	100	0	100
		0.0	100.0	0.0	16.7
		0.0	50.0	0.0	
		0.0	16.7	0.0	

COLUMN		200	200	200	600
TOTAL		33.3	33.3	33.3	100.0

RAW CHI SQUARE = 600.00000 WITH 4 DEGREES OF FREEDOM. SIGNIFICANCE = 0.00000
 CRAMER'S V = 0.70711
 CONTINGENCY COEFFICIENT = 0.70711
 LAMBDA (ASYMMETRIC) = 0.50000 WITH VA05 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.50000
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.73368 WITH VA05 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.64746
 KENDALL'S TAU B = 0.00000. SIGNIFICANCE = 0.5000
 KENDALL'S TAU C = 0.00000. SIGNIFICANCE = 0.5000
 GAMMA = 0.00000
 SOMERS'S D (ASYMMETRIC) = 0.00000 WITH VA05 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.00000
 ETA = 0.00000 WITH VA05 DEPENDENT.
 PEARSON'S R = 0.00000 SIGNIFICANCE = 0.5000

FILE ECOLETE (CREATION DATE = 08/29/80)

***** C R O S S T A B U L A T I O N O F *****
VA01 BY VA02 ***** PAGE 1 OF 1

VA01	VA02	COUNT	ROW PCT	COL PCT	TOT PCT	ROW TOTAL
100.	150	7	3.2	28.6	63	220
	68.2	3.2	2.4	56.8	37.0	
	75.8	1.2	10.6			
	25.2	21	32		87	
	34	24.1	36.8		14.6	
	39.1	7.3	28.8			
	17.2	3.5	5.4			
	5.7					
100000.	14	258	16		288	
	4.9	89.6	5.6		48.4	
	7.1	90.2	14.4			
	2.4	43.4	2.7			
COLUMN	198	286	111		595	
TOTAL	33.3	48.1	18.7		100.0	

CHI SQUARE = 410.71751 WITH 4 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000

CRAMER'S V = 0.58749

CONTINGENCY COEFFICIENT = 0.63905

LAMBDA (ASYMMETRIC) = 0.59609 WITH VA01 DEPENDENT. = 0.50485 WITH VA02 DEPENDENT.

LAMBDA (SYMMETRIC) = 0.55032

UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.40768 WITH VA01 DEPENDENT. = 0.39531 WITH VA02 DEPENDENT.

UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.40140

KENDALL'S TAU B = 0.26713 SIGNIFICANCE = 0.0000

KENDALL'S TAU C = 0.24662 SIGNIFICANCE = 0.0000

GAMMA = 0.32710

SOMERS'S D (ASYMMETRIC) = 0.26373 WITH VA01 DEPENDENT. = 0.27058 WITH VA02 DEPENDENT.

SOMERS'S D (SYMMETRIC) = 0.26711

ETA = 0.80734 WITH VA01 DEPENDENT. = 0.27090 WITH VA02 DEPENDENT.

PEARSON'S R = 0.21132 SIGNIFICANCE = 0.0000

***** C R O S S T A B U L A T I O N O F *****
 VA03 VA04
 ***** PAGE 1 OF 1

		VA04				
COUNT	ROW	1.1	2.1	3.1	4.1	TOTAL
3000.	1	170	0	0	0	170
	2	150.0	0.0	0.0	0.0	28.5
	3	100.0	0.0	0.0	0.0	
	4	28.6	0.0	0.0	0.0	
5000.	1	0	0	0	158	158
	2	0.0	0.0	0.0	100.0	26.6
	3	0.0	0.0	0.0	100.0	
	4	0.0	0.0	0.0	26.6	
10000.	1	0	237	50	0	267
	2	0.0	88.8	11.2	0.0	44.9
	3	0.0	100.0	100.0	0.0	
	4	0.0	39.8	5.0	0.0	
COLUMN		170	237	30	158	595
TOTAL		28.6	39.8	5.0	26.6	100.0

CHI SQUARE = 1189.99998 WITH 6 DEGREES OF FREEDOM SIGNIFICANCE = 0.0000
 CRAMER'S V = 1.00000
 CONTINGENCY COEFFICIENT = 0.81650
 LAMBDA (ASYMMETRIC) = 1.00000 WITH VA03 DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.95627
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 1.00000 WITH VA03 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.93134
 KENDALL'S TAU B = 0.25491 SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.25476 SIGNIFICANCE = 0.0000
 GAMMA = 0.26271
 SOMERS'S D (ASYMMETRIC) = 0.24735 WITH VA03 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.25480
 ETA = 1.00000 WITH VA03 DEPENDENT.
 PEARSON'S R = 0.09288 SIGNIFICANCE = 0.0117

AGE7

COUNT	1	2	3	4	5	6	7	TOTAL
ROW PCT	0.0	0.0	1.4	1.4	1.4	4.2	91.7	72
COL PCT	0.0	0.0	1.9	3.2	2.3	10.3	59.2	23.2
TOT PCT	0.0	0.0	0.3	0.3	0.3	1.0	21.3	
MECHEL	0	1	0	1	3	22	3	30
	0.0	3.3	0.0	3.3	10.0	73.3	10.0	9.7
	0.0	2.3	0.0	3.2	4.8	75.9	4.1	
	0.0	0.3	0.0	0.3	1.0	7.1	1.0	
	1	0	1	5	39	3	1	46
	2.2	0.0	2.2	8.7	78.3	6.5	2.2	14.8
	2.8	0.0	1.9	12.9	81.2	10.2	1.4	
	0.3	0.0	0.3	1.3	11.6	1.0	0.3	
	1	1	4	20	2	0	1	29
	3.4	3.4	13.8	69.0	6.9	0.0	3.6	9.4
	2.8	2.3	7.5	64.5	4.5	0.0	1.4	
	0.3	0.3	1.3	6.5	0.6	0.0	0.3	
	1	5	40	3	1	0	1	51
	2.0	9.8	78.4	5.9	2.0	0.0	2.0	16.5
	2.8	11.6	75.5	9.7	2.3	0.0	1.4	
	0.3	1.6	12.9	1.0	0.3	0.0	0.3	
	5	34	5	1	0	1	1	47
	10.6	72.3	10.6	2.1	0.0	2.1	2.1	15.2
	13.9	79.1	9.4	3.2	0.0	3.4	1.4	
	1.6	11.0	1.6	0.3	0.0	0.3	0.3	
	28	2	2	1	1	0	1	35
	80.0	5.7	5.7	2.9	2.9	0.0	2.9	11.3
	77.8	4.7	3.8	3.2	2.3	0.0	1.4	
	9.0	0.6	0.6	0.3	0.3	0.0	0.3	
COLUMN	36	43	53	31	64	29	74	310
TOTAL	11.6	13.9	17.1	10.0	14.2	9.4	23.9	100.0

PAV CHI SQUARE = 1038.50659 WITH 36 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.74722
 CONTINGENCY COEFFICIENT = 0.87756
 LAMBDA (ASYMMETRIC) = 0.72996
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.59197 WITH MECHEL DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.59228
 KENDALL'S TAU A = -0.84769. SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = -0.83335. SIGNIFICANCE = 0.0000
 CONDITIONAL GAMMA = -0.90183
 SOMERS'S D (ASYMMETRIC) = -0.84817 WITH MECHEL DEPENDENT.
 SOMERS'S D (SYMMETRIC) = -0.84769
 ETA = 0.89878 WITH MECHEL DEPENDENT.
 PERSONS R = 0.89844 SIGNIFICANCE = 0.0000
 = 0.72881 WITH AGE7 DEPENDENT.
 = 0.59260 WITH AGE7 DEPENDENT.
 = -0.84721 WITH AGE7 DEPENDENT.
 = 0.89995 WITH AGE7 DEPENDENT.

SUMMARY GAMMAS FOR CROSSTABULATION OF MECHEL BY AGE7
 ZFPO-ORDER GAMMA = -0.00234
 FIRST-ORDER PARTIAL GAMMA = -0.03342

EXERCICE SUR LES LIAISONS

MECHEL ECHELLE DE PERMISSIVITE CROSSTABULATION OF BY AGE7

AGE7	1	2	3	4	5	6	7	TOTAL
MECHEL	27	34	40	3	0	0	0	104
ROW PCT	93.1	97.1	100.0	100.0	0.0	0.0	0.0	100.0
COL PCT	79.4	82.4	81.6	11.1	0.0	0.0	0.0	100.0
TOT PCT	9.5	11.9	14.0	1.1	0.0	0.0	0.0	100.0
MECHEL	5	34	4	1	0	0	0	44
ROW PCT	11.4	77.3	9.1	2.3	0.0	0.0	0.0	100.0
COL PCT	14.7	87.2	8.2	3.7	0.0	0.0	0.0	100.0
TOT PCT	1.8	11.9	1.4	0.4	0.0	0.0	0.0	100.0
MECHEL	2	4	40	3	0	0	0	49
ROW PCT	2.1	8.3	83.3	6.3	0.0	0.0	0.0	100.0
COL PCT	2.9	10.3	81.6	11.1	0.0	0.0	0.0	100.0
TOT PCT	0.4	1.4	14.0	1.1	0.0	0.0	0.0	100.0
MECHEL	1	0	4	20	2	0	1	28
ROW PCT	3.6	0.0	14.3	71.4	7.1	0.0	3.6	100.0
COL PCT	2.9	0.0	8.2	74.1	4.9	0.0	1.4	100.0
TOT PCT	0.4	0.0	1.4	7.0	0.7	0.0	0.4	100.0
MECHEL	0	0	0	3	35	3	0	41
ROW PCT	0.0	0.0	0.0	7.3	85.4	7.3	0.0	100.0
COL PCT	0.0	0.0	0.0	11.1	85.4	11.5	0.0	100.0
TOT PCT	0.0	0.0	0.0	1.1	12.3	1.1	0.0	100.0
MECHEL	0	0	0	0	3	21	2	26
ROW PCT	0.0	0.0	0.0	0.0	11.5	80.8	7.7	100.0
COL PCT	0.0	0.0	0.0	0.0	7.3	80.8	2.9	100.0
TOT PCT	0.0	0.0	0.0	0.0	1.1	7.4	0.7	100.0
MECHEL	0	0	0	0	1	2	66	69
ROW PCT	0.0	0.0	0.0	0.0	1.4	2.9	95.7	100.0
COL PCT	0.0	0.0	0.0	0.0	2.4	7.7	95.7	100.0
TOT PCT	0.0	0.0	0.0	0.0	0.4	0.7	23.2	100.0
COLUMN TOTAL	34	39	49	27	41	26	69	285
TOTAL	11.9	13.7	17.2	9.5	14.4	9.1	24.2	100.0

RAY CHI SQUARE = 1142.16772 WITH 36 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.81727
 CONTINGENCY COEFFICIENT = 0.89460
 LAMBDA (ASYMMETRIC) = 0.80556 WITH MECHEL DEPENDENT.
 LAMBDA (SYMMETRIC) = 0.80556
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.72461 WITH MECHEL DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.72407
 KENDALL'S TAU B = 0.93580. SIGNIFICANCE = 0.0000
 KENDALL'S TAU C = 0.91765. SIGNIFICANCE = 0.0000
 CONDITIONAL GAMMA = 0.97760
 SOMERS'S D (ASYMMETRIC) = 0.93540 WITH MECHEL DEPENDENT.
 SOMERS'S D (SYMMETRIC) = 0.93620 WITH AGE7 DEPENDENT.

NECHEL SCHELLE DE DEMISSIVITE C R O S S T A B U L A T I O N O F AGE7 BY AGE7 PAGE 1 OF 1

NECHEL	COUNT	AGE7							TOTAL
		1	2	3	4	5	6	7	
1	27	1	1	2	1	1	1	44	101
	26.7	1.0	1.0	2.0	1.0	1.0	3.0	45.3	17.0
	38.6	1.2	2.0	2.0	1.7	1.2	5.5	45.2	
	4.5	0.2	0.2	0.2	0.2	0.2	0.5	11.1	
2	5	35	4	4	2	3	22	3	74
	6.9	47.3	5.4	2.7	5.1	24.7	4.1	4.1	12.4
	7.1	42.7	3.9	3.4	3.5	40.0	2.1	2.1	
	0.5	5.0	0.7	0.3	0.5	3.7	0.5	0.5	
3	2	4	41	7	7	35	2	1	94
	2.1	4.3	43.6	7.4	7.4	38.3	3.2	1.1	15.8
	2.5	4.9	40.2	12.1	4.2	5.5	0.7	0.7	
	0.3	0.7	6.9	1.2	5.1	0.5	0.5	0.2	
4	2	1	8	4.0	4	2	0	2	57
	3.5	1.8	14.0	70.2	7.0	0.0	0.0	3.5	4.6
	2.9	1.2	7.8	65.0	4.7	0.0	0.0	1.4	
	0.3	0.2	1.3	4.7	0.7	0.0	0.0	0.3	
5	1	5	4.0	5	34	7	1	1	92
	1.1	5.4	43.5	6.5	35.1	3.3	1.1	1.1	15.5
	1.4	6.1	39.2	10.3	42.4	5.5	0.7	0.7	
	0.2	0.8	4.7	1.0	5.1	0.5	0.2	0.2	
6	5	34	5	1	22	3	3	3	73
	6.8	46.6	6.8	1.4	30.1	4.1	4.1	4.1	12.3
	7.1	41.5	4.9	1.7	3.5	40.0	2.1	2.1	
	0.8	5.7	0.8	0.2	0.5	3.7	0.5	0.5	
7	29	2	2	2	1	2	2	47	104
	26.9	1.9	1.9	1.0	1.9	1.9	1.9	64.4	17.5
	40.0	2.4	2.0	1.7	2.4	3.4	46.9		
	4.7	0.3	0.3	0.2	0.3	0.3	11.3		
COLUMN TOTAL	70	82	102	58	45	55	143	595	
TOTAL	11.8	13.8	17.1	9.7	14.3	9.2	24.0	100.0	

PAW CHI SQUARE = 995.75635 WITH 36 DEGREES OF FREEDOM. SIGNIFICANCE = 0.0000
 CRAMER'S V = 0.52813
 CONTINGENCY COEFFICIENT = 0.79118
 LAPLACE (ASYMMETRIC) = 0.33605 WITH NECHEL DEPENDENT. = 0.39823 WITH AGE7 DEPENDENT.
 LAPLACE (SYMMETRIC) = 0.36585
 UNCERTAINTY COEFFICIENT (ASYMMETRIC) = 0.37075 WITH NECHEL DEPENDENT. = 0.37671 WITH AGE7 DEPENDENT.
 UNCERTAINTY COEFFICIENT (SYMMETRIC) = 0.37350
 MENDEL'S TAU B = -0.00208. SIGNIFICANCE = 0.4737
 MENDEL'S TAU C = -0.00206. SIGNIFICANCE = 0.4737
 GAMMA = -0.00234
 SOMERS'S D (ASYMMETRIC) = -0.00209 WITH NECHEL DEPENDENT. = -0.00207 WITH AGE7 DEPENDENT.
 SOMERS'S D (SYMMETRIC) = -0.00208
 ETA = 0.01471 WITH NECHEL DEPENDENT. = 0.32043 WITH AGE7 DEPENDENT.
 PEARSON'S R = -0.00239. SIGNIFICANCE = 0.4769

ANNEXE 2

Nous présentons dans cette annexe une analyse en composantes principales d'un tableau constitué de l'observation de 23 indices d'association calculés sur 500 tableaux de contingence 2 x 2 simulés selon une méthode de congruence multiplicative.

$$u_{n+1} = a u_n \pmod{2^m}$$

avec ici $m = 36$, $a = 64\ 155\ 708\ 247$ et $u_0 = a$

Cette étude s'inspire d'un article de WEISBERG [78] qui propose d'effectuer un classement des indices selon le modèle de relation parfaite pour lequel ils sont maximum et le modèle de non relation pour lequel ils sont nuls.

Nous distinguons deux modèles de relation parfaite suivant que la variable dépendante est une fonction monotone stricte ou non de la variable indépendante, appelés modèle 1 et 2.

Nous distinguons quatre modèles de non relation :

- a : le mode conditionnel est constant .
- b : l'effectif sur la diagonale principale est égal à l'effectif sur l'autre diagonale .
- c : l'indépendance statistique .
- d : les distributions conditionnelles sont uniformes (c'est le modèle le plus strict) .

Les différents indices sont :

Modèle	Indices	Formule
1 c	ϕ^2	
1 c	V	$\sqrt{\phi^2}$
1 a	LAYX	$\lambda_{Y/X}$ de GUTTMAN
1 a	LAXY	$\lambda_{X/Y}$ de GUTTMAN
1 a	LAMD	λ symétrisé
	ALPH	$\frac{P_{11} P_{22}}{P_{12} P_{21}}$ appelé odd ratio
2 c	Q	$\frac{\alpha - 1}{\alpha + 1} = \gamma$ (YULE)
2 c	Y	$\frac{\sqrt{\alpha - 1}}{\sqrt{\alpha + 1}}$ (YULE)
1 c	KAPP	$\frac{2(P_{11}P_{22} - P_{21}P_{12})}{P_{1.} P_{2.} + P_{.1} P_{.2}}$
1 b	G	$ P_{11} + P_{22} - P_{12} - P_{21} $
1 d	O	$(P_{11} - P_{21})(P_{22} - P_{12}) / P_{.1}P_{.2}$
1 b	CC	$1 - 4 (P_{11} - P_{22})(P_{12} + P_{21})$
2 d	CmA	$\text{Max} (P_{11} - P_{21} / P_{.1}) P_{12} - P_{22} / P_{.2} $
2 d	Cmo	$\frac{1}{2} [P_{11} - P_{21} / P_{.1} + P_{12} - P_{22} / P_{.2}]$
2 d	TET	$2 [\text{Max} (P_{11}, P_{21}) + \text{Max} (P_{12}, P_{22})] - 1$
2 c	AR	$\sin(\frac{\pi}{2} Y)$ (approximation du tétrachorique)
2 c	DYX	coefficient $d_{Y/X}$ de SOMERS
2 c	DXY	coefficient $d_{X/Y}$ de SOMERS
2 c	UXY	
2 c	UYX	coefficients d'incertitude.
1 c	U	
1 c	LOGA	$\log \alpha$

L'étude du listing montre que ces indices d'inégales sensibilité, se classent du plus sensible : Q d'écart-type . 659, au moins sensible : CC d'écart-type .043 ; ce dernier indice étant toujours fort puisque sa moyenne est de .933 .

Les indices α et $\log \alpha$ bien que non comensurables sont fortement sensibles - cette comparaison doit être lue avec prudence puisque certains indices varient entre 0 et 1 alors que d'autres varient entre -1 et +1 (il nous semble à posteriori nécessaire de transformer les données afin de faire disparaître ce facteur d'hétérogénéité). L'étude de la matrice de corrélation met en évidence l'existence de deux grands groupes d'indices peu corrélés l'un à l'autre. On retrouve le partage donné par WEISBERG entre modèle 1 de stricte monotonie et modèle 2 de monotonie faible ; on trouve un troisième groupe moins homogène, mais distincts des deux grands groupes contenant les trois coefficients qui testent la relation nulle d'uniformité des conditionnelles (TET, CmA, CmO) - le coefficient KAPPA reste très particulier.

Ces regroupements sont présentés dans le tableau suivant :

$r \geq .980$	φ^2 . $u_{Y/X}, u_{X/Y}, u$	φ, Y, AR	CmO, CmA, TETA
$\geq .960$		$D_{Y/X}, D_{X/Y}$	
$\geq .940$	V, λ	Log A	
$\geq .920$	$\lambda_{X/Y}$		
$\geq .860$	$\lambda_{X/Y}$		
.			
$\geq .780$	0		
$\geq .700$	G		
$\geq .600$		K	
$\geq .500$	CmA, CmO, TETA	CC	

↑
NIVEAU DE REGROUPEMENT RETENU POUR r :

TABLEAU DES REGROUPEMENTS

L'analyse centrée réduite montre que la dichotomie précédente détermine les deux premiers axes. Le groupe associé au modèle 1 différencie les tableaux suivant le premier axe : les tableaux bien représentés sont caractérisés par l'existence d'une diagonale fortement dominante. Le groupe associé au modèle 2 différencie les tableaux suivant le deuxième axe : les tableaux bien représentés chargent soit trois cases, soit une diagonale qui est alors déséquilibrée. Le troisième axe traduit l'opposition entre le groupe bien représenté CmA, CmO, TET à l'indice 0 : les tableaux bien représentés présentent une forte ligne ou colonne, l'opposition se faisant suivant le degré d'équilibre entre les éléments de cette ligne ou colonne, parce que 0 est une moyenne géométrique.

Les tableaux bien représentés sur l'axe 1 (resp. 2) sont caractérisés par un coefficient α faible (resp. fort) montrant ainsi que le modèle de monotonie fort peut être interprété en termes d'écart à l'Indépendance alors que le modèle de monotonie faible nécessite l'utilisation du codage pour être modélisé et est donc interprétable en temps que modèle de liaison.

Nous donnons dans cette annexe quelques uns des tableaux les mieux représentés.

A_1 : sur l'axe 1

A_2 : sur l'axe 2

A_3 : sur l'axe 3

Ainsi que le listage des résultats.

On a signalé par :

C.1. les tableaux ayant une forte contribution à l'inertie de l'axe précisé,

C.0. les tableaux fortement corrélés à l'axe spécifié.

MATRICE DES CORRELATIONS

	PHI2	V	LAYX	LAYZ	LAYS	U	Y	Y	KAPP	S	O	CC	
PHI2	1.000												
V	0.953	1.000											
LAYX	0.490	0.824	1.000										
LAYZ	0.878	0.302	0.752	1.000									
LAYS	0.946	0.885	0.936	0.928	1.000								
U	-0.067	-0.068	0.008	-0.105	-0.059	1.000							
Y	-0.069	0.083	-0.013	-0.123	-0.082	0.935	1.000						
KAPP	0.301	0.265	0.272	0.311	0.301	0.600	0.615	1.000					
S	0.761	0.774	0.635	0.632	0.686	-0.111	-0.124	0.232	1.000				
O	0.732	0.679	0.574	0.600	0.732	-0.082	-0.095	0.343	0.494	1.000			
CC	0.151	0.126	0.099	0.071	0.068	0.524	0.561	0.437	0.358	0.272	1.000		
CMA	0.573	0.582	0.491	0.421	0.497	-0.091	-0.101	0.165	0.772	0.019	0.425	1.000	
CHC	0.591	0.599	0.482	0.498	0.536	-0.081	-0.095	0.045	0.584	0.006	0.313	0.425	1.000
TET	0.584	0.679	0.482	0.324	0.454	-0.050	-0.063	0.029	0.617	0.005	0.325	0.425	1.000
AR	-0.074	-0.075	-0.050	-0.112	-0.067	0.988	0.994	0.609	-0.115	-0.067	0.571	0.425	1.000
DYX	-0.097	-0.096	-0.043	-0.137	-0.086	0.959	0.969	0.655	-0.122	-0.128	0.544	0.425	1.000
DYU	-0.091	-0.067	-0.029	-0.120	-0.084	0.951	0.972	0.644	-0.107	-0.077	0.567	0.425	1.000
UYX	0.989	0.944	0.932	0.673	0.917	-0.090	-0.113	0.291	0.766	0.745	0.172	0.425	1.000
UYU	0.965	0.947	0.886	0.513	0.916	-0.068	-0.093	0.273	0.754	0.646	0.180	0.425	1.000
U	0.996	0.950	0.869	0.853	0.926	-0.075	-0.096	0.286	0.761	0.708	0.170	0.425	1.000
LOGA	-0.116	-0.112	-0.051	-0.138	-0.108	0.936	0.979	0.600	-0.138	-0.105	0.527	0.425	1.000
ALPH	0.304	0.225	0.252	0.246	0.241	0.194	0.238	0.260	0.168	0.184	0.124	0.425	1.000
AD	0.441	0.401	0.326	0.376	0.372	-0.625	-0.667	-0.240	0.538	0.316	-0.103	0.425	1.000

	CMA	CEO	TET	AR	DYX	DYU	UYX	UYU	KAPP	S	O	LOGA	ALPH
CMA	1.000												
CEO	0.285	1.000											
TET	0.217	0.277	1.000										
AR	-0.095	-0.086	-0.054	1.000									
DYX	0.086	0.086	0.054	0.566	1.000								
DYU	0.101	0.098	0.060	0.962	0.957	1.000							
UYX	0.575	0.602	0.597	-0.075	-0.123	-0.106	1.000						
UYU	0.609	0.616	0.633	-0.076	-0.089	-0.098	0.973	1.000					
U	0.583	0.607	0.613	-0.062	-0.104	-0.094	0.954	0.991	1.000				
LOGA	-0.411	-0.409	-0.084	0.954	0.940	0.955	-0.140	-0.126	-0.128	1.000			
ALPH	0.169	0.188	0.139	0.613	0.640	0.625	0.320	0.326	0.325	0.326	1.000		

DIAGRAMME DES VALEURS PROPRES DE LA MATRICE

NOUVEAU	VAL. PROPRE	POURCENT	CUMULE	VARIATION
1	0.9896335	0.1	47.178	0.000
2	0.6503438	0.1	71.517	15.725
3	0.2232668	0.1	85.145	20.750
4	0.6854285	0.1	92.440	7.371
5	0.4026098	0.1	94.225	1.346
6	0.3265748	0.1	95.340	0.363
7	0.2563965	0.1	97.263	0.172
8	0.2347718	0.1	98.379	0.266
9	0.9222645	0.1	98.819	0.672
10	0.7123366	0.1	99.148	0.607
11	0.5191368	0.1	99.463	0.675
12	0.4226147	0.1	99.666	0.666
13	0.3256065	0.1	99.764	0.666
14	0.2643572	0.1	99.840	0.025
15	0.1335322	0.1	99.994	0.072
16	0.5021566	0.2	99.978	0.033
17	0.3278336	0.2	99.994	0.006
18	0.1826525	0.2	99.992	0.007
19	0.1385545	0.2	99.995	0.002
20	0.1900089	0.3	100.000	0.006
21	0.3289752	0.4	100.000	0.001

QUALITE DE LA REPRESENTATION DES VARIABLES (TRANSFORMEES) EN COSINUS CARRES = COEFFICIENTS DE DETERMINATION

	1	2	3	4	5	6
PHI2	0.960	0.000	0.013	0.002	0.002	0.001
V	0.956	0.007	0.003	0.007	0.004	0.002
LAYX	0.750	0.019	0.025	0.048	0.013	0.018
LAXY	0.755	0.001	0.009	0.000	0.029	0.008
LAMD	0.873	0.007	0.056	0.014	0.000	0.034
Y	0.623	0.003	0.000	0.008	0.005	0.001
KAPP	0.654	0.047	0.000	0.007	0.001	0.000
G	0.635	0.027	0.063	0.070	0.218	0.027
O	0.438	0.001	0.043	0.155	0.013	0.001
CC	0.026	0.436	0.136	0.224	0.038	0.040
CHA	0.470	0.002	0.440	0.010	0.006	0.018
CMQ	0.474	0.002	0.393	0.028	0.054	0.017
TEI	0.446	0.004	0.425	0.035	0.000	0.072
AR	0.031	0.945	0.000	0.003	0.004	0.000
OYX	0.037	0.912	0.000	0.010	0.000	0.000
OXY	0.035	0.924	0.001	0.002	0.000	0.001
OYA	0.951	0.009	0.013	0.000	0.001	0.005
UXY	0.939	0.008	0.003	0.006	0.006	0.005
U	0.957	0.007	0.009	0.002	0.003	0.005
LQ6A	0.046	0.393	0.001	0.006	0.000	0.000
ALPH1	0.067	0.086	0.003	0.003	0.004	0.000
AD	0.274	0.306	0.015	0.125	0.016	0.002

ANALYSE EN COMPOSANTS PRINCIPAUX (CATHERINE)

9
YAK
DXY
DXY

KADD

ALPH

PHI 6 LUY LAYX

LI LUY Y C CMB IET

LXYY C CMA O

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

+++++

ANALYSE EN COMPOSANTES PRINCIPALES (CATHERINE)

AXE 1-ET-3

0

LAXY

LAND

LAXX

PHI+LIXX

U

LXY + V

AD

6

CC

ANALYSE EN COMPOSANTES PRINCIPALES (CATHERINE)

AKS 2-87-3

KADR

G Y
CGA-DY-AR

DYX

AD

CC

CMR TET

ANALYSE EN COMPOSANTES PRINCIPALES (CATHERINE)

AXE 1 ET 2

CC

S

AD

0

MA

LAXY

UYX

+++++

LAMP

CO

137

LAYX

tableau numero : 40

PHI2	0.37546	0	0	0.22678
Y	0.51642	CC	0	0.92987
LAMBDA	0.52291	TETA	0	0.51356
LAMBDA	0.35764	CMO	0	0.52230
LAMBDA	0.44571	CMA	0	0.22792
ALPHA	0.4635233	R	0	0.92026
Y	0.51642	DYX	0	0.59230
Y	0.51642	DYX	0	0.64090
KAPPA	0.51642	UYX	0	0.22427
B	0.22322	UYX	0	0.34331
G	0.51642	U	0	0.3352

tableau numero : 41

PHI2	0.68603	0	0	0.20195
Y	0.27671	CC	0	0.82288
LAMBDA	0.19235	TETA	0	0.50150
LAMBDA	0.40000	CMO	0	0.49391
LAMBDA	0.22740	CMA	0	0.69884
ALPHA	0.32737	R	0	0.42072
Y	0.51352	DYX	0	0.20493
Y	0.27341	UYX	0	0.27341
KAPPA	0.44158	UYX	0	0.04110
B	0.57474	UYX	0	0.05969
G	0.22322	U	0	0.04540

tableau numero : 42

PHI2	0.66600	0	0	0.02619
Y	0.87746	CC	0	0.92842
LAMBDA	0.40000	TETA	0	0.24925
LAMBDA	0.00000	CMO	0	0.18845
LAMBDA	0.00000	CMA	0	0.28501
ALPHA	1.49674	R	0	0.15669
Y	0.19231	DYX	0	0.09656
Y	0.10015	DYX	0	0.06214
KAPPA	0.42227	UYX	0	0.00615
B	0.64569	UYX	0	0.00446
G	0.21522	U	0	0.00517

Cor. A4

tableau numero : 41

PH12	0.00395	0	0.00216
V	0.06282	CC	0.93396
LAMBDA	0.00000	TETA	0.15115
LAMBDA	0.00000	CMO	0.09546
LAMBDA	0.00000	CMA	0.17887
ALPHA	1.40142	R	0.13187
Y	0.16712	DX	0.08341
Y	0.08418	DX	0.04731
KAPPA	0.05447	UY	0.00435
Y	0.74346	UY	0.00287
G	0.14715	U	0.00346

tableau numero : 62

PH12	0.35900	0	0.35894
V	0.62378	CC	0.91974
LAMBDA	0.46099	TETA	0.69570
LAMBDA	0.45734	CMO	0.62235
LAMBDA	0.45907	CMA	0.79079
ALPHA	0.04385	R	0.85586
Y	0.91593	DX	0.62235
Y	0.65361	DX	0.62505
KAPPA	0.03230	UY	0.31462
Y	5.57227	UY	0.31562
G	0.68570	U	0.31512

tableau numero : 63

PH12	0.17367	0	0.02366
V	0.42250	CC	0.96663
LAMBDA	0.19647	TETA	0.57958
LAMBDA	0.03670	CMO	0.40265
LAMBDA	0.12500	CMA	0.77476
ALPHA	8.37587	R	0.69192
Y	0.79660	DX	0.40265
Y	0.40660	DX	0.45665
KAPPA	0.42205	UY	0.14482
Y	0.20616	UY	0.15883
G	0.57958	U	0.15150

Con. A

tableau numero: 164

PHI2	0.00202	0.13340
V	0.04508	0.94599
LAMBDA	0.00000	0.37934
LAMBDA	0.00000	0.36888
LAMBDA	0.00000	0.42056
ALPHA	1.27086	0.09189
Y	0.11927	0.05168
KAPPA	0.05945	0.03933
B	0.02097	0.00199
B	1.37852	0.00163
B	0.15939	0.00179

CoA AA

tableau numero: 165

PHI2	0.00369	0.13255
V	0.04874	0.89982
LAMBDA	0.00000	0.33313
LAMBDA	0.00000	0.37040
LAMBDA	0.00000	0.42860
ALPHA	0.72224	0.12396
Y	0.15723	0.06620
KAPPA	0.07911	0.03405
B	0.07299	0.00350
B	0.74502	0.00296
B	0.13317	0.00321

tableau numero: 166

PHI2	0.03720	0.11043
V	0.12287	0.83235
LAMBDA	0.00000	0.33934
LAMBDA	0.00000	0.38089
LAMBDA	0.00000	0.56701
ALPHA	2.43914	0.33773
Y	0.41246	0.16612
KAPPA	0.21929	0.19986
B	0.20164	0.02280
B	0.41636	0.03033
B	0.10110	0.02955

tableau numero : 77

PHI2	0.07327	0.00000	-0.29584
LAMBDA	0.24971	0.00000	0.95663
LAMBDA	0.14745	0.00000	0.64164
LAMBDA	0.00000	0.172	0.58729
LAMBDA	0.09964	0.00000	0.80192
ALPHA	4.15751	0.821	0.51165
Y	0.61222	0.000	0.21463
Y	0.34189	0.000	0.34144
KAPPA	-0.37518	0.000	0.05376
D	0.46696	0.000	0.07536
G	0.36336	0.000	0.06282

tableau numero : 78

PHI2	0.00382	0.00000	-0.00000
LAMBDA	0.64179	0.00000	0.99639
LAMBDA	0.00000	0.00000	0.52332
LAMBDA	0.00000	0.00000	0.26441
LAMBDA	0.00000	0.00000	0.52882
ALPHA	0.30320	0.00000	-0.43440
Y	0.52882	0.00000	0.26441
Y	0.28604	0.00000	-0.01477
KAPPA	0.00000	0.00000	0.02895
D	0.19743	0.00000	0.00295
G	0.52332	0.00000	0.00336

Con. A4

tableau numero : 79

PHI2	0.01014	0.00000	0.00078
LAMBDA	0.12704	0.00000	0.98365
LAMBDA	0.00000	0.00000	0.12518
LAMBDA	0.00000	0.00000	0.12518
LAMBDA	0.00000	0.00000	0.26205
ALPHA	0.58125	0.00000	-0.21022
Y	0.34493	0.00000	-0.13252
Y	0.15482	0.00000	-0.12130
KAPPA	0.00000	0.00000	0.01257
D	0.24766	0.00000	0.01183
G	0.17518	0.00000	0.01019

tableau numero : 94

PHI2	0.13728	0	0.13728
V	0.27045	CC	0.9958
LAMBDA	0.302	IETA	0.5940
LAMBDA	0.302	CNO	0.5486
LAMBDA	0.302	CMA	0.7408
ALPHA	0.698	F	0.3535
Y	0.698	PYX	0.3896
X	0.698	UYX	0.2438
KAPPA	0.698	UXY	0.1321
D	0.698	U	0.1918
U	0.698		

tableau numero : 95

PHI2	0.25357	0	0.25357
V	0.50714	CC	0.98365
LAMBDA	0.29050	IETA	0.49149
LAMBDA	0.29050	CNO	0.52422
LAMBDA	0.29050	CMA	0.64246
ALPHA	0.511	R	0.74509
Y	0.511	PYX	0.52482
X	0.511	UYX	0.48337
KAPPA	0.511	UXY	0.20602
D	0.511	U	0.19406
U	0.511		0.19986

tableau numero : 96

PHI2	0.01529	0	0.00575
V	0.18780	CC	0.96608
LAMBDA	0.00000	IETA	0.49550
LAMBDA	0.00395	CNO	0.27849
LAMBDA	0.00287	CMA	0.54646
ALPHA	0.28717	R	0.45710
Y	0.55380	PYX	0.27849
X	0.30219	UYX	0.12672
KAPPA	0.00059	UXY	0.04545
D	2.97429	U	0.02745
U	0.49550		0.03531

Con. Au

tableau numero : 106

PHI2		0.13547	0	0.00000
V		0.36806	CC	0.95320
LAMBDA	0.234	0.19490	TETA	0.30531
LAMBDA	0.653	0.00000	CMO	0.35383
LAMBDA		0.10797	CMA	0.70766
ALPHA	0.347	0.24127	R	0.60834
		0.70766	DYX	0.35383
	0.451	0.41467	DXY	0.38286
KAPPA		0.33322	UYX	0.10561
		0.53221	UXY	0.11181
		0.30531	U	0.10862

tableau numero : 107

PHI2		0.48458	0	0.54988
V		0.83324	CC	0.99595
LAMBDA	0.150	0.421	TETA	0.68368
LAMBDA		0.62470	CMO	0.26299
LAMBDA	0.003	0.54857	CMA	0.94265
ALPHA	0.571	128.72496	R	0.96784
		0.98458	DYX	0.76299
	0.279	0.83800	DXY	0.62986
KAPPA		0.64563	UYX	0.46531
		0.18787	UXY	0.40484
		0.48368	U	0.43297

C.I. A2

tableau numero : 108

PHI2		0.00933	0	0.00632
V		0.09687	CC	0.89395
LAMBDA		0.00000	TETA	0.17317
LAMBDA	0.263	0.00000	CMO	0.13141
LAMBDA		0.00000	CMA	0.23538
ALPHA	0.433	0.65309	R	0.16593
		0.20986	DYX	0.10427
	0.702	0.10611	DXY	0.09000
KAPPA		0.09458	UYX	0.00764
		1.37292	UXY	0.00687
		0.15716	U	0.00724

tableau numero : 109

PHI2	0	0.4205	0	0	0.45082
V	0	0.20505	CC	0	0.95570
LAMBDA	0.248	0.0000	TETA	0	0.61562
LAMBDA	0.540	0.0000	CMO	0	0.70248
LAMBDA	0.015	0.0000	CMA	0	0.88593
ALPHA	0.192	0.0000	R	0	0.57229
Y	0.252	0.67923	DYX	0	0.18345
Y	0.252	0.39172	DYX	0	0.22919
KAPPA	0	0.12789	LYX	0	0.04404
U	0	1.35059	UXY	0	0.05188
G	0	0.14	U	0	0.04764

tableau numero : 110

PHI2	0	0.08687	0	0	0.10308
V	0	0.29475	CC	0	0.87634
LAMBDA	0.539	0.0000	TETA	0	0.29730
LAMBDA	0.193	0.23861	CMO	0	0.32860
LAMBDA	0.276	0.14925	CMA	0	0.39855
ALPHA	0.481	0.25331	R	0	0.49608
Y	0.726	0.54578	DYX	0	0.32860
Y	0.726	0.33041	DYX	0	0.26438
KAPPA	0	0.11422	UYX	0	0.07478
U	0	1.84615	UXY	0	0.06386
G	0	0.29730	U	0	0.06889

tableau numero : 111

PHI2	0	0.04071	0	0	0.04931
V	0	0.20177	CC	0	0.97035
LAMBDA	0.753	0.0000	TETA	0	0.51952
LAMBDA	0.036	0.02834	CMO	0	0.32256
LAMBDA	0.043	0.02147	CMA	0	0.55652
ALPHA	0.267	0.23853	R	0	0.51413
Y	0.921	0.61481	DYX	0	0.32256
Y	0.921	0.34372	DYX	0	0.12622
KAPPA	0	0.07401	UYX	0	0.06339
U	0	3.16250	UXY	0	0.03134
G	0	0.51952	U	0	0.04194

Cor. A4

tableau numero : 165

PHI2	0.04789	0.05086
V	0.21885	0.91404
LAMBDA	0.464	0.22923
LAMBDA	0.17026	0.22590
LAMBDA	0.09461	0.23885
ALPHA	2.56834	0.34754
Y	0.42993	0.22590
Y	0.22594	0.21202
KAPPA	0.12546	0.03648
U	0.22704	0.03484
G	0.22923	0.03564

tableau numero : 170

PHI2	0.01147	-0.69391
V	0.10708	0.92684
LAMBDA	0.061	0.87288
LAMBDA	0.00000	0.35466
LAMBDA	0.00000	0.90208
ALPHA	0.39468	-0.3105
Y	-0.47402	-0.06643
Y	-0.22232	-0.17262
KAPPA	-0.30525	0.01030
U	2.97015	0.02122
G	0.59760	0.01388

Cor. A 3

tableau numero : 171

PHI2	0.01424	-0.07326
V	0.11934	0.89125
LAMBDA	0.00943	0.31131
LAMBDA	0.00000	0.29397
LAMBDA	0.07521	0.40870
ALPHA	0.60309	-0.19627
Y	-0.24759	-0.11473
Y	-R.12575	-0.12413
KAPPA	0.04946	0.01040
U	1.37857	0.01101
G	0.11916	0.01070

tableau numero 175

PHI2 C. 11579
 LAMBDA C. 12554
 LAMBDA C. 11110
 LAMBDA C. 05057
 LAMBDA C. 07427
 ALPHA C. 36401
 Y C. 45408
 KAPPA C. 26062
 G C. 11127
 U C. 43859
 U C. 17537

C C. 05718
 CC C. 96214
 TETA C. 17317
 CMG C. 25800
 CMA C. 35484
 F C. 39808
 DYA C. 25800
 DXY C. 06109
 DYY C. 03382
 UXY C. 01149
 U C. 01715

tableau numero 176

PHI2 C. 12241
 LAMBDA C. 14911
 LAMBDA C. 01000
 LAMBDA C. 08516
 LAMBDA C. 01000
 ALPHA C. 23837
 Y C. 17249
 KAPPA C. 04047
 G C. 92857
 U C. 03704

C C. 00110
 CC C. 90163
 TETA C. 03704
 CMG C. 05230
 CMA C. 09508
 R C. 08379
 DYA C. 05330
 DXY C. 04524
 DYY C. 00196
 UXY C. 00174
 U C. 00185

tableau numero 177

PHI2 C. 09479
 LAMBDA C. 03354
 LAMBDA C. 07550
 LAMBDA C. 21936
 LAMBDA C. 71191
 ALPHA C. 00222
 Y C. 09159
 KAPPA C. 01669
 G C. 12374
 U C. 01100

C C. 02717
 CC C. 97705
 TETA C. 90190
 CMG C. 90958
 CMA C. 92051
 F C. 99069
 DYA C. 90956
 DXY C. 78388
 DYY C. 67825
 UXY C. 59566
 U C. 23428

Con A1

tableau numero : 247

PHI2	0.0065	0	-0.0064
V	0.0253	CC	0.9120
LAMBDA YX	0.0000	TETA	0.05706
LAMBDA X	0.0000	CMO	0.03985
LAMBDA	0.0000	CMA	0.07060
ALPHA	1.13118	R	0.04838
Y	0.06155	DYX	0.03076
X	0.03081	DX	0.02120
KAPPA	0.02482	UYX	0.00062
B	0.89924	UXY	0.00047
G	0.05305	U	0.00053

tableau numero : 248

PHI2	0.01423	0	-0.58714
V	0.11929	CC	0.98223
LAMBDA YX	0.0000	TETA	0.82783
LAMBDA X	0.0000	CMO	0.77125
LAMBDA	0.0000	CMA	0.85888
ALPHA	0.40398	R	-0.34291
Y	0.42452	DYX	-0.08763
X	0.22280	DX	-0.16238
KAPPA	0.27476	UYX	0.01313
B	3.82609	UXY	0.02090
G	0.58559	U	0.01613

Con. A's

tableau numero : 249

PHI2	0.04718	0	-0.00078
V	0.21720	CC	0.91602
LAMBDA YX	0.04417	TETA	0.19319
LAMBDA X	0.0000	CMO	0.21615
LAMBDA	0.04947	CMA	0.43049
ALPHA	2.50274	P	0.34676
Y	0.42302	DYX	0.21434
X	0.22541	DX	0.22010
KAPPA	0.20809	UYX	0.03488
B	0.67899	UXY	0.03555
G	0.19119	U	0.03521

tableau numero : 265

PHI2	0	0.3215
V	CC	0.98869
LAMBDA	TETA	0.91592
LAMBDA	CMO	0.6806
LAMBDA	CMA	0.9759
ALPHA	R	0.97443
K	DYX	0.6906
Y	DX	0.87784
KAPPA	UYX	0.4653
D	UXY	0.54872
G	U	0.50313

C.I. A₂

tableau numero : 266

PHI2	0	0.6540
V	CC	0.94784
LAMBDA	TETA	0.85185
LAMBDA	CMO	0.81090
LAMBDA	CMA	0.92117
ALPHA	R	0.90992
G	DYX	0.81090
Y	DX	0.84150
KAPPA	UYX	0.57731
D	UXY	0.59308
G	U	0.58509

C.I. A₁ A₂

tableau numero : 267

PHI2	0	0.22090
V	CC	0.58159
LAMBDA	TETA	0.60360
LAMBDA	CMO	0.58705
LAMBDA	CMA	0.93881
ALPHA	R	0.97676
G	DYX	0.58705
Y	DX	0.68424
KAPPA	UYX	0.33348
D	UXY	0.37261
G	U	0.35196

A. 61

Con. A.1

tableau numero : 301

PHI2	0.60674	0	0.69376
V	0.82870	CC	0.93753
LAMBDA	0.76142	TETA	0.83984
LAMBDA	0.78723	CMO	0.83328
LAMBDA	0.73437	CMA	0.85782
ALPHA	0.08809	R	-0.96662
Y	-0.90395	DYX	-0.83328
Y	-0.83493	DYX	-0.82414
KAPPA	-0.05006	UYX	0.58483
D	11.48750	UXY	0.32017
G	0.83964	U	0.58249

tableau numero : 302

PHI2	0.01572	0	-0.00137
V	0.12540	CC	0.89840
LAMBDA	0.00000	TETA	0.16316
LAMBDA	0.00000	CMO	0.13268
LAMBDA	0.00000	CMA	0.26010
ALPHA	1.68522	R	0.20241
Y	0.25513	DYX	0.12742
Y	0.12974	DYX	0.12341
KAPPA	0.12691	UYX	0.01180
D	0.72539	UXY	0.01153
G	0.15916	U	0.01166

tableau numero : 303

PHI2	0.60000	0	-0.00248
V	0.82203	CC	0.98177
LAMBDA	0.00000	TETA	0.04505
LAMBDA	0.00000	CMO	0.05010
LAMBDA	0.00000	CMA	0.05356
ALPHA	0.97637	R	-0.00839
Y	-0.01093	DYX	-0.00545
Y	-0.00347	DYX	-0.00076
KAPPA	0.00306	UYX	0.00001
D	0.92115	UXY	0.00000
G	0.04104	U	0.00000

tableau numero 230

PHI2	0.01573	0	0.13737
Y	0.12563	CC	0.42210
LAMBDA	0.02000	TEA	0.36537
LAMBDA	0.00000	CMO	0.38939
LAMBDA	0.00000	CMA	0.50877
ALPHA	1.76547	R	0.21994
Y	0.27679	DYX	0.11939
Y	0.14115	DYX	0.13220
KAPPA	0.12208	UYX	0.01195
Y	0.97115	UYX	0.01267
Y	0.04104	U	0.01239

tableau numero 231

PHI2	0.04521	0	0.05765
Y	0.25732	CC	0.92954
LAMBDA	0.09290	TEA	0.40741
LAMBDA	0.00000	CMO	0.34223
LAMBDA	0.05136	CMA	0.58610
ALPHA	3.14572	R	0.42430
Y	0.51257	DYX	0.24387
Y	0.22892	DYX	0.27151
KAPPA	0.02123	UYX	0.04938
Y	0.49773	UYX	0.05338
Y	0.33534	U	0.05130

tableau numero 232

PHI2	0.00557	0	0.10600
Y	0.07466	CC	0.99410
LAMBDA	0.00000	TEA	0.38739
LAMBDA	0.00000	CMO	0.33070
LAMBDA	0.00000	CMA	0.38865
ALPHA	0.25156	R	0.49817
Y	0.50000	DYX	0.33070
Y	0.34195	DYX	0.01686
KAPPA	0.00264	UYX	0.04154
Y	0.26471	UYX	0.00407
Y	0.37739	U	0.00741

Cor. A4

tableau numero : 319

PH12	0.00000	0	0.16020
CC	0.26003	CC	0.95260
V	0.14333	TETA	0.56557
LAMBDA YX	0.00000	CMO	0.45774
LAMBDA XZ	0.00000	CMA	0.62899
LAMBDA	0.00000	R	0.33716
ALPHA	0.41664	DYA	0.17125
Y	0.41779	DXY	0.15197
Z	-0.21891	UYX	0.02482
KAPPA	0.02365	UXY	0.02273
G	2.70090	U	0.02373
	0.45846		

tableau numero : 320

PH12	0.00092	0	0.00136
CC	0.03033	CC	0.96076
LAMBDA YX	0.00000	TETA	0.10911
LAMBDA XZ	0.00000	CMO	0.06696
LAMBDA	0.00000	CMA	0.11842
ALPHA	1.23282	R	0.08425
Y	0.10707	DYA	0.05366
Z	0.05369	DXY	0.01721
KAPPA	-0.00225	UYX	0.00155
G	0.00651	UXY	0.00067
	0.10711	U	0.00093

tableau numero : 321

PH12	0.15698	0	0.06386
CC	0.39608	CC	0.95387
LAMBDA YX	0.21103	TETA	0.34134
LAMBDA XZ	0.14099	CMO	0.39052
LAMBDA	0.17750	CMA	0.62825
ALPHA	6.52303	R	0.63417
Y	0.74415	DYA	0.39052
Z	0.43726	DXY	0.40172
KAPPA	0.07336	UYX	0.12326
G	0.49104	UXY	0.12580
	0.34134	U	0.12452

Col A

tableau numero : 328

PHI2	0.432	0.0000	V	0.0000
V	0.20203	0.93666	CC	0.93666
LAMBDA Y X	0.0000	0.3536	TEIA	0.3536
LAMBDA Y X	0.0000	0.22786	CMO	0.22786
LAMBDA	0.0000	0.45571	CMA	0.45571
ALPHA	0.372	0.36978	R	0.36978
Y	0.45571	0.22786	DYX	0.22786
Y	0.25110	0.17913	DXY	0.17913
KAPPA	0.19516	0.01699	UYX	0.01699
B	0.47563	0.03101	UXY	0.03101
B	0.35535	0.03374	U	0.03374

tableau numero : 329

PHI2	0.0000	0.00716	0	0.00716
V	0.02312	0.88165	CC	0.88165
LAMBDA Y X	0.0000	0.09309	IEIA	0.09309
LAMBDA Y X	0.0000	0.08787	CMO	0.08787
LAMBDA	0.0000	0.11145	CMA	0.11145
ALPHA	0.43922	0.03736	R	0.03736
Y	0.4755	0.02360	DYX	0.02360
Y	0.02329	0.22265	DXY	0.22265
KAPPA	0.05525	0.00040	UYX	0.00040
B	0.08996	0.00039	UXY	0.00039
B	0.04304	0.00019	U	0.00019

Con. A4

tableau numero : 330

PHI2	0.30229	0.29894	0	0.29894
V	0.60191	0.37243	CC	0.37243
LAMBDA Y X	0.51717	0.58759	IEIA	0.58759
LAMBDA Y X	0.47449	0.58780	CMO	0.58780
LAMBDA	0.51760	0.60361	CMA	0.60361
ALPHA	0.04934	0.84015	R	0.84015
Y	0.32505	0.58280	DYX	0.58280
Y	0.63492	0.61635	DXY	0.61635
KAPPA	0.01654	0.28561	UYX	0.28561
B	0.84351	0.29557	UXY	0.29557
B	0.51750	0.29051	U	0.29051

tableau numero : 421

PHI2	0.0921	0	0.6682
V	0.09595	CC	0.99380
LAMBDA	0.110	TETA	0.76376
LAMBDA	0.111	CMO	0.81962
LAMBDA	0.725	CMA	0.90303
LAMBDA	0.882	F	0.40659
ALPHA	0.125	DYX	0.08341
Y	0.100	DYX	0.11041
Y		DYX	0.01238
KAPPA		UYX	0.01528
B		UYX	0.01368
G		U	

tableau numero : 422

PHI2	0.12820	0	0.00000
V	0.35805	CC	0.96515
LAMBDA	0.00000	TETA	0.57958
LAMBDA	0.00000	CMO	0.36324
LAMBDA	0.00000	CMA	0.72647
LAMBDA	0.31193	R	0.62602
ALPHA	0.202	DYX	0.36324
Y	0.798	DYX	0.35294
Y	0.000	DYX	0.11068
KAPPA		UYX	0.10836
B		UYX	0.10836
G		U	0.10951

tableau numero : 423

PHI2	0.06471	0	0.01544
V	0.25439	CC	0.89874
LAMBDA	0.00000	TETA	0.36136
LAMBDA	0.02744	CMO	0.26346
LAMBDA	0.1459	CMA	0.49577
LAMBDA	0.31674	F	0.42549
ALPHA	0.1590	DYX	0.26346
Y	0.27976	DYX	0.24562
Y	0.00564	UYX	0.05205
KAPPA	2.13166	UYX	0.04946
B	0.36170	U	0.05072
G			

tableau numero : 424

PH12	0.64103	0	
V	0.98107	CC	
LAMBDA	0.77972	TETA	
LAMBDA	0.80223	CRD	
LAMBDA	0.86047	CRD	
ALPHA	0.25609	R	
Y	0.85774	UYX	
Y	0.12405	UYX	
KAPPA	0.02432	UYX	
B	0.03115	UYX	
	0.00897		

C.I. A3

tableau numero : 425

PH12	0.11234	0	
V	0.95424	CC	
LAMBDA	0.31532	TETA	
LAMBDA	0.39361	CRD	
LAMBDA	0.60000	R	
ALPHA	0.60641	R	
Y	0.39361	UYX	
Y	0.33806	UYX	
KAPPA	0.11357	UYX	
B	0.10172	UYX	
G	0.10732	U	

tableau numero : 426

PH12	0.00291	0	
V	0.95287	CC	
LAMBDA	0.40340	TETA	
LAMBDA	0.43161	CRD	
LAMBDA	0.49767	CRD	
ALPHA	0.11657	R	
Y	0.06006	UYX	
Y	0.64846	UYX	
KAPPA	0.00286	UYX	
B	0.00245	UYX	
G	0.00254	U	

A. 67

tableau numero : 430

PHI2	C. 21050	0	- C. 00171
V	C. 45839	CC	C. 97003
LAMBDA	C. 41068	TEIA	C. 42743
LAMBDA	C. 00000	CMO	C. 41704
LAMBDA	C. 25873	CMA	C. 83203
ALPHA	C. 86227	R	C. 74437
V	C. 81140	PYX	C. 41499
Y	C. 53443	DXY	C. 50743
KAPPA	C. 41939	UYX	C. 16322
.....	C. 40309	UXY	C. 18886
G	C. 42543	U	C. 17511

Con. A

tableau numero : 431

PHI2	C. 62621	0	C. 61405
V	C. 79133	CC	C. 91875
LAMBDA	C. 76957	TEIA	C. 79379
LAMBDA	C. 75708	CMO	C. 78660
LAMBDA	C. 76349	CMA	C. 85507
ALPHA	C. 61282	R	C. 94944
V	C. 97469	PYX	C. 78660
Y	C. 74660	DXY	C. 79610
KAPPA	C. 11732	UYX	C. 51784
.....	C. 69903	UXY	C. 52235
G	C. 76379	U	C. 52008

tableau numero : 432

PHI2	C. 30211	0	C. 02958
V	C. 54965	CC	C. 98368
LAMBDA	C. 42033	TEIA	C. 61161
LAMBDA	C. 05346	CMO	C. 46949
LAMBDA	C. 24413	CMA	C. 90634
ALPHA	C. 72668	R	C. 85801
V	C. 91200	PYX	C. 46949
Y	C. 64673	DXY	C. 64348
KAPPA	C. 03642	UYX	C. 23216
.....	C. 26699	UXY	C. 29279
G	C. 61161	U	C. 25916

tableau numero 1.477

PHI2	0.20804	0.00000	0.00000
V	0.45191	0.00000	0.00000
LAMBDA1	0.21543	0.00000	0.00000
LAMBDA2	0.24490	0.00000	0.00000
LAMBDA3	0.21532	0.00000	0.00000
ALPHA	0.50230	0.00000	0.00000
Y	0.07459	0.00000	0.00000
KAPPA	0.51081	0.00000	0.00000
B	0.29704	0.00000	0.00000
U	0.43744	0.00000	0.00000
U	0.27159	0.00000	0.00000

Cont
A

tableau numero 1.478

PHI2	0.00703	0.00000	0.00000
V	0.08354	0.00000	0.00000
LAMBDA1	0.05241	0.00000	0.00000
LAMBDA2	0.08000	0.00000	0.00000
LAMBDA3	0.03310	0.00000	0.00000
ALPHA	0.68459	0.00000	0.00000
Y	0.11758	0.00000	0.00000
KAPPA	0.09463	0.00000	0.00000
B	0.0229	0.00000	0.00000
U	0.21010	0.00000	0.00000
U	0.00510	0.00000	0.00000

tableau numero 1.479

PHI2	0.09346	0.00000	0.00000
V	0.20577	0.00000	0.00000
LAMBDA1	0.00100	0.00000	0.00000
LAMBDA2	0.00100	0.00000	0.00000
LAMBDA3	0.00100	0.00000	0.00000
ALPHA	0.08334	0.00000	0.00000
Y	0.04648	0.00000	0.00000
KAPPA	0.55238	0.00000	0.00000
B	0.00100	0.00000	0.00000
U	0.29882	0.00000	0.00000
U	0.00501	0.00000	0.00000

0.11874	0.00000	0.00000	0.00000
0.96459	0.00000	0.00000	0.00000
0.39139	0.00000	0.00000	0.00000
0.45326	0.00000	0.00000	0.00000
0.74019	0.00000	0.00000	0.00000
0.71643	0.00000	0.00000	0.00000
0.45320	0.00000	0.00000	0.00000
0.45056	0.00000	0.00000	0.00000
0.16635	0.00000	0.00000	0.00000
0.16564	0.00000	0.00000	0.00000
0.16599	0.00000	0.00000	0.00000

0.20804	0.00000	0.00000	0.00000
0.89383	0.00000	0.00000	0.00000
0.46547	0.00000	0.00000	0.00000
0.46212	0.00000	0.00000	0.00000
0.53640	0.00000	0.00000	0.00000
0.14812	0.00000	0.00000	0.00000
0.07425	0.00000	0.00000	0.00000
0.09463	0.00000	0.00000	0.00000
0.00505	0.00000	0.00000	0.00000
0.00605	0.00000	0.00000	0.00000
0.00552	0.00000	0.00000	0.00000

0.09346	0.00000	0.00000	0.00000
0.20577	0.00000	0.00000	0.00000
0.00100	0.00000	0.00000	0.00000
0.00100	0.00000	0.00000	0.00000
0.00100	0.00000	0.00000	0.00000
0.08334	0.00000	0.00000	0.00000
0.04648	0.00000	0.00000	0.00000
0.55238	0.00000	0.00000	0.00000
0.00100	0.00000	0.00000	0.00000
0.29882	0.00000	0.00000	0.00000
0.00501	0.00000	0.00000	0.00000

tableau numero : 481

PHI2	0.14254	0	0.00114
V	0.37754	CC	0.85792
LAMBDA	0.00000	TETA	0.23323
LAMBDA	0.00260	CMO	0.40630
LAMBDA	0.00149	CMA	0.81119
ALPHA	0.46396	R	-0.72079
R	-0.81167	DYX	-0.40630
Y	-0.51238	DXY	-0.35083
KAPPA	-0.36734	UYX	0.13688
U	1.60836	UXY	0.12304
G	0.23323	U	0.12959

tableau numero : 482

PHI2	0.71719	0	0.70356
V	0.84687	CC	0.93617
LAMBDA	0.87186	TETA	0.84785
LAMBDA	0.82160	CMO	0.84146
LAMBDA	0.82688	CMA	0.90859
ALPHA	0.00609	R	-0.97429
R	-0.87789	DYX	-0.84146
Y	-0.85211	DXY	-0.85231
KAPPA	-0.10019	UYX	0.61163
U	12.14474	UXY	-0.61731
G	0.84785	U	0.61446

C.I. A4

tableau numero : 483

PHI2	0.59424	0	0.59180
V	0.77087	CC	0.98853
LAMBDA	0.76571	TETA	0.76977
LAMBDA	0.76240	CMO	0.77064
LAMBDA	0.76356	CMA	0.81633
ALPHA	0.20174	R	0.93806
R	0.96826	DYX	0.77064
Y	0.77465	DXY	0.77110
KAPPA	0.65775	UYX	0.48800
U	0.13009	UXY	0.48821
G	0.76977	U	0.48811

tableau numero : 490

PHI2	0.1536	0	0.0268
V	0.43053	CC	0.9276
LAMBDA X	0.0974	TETA	0.40741
LAMBDA Y	0.00000	CMO	0.49490
LAMBDA Z	0.0577	CMA	0.89770
ALPHA	35.62308	R	0.80358
Y	0.27621	DYX	0.40280
X	0.59407	UYX	0.46017
KAPPA	0.07650	UXY	0.16345
B	0.54405	UY	0.18003
G	0.27530	U	0.17134

tableau numero : 491

PHI2	0.0113	0	0.00805
V	0.03363	CC	0.94336
LAMBDA X	0.00000	TETA	0.14114
LAMBDA Y	0.00000	CMO	0.10222
LAMBDA Z	0.00000	CMA	0.15447
ALPHA	0.82254	R	0.07659
Y	0.0737	DYX	0.04828
X	0.04880	UYX	0.02344
KAPPA	0.03531	UY	0.01140
B	0.28604	UXY	0.00082
G	0.12513	U	0.00104

tableau numero : 492

PHI2	0.91840	0	0.91833
V	0.95833	CC	0.9961
LAMBDA X	0.95000	TETA	0.94000
LAMBDA Y	0.95000	CMO	0.92833
LAMBDA Z	0.95000	CMA	0.96667
ALPHA	2301.00095	R	0.99796
Y	0.99913	DYX	0.95833
X	0.95916	UYX	0.95833
KAPPA	0.94250	UY	0.85495
B	0.02061	UXY	0.85495
G	0.00000	U	0.85495

C.I

A1 Az

tableau numero : 493

*

PHI2	0.0000	0	0.91833
V	0.02916	CC	0.99064
LAMBDA YX		TEIA	0.96000
LAMBDA YX	0.590	CMO	0.95833
LAMBDA	0.0000	CMA	0.96667
ALPHA	0.070	R	0.16130
W	0.51282	DYX	0.00833
Y	0.20408	DY	0.10204
KAPPA	0.10313	UYX	0.00062
W	0.02871	UY	0.00425
G	0.66667	U	0.00108
G	0.20000		

PHI2	0.0000
V	0.02916
LAMBDA YX	
LAMBDA YX	0.590
LAMBDA	0.0000
ALPHA	0.070
W	0.51282
Y	0.20408
KAPPA	0.10313
W	0.02871
G	0.66667
G	0.20000

tableau numero : 494

PHI2	0.05202	0	-0.62877
V	0.22809	CC	0.92000
LAMBDA YX	0.00000	TEIA	0.78000
LAMBDA YX	0.590	CMO	0.80612
LAMBDA	0.00000	CMA	0.95122
ALPHA	0.12250	R	-0.68631
W	-0.78174	DYX	-0.14510
Y	-0.48143	DY	-0.35853
KAPPA	0.01815	UYX	0.04581
W	1.00000	UY	0.08948
G	0.00000	U	0.06060

PHI2	0.05202
V	0.22809
LAMBDA YX	0.00000
LAMBDA YX	0.590
LAMBDA	0.00000
ALPHA	0.12250
W	-0.78174
Y	-0.48143
KAPPA	0.01815
W	1.00000
G	0.00000

tableau numero : 495

PHI2	0.01042	0	0.33333
V	0.10206	CC	0.92800
LAMBDA YX	0.00000	TEIA	0.60000
LAMBDA YX	0.590	CMO	0.58333
LAMBDA	0.00000	CMA	0.66667
ALPHA	0.00000	R	0.19822
W	0.66667	DYX	0.06333
Y	0.25000	DY	0.12500
KAPPA	0.12702	UYX	0.00762
W	0.18182	UY	0.01026
G	0.66667	U	0.00875
G	0.20000		

PHI2	0.01042
V	0.10206
LAMBDA YX	0.00000
LAMBDA YX	0.590
LAMBDA	0.00000
ALPHA	0.00000
W	0.66667
Y	0.25000
KAPPA	0.12702
W	0.18182
G	0.66667
G	0.20000

B I B L I O G R A P H I E

- [1] ADHIKARI B.P. Distance, discrimination et résumé exhaustif
 JOSHI D.D. (1956) Publ. Inst. Statis. Univ. Paris 5,57-74
- [2] AHMED A.H.N. Partial Ordering of positive quadrant dependence, with
 LANGBERG N.A. applications (1979)
 LEON R.V. Technical report AFOSR 78,3. Department of Statistics
 PROSCHAN F. Florida State University
- [3] ALI S.M. Association between random variables and the dispersion
 SILVEY S.D. of a Radon-Nikodym derivative (1964) JRSS B, 100-107
- A further result on the relevance of the dispersion of
 a Radon Nikodym derivative to the problem of measuring
 association
 (1964) JRSS B, 28, 131-142
- A general class of coefficient of divergence of on
 distribution from another (1966) JRSS B, 28, 131-142
- [4] ANDREWS F.M. A guide for selecting statistical techniques for analysing
 KLEM L. social science data (1978)
 DAVIDSON T.N. Survey research center - Institute for social research.
 O'MALLEY P.M. The University of Michigan
 RODGERS W.L.
- [5] ARMSTRONG R.D. Least absolute value estimators for one-way and two-way
 FROME E.L. tables (1978)
Communications in Statistics, Theory and Methods 79-96
- [6] d'AUBIGNY G. Fiches de mathématiques et de statistiques (1980)
 BON F. Ecole d'été d'informatique et mathématiques en sciences
 ROMIER Guy sociales - Multigraphié C.N.R.S. Grenoble

- [7] d'AUBIGNY G. Bases mathématiques de l'analyse linéaire des données (198)
Cours de D.E.A. - IRMA GRENOBLE
- [8] d'AUBIGNY G. Description statistique des données ordinales. Analyse
multidimensionnelle (1975) Thèse GENOBLE
- [9] BARTLETT M.S. Multivariate analysis (1947) J. Roy Stat. Soc. Suppl., B, 9, 1
- [10] BELL C.B. Mutual information and maximal correlations as measures
of dependance (1963) Ann math Stat 587-595
- [11] BERKSON J. Minimum chi square, not maximum likelihood (1980).
Ann. Stat 8 n° 3, 457-487
- [12] BISHOP Y.M.M. Discrete multivariate analysis (1976) The M.I.T. Press
FIENBER S.E. Cambridge - Massachussetts
HOLLAND P.W.
- [13] BLUMQUIST N. On a measure of dependence between two random variables
Ann. math. statist. 21, 593-600
- [14] BRONS H.K. Generalized means and associated families of distribution
BRUNK H.D. (1969) Ann Math Stat 40, n° 2, 339-355
FRANCK W.E.
HANSON D.L.
- [15] CAILLEZ F. Introduction à l'analyse des données (1976)
PAGES J.P. S.M.A.S.H. 9, rue Duban 75016 Paris
- [16] CARLIER A. Tables de contingence multidimensionnelles (1977) - Thèse
de Spécialité - Université Paul Sabatier Toulouse
- [17] CAZES P. Codage et analyse de tableaux logiques
BEAUMERDER A. Introduction à l'analyse des variables qualitatives (1977)
BONNEFOUS S. Cahiers du Bureau Universitaire de recherche opération-
nelle, cahier n° 27
- [18] COLLOMBIER D. Analyse des tables de contingence multidimensionnelles (19
Thèse d'état - Université Paul Sabatier TOULOUSE
- [19] COSTNER H.L. "Criteria for measures of association" (1965) Am. Sco.
Rev. 30, 341-353

- [20] COXHEAD P. Measuring the relationship between two sets of variables (1974) Brit. J. of Math. and Stat. psy. 27, 205-212
- [21] CRAMER E.M. Some symmetric, invariant measures of multivariate association (1979). Psychometrika 44,1, 43-54
NICEWANDER
- [22] CRAMER H. Mathematical methods of Statistics (1946)
Princeton Univ. Press Princeton U.S.A.
- [23] DANTELS H.E. The relation between measure of correlation in the universe of sample permutations (1944)
Biometrika, Vol. 33, 129-135
- [24] DAUDIN I.J. Etude de la liaison entre variables aléatoires
Regression sur variables qualitatives (1978) Thèse Paris XI.
- [25] DEMPSTER A.P. Covariance selection (1972) Biometrics 28, 157-175
- [26] DEVILLE J.C. Exhaustivité approchée dans les tableaux statistiques(1976)
Annales de l'INSEE n° 22-23, 191-211
- [27] DOMENGES D. Analyse factorielle sphérique : une exploration (1979)
VOLLE M. Annales de l'INSEE n° 35
- [28] EFRON B. Regression and ANOVA with zero-one data : Measures of residual variation (1978)
J. Am. Stat. Ass. Vol 73 n° 361, 113-121
- [29] ESCOUFIER Y. Le traitement des variables vectorielles (1973)
Biometrics 29, 751-760
- [30] ESCOUFIER Y. Echantillonnage dans une population de variables aléatoires réelles (1970) Publ. Inst. Stat. Univ. Paris 19
(fascicule 4) 1-47

- [31] FERON R. Information, regression (1954) Thèse d'état Faculté des sciences de Paris
- [32] FRECHET M. Sur les tableaux de corrélation dont les marges sont données (1951) Ann Univ. Lyon Section A Série 3,14, 53-77
- [33] GARNER W.R. The relation between Information and variance analysis
Mc GILL W.J. (1956) Psychometrika vol. 21 n° 3, 219-228
- [34] GINI C. Indici di concordanza (1915-16) Atti del reale istituto veneto di scienza lettere ed arti series 8, 75(2),1419-61
- [35] GINI C. Sul criterio di concordanza tra due caratteri (1915-16) Atti del reale istituto Veneto di scienza lettere ed arti series 8, 74 (2), 1903-42
- [36] GINI C. Le Medie (1958)
Unione Tipografico Editrice Torinese
- [37] GOODMAN L.A. Measures of association for cross - classification (1979)
° KRUSKAL W. Springer Series in Statistics - Springer Verlag Heidelberg
- [38] GRAYBILL E.A. Theory and application of the linear model (1976)
Duxbury Press. North Scituate Massachussetts
- [39] GREUB W.H. Multilinear Algebra (1967)
Springer Verlag Heidelberg
- [40] HABERMAN S.J. The analysis of frequency data (1974) Chicago University Press Chicago
- [41] HAJEK J., SIDAK Z. Theory of rank tests (1967) Academic Press New York
- [42] HILDEBRAND D.K. Prediction analysis of cross classification (1977)
LAING
ROSENTHAL H. Wiley

- [43] HOTELLING H. Relations between two sets of variates (1936)
Psychometrika 28, 321-377
- [44] HUBER P.J. Robust estimation of a Location parameter (1964)
Ann. Math. Stat 35 (73-101)
- [45] HUBERT L.J. Matching models in the analysis of cross-classifications(1979)
- [46] KENDALL M.G. Rank Correlation methods (1962)
Charles Griffin and Company London (third. ed.)
- [47] KIM J. Predictive measures of ordinal association (1971)
J. Ann. Stat. Ass. 53, 891-907
- [48] KIMELDORF G. Monotone dependance (1978)
SAMPSON A. Ann. Statist 6, 895-903
- [49] KRUSKAL W.H. Ordinal measures of association (1958)
J. Ann. Stat. Ass 53, 814-861
- [50] KULLBACK S. Information Theory and Statistics (1959)
Wiley
- [51] LANCASTER H.O. Corrélation and complete dependance random variables (1963)
Ann. Math. Statist. 34, 1315-1321
- [52] LANCASTER H.O. The chi square distribution, (9169) Wiley
- [53] LAWLEY D.N. A generalization of Fisher's Z-test (1938) *Biometrika* 30,180-
- [54] LE CALVE G. Systèmes aléatoires à liaison complète (1976) 187
Thèse d'état - Université de Rennes
- [55] LEHMAN E.L. Testing statistical Hypothèses (1959) Wiley
- [56] LEHMAN E.L. Same concepts of dependance (1966)
Ann. math. statist. 37 1137-1153

- [57] LERMAN J.C. Etude distributionnelle de statistiques de proximité entre structures finies de même type. Application à la classification automatique (1973) Cahiers du bureau universitaire de recherche opérationnelle, 19, I.S.U.P.
- [58] LORD F.M. Statistical theories of mental test scores (1968)
NOVICK M.R. Addison Wesley Publishing Company, Reading
- [59] MARGOLIN B.H. An Analysis of variance for categorial data. (1971)
LIGHT R.J. J. Ann. Stat. Ass. 69 755-756
- [60] MARGOLIN B.H. An Analysis of variance for categorial data. Small sample comparison with chi square and other competitions (1974)
LIGHT R.J. J. Ann Stat. Ass. 69 755-764
- [61] MATUSITA K. On the theory of Statistical decision functions" (1951)
Ann. Inst. Stat. Math. Tokyo Vol III 17-35
- [62] MULAİK S.A. The foundations of factor analysis (1972)
Mc GRAW-HILL Book C^{NY} New York
- [63] NELDER Generalized linear models (1972)
WEDDERBURN R.W.M.. J.R. Statist. Soc. A. 135, 370-384
- [64] NIE et Alii SPSS PRIMER (2nd edition) Mc GRAW-HILL BOOK Company New Yor
- [65] PFANZAGL J. Theory of Measurement (1971) (2ème édition)
Physica Verlag Würzburg Wien
- [66] POON D.A. Robust measures of association (1977)
Phil. doc. Thesies . University of California - Berkeley
- [67] RAO C.R. Linear Statistical Inference and Its Applications (1965)
(2ème édition) WILEY
- [68] RENYI A. Calcul des probabilités (1966)
Dunod

- [69] RENYI A. On measures of dependance (1959)
Act. Math. Budapest. 10 441-451
- [70] SHARMA B.D. Entropy as an optimal measure (1978)
 GUPTA H.C. Colloque International du CNRS n° 276
 Thème de l'information 151 159
 Editions du CNRS Paris
- [71] SOMERS R.H. "Analysis of partial rank correlation measures based on
 the product moment model : Part one" (1974)
 Social forces 53 229-246
- [72] STEVENS S.S. Mathematics, measurements and psychophysics (1951)
 Hand book Exp. Psychologie, New York, 1-49
- [73] SUPPES P. Basic measurement Theory (1963)
- ZINNES J.L. Handbook of Mathematical Psychology
 Ed. R.D. LUCE, R.R. BUSH et E. GALANTER
 Wiley New York Vol. 1 chap. 1-8
- [74] TOGERSON Theory and Methods of Scaling (1958) Wiley
- [75] VAN DER PYL Propriétés de l'Information d'ordre α et de type β (1978)
 Colloque International du CNRS n° 276 - Thème de
 l'Information, 161-171 - Editions du CNRS - Paris
- [76] WALD A. Selected Papers in Statistics and probability
 (1969) Stanford University Press, Stanford California
- [77] WALLIS W.A. The correlation ratio for ranked data (1948)
 J. Ann. Stat. Ass. 33 533-538
- [78] WEISBERG H.F. Models of Statistical Relationships (1974)
 The America Political Science Review Vol. 68 - 1638-1655
- [79] WERMUTH N. Analogies between multiplicative models in contingency
 tables and covariance selection (1976)
 Biometrics 32 95-108

Dernière page d'une thèse

VU

Grenoble, le 19.9.80

Le Président de la thèse

Vu, et permis d'imprimer,

Grenoble, le 19.9.80

Le Président de l'Université Scientifique et Médicale

G. Cau

G. CAU

