



HAL
open science

Formation à la statistique par la pratique d'enquêtes par questionnaires et la simulation : étude didactique d'une expérience d'enseignement dans un département d'IUT

Jean-Claude Oriol

► To cite this version:

Jean-Claude Oriol. Formation à la statistique par la pratique d'enquêtes par questionnaires et la simulation : étude didactique d'une expérience d'enseignement dans un département d'IUT. Education. Université Lumière - Lyon II, 2007. Français. NNT : . tel-00191166

HAL Id: tel-00191166

<https://theses.hal.science/tel-00191166>

Submitted on 27 Nov 2007

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

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

Université Lumière Lyon 2

Ecole Doctorale EPIC

Institut de Sciences et Pratiques de l'Éducation et de la Formation

Formation à la statistique par la pratique d'enquêtes par questionnaires
et la simulation : étude didactique d'une expérience d'enseignement
dans un département d'IUT

Par Jean-Claude ORIOL

Thèse de Doctorat en Sciences de l'Éducation

sous la direction de Jean-Claude RÉGNIER

présentée et soutenue publiquement le 17 novembre 2007

Devant un jury composé de :

Abdelaziz BOURAS, Professeur, Université de Lyon 2

Régis GRAS, Professeur Émérite, Université de Nantes, Rapporteur

Michel LE NIR, Maître de Conférences, Université Lyon 2

Jean-Louis PIEDNOIR, Inspecteur Général Honoraire de Mathématiques

Jean-Claude RÉGNIER, Professeur, Université Lyon 2

Gérard VERGNAUD, Directeur de Recherche Émérite, C.N.R.S., Rapporteur

Remerciements

Analyser de quelles façons la pratique d'enquête par questionnaires et la simulation sont des composantes incontournables d'un apprentissage de la statistique emmène ceux qui conduisent de telles études de la réalité rassurante des pédagogies innovantes vers les pentes abruptes de la recherche.

Nul ne doute que cette distanciation ne puisse se faire avec profit que si l'on appartient à une communauté réunie par la volonté de certains et autour de sujets semblables et c'est ce que j'ai trouvé dans le groupe de recherche ADATIC fondé et animé par Jean-Claude Régnier. De ma rencontre avec lui est né le projet de ce travail de thèse. J'avais des idées, trop disent certains et je crois que cela n'est pas toujours un compliment, sur l'enseignement en général et sur l'enseignement de la statistique en particulier. Jean-Claude Régnier a su m'aider à serrer un sujet et m'empêcher de passer d'une innovation à l'autre dans un système de « fuite en avant ». C'est une véritable association qu'il a su installer entre nous faite de respect, de chaleur et d'amitié. Depuis notre première rencontre nous avons cosigné quelques articles et j'espère que ce travail en commun se prolongera longtemps. Travailler avec JC Régnier sur l'enseignement de la statistique est pour moi une chance évidente, et si les relations amicales qu'il aime à cultiver prennent souvent les chemins de discussions passionnées, elles se doublent, et c'est sans doute le plus important, d'un regard scientifiquement aiguisé allant du particulier à l'essentiel, et ainsi, directement au cœur des problématiques. Je le remercie vivement et ceci n'est pas une formule.

Je remercie Régis Gras et Gérard Vergnaud d'avoir été les rapporteurs de ce travail. Je suis sûr que leurs remarques et leurs commentaires enrichiront mes réflexions sur les sujets

traités ou sur des points annexes. Je remercie Abdelaziz Bouras d'avoir accepté d'être membre de ce jury ; sa présence bienveillante à la direction du Cerral¹ nous a permis de discuter de sujets variés, comme la simulation par exemple, et sa vision globale et non sectaire des activités de recherche m'a donné l'occasion de participer à plusieurs colloques. Je remercie Michel Le Nir d'avoir été en quelque sorte un des catalyseurs de ce projet : il m'a demandé il y a plusieurs années d'encadrer avec lui des étudiants menant un projet d'enquête par questionnaires en partenariat avec la TCL (Transports en Commun Lyonnais) ; il s'agissait d'une enquête de satisfaction des usagers, les étudiants devant administrer 1200 questionnaires en une journée, puis analyser, rédiger et présenter les résultats de cette enquête ; nous avons d'ailleurs, présenté et cosigné un article décrivant cette activité. C'était la première enquête réelle que je mettais au service de l'enseignement de la statistique et j'en ai immédiatement compris les retombées en termes pédagogiques. En tant que directeur de l'IUT il a essayé, en ces derniers temps de rédaction, de protéger des espaces dévolus à l'écriture, et son intérêt pour les enquêtes ne pouvait être qu'un lien de plus entre nous ; connaissant la surcharge de son emploi du temps, je le remercie particulièrement d'avoir accepté d'être dans ce jury. Je remercie Jean-Louis Piednoir dont le regard critique sur l'enseignement de la statistique et l'expertise dans le champ de l'enseignement m'ont toujours apporté idées et éclaircissements.

Je remercie mes collègues du département STID (Statistique et Traitement Informatique de Données) surtout Hélène et Anne d'avoir eu à supporter quelques contraintes d'emploi du temps, les étudiants qui ont mené à bien un grand nombre d'enquêtes avec sérieux montrant leurs aptitudes de statisticiens, et bien entendu la sphère de mes amis et de ma famille pour leur présence à mes côtés.

¹ Le **CERRAL** est un centre d'études et de recherche-action de l'IUT Lumière dont les activités sont liées aux formations spécifiques de L'IUT Lumière et accompagnent l'évolution de certaines de leurs facettes : les études en alternance, l'interaction entre disciplines des Sciences de l'Ingénieur et des Sciences humaines et Sociales, le transfert de technologie, etc.. Le centre CERRAL héberge actuellement, entre autre, deux équipes de recherche : [l'équipe de recherche PRISMa Lyon 2](#) (EA 2058, Ecole doctorale EDIIS) et [l'équipe de recherche sur l'Alternance et l'Apprentissage](#). De part sa proximité des formations de l'IUT, le centre offre un cadre de coopération et d'échange indispensable à l'ensemble des intervenants de l'IUT et apporte une aide continue aux projets et équipes de recherche qu'il héberge. Le CERRAL est le lieu par excellence pour expérimenter des recherches pluridisciplinaires à travers des projets communs et confronter des points de vue différents (in http://iut.univ-lyon2.fr/article.php?id_article=300)

Introduction

La recherche et la découverte sont généralement considérées comme des moteurs pour l'apprentissage et paradoxalement le dispositif d'enquête (l'origine de ce mot provient en effet du début du XII^e enquête, du latin populaire inquæsita, du latin classique inquisita, ayant comme substantif inquirere qui signifie «chercher à découvrir») n'est au centre que d'un petit nombre de cours de statistique. La "littérature didactique" apparaît, en outre, relativement pauvre sur ce sujet comme si le processus puisait son importance dans son déroulement même.

Peut-être cette singulière mise en retrait s'enracine-t-elle dans la double nature de la statistique et/ou des statistiques ?

En effet, un certain nombre d'enseignants chargés d'un cours de statistique mettent l'accent sur la différence entre "la statistique" et "les statistiques"; ces dernières auraient comme objet de présenter des résultats, tandis que la statistique aurait le statut d'une science. Cette question a en particulier été abordée par J-C Régnier (2002, 2006) qui considère qu'il y a là une caractérisation dont la didactique de la statistique doit se saisir.

Comme on peut le remarquer dans l'ouvrage "Éléments d'histoire des sciences" dirigé par Michel Serres, c'est dans le passage de la table des relevés des observations à la loi que se construit une science. Les exemples sont nombreux et les noms sont célèbres : Newton et la gravitation, Berthollet et Lavoisier et l'affinité en chimie, Galilée et le mouvement des planètes, Mendeleïev et le classement périodique des éléments, etc. (SERRES, 1989).

Ce constat est décomposé pour la statistique par Colin Ronan dans "Histoire mondiale des sciences" quand, d'une part il relève "*On peut dire que les statistiques débutèrent au XVII^e siècle, lorsqu'une série de relevés de noyades ayant abouti à la mort fut effectuée à Londres, en 1662, par John Gaunt et William Petty*"(RONAN, 1988, p.604), et que d'autre part il souligne que les travaux de Quételet, Poisson, Galton et Pearson fondèrent la statistique : "*Nommé professeur de mathématiques au University College en 1884 il [Pearson] accepta en 1911 la chaire fondée par Galton et l'occupa avec distinction pendant les vingt-deux années qui suivirent, posant les solides fondations mathématiques de toute la science statistique du XX^e siècle...*" (RONAN, 1988, p.608).

Cette différence se retrouve dans les ouvrages d'histoire des mathématiques qui souvent ne citent que "Calcul de probabilités" ou parfois "statistique mathématique" dans leurs index.

Partant d'un autre point de vue, Adolphe Quételet écrit dans une lettre intitulée "La statistique est-elle un art ou une science" "*...toutes les sciences d'observation, à leurs début, ont subit les mêmes phases ; c'était des arts, car elles se bornaient à grouper d'une manière plus ou moins heureuse des collections de faits appartenant à un même ordre des choses ; et c'est par le rapprochement et l'étude de ces faits qu'elles se sont élevées ensuite au rang où on les voit briller aujourd'hui.*" (Dans la Lettre au Duc de Saxe-Cobourg et Gotha, cité par Droysbeke J.-J. et Tassi Ph. in Histoire de la statistique, PUF, 1997, p. 5).

Et l'on voit encore cette double origine dans l'enseignement de la statistique : d'un côté les tenants d'une approche liée à l'histoire et à l'épistémologie de cette science et privilégiant les études de cas, le terrain, etc., de l'autre, les partisans d'une approche plus formelle faisant généralement précéder la statistique par le calcul des probabilités. On retrouve alors les mêmes principes pour la statistique et son enseignement que ceux énoncés par André Revuz pour les mathématiques dans "Est-il impossible d'enseigner les mathématiques ?" avec en particulier "*ne pas massacrer les mathématiques au nom des applications, ne pas négliger ces applications au nom d'une pureté mal comprise*" (REVUZ, 1980, p.142).

Nous n'interviendrons pas ici dans ce débat et nous nous placerons a priori du côté de ceux qui font d'abord faire des statistiques comme un élément à l'intérieur du cours de statistique. Notre projet a été plutôt de décrire et d'analyser des travaux d'enquêtes par questionnaire réalisés par des étudiants en situation professionnelle, leurs conditions de réalisations, leurs places dans les pratiques d'enseignement de la statistique. Il s'agissait de voir comment le travail d'enquête par questionnaires facilite la construction des connaissances de l'apprenant, donne du sens à ses apprentissages et augmente sa capacité à utiliser les outils statistiques dans un cadre professionnel et de quelle façon complémentaire la simulation vient améliorer la compréhension des phénomènes statistiques.

Notre propos sera ici centré sur l'apprentissage de la statistique par des étudiants dans un IUT, i.e. quels principes viennent étayer notre réflexion, quelles sont les méthodes que nous avons mises en place dans ce cadre et quels sont les résultats que nous avons obtenus ?

Afin d'essayer d'atteindre cet objectif nous avons structuré notre travail de la façon suivante :

- une première partie considérant l'histoire de la construction de la statistique en privilégiant la statistique et en se gardant de glisser vers une histoire du calcul des probabilités,
- une deuxième partie donnant des éclaircissements sur le cadre conceptuel et le contexte de la recherche,
- une troisième partie montrant des exemples d'enquêtes par questionnaires réalisées par les étudiants,
- une quatrième partie développant la place de la simulation dans la construction des connaissances statistiques ; nous essayerons de montrer comment grâce à elle, les étudiants peuvent dégager les invariants de la variabilité, et atteindre à une pensée statistique.
- une cinquième partie décrivant deux enquêtes effectuées par des étudiants de deuxième année du département Statistique et Traitement Informatique des Données (STID) de l'IUT Lumière (Lyon 2), ces enquêtes utilisent l'Analyse Statistique Implicative (ASI) et ont pour objet les représentations de la statistique chez un sous-ensemble des étudiants d'IUT.

En résumé ce travail de thèse essaiera de montrer comment l'enquête par questionnaire et la simulation participent à la construction de l'esprit statistique dans un enseignement s'adressant à des étudiants d'un département d'IUT.

Pour tenter de répondre à cette question, nous avons analysé les données construites à partir des observations sur le terrain pour mieux comprendre les avantages et les obstacles liés à cette ingénierie pédagogique, en ramenant notre propos dans le temps et dans l'espace des apprentissages : i.e. quelles sont les statistiques que l'on voudrait que soient capables de manipuler nos étudiants, avec quelle maîtrise, et dans quel dessein ?

1. Partie I

Une approche historique de la statistique : des premiers pas à la « géométrie du hasard » et aux « maîtres de l'erreur »

Aborder ce travail par une étude de l'histoire de la statistique peut apparaître au premier abord une approche banale et sans grand intérêt ; banale car le déploiement d'une question dans l'espace et dans le temps est une approche traditionnelle, et sans grand intérêt en raison des ouvrages que l'on peut citer pour circonscrire l'histoire de la statistique.

Nous nous inscrirons en faux avec ce liminaire car loin de nous laisser guider par cette première partie vers une chronogénèse des concepts statistiques nous marquerons les différences entre une approche que l'on peut lier à l'histoire de l'apparition des outils statistiques et une démarche plus proche d'une topogénèse des concepts statistiques.

Ainsi notre travail donnera une situation privilégiée à l'enquête comme lieu contribuant à la fois aux interrogations précédant la mise en place d'outils théoriques mais également comme sujet et occasion de justifier, comme a posteriori, la mise en place de théories novatrices en leurs temps.

Bien entendu étudier l'histoire de la statistique en tant que mode de connaissance du monde ne peut se faire sans approfondir les outils formels développés par le « calcul des probabilités » puis par la « statistique mathématique » qui en constituent à la fois la preuve devant la communauté scientifique mais aussi qui nourrissent un ensemble de techniques, de résultats et de théorèmes. Nous rencontrerons également dans cette étude les fondements matériels et institutionnels qui ont accompagné le développement de la statistique, justifiant son existence dans le cadre des états où elle assure une fonction à la fois utilitaire et symbolique.

Nous aurons donc à parcourir ce champ sous un triple éclairage : celui de l'évolution de la science, celui de l'organisation des États et celui relatif à la cité, à la chose publique, au gouvernement de l'état ou plus simplement à la politique.

1.1. Des débuts prometteurs

Il semble qu'à ses débuts la statistique soit à la fois le signe de la naissance d'un pouvoir que ce dernier ait été le fait d'une cité ou du moins d'un état centralisé et d'un pouvoir religieux les deux étant souvent confondus.

1.1.1. Les systèmes artificiels à mémoire (SAM)

A l'origine du stockage de l'information on a trouvé depuis une centaine d'années un certain nombre d'objets gravés, principalement sur des os ou des bois de rènes au Paléolithique supérieur (environ -35 000 ans en Europe et -60 000 ans en Afrique).

Souvent les premières interprétations de l'utilisation de ces objets étaient qu'ils étaient dans le registre du rite, i.e. fréquemment quand on ne comprend pas la signification ou l'utilisation d'un objet on décide que c'est un objet appartenant au registre du religieux. C'est ce que souligne Francesco d'Errico² lorsqu'il met en évidence la difficulté de l'interprétation de ces objets archéologiques. Ainsi c'est à partir de ces objets que ce chercheur construit des modèles ethnographiques qui l'amènent à répondre à la question de la « modernité de l'homme » et à substituer au modèle du « big bang culturel » une évolution plus régulière entre -200 000 et -20 000 avant JC³.

1.1.2. Des actes fondateurs : les listes, les nombres

1.1.2.1. L'usage de listes

C'est dans les premières listes que l'on reconnaît à la fois les premières manifestations de désignation d'objets et du concept de nombre. Cela ressemble fort à une naissance d'un

² Francesco d'Errico, directeur de recherche au CNRS, Institut de Préhistoire et de Géologie du Quaternaire, CNRS UMR 5199 PACEA

³ « Un modèle a longtemps été admis pour rendre compte de cette grande transformation. La modernité de l'homme serait associée à une brusque révolution culturelle ayant eu lieu il y a 40 000 ans environ, soit au début du Paléolithique supérieur. Cette mutation culturelle aurait eu lieu en Europe et coïnciderait avec l'arrivée des hommes anatomiquement modernes, des hommes comme nous en somme, sur le vieux continent. Ce changement a longtemps été considéré comme soudain et explosif. Contre le modèle du big bang culturel du Paléolithique supérieur, un autre modèle a été récemment proposé. Selon ce scénario la modernité culturelle aurait débuté en Afrique, le continent où selon la génétique notre espèce a eu son origine il y a environ 200 000 ans, et se serait déployée par étapes entre 200 000 et 20 000 ans BP (au cours de l'époque dite du Middle Stone Age africain). Ce modèle suppose donc une évolution beaucoup plus graduelle et non européenne. En particulier le deuxième modèle, qui est en train de s'imposer comme le paradigme dominant, lié directement à l'origine biologique de notre espèce à l'origine des comportements modernes. L'idée est simple : le processus qui a produit notre espèce en Afrique a du octroyer à cette dernière certains avantages (langage, pensée symbolique, capacités cognitives supérieures) qui ont favorisé sa colonisation de l'Eurasie et le remplacement par celle-ci des populations humaines vivant dans ces régions » in Conférence de Francesco d'Errico au Cercle Genevois d'Archéologie, 4 avril 2006.

couple : l'écriture et les nombres avec comme première conséquence un forme de statistiques correspondant au classement et à la présentation des données.

C'est d'ailleurs ce que fait remarquer Jack Goody (Goody J., 1993, p.167) dans un entretien ayant pour sujet « Ecritures et sociétés » :

« Les études menées récemment sur les débuts de l'écriture indiquent en effet qu'elle s'est développée à Sumer surtout pour des raisons de comptabilité : les marchands utilisaient ce moyen pour envoyer des messages à d'autres marchands. Il s'agissait de brèves indications (de mots isolés ou numéros) portées sur des pots ou des tablettes, de listes de mots apparemment destinées à instruire les élèves dans les écoles. Ces premières écritures ne comportent ni phrase ni paragraphe. »

Evidemment nous savons qu'avec le temps l'écriture va remplir des fonctions fort différentes, passant de la liste de couples (description d'un objet, nombre) à des fonctions plus transcendantes citons par exemple la description du cadastre, algorithmes de calcul d'aires, contrats de commerce, etc. et surtout transcription du discours ou du langage parlé. Ainsi Goody (Goody J., 1993, p.167) poursuit :

« Des écrits plus proches de la parole se développèrent bien entendu sous forme de contrats, de codes, et même de documents historiques et littéraires. Mais les types d'écritures non liées au discours jouèrent un rôle particulièrement important, favorisant le développement des mathématiques et de ses applications (comme par exemple le calcul de l'aire d'un champ). C'est dans la période babylonienne du premier millénaire qu'apparaissent des notations liées au cadastre ainsi que des listes concernant le lever du soleil et de la lune, permettant ainsi l'apparition d'une vraie astronomie prévisionnelle. »

On voit donc qu'en quelques pas nous sommes passés de listes d'objets et de nombres, c'est-à-dire un début des statistiques descriptives à une véritable approche de statistiques prévisionnelles.

Ce double aspect de ces listes, d'une part le codage réservé à ceux qui en connaissent les clés, et d'autre part la prévision d'un futur par essence imprévisible, va installer pour longtemps respect et défiance vis-à-vis du statistique. Comment en effet ne pas être empli de respect pour celui qui sait à l'avance ce que l'on va trouver dans une jarre ou dans une tablette d'argile et ne pas craindre celui qui par la lecture des tables arrive à prévoir les phases de la

lune ou l'arrivée de Sirius permettant de prévoir les crues du Nil (même si ces dernières comportent, on s'en doute, quelques imprécisions) ?

Tout au long de notre travail nous rencontrerons cette facette de la statistique comme une entrave à son évolution, mais également dans la partie didactique de ce travail comme un obstacle au développement d'une pensée statistique chez l'individu.

1.1.2.2. Une invention de comptables

Georges Ifrah (Ifrah, 1981, p. 160) intitule sous forme interrogative « *Les inventeurs de l'écriture : une invention de comptables ?* » le chapitre 10 de son ouvrage « Histoire universelle des chiffres ». Notons par anticipation que l'on retrouvera sous forme de quolibet le nom d'actuaire dans la bouche des bourbakistes pour désigner les statisticiens (INSEE, 1977).

Mais revenons à Georges Ifrah qui poursuit l'opinion contenue dans le titre du chapitre :

« *L'idée de l'écrit, en Mésopotamie comme en Elam, serait donc bien née de besoins strictement utilitaires dans un contexte essentiellement économique.* » (Ifrah G., 1981, p.178).

Et donc l'écriture en tant que version du langage articulé ne s'est imposée que plus tard, c'est ce que l'auteur renforce quelques lignes après :

« *Il ressort, en tout cas très nettement de ce qui précède que la comptabilité écrite a très vraisemblablement précédé, en Elam et sans doute aussi en pays de Sumer, la transcription du langage articulé.* » (Ifrah G., 1981, p.178).

Et pour résumer nous pouvons dire que les besoins statistiques et comptables ont entraîné la naissance de l'écrit, et que cet écrit a ensuite rempli les fonctions qu'on lui connaît.

1.1.3. Une origine de la statistique descriptive

Nul besoin n'est ici de rappeler le « miracle grec » à partir du VII^{ème} siècle avant JC qui permit l'instauration de la logique, de la démocratie et de la méthode scientifique.

Le philosophe Aristote⁴ (Stagire, Macédoine, aujourd'hui Stavro, - 384 - Chalcis, Eubée, -322) est un de ceux qui dans un traité de politique intitulé « Constitution d'Athènes »⁵ aborde une forme de statistique descriptive, telle qu'on peut le constater dans les quelques lignes ci-dessous tirées du chapitre XXV⁶ :

« En même temps, suivant la politique inaugurée par Aristide, ou assura à la multitude largement sa subsistance. Il arriva que par les contributions extraordinaires, par les droits et impôts, par les alliés, plus de vingt mille hommes étaient nourris. Il y avait en effet six mille juges, mille six cents archers et en outre douze cents cavaliers ; le Conseil comptait cinq cents membres, les gardes des arsenaux étaient au nombre de cinq cents, et les gardes en ville au nombre de cinquante ; environ sept cents hommes exerçaient des magistratures dans le pays ; environ autant, en dehors du pays. Plus tard, quand Athènes eut entrepris la guerre, il y eut deux mille cinq cents hoplites, vingt vaisseaux croiseurs, d'autres vaisseaux pour la perception des tributs ayant à bord les deux mille hommes désignés par le sort. Ajoutons le prytanée, les orphelins, les géôliers. Tout ce monde tirait sa subsistance des revenus publics. »

Thomas d'Aquin au XIII^{ème} siècle donne aux idées d'Aristote un statut de doctrine officielle de l'église, cette ligne de l'église romaine a pour conséquence immédiate une influence considérable des idées d'Aristote et ses textes ont inspiré les premières études statistiques en Europe en Occident, jusqu'au XVI^{ème} siècle : en effet à cette période, Galilée, Torricelli et Pascal remettent en cause le système aristotélicien à partir de données expérimentales.

⁴ en grec Aristotelês, dit le Stagirite

⁵ Le texte en question provient d'un papyrus acheté en Egypte en 1879, une querelle est toujours en cours concernant le fait qu'Aristote soit l'auteur ou non de ce texte, cependant il existe de nombreux témoignages de l'antiquité attestant de la réalité d'un recueil d'un grand nombre de constitutions des cités grecques par les élèves d'Aristote.

⁶ Aristote. Constitution d'Athènes, traduite par B. Haussoullier, Paris : E. Bouillon, 1891, Bib. de l'École des hautes-études. Sciences philologiques et historiques. Fascicule 89

1.1.4. D'autres approches de la statistique

Si l'on cherche ailleurs que dans une filiation directe on trouve plusieurs exemples montrant l'universalité de la démarche statistique

Le cas chinois : l'empereur chinois Yao, ordonne le recensement des productions agricoles en 2238 avant J.-C. in Gerbaud (Gerbaud, 2006, p. 3) ; l'Egypte : dès 1700 avant J.-C. le pharaon Amasis décrète la peine de mort contre celui qui ne se plie pas aux déclarations de son nom, sa profession et ses moyens de subsistance ; l'exemple indien de l'Arthasastra : au IV^{ème} siècle avant J.-C. Kautilya, ministre du roi Candragupta (Empire indien des Maurya) rédige un traité de science politique et économique qui décrit les techniques de recensement des populations ; et enfin citons les quipus des Incas qui étaient des systèmes fondés sur des cordelettes, des nœuds et des couleurs permettant d'avoir des statistiques sur les récoltes.(Morlat, 2000, p. 3).

Ces divers exemples brièvement abordés montrent bien l'universalité de la démarche statistique. Cet aspect global de l'approche statistique par les états se retrouve dans un autre élément présent dans les débuts de la statistique à savoir la mesure des sols et des propriétés tels qu'ils apparaissent dans la constitution des divers cadastres, c'est ce que nous allons aborder maintenant.

1.2. Le cadastre

1.2.1. Liminaire lapidaire

L'établissement du cadastre⁷ est dans un premier temps une des pierres sur laquelle nous avons construit la statistique et ses outils. Sans vouloir en faire une étude historique exhaustive donnons quelques points de repères tels que les présente Jean-Paul Miserez⁸, cette chronologie étant reprise par l'Institut de la gestion publique et du développement économique (IGPDE). Nous distinguerons trois périodes : une première époque couvrant l'Antiquité et le Moyen Âge, puis de la Renaissance à la Révolution française et enfin de cette dernière à la fin du XIX^{ème}.

⁷ Les données de ce paragraphe proviennent des documents de l'Institut de la gestion publique et du développement économique (IGPDE)

1.2.2. Les périodes antique et médiévale

J.P. Miserez nous donne un aperçu des pratiques concernant le cadastre et le cens depuis l'Antiquité jusqu'au Moyen Âge. Nous avons résumé dans le tableau suivant :

Période	Lieu	Souverain	Evènement / découverte
Env. 4000 av JC	Chaldée		Tablette donnant le plan coté, la superficie, la description d'un groupe de parcelles de la ville de Dunghi.
3200-2800 av JC	Égypte		Recensement foncier en en vue de la perception de l'impôt, basé sur la surface des propriétés et leur rendement
2238 av JC	Chine	Empereur Yao	Recensement des productions agricoles.
2000 av JC	Égypte	Sésostris	Les terres sont divisées en parcelles qui déterminent l'assiette de l'impôt foncier.
1600-1400 av JC	Italie		Plan gravé sur un rocher plat. Les lignes figurent les ruisseaux, les canaux d'irrigation et les routes.
Env. 1200 av JC.	Israël	Josué	Organisation territoriale et tribale d'Israël. Etablissement d'un cadastre et le partage des sols entre les tribus (Josué, XVIII, 4-9).
Env. 700 av JC	Lucanie		Cadastre sur 10000 hectares divisés en lots rectangulaires de 6 hectares chacun.
Vers 578-535 av JC	Etrurie	Servius Tullius	Cadastre de Rome.
63 av JC - 14 ap. JC	Gaule	Auguste	Etablissement du cadastre.
77 après JC	Gaule		Gravure du cadastre d'Orange.
380 après JC	Rome	Théodose I le Grand	Registre public : contenance, nature et qualité des biens par déclaration des propriétaires.
645-649	Japon	Ere de Taïka	Confection de registres d'état-civil et du cadastre.
Env. 1000	Inde	Raja le Grand	Confection d'un cadastre
1086	Angleterre	Guillaume le Conquérant	Fin de l'établissement du Domesday Book
1303 et 1493	France		. La royauté tente « à plusieurs reprises de faire dresser, çà et là, un "cadastre, compoix ou estime", local d'abord, puis régional et enfin général, qui servît d'assiette équitable des subsides demandés au royaume ⁹ . »
1364-1380	France	Charles V	Organisation des « parcelles »
1368-1398	Chine	Hongwu	Recensement de la population et cadastre général

⁸ ingénieur-géomètre, chargé de cours à l'École polytechnique de Lausanne

⁹ DUPONT-FERRIER (G.). – Études sur les institutions financières de la France. – Paris, Librairie Firmin-Didot, 1930-1932, 2 vol., I, p. 10 et II, p. 48, 136, 140, 141 et 246. [BH : 8° 14207.]

On le voit dans cette énumération le cadastre, le cens sont les outils qui permettent au pouvoir de l'état de dégager les ressources nécessaires à son fonctionnement.

1.2.3. De la Renaissance à la Révolution française

La maîtrise et la connaissance des outils scientifiques se propagent durant cette période traversée par de grands changements et on essaie d'installer des réformes

Période	Lieu	Souverain	Evènement / découverte
1427	Florence		Cadastre
1467	Tabriz (Iran)	Uzun Hasan	Cadastre agraire
A partir de 1599	Vietnam du Nord	Trinh- Cuong	Elaboration d'un cadastre

Les plans cadastraux se succèdent alors, on a ainsi et parmi d'autres : 1664.Montauban (sous l'impulsion du ministère Colbert), 1670 Lausanne, 1713-1740 Prusse (Frédéric-Guillaume Ier), 1716-1718 Catalogne (Philippe V), 1730 Savoie (Charles-Emmanuel III), 1740 Bâle (prince évêque), 1781-1786 Sicile (le vice-roi Caracciolo n'arrive pas à établir le cadastre).

1.2.4. De la Révolution française à la fin du XIX^{ème}

Cette période nous est plus proche et nous en trouvons encore les traces dans les mairies des villes et des villages¹⁰.

Beaucoup de débats ont eu lieu pendant la révolution française : fallait-il un cadastre général ? Charles Gomel (GOMEL, 1896, p.130-131) rapporte que « (...) *par suite d'illusions qui parfois se rencontrent encore aujourd'hui dans les discussions des Chambres* [il écrit en 1896 sur 1789], *plusieurs cahiers exprimèrent l'espoir qu'un cadastre général pourrait servir de base à la répartition de l'impôt foncier entre les différentes provinces. L'expérience a prouvé depuis que le cadastre parcellaire n'est utile que pour la répartition entre propriétaires d'une même commune.* »

Jean Pigeire rapporte qu'un peu plus tard en nivôse an XI (décembre 1800) Chaptal¹¹, ministre de l'intérieur, envoie « *les instructions successives pour l'établissement d'un*

¹⁰ Il existe des mairies où dans certains cas on ressort le « cadastre napoléon »

cadastre national, garantie des cultivateurs, base des jugements et des contributions », (PIGEIRE, 1932).

En 1807, Napoléon décrète une loi sur le cadastre : ce dernier devait être un instrument fiscal et administratif, mais aussi, fait nouveau, juridique.

Il fallut presque cent ans pour qu'entre 1891 et 1898, les chambres décident le renouvellement du cadastre à la suite de l'exposé du budget pour 1892¹², afin qu'il serve aussi d'instrument juridique.

Durant ce siècle le cadastre avait été généralisé dans tous les états modernes, de la Prusse à l'Italie et de la Belgique à l'Australie (Act Torrens 1858).

1.2.5. En guise de conclusion de cet aperçu de l'histoire du cadastre

De cette rapide histoire du cadastre nous tirerons quelques conclusions se rapportant à notre travail :

- Le cadastre a été en premier lieu un instrument fiscal et/ou administratif
- Les perfectionnements scientifiques ont permis une plus grande précision et une fiabilité accrue des cadastres
- Au fil des siècles les sociétés en ont aussi fait un instrument juridique

La statistique fonctionne de manière homéomorphe au cadastre : les ingrédients sont différents mais sa structuration au fil des siècles lui a fait également changer de statut.

¹¹ Ministre de l'intérieur de 6 novembre 1800 au 4 août 1804

¹² ROUVIER (Maurice), ministre des Finances. – Discours prononcé par M. Maurice Rouvier, ... séance du 24 octobre 1891. Discussion du projet de loi portant fixation du budget général de l'exercice 1892. – *Paris, impr. des journaux officiels*, 1891. « Le cadastre perpétué à l'aide d'un système permanent de conservation, ne serait pas seulement un instrument fiscal et administratif : il devrait satisfaire à d'autres besoins. Des abornements généraux et une triangulation rigoureuse précéderaient le renouvellement des opérations : le cadastre constituerait la base de la propriété foncière ; il assurerait la sécurité des hypothèques et la régularité des transactions immobilières ; il fournirait enfin à l'agriculture, par le développement des institutions de crédit, les moyens d'action qui lui font défaut aujourd'hui. En un mot, il deviendrait le Grand Livre terrier de la France ; » (§ 26 et 27.) [BnF : 8° Le⁹⁰. 957.]

En effet elle est passée d'un instrument essentiellement destiné à dénombrer les richesses, les sujets et à prélever des impôts au statut d'instrument prévisionnel, on peut dire qu'elle a outrepassé les statistiques pour atteindre à la statistique.

On peut penser sur la base de certaines pratiques que ces outils, qui appartiennent assurément à la statistique descriptive, ont eu une certaine fonction prévisionnelle et annoncent, de ce fait, une statistique comportant une conjecture sur le futur une espèce de statistique inférentielle « en acte ».

1.3. De la théocratie

Il y a une certaine légitimité à supposer que la prise de conscience par l'homme de son humanité soit arrivée très tôt, et qu'elle précède puis coïncide avec des productions symboliques. Parmi toutes les activités symboliques possibles, celles dont on a trouvé les traces les plus anciennes consistent en l'organisation de sépultures. On conviendra qu'un ensemble d'ossements humains fossilisés n'est pas forcément une sépulture, et il semble que le meilleur critère soit « la préservation du squelette en connexion » (VANDERMEERSCH, 1986), ce qui, au passage, fait appel à une preuve d'ordre statistique.

La datation des plus anciennes sépultures découvertes est évaluée à 100 000 ans. Il s'agit du site de Skhül en Israël fouillé depuis 1930 par Dorothy Garrod (1892-1968, professeur à l'Université de Cambridge), puis par d'autres chercheurs. Cette prise de conscience de sa condition par l'homme l'a immanquablement conduit au religieux et de là, aux religions. Il était alors naturel que s'installent des théocraties dans divers groupes sociaux, certaines étant d'actualité. Leurs positions par rapport à la science et donc à la statistique sont nombreuses, diverses et variées, parfois en totale opposition et souvent ambiguës.

Dans des hypothèses selon lesquelles une entité supérieure domine le futur il est évidemment sacrilège de faire des calculs sur ce futur. Ainsi dans la Bible selon que le recensement est l'ordre de Dieu ou du roi, il connaît des fortunes opposées. Dans le Premier livre des chroniques (Bible, p. 1288) on a une statistique des enfants d'Israël ordonnée par Dieu :

« Tous les Israélites avaient été enregistrés et voilà qu'ils étaient inscrits sur le Livre des rois d'Israël.

Quant aux Judéens, ils avaient été déportés à Babel à cause de leur infidélité.

Les premiers habitants qui réoccupèrent leurs propriétés et leurs villes furent ceux d'Israël : les prêtres, les Lévites, les oblats.

À Jérusalem habitèrent des fils de Juda, des fils de Benjamin, des fils d'Ephraïm et de Manassé. ».

Et il semble que cet inventaire ne pose aucun problème.

Par contre, toujours dans le Premier livre des chroniques chapitre XXI (Bible, p.1325) lorsque le roi David prend l'initiative de faire de même pour son propre compte pourrait-on dire :

« Satan se dressa contre Israël et il incita David à dénombrer Israël. David dit à Joab et aux chefs du peuple : « Allez compter Israël, de Bersabée à Dan, puis revenez m'en faire connaître le chiffre. Joab répondit : « Que Yavé accroisse son peuple de cent fois autant ! Monseigneur le roi, ne sont-ils pas tous les serviteurs de Monseigneur ? Pourquoi Monseigneur fait-il cette enquête ? Pourquoi Israël deviendrait-il coupable ? » Cependant l'ordre du roi s'imposa à Joab. Joab partit et parcourut tout Israël puis rentra à Jérusalem. [...] L'ordre du roi avait tant répugné à Joab qu'il n'avait recensé ni Lévy ni Benjamin. »

La Bible nous raconte que Dieu se fâche et demande à David de choisir entre plusieurs punitions toutes plus abominables les unes que les autres et frappant tout le peuple :

« Dieu vit avec déplaisir cette affaire et il frappa Israël. David dit alors à Dieu : « C'est un grand péché que j'ai commis en cette affaire ! Maintenant veuille pardonner à ton serviteur, car j'ai commis une grande folie » [...] Yavé envoya donc la peste en Israël... À la suite de quoi David érigea le fameux temple à Yavé. »

On le voit à cet exemple, une pratique statistique, même simple, est un domaine où les hommes doivent s'aventurer avec prudence.

Si l'on fait un énorme saut dans l'espace et dans le temps on rapprochera l'épisode précédent à celui de Staline déportant les statisticiens qui avaient fait des sondages, seuls les recensements ayant un statut de preuve à ses yeux. Lors d'un « Café de la Statistique »

(toujours à propos de Staline) ayant pour thème « *L'indépendance de la statistique à l'égard du pouvoir politique* » organisé par la SFDS (Société Française de Statistique) le 16 mai 2006 et animé par Jean-Marie Delarue¹³, un des participants raconte que « *Staline avait anticipé le résultat du recensement démographique ; l'office statistique a trouvé un chiffre inférieur ; son directeur n'a pas d'emblée été sanctionné, mais on a nommé une commission d'enquête ; comme il advient pour tout recensement, elle a trouvé des omissions (et, sans doute aussi, des doubles comptes) : on a conclu au sabotage et le directeur a été fusillé. Puis, le président de la commission d'enquête a été nommé à sa place ; il a alors mieux compris les conditions de l'opération et en est venu à partager l'avis de ses collaborateurs : il a été fusillé à son tour* ».

Nathalie Moine (MOINE, 2002) a développé dans un article « Le miroir des statistiques, Inégalités et sphère privée au cours du second stalinisme » une étude fort complète des rapports pour le moins complexes entre les statistiques et le pouvoir dans la société soviétique sous Staline.

On pourrait penser qu'une société laïque, débarrassée en partie du poids de la religion, puisse plus facilement accéder à une pensée statistique mais nous allons voir au paragraphe suivant qu'il reste encore quelques obstacles.

1.4. Calculer le futur se pose comme un obstacle épistémologique

1.4.1. Une pensée débarrassée des superstitions ?

Dans l'ouvrage *Psychopathologie de la vie quotidienne* (Freud, 1969), l'auteur nous fait rencontrer la superstition comme une confirmation de « *l'existence d'une connaissance inconsciente et refoulée de la motivation des actes manqués et accidentels* ».

A partir de là on peut émettre l'hypothèse que les superstitions seront toujours constituantes de l'appareil psychique de l'individu, tout au plus faudra-t-il apprendre à vivre avec, et l'on ne s'étonnera pas que tel président ait consulté régulièrement une célèbre et médiatique astrologue, que le suivant ait vu, dit on, des marabouts africains, et que les jeux de la « Française des jeux » aient autant de succès.

¹³ ancien vice-président du Conseil national de l'information statistique

Et donc, en ce qui concerne les apprentissages statistiques ces derniers ne seront, dans un premier temps, qu'un cosmétique sur les superstitions qu'à des degrés divers chaque individu garde en lui. C'est à force de persévérance et par la multiplication des situations qu'une pensée statistique s'installera à côté de « petites superstitions » fonctionnant comme des addictions marqueurs de névroses plus ou moins handicapantes.

En revanche, pour le même individu, ils viendront compléter l'outillage logique et l'« habitus », (BOURDIEU, 1980, pp. 87), tout en portant un certain nombre de singularités et de contradictions que nous abordons ci après.

1.4.2. Un paradoxe par rapport à la pensée d'Aristote : prendre et rejeter à la fois

Mais revenons au développement d'une pensée ayant pour sujet le statistique. La méthode scientifique se nourrit au sein de la pensée d'Aristote et c'est également vrai de la statistique.

Nous avons déjà rencontré les rapports d'Aristote et de la pensée officielle chrétienne par l'intermédiaire de Thomas d'Aquin (1225-1274). Ce dernier trouve dans les écrits d'Aristote un système qu'il pense le meilleur pour développer une doctrine de l'église. Dans ce renouveau de la scolastique il est en conflit avec les partisans d'un courant traditionnel, en particulier les franciscains, qui s'appuient sur la glose d'Augustin (354-430) inspirée du système néo platonicien. Notons qu'Averroès¹⁴ se réclame aussi d'Aristote et qu'il est sans doute un des principaux enseignants de sa pensée à l'Occident.

Mais ce qui a permis le développement de la pensée et de la méthode scientifique devient un obstacle pour développer une pensée statistique, ainsi lorsqu'on lit chez Aristote (Aristote, Derniers analytiques) dans le début du chapitre VIII des « Derniers analytiques » :

« Toute conclusion démontrée est éternelle : il n'y a donc pas de démonstration pour les choses périssables, de même qu'il n'y a pour elles que science d'accident. - Les définitions sont éternelles comme les démonstrations, dont elles ne sont qu'une forme. - La

¹⁴ Averroès (1126-1198) a commenté pendant une grande partie de sa vie les écrits d'Aristote

démonstration peut s'appliquer à certaines choses passagères, mais dont l'essence est éternelle, par exemple certains phénomènes naturels. »¹⁵

On retient qu' « il n'y a pas de démonstration pour les choses périssables ».

Et dans le paragraphe 10 de ce chapitre Aristote rajoute :

§ 10. Τῶν δὲ συμβεβηκότων μὴ καθ' αὐτά, ὃν τρόπον διωρίσθη τὰ καθ' αὐτά, οὐκ ἔστιν ἐπιστήμη ἀποδεικτική. οὐ γὰρ ἔστιν ἐξ ἀνάγκης δεῖξαι τὸ συμπέρασμα· τὸ συμβεβηκὸς γὰρ ἐνδέχεται μὴ ὑπάρχειν· περὶ τοῦ τοιούτου γὰρ λέγω συμβεβηκός.

Ce qui peut se traduire par :

« § 10. Il est impossible de savoir par démonstration les accidents qui ne sont pas essentiels dans le sens même de la définition que nous avons donnée de ce mot : c'est qu'en effet on ne peut jamais pour les accidents démontrer que la conclusion est nécessaire, puisqu'un accident est ce qui peut ne pas être, seule espèce d'accident dont je veuille ici parler. »¹⁶

Ceux qui voulaient développer une pensée statistique étaient obligés de faire un grand écart : d'une part accepter la partie du système qui leur permettait d'avancer (la logique, etc.), mais d'autre part transgresser le système mis en place par Aristote lui-même.

1.4.3. Franchir l'obstacle épistémologique

Les difficultés précédentes montrent bien qu'il existe un obstacle à la construction de la pensée statistique ; en ce sens nous suspectons ici un obstacle de nature épistémologique et nous devons sans doute ne pas le sous-estimer dans le cadre des apprentissages statistiques.

Devant les affirmations contenues dans les « Derniers analytiques », qui plus est appuyées par les thuriféraires d'une doxa thomiste, les contradicteurs ont dû :

- temporiser afin d'attendre que les forces ennemies soient moins puissantes ;

¹⁵ Traduction de Jules Barthélemy-Saint-Hilaire

¹⁶ ibidem

- attendre que certaines théories d'Aristote soient contredites par l'expérience ;
- patienter jusqu'à ce que la doctrine de l'église s'émancipe par rapport aux théories aristotéliennes ;
- trouver un terrain d'exercice, autrement dit un « angle d'attaque », ce sera essentiellement le jeu ;
- construire des outils appropriés.

Et on le constate, les premiers développements vont se dérouler dans le cadre des jeux de hasard (interdits puis tolérés par l'église), ces situations étant investies à l'aide d'outils construits pour la circonstance : la combinatoire et le calcul des probabilités naissant.

1.5. Aleae Geometria

1.5.1. Introduction

C'est de ce titre inouï, donné par Pascal (Pascal, 1954, p.75) dans le texte adressé en 1654 à l'académie parisienne des sciences « CELEBERRIMÆ MATHESEOS ACADEMIÆ PARISIENSI » ou encore « A LA TRES ILLUSTRÉ ACADEMIE PARISIENNE DE SCIENCE » faisant référence à ce qui va devenir le calcul des probabilités que nous engagerons cette partie de notre étude. En effet si on ne peut difficilement parler de statistique sans parler de calcul de probabilités, l'inverse devrait être également vrai.

1.5.2. La géométrie du hasard

Dans ce texte, Pascal transmet plusieurs « travaux » différents et il introduit et décrit ainsi la « Géométrie du hasard » :

« Et puis un traité tout à fait nouveau d'une matière absolument inexplorée jusqu'ici, savoir : la répartition du hasard dans les jeux qui lui sont soumis, ce qu'on appelle en Français faire les partis des jeux »

En fait ce que l'on présente quelquefois comme une découverte est un sujet à la mode à l'époque. Il s'agit de savoir comment répartir les gains en cas d'arrêt non prévu d'un jeu.

D'une manière indépendante Fermat (1601-1665), magistrat à Toulouse qui travaille la théorie des nombres, l'optique, le calcul infinitésimal, la géométrie analytique et le calcul des

probabilités a trouvé une solution au « problème des partis » et il pense que quelqu'un pourrait s'occuper à publier l'ensemble de ses travaux et il pense à Carcavi, à Huygens et à Pascal (1623-1662) déjà connu pour ses travaux en mécanique et en mathématiques (en fait aucune des trois personnes ne fit ce travail et c'est le fils aîné de Fermat, Samuel, qui le réalisa après la mort de son père).

Le contact entre Fermat et Pascal a lieu en 1654 par le biais d'une correspondance dont nous avons heureusement une trace au moins partielle.

Privilège nous est donné d'avoir d'une part deux solutions différentes développées l'une par Pascal et l'autre par Fermat, et d'autre part une partie de la correspondance (certaines lettres sont perdues) entre les deux brillants esprits. On trouve une analyse de cette correspondance sous la plume de Nicolas Trotignon (TROTIGNON, 2006). Pascal va publier une solution dans la « Règle des partis » (PASCAL, 1954, p.115), Fermat se contentera d'essais restés manuscrits de son vivant.

Mais revenons à cette correspondance intéressante à plus d'un titre.

En effet elle nous donne à voir un archétype, celui d'un échange entre deux façons d'envisager le calcul des probabilités. Les solutions avancées par Fermat qui utilise l'analyse combinatoire et le principe des probabilités composées semblent plus magistrales et plus fécondes que la méthode énoncée par Pascal.

Elle peut également alimenter notre réflexion sur la « validation entre pairs »¹⁷, bien que leurs approches soient différentes ; mais plus fondamentalement elle sonne la fin d'une croyance en un futur contingent relevant de la simple supposition, En effet sur ce dernier point Pascal et Fermat donnent au calcul la capacité de pouvoir renseigner sur le futur contingent : l'idée nouvelle est de partir de l'ensemble des dénouements envisageables et par une pensée rétroactive, donner en remontant au présent, in fine, la connaissance d'un futur « calculé ».

¹⁷ idée reprise par un certain nombre de didacticiens des mathématiques tel Nicolas Balacheff écrivant : « La nécessité de prouver est liée à la situation dans laquelle on se trouve; la preuve est un acte social, elle s'adresse à un individu (éventuellement soi-même) qu'il faut convaincre. » in Encyclopaedia Universalis.

Mais si Fermat reste un magistrat de province pour qui les mathématiques sont un passe temps, Pascal essaie de « faire carrière » et pour cela fréquente les salons parisiens où gravitent toutes sortes d'idées et de gens. Parmi ceux-ci, le chevalier de Méré qui se demande à partir de combien de « coups » on peut parier, en jouant avec deux dés, avoir un double six, et avoir une espérance supérieure à $\frac{1}{2}$. Pascal donne une solution et trouve là une certaine notoriété.

Soulignons, in fine, une petite note que Pascal (PASCAL, 1665, p.219) rajoute à une lettre du 29 août 1654 de Fermat à lui-même (publiée après la mort de Pascal en 1665) « *il écrivait dans sa province ce que j'inventais à Paris, heure pour heure, comme nos lettres écrites et reçues en même temps en témoignent.* ». Or la note (de Pascal) est complétée par un renseignement concernant un travail de Pierre Boutroux qui « *a établi [...] dans l'édition des Grands Ecrivains de la France (t. III, p.334) [que] les découvertes de Fermat et de Pascal ne furent pas simultanées. Fermat avait dix-huit ans de priorité.* »

Nul doute que l'insistance de Pascal ne soit, au minimum, le reflet d'un doute sur la simultanéité de leurs découvertes.

1.5.3. Cardano et Ludo Aleae (1663) ; Galilée et le « Problème du Duc de Toscane » (1620)

Notons au passage que si l'acte « officiel » de naissance du calcul de probabilité est, tel que nous l'avons vu ci-dessus, attribué à Fermat et Pascal, on donne à Cardano et à Galilée les prémisses de cette naissance.

Ainsi Cardano (CARDANO, 1663, p. 273) développe au chapitre XXVIII dans « De ludo Aleae Liber », écrit aux environs de 1564 mais publié en 1663 (HALD, 2003), une série de textes concernant divers jeux.

•••••

CAPVT XXVIII.

*De longo consilio, & iudicio,
procedentiaque.*

Maximum autem in ludo Aleæ præcipuè est prouidentia, & iudicium ad multos iactus, vt in alueare tabulæ sic disponantur, vt cum duplex sit via ad finem, vna, vt celerius progrediamur, altèra, vt collusorem retardemus: celerius progredimur occupando meliora loca, ac citius, vt contra retardamus, vel vt deteriora loca habeat, & pauciora, vel vt lentius ingrediatur vltimam fidem. Cum ergo dictæ viz sint vtraque

Ce qui peut se traduire par :

« Plans de grande envergure, Jugement, et Procédé

DE LA PLUS GRANDE IMPORTANCE dans le jeu sont la prévoyance et le jugement concernant les nombreux jets, car le jeu de plateau est tel qu'il y a deux directions à notre but, l'une qui est d'avancer le plus rapidement, l'autre qui est de retarder l'adversaire ; nous avançons plus rapidement en occupant de meilleures positions, et plus rapidement, si au contraire, nous retardons notre adversaire de manière à ce qu'il ait de plus mauvaises positions ou moins de choix ou qu'il aille en arrière en repartant de la dernière rangée. »¹⁸

Quant à Galilée il traite d'un problème (GALILEO GALILEI, 1718) connu sous le nom « Problème du Duc de Toscane ». Galilée met en scène un haut personnage lui demandant d'« expliquer » la différence lors d'un lancer de trois dés entre la fréquence obtenue d'une somme égale à 9 et celle d'une somme égale à 10, alors que ces nombres 9 et 10 sont obtenus « d'autant de façons » (sous entendu non ordonnées soit 6 façons). Galilée affirme que c'est une « lunga osservazione » qui aurait permis de constater une différence entre les fréquences mais nous savons que les probabilités ne diffèrent que de 1/108 ce qui

¹⁸ la traduction de Sydney Henry Gould « Cardano, The Gambling Scholar » de Ore Oystein (1953)

rend difficile un tel constat. Galilée fournit une réponse tout à fait satisfaisante à la question initiale dans des pages publiées seulement en 1718.

Remarquons bien que dans les deux cas soit de Galilée et du Duc de Toscane, soit de Pascal et du chevalier de Méré le calcul des probabilités est convoqué en réponse à une question issue de la statistique, mais que les statistiques sont elles même fournies comme attestant les résultats trouvés par la théorie des probabilités.

1.6. Les maîtres de l'erreur

Stéphane Callens (CALLENS, 1997) rapporte le mot (péjoratif) d'Auguste Comte à propos des statisticiens comme étant les « maîtres de l'erreur ». Cette maîtrise de l'erreur a commencé bien avant cet épisode, et c'est dans la ligne des travaux de De Moivre (De MOIVRE, 1733) et de Bernouilli que l'analyse s'adapte à la mesure des erreurs (essentiellement des erreurs de mesure en astronomie). La fin du XVIII^{ème} siècle et le XIX^{ème} sont riches d'évènements où statistique et probabilités sont mêlés : Lagrange par exemple conçoit la méthode des moindres carrés (paternité contestée par Gauss) et Laplace (LAPLACE, P.S., 1810) obtient les succès que l'on sait. Cet écart important dans le temps entre De Moivre et Laplace, souligné par Bernard Ycart¹⁹, recouvre une transformation du rôle de la science entre le XVIII^{ème} siècle et le XIX^{ème}. Ainsi dans la préface à *l'Essai philosophique sur les probabilités*, de Laplace, R. Thom écrit en parlant du texte de Laplace :

« C'est un texte révélateur d'une transition essentielle, la charnière qui relie l'humanisme éclairé du XVIII^{ème} siècle au scientisme autoritaire et dogmatique qui dominera à la fin du XIX^{ème} siècle - et qui domine encore une bonne part de la société contemporaine... Le XVIII^{ème} siècle fut celui des mathématiques de l'intelligibilité, le XIX^{ème} celui des mathématiques du contrôle. »

Mais la statistique se veut également instrument social de Condorcet à plus tard Cournot ; Quételet va établir des premières bases de données, par ailleurs la statistique mathématique anglaise va prolonger les travaux de Darwin concernant la biologie, Galton s'appuyant sur la variabilité et constatant la « régression » des tailles vers la moyenne. Puis

¹⁹ B. Ycart http://www-lmc.imag.fr/lmc-sms/Bernard.Ycart/smel/articles/etoiles/cadre_etoiles.html

les tests, déjà en germe dans les travaux de Laplace vont faire une entrée en force dans le XX^{ème} avec le test du χ^2 en 1900, suivi par les statistiques inférentielles (Fischer, Student, Pearson Karl et Egon, Neymann) en ce qui concerne les tests d'hypothèses et les intervalles de confiance, et complété par les travaux de Wald (théorie des décisions statistiques) et de Shewhart (cartes de contrôle). Malgré les travaux de Borel (1924), Lévy et quelques autres, la France est en retard sur ce plan, et ce retard va s'accroître avec l'attitude du groupe Bourbaki très anti statisticiens qu'ils traiteront d'actuares.

C'est l'axiomatique de Kolmogorov qui donne une unité à l'ensemble en 1933, la fin du XX^{ème} voyant fleurir de nouvelles méthodes d'analyse des données : l'analyse en composantes principales, l'analyse factorielle des correspondances qui ne peuvent s'épanouir que grâce aux progrès considérables faits par les instruments de calcul.

1.7. Conclusion

On le voit nous avons donné dans cette partie un aperçu sur l'histoire de la statistique et du calcul des probabilités. Cette césure historique se retrouve dans l'étymologie des mots que l'on utilise dans le champ étudié.

Ainsi si l'on attribue souvent l'introduction du terme « statistique » à un professeur de Göttingen, G. Achenwall, qui aurait en 1746 créé le mot Statistik, dérivé de la notion Staatskunde, l'étymologie nous donne une histoire un peu plus complexe (Le ROBERT, 1993, p.2016) dont voici un résumé :

Latin : **STATUS** : *état*



Italien : **STATO** : *état*

↪ **XVI^e STATISTA** : *homme d'état*

↪ **1633 : STATISTICA** : *relatif à l'état*



Latin moderne : 1672 : **STATISTICUS**



Allemand : STATISTIK Français : STATISTIQUE Anglais : 1798 : STATISTICS
(avant : political arithmetic)

Le nom signifie : « étude méthodique par des procédés numériques des faits sociaux qui définissent un état. »

Le nom change de valeur, d'abord en Angleterre, en devenant « statistics » (1798), puis en France en 1832, et il prend le sens de : « ensemble de techniques d'interprétation mathématique appliquées à des phénomènes. »

Le nom désigne ensuite (1862) l'objet des statistiques : « ensemble de données numériques concernant une même catégorie de faits »

Alors que l'origine du mot hasard (Le ROBERT, 1993, p.946) est la suivante :

Arabe : **ZAHR** = *fleur*



AZ-ZAHR = *jeu de dés* parce que les dés portaient une fleur sur l'une des faces



Espagnol : **AZAR** = *jeu de dés, coup défavorable au jeu de dés*



Français : **HASART** (vers 1150)

↳ **HASARD** (1200) = *un jeu de dés*



coup heureux à ce jeu (6)

(« jeu de hasard » : 1538)



XIII^e Sens figuré : mauvais coup



XV^e : risque, danger

(seule acception qui reste aujourd'hui : « les hasards de la guerre »)



Début XVI^e : cas, évènement fortuit

Milieu XVI^e : cause qu'on attribue à ce qui arrive sans raison apparente (cf en sciences « les lois du hasard »)

Si la preuve est du côté de la probabilité, la collecte et le traitement des données sont bien recouverts par le mot statistique et cela montre la nécessité d'avoir deux cordes (au moins) dans l'apprentissage de la statistique, l'une utilisant la pratique d'enquête par questionnaires participe au sens de l'apprentissage, l'autre à travers la simulation permet de multiplier le nombre d'essais réalisant en cela les affirmations de Galilée et Pascal concernant la lettre au Duc de Toscane pour l'un et la question du Chevalier de Méré pour l'autre.

2. Partie II

Cadre conceptuel et contexte de la recherche

2.1. Cadre conceptuel

Dans cette partie notre propos est d'assembler des propositions pour un apprentissage de la statistique s'appuyant sur une situation d'enseignement. Nous déploierons le cadre conceptuel dans lequel nous faisons nos propositions, construisant et exposant la problématique pour aboutir à la question centrale qui nous préoccupe et finirons par les hypothèses dans lesquelles s'inscrit cette recherche.

Ceux qui désirent explorer et améliorer les conditions de l'apprentissage d'une discipline peuvent prendre différents chemins :

- avoir une démarche « empirique », à savoir une méthode selon laquelle toutes nos théories proviennent de l'expérience, permettant de déceler sur le plan pédagogique ce qui marche et ce qui ne marche pas,
- s'appuyer sur les sciences de l'éducation domaine scientifique multi référentiel,
- construire un ensemble de situations permettant une véritable topogénèse à partir des difficultés relevées dans le champ disciplinaire,
- fonder une ingénierie pédagogique construite à partir de la connaissance épistémologique et historique de la discipline.
- considérer l'enseignement de la discipline comme un objet particulier,
- s'inscrire dans une approche praxéologique c'est-à-dire une science portant ou voulant porter sur les différentes manières d'agir.

Apprendre est tellement intrinsèque de l'espèce humaine que derrière un grand nombre de philosophes pointe une théorie de l'apprentissage (Platon, Aristote, Rousseau, etc...). Jusqu'à des périodes relativement récentes nous sommes restés sur une appropriation de la connaissance à la manière du voyage d'Ulysse. Ce voyage, on l'aura deviné, n'est pas seulement une pérégrination dans l'espace et dans le temps, mais c'est surtout un passage initiatique dans l'univers des connaissances, propulsé par une envie, un moteur : découvrir l'inconnu, le sens des choses, atteindre la *terra incognita*.

2.1.1. Où l'on expose que voyager peut permettre d'apprendre, mais que l'on ne sort pas indemne d'une initiation

Notre première approche s'appuie sur une représentation de l'enseignement illustrée par une métaphore plaçant notre ingénierie pédagogique dans une perspective anthropologique et culturelle. Ce fonds culturel va construire les représentations sociales de l'enseignement la statistique et va interagir sur son enseignement.

Nous venons d'évoquer le voyage d'Ulysse et son importance dans les représentations sociales des apprentissages et, curieusement, nous retrouvons cette façon d'apprendre, en quelque sorte modernisée, dans le roman de Jules Verne « Les enfants du Capitaine Grant ».

Afin de mieux illustrer notre propos rappelons en les grandes lignes.

L'équipage d'un yacht harponne un requin, le dépèce, et on distingue un objet dans son estomac, et après diverses hypothèses le second du yacht trouve quel est cet objet :

« -taisez-vous donc, vous autres, répliqua Tom Austin, le second du yacht, ne voyez-vous pas que cet animal était un ivrogne fieffé, et que pour n'en rien perdre il a bu non seulement le vin, mais encore la bouteille ? ».

Bien entendu la bouteille contient un message en mauvais état que l'on va essayer de déchiffrer :

« lord Glenarvan se décida à briser le goulot de la précieuse bouteille. Il fallut employer le marteau, car l'enveloppe pierreuse avait acquis la dureté du granit. Bientôt ses débris tombèrent sur la table, et l'on aperçut plusieurs fragments de papier adhérents les uns aux autres. »

Et les fragments révèlent un billet écrit en trois langues (anglais, français et allemand) d'un même message (cette situation est bien entendu semblable à celle de la pierre de Rosette dont une copie avait permis à Champollion, quelques dizaines d'années auparavant, de déchiffrer les hiéroglyphes égyptiens).

Ainsi voilà la présentation du contenu de la bouteille :

" il y a là, dit-il, trois documents distincts, et vraisemblablement trois copies du même document traduit en trois langues, l'un anglais, l'autre français, le troisième allemand. Les quelques mots qui ont résisté ne me laissent aucun doute à cet égard. »

Et se pose immédiatement la question de la signification de ce document :

« -mais au moins, ces mots présentent-ils un sens ? Demanda lady Glenarvan. »

L'histoire peut prendre corps : on est sans nouvelles du Capitaine Grant, et ce message provient de lui.

Mais comment décoder ce message, faire des hypothèses, quel système utiliser ? La méthode est donnée par un des protagonistes :

« tout d'abord, reprit Glenarvan, il faut considérer trois choses bien distinctes dans ce document : 1) les choses que l'on sait ; 2) celles que l'on peut conjecturer ; 3) celles qu'on ne sait pas. »

Ceux qui ont découvert ce message (Lady et Lord Glenarvan) vont entrer en contact avec les deux enfants du capitaine Grant, Mary une fille de seize ans et Robert un garçon de douze ans qui sont orphelins de mère. Ils décident de monter une expédition maritime en direction des mers du sud afin de retrouver le capitaine Grant. C'est ce que Michel Serres, fort justement, a relevé comme étant la recherche du père. Le roman et l'aventure auraient pu continuer ainsi mais Jules Verne introduit un nouveau personnage possédant les qualités et les défauts attribués aux savants.

On sait la figure d'Archimède voulant finir sa démonstration et tué par un soldat romain lors de la prise de Syracuse par les armées romaines (Ronan, P. 143), épisode figé dans les esprits l'image du savant. On connaît, plus proche de J. Verne, la grande expédition scientifique et militaire qu'a été la campagne d'Egypte sous le commandement de Bonaparte où avait cours une « plaisanterie courante [qui était :] "Formez le carré ! Les savants et les ânes au centre!", on appelait encore les ânes les « demi-savants »²⁰, on retrouve cette dérision de l'activité des savants, pourtant modernes argonautes dans cette expédition, pendant la campagne par exemple dans « la scène où Junot, ennuyé de devoir assister par ordre à une conférence scientifique, fait mine de ronfler, pour se réveiller en maugréant contre ce « sacré Institut », dans lequel, grogne-t-il, Lannes devrait être reçu sur la seule foi de son nom. »²¹.

²⁰ Article sur Monge in www.anales.org/archives/x/monge7.html

²¹ Article sur Monge in www.anales.org/archives/x/monge7.html

Enfin le roman de Verne paraît en 1867 treize années avant la mort (en 1880) du mathématicien Michel Chasles « célèbre pour avoir été la victime d'une énorme escroquerie: il a payé près de 200 000 francs pour diverses fausses lettres d'hommes de science et pour une lettre de Marie-Madeleine à Lazare »²².

Imprégné par les idées de son temps, J. Verne, dont Michel Clamen a dit qu'il « faisait de la veille technologique »²³, construisait une véritable projection des inventions de son temps pour donner de la vraisemblance au récit.

Donc le nouveau personnage qui est un savant, on saura bientôt qu'il est géographe (la géographie étant l'étude de la surface de la terre et de ses habitants, on peut supposer que ce choix n'est pas un hasard), monte à bord d'un yacht, prend une cabine, s'endort et se réveille en pleine mer s'apercevant alors qu'il s'est trompé de navire ; il pense descendre à la prochaine escale mais il va aller d'escale en escale cherchant à vérifier « sur le terrain » ce qu'il a appris dans les livres.

Et il se présente ainsi : « Jacques-élicin-François-Marie Paganel, secrétaire de la société de géographie de Paris, membre correspondant des sociétés de Berlin, de Bombay, de Darmstadt, de Leipzig, de Londres, de Pétersbourg, de Vienne, de New-York, membre honoraire de l'institut royal géographique et ethnographique des Indes orientales ».

Outre l'effet ironique et comique de cet énoncé, on voit par cette accumulation l'aspect international de la science, mais Verne reviendra à cet aspect par la suite.

Cette idée d'aller vérifier sur le terrain l'exactitude des théories contenues dans les livres a été soulignée par Michel Tournier : l'homme de cabinet doit se faire un homme de terrain et nous verrons qu'il n'en sortira pas indemne.

Tout au long du récit la distraction constante productrice d'erreurs et l'espèce de naïveté de Paganel vont servir de deus ex machina au récit et c'est une dernière erreur qui permettra à la recherche d'aboutir.

²² Article sur Chasles Michel, services rédactionnels de l'Encyclopædia Universalis

²³ Entretien avec Michel Clamen « Cent ans après : les anticipations de Verne face à la science », on peut trouver également ce texte sur le site :

http://www.cite-sciences.fr/francais/ala_cite/expositions/jules_verne/inventions/clamen.html le 01/10/2007

Nous avons d'ailleurs souvent observé dans certains aspects de l'attitude de chercheurs encadrant des groupes d'élèves ou d'étudiants, par exemple dans l'opération « 1000 classes 1000 chercheurs »²⁴ ou encore dans « Maths en Jeans »²⁵, les comportements que l'on trouve dans le personnage de Paganel : distraction cultivée, apprentissage de concepts nomades, maniement érudit de paradoxes savants, mépris plus ou moins affiché des apprentissages effectués dans les systèmes scolaires, etc. Ces constats m'ont tout naturellement amené à dire et à écrire que parfois ces chercheurs devenus enseignants par hasard et pour un temps limité étaient atteints avec plus ou moins de gravité du « syndrome de Paganel ».

Mais revenons à notre propos concernant les apprentissages improvisés pour les enfants. On parle très peu de Mary, elle va simplement séduire par sa beauté, son caractère et toutes ses autres qualités le capitaine, et la fin nous promet un mariage. En revanche, Robert va faire l'objet d'un programme de formation des plus complets. Ainsi J. Paganel propose :

« Un rude petit bonhomme, dit-il, je lui apprendrai la géographie. » Et J. Verne ajoute : *« Or, comme John Mangles se chargeait d'en faire un marin, Glenarvan un homme de cœur, le major un garçon de sang-froid, lady Helena un être bon et généreux, Mary Grant un élève reconnaissant envers de pareils maîtres, Robert devait évidemment devenir un jour un gentleman accompli. »*

Le premier but du voyage est la Patagonie, et personne ne connaissant l'espagnol notre savant l'apprend tout seul sur des livres trouvés à bord, pendant que commence l'instruction du jeune Robert : *« Aussi étudiait-il [le narrateur parle de Paganel] avec acharnement, et on l'entendait marmotter incessamment des syllabes hétérogènes. Pendant ses loisirs, il ne manquait pas de donner une instruction pratique au jeune Robert, et il lui apprenait l'histoire de ces côtes dont le Duncan s'approchait si rapidement. »*

Il s'apercevra bientôt en essayant de communiquer, et nous avec lui, qu'il a appris le portugais ce qui évidemment n'a pas grand chose à voir avec l'espagnol. Mais Verne laisse

²⁴ AUDIN (Pierre), DUCHET (Pierre) *1000 classes - 1000 chercheurs, 1 millième*, brochure n°74, APMEP, Paris 1989, 85 pp.

²⁵ <http://mathenjeans.free.fr/amej/accueil.htm>

supposer que cet apprentissage lui permet de posséder plus rapidement les rudiments de la langue nécessaire pour communiquer avec les habitants finalement rencontrés.

Quand il s'agit d'entreprendre une expédition terrestre le jeune Robert est volontaire « *et moi ! S'écria le jeune Grant* », sa sœur veut l'en empêcher « *-Robert ! Robert ! Dit Mary* » et Paganel enfonce le clou de ses certitudes pédagogiques « *-et pourquoi pas ? Répondit Paganel. Les voyages forment la jeunesse.* »

D'erreur en égarement, de malentendu en confusion, les chercheurs trouvent enfin le capitaine Grant qui constate les effets de l'instruction (et du temps) sur son fils en ces mots : « *Puis, se tournant vers son fils : " comme il a grandi ! C'est un homme ! "* ».

Les diverses hypothèses successives de Paganel émises pour interpréter les fragments du document et qui ont emmenés la petite équipe dans des aventures « à la Jules Verne » vont pouvoir être éclairées par l'auteur des lignes que l'on a essayé de déchiffrer. :

« *Paganel n'y tint plus, et, saisissant la main d'Harry Grant : " capitaine, s'écria-t-il, me direz-vous enfin ce que contenait votre indéchiffrable document ? " à cette demande du géographe, la curiosité fut générale, car le mot de l'énigme, cherché depuis neuf mois, allait être prononcé ! " eh bien, capitaine, demanda Paganel, vous souvenez-vous des termes précis du document ? -exactement, répondit Harry Grant, et pas un jour ne s'est écoulé sans que ma mémoire ne m'ait rappelé ces mots auxquels se rattachait notre seul espoir. -et quels sont-ils, capitaine ? Demanda Glenarvan. Parlez, car notre amour-propre est piqué au vif.* »

On va enfin connaître la solution de l'énigme « *-je suis prêt à vous satisfaire, répondit Harry Grant, mais vous savez que, pour multiplier les chances de salut, j'avais renfermé dans la bouteille trois documents écrits en trois langues. Lequel désirez-vous connaître ? -ils ne sont donc pas identiques ? S'écria Paganel. -si, à un nom près. -eh bien, citez le document français, reprit Glenarvan ; c'est celui que les flots ont le plus respecté, et il a principalement servi de base à nos interprétations. -mylord, le voici mot pour mot, répondit Harry Grant. " le 27 juin 1862, le trois-mâts Britannia, de Glasgow, s'est perdu à quinze cents lieues de la Patagonie, dans l'hémisphère austral. Portés à terre, deux matelots et le capitaine Grant ont atteint à l'île Tabor... -hein ! Fit Paganel. -" là, reprit Harry Grant, continuellement en proie à une cruelle indigence, ils ont jeté ce document par 153 de longitude et 3711 de latitude. Venez à leur secours, ou ils sont perdus. " à ce nom de Tabor, Paganel s'était levé brusquement ; puis, ne se contenant plus, il s'écria : " comment, l'île Tabor ! Mais c'est l'île*

Maria-Thérèse ? -sans doute, Monsieur Paganel, répondit Harry Grant, Maria-Thérèse sur les cartes anglaises et allemandes, mais Tabor sur les cartes françaises ! " »

On le voit, l'internationalisation de la science, que les titres honorifiques de Paganel annonçaient, se brise sur les récifs d'une simple affaire de dénomination, une banale histoire de signifiant, et avec elle l'hypothèse, moteur de l'intrigue, d'un texte identique écrit dans trois langues différentes.

Mais on ne sort pas indemne d'une telle recherche et ce voyage initiatique se devait de trouver une fin symbolisant les atteintes faites à Paganel. L'ultime chapitre s'intitule « *la dernière distraction de Jacques Paganel*. ». En fait Paganel doit se marier et n'arrive pas à se décider sans que personne ne comprenne les raisons de cette indécision jusqu'à ce que le narrateur n'en donne l'argument pendant un épisode comportant trois jours de captivité chez les maoris lors de laquelle le géographe a été tatoué : « *mais tatoué des pieds aux épaules, et il portait sur sa poitrine l'image d'un kiwi héraldique, aux ailes éployées, qui lui mordait le cœur. Ce fut la seule aventure de son grand voyage dont Paganel ne se consola jamais et qu'il ne pardonna pas à la Nouvelle-Zélande* »

Voyage initiatique, recherche du père, application de raisonnements scientifiques, erreurs motrices de la quête de la connaissance, recherche du sens des apprentissages, écueil des signifiants, les héros de Jules Verne ont été transformés par leurs explorations et Paganel, en tant que maestro de l'équipe en a payé le prix sur et dans sa peau.

Si nous sommes restés un certain temps sur cette histoire c'est qu'elle est symbolique à plus d'un sens : l'accès aux connaissances y est une question de classe sociale, il y a une inégalité de traitement entre les garçons et les filles, et enfin l'enjeu véritable des apprentissages est dans la découverte à partir de l'exploration du monde et non dans l'école. Ce dernier aspect se trouve souvent projeté sous les feux de l'actualité quand des chercheurs renommés, sortant de grandes écoles, couverts de diplômes, affirment sereinement qu'ils se sont construits ailleurs que dans les apprentissages scolaires, ce qui est peut-être vrai mais véritablement incomplet.

Et bien qu'elle soit dominante dans une partie de la société, séduisante à plus d'un titre, et même s'il y a peut-être en chacun de nous un Ulysse de banlieue ou un Paganel en herbe qui s'ignore, nous ne partageons pas cette vision des apprentissages, lui préférant des approches plus pragmatiques, plus réalistes et plus utilitaristes.

2.1.2. Désir de savoir : une bien mauvaise traduction

Une autre approche de l'étude des apprentissages pourrait prendre appui sur la psychanalyse et en particulier sur les écrits de Freud. Ce dernier dans « Sigmund Freud présenté par lui-même », parle du « désir de savoir » (FREUD, 1925, p.14) qu'il dit avoir eu très tôt et qui l'a conduit à des lectures précoces de la Bible en particulier. Un éclairage de cette affirmation de Freud nous est donné par Lacan qui rappelle dans son séminaire que « le désir de savoir n'a aucun rapport avec le savoir » (LACAN, 1991, p.22). Dans cette phrase rappelons-nous que Lacan parle du « désir de savoir » présent dans les textes de Freud. Certains exégètes des textes freudiens regrettent que le texte original en allemand n'ait pas été traduit en français par « désir de rechercher », cela eut été plus proche de l'original et aurait permis d'éviter bien des contresens.

Malgré cette réserve, la théorie freudienne nous permet de mieux comprendre comment en statistique, comme en psychanalyse, il y a eu déplacement des enjeux de la connaissance du savoir sur l'objet. Il faut dans les deux cas faire une place au non savoir dans le savoir lui-même, et cela contraint le statisticien à une écriture corsetée par des règles de codage scrupuleuses dont la symbolique permet une différenciation des objets manipulés. Par exemple si l'on considère la situation du test de comparaison de deux moyennes expérimentales dans le cas de deux échantillons indépendants il faudra distinguer par l'écriture :

- quatre moyennes soit les deux moyennes connues des échantillons et les deux moyennes inconnues des deux populations,
- cinq écarts types, les écarts types inconnus des deux populations les deux écarts types connus des échantillons et l'écart type estimé à partir de ces deux derniers.

Certes le mathématicien apprenti ou confirmé manipule aussi des objets inconnus par exemple lorsqu'il résout des équations, mais il espère en déterminer des solutions et trouver les valeurs des variables inconnues, à la différence du statisticien qui sait au départ qu'il ne connaîtra pas les valeurs de certaines variables et qu'elles resteront dans le signifié, en

quelque sorte inter dites²⁶, cette place dans le signifié étant en général spécifiée par une marque dans le signifiant.

Et pour revenir à notre exemple du test des moyennes de deux populations à partir de deux échantillons indépendants, il existera sur ces neuf objets un choix de signifiants étendu et élaboré entre les lettres grecques, romaines, majuscules, minuscules et agrémentées d'un accent circonflexe dans le cas de variables estimées. On voit que dans cette situation sur neuf objets, quatre seulement sont connus et cinq appartiennent à ce que j'ai appelé ici le non-savoir mais font partie malgré tout du savoir lui-même, et nul ne doute, qu'ici en particulier, le jeu sur les signifiants, figure imposée sans laquelle l'apprentissage ne pourrait s'effectuer, ne rende difficile l'émergence du signifié.

Cette difficulté à faire « rentrer » le non savoir dans le savoir lui-même fait partie intégrante de la pensée statistique, en particulier quand, après avoir fait un cours de statistique descriptive, on aborde la statistique inférentielle et l'échantillonnage. Et ce n'est pas par hasard que l'on va retrouver des obstacles similaires en mécanique quantique quand elle viendra au devant de la scène, celle-ci empruntant largement à la statistique et aux probabilités.

2.1.3. Apprentissage de la statistique et conceptualisation

Nous venons de voir deux pistes sur lesquelles nous pourrions avancer, prenons à présent un chemin plus classique dans la ligne des travaux de Piaget (1896-1980) et de Vygotski (1896-1934), faisant de l'assimilation et de l'accommodation les outils fondamentaux de l'adaptation. Piaget influence toutes les théories concernant les apprentissages du XX^{ème} siècle à partir de la construction d'une épistémologie constructiviste de la connaissance contre des théories plus empiriques. A partir d'une méthode génétique il situe la connaissance dans l'activité pratique et cognitive du sujet. Vygotski est redécouvert depuis quelques années. Il se situe également dans une perspective constructiviste de l'apprentissage mais il laisse une place considérable au langage dit égoцентриque constitué de monologues qui accompagnent l'action. Il pense que ce langage s'intériorise et qu'il reste un élément important de la régulation de l'action alors que Piaget pense que ce langage est

²⁶ Ici il n'y a pas de lapsus toccati : c'est bien inter dites.

remplacé par un langage à fonction socialisante, il développe par ailleurs la notion de zone de développement proximal ou zone proximale de développement.

Dans cette partie nous serons dans ce contexte théorique concernant l'apprentissage.

2.1.3.1. Conceptualisation selon Gérard Vergnaud

La vision des apprentissages de la statistique nous permet de rebondir sur la question fondamentale : « Comment apprenons-nous » ?

Dans la lignée de Piaget et de Vygotski, dont les travaux ont été évoqués ci-dessus, Gérard Vergnaud donne une réponse qui nous semble convenir au travail des apprentissages de la statistique en définissant la théorie des champs conceptuels : « *La théorie des champs conceptuels est une théorie cognitiviste, qui vise à fournir un cadre cohérent et quelques principes de base pour l'étude du développement et l'apprentissage des compétences complexes, notamment celles qui relèvent des sciences et des techniques.(...) [et] pour la compréhension des filiations et des ruptures entre connaissances* ».

La conceptualisation, élément central de l'apprentissage, prend appui sur la notion de concept, qui « est un triplet de trois ensembles : »

« l'ensemble des situations qui donnent du sens au concept (la référence), »

« l'ensemble des invariants sur lesquels repose l'opérationnalité des schèmes (le signifié) »

« ..l'ensemble des formes langagières et non langagières qui permettent de représenter symboliquement le concept, ses propriétés, les procédures de traitement (le signifiant) » (VERGNAUD, 1990, p. 135).

Suivons G. Vergnaud (1994, p. 180) qui pour affiner

« progressivement la définition d'un schème ... [dit]... d'abord que c'est une totalité dynamique fonctionnelle, c'est-à-dire quelque chose qui fonctionne comme une unité ; en second lieu que c'est une organisation invariante de la conduite pour une classe de situations données (l'algorithme est un cas particulier du schème) ; et en troisième lieu qu'un schème est composé de quatre catégories d'éléments :

- des buts, intentions et anticipations

- des règles d'action
- des invariants opératoires
- des possibilités d'inférences en situation. »

Mais G. Vergnaud est revenu sur ce qu'il appelle lui-même le « triangle des linguistes » (VERGNAUD, 1994, p. 189-190) pour le transformer ainsi :

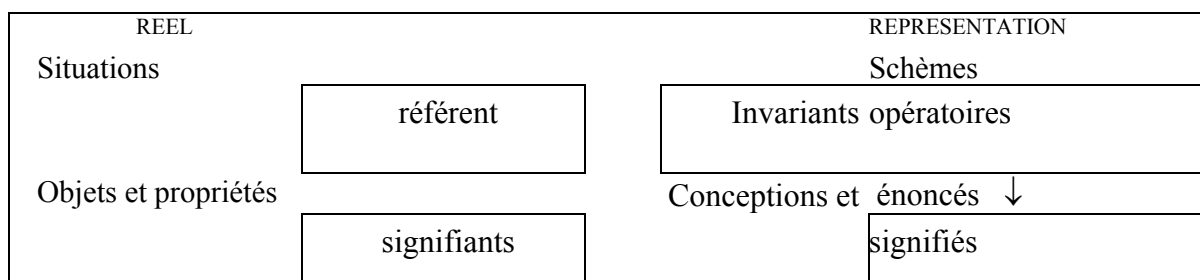


Figure 1 : Schème - situation selon Vergnaud

Rappelons enfin que pour G. Vergnaud un champ conceptuel est un ensemble de situations renvoyant à l'idée de procédure.

2.1.3.2. Le couple schème situation

Pour G. Vergnaud c'est le couple schème situation qui est porteur des apprentissages. Essayons de comprendre sur un exemple comment fonctionnent ces diverses notions. En statistique une notion simple comme la moyenne arithmétique d'une série de données est un concept puisqu'elle comporte à la fois un signifiant, un ensemble de situations (la référence) et un ensemble d'invariants sur lesquels repose l'opérationnalité des schèmes (le signifié). Mais un concept n'existe pas tout seul et G. Vergnaud a développé la notion de champ conceptuel. Ainsi la moyenne précédente est en relation avec la distance quadratique et la variance ; la moyenne d'une série de valeurs x_i est la valeur de x pour laquelle la fonction :

$$f(x) = \sqrt{\sum_{i=1}^{i=N} (x - x_i)^2}$$

atteint son minimum.

On retrouve le même rapport entre la médiane et la norme valeur absolue.

On perçoit à cet exemple simplifié à l'extrême de quelle façon les concepts sont en tension les uns avec les autres à la fois s'expliquant et s'impliquant, afin de former un champ conceptuel.

2.2. Enseignement et didactique de la statistique

2.2.1. La théorie des situations didactiques de Guy Brousseau appliquée à la statistique

La théorie des situations didactiques a été développée par Guy Brousseau à partir des années 70 lors de ses nombreux travaux dans le champ de la didactique des mathématiques.

Il est revenu sur un travail sur les situations didactiques en statistique fait entre 1971 et 1973 et laissé sans prolongement, lors de l'école d'été de didactique des mathématiques (BROUSSEAU, 2003) plus connu sous le nom de " la bouteille " ou encore « la bouteille de Brousseau ». Dans cette expérimentation on donne aux élèves de classe primaire une bouteille opaque dont le bouchon est soudé et qui contient cinq billes blanches ou noires. Un dispositif permet en renversant la bouteille de faire apparaître une bille (et une seule) dans le bouchon transparent. On demande aux élèves quelle est la composition de l'ensemble des billes contenu dans la bouteille.

En conclusion de son cours de l'école d'été de didactique Guy Brousseau dresse une typologies des situations didactiques en statistique (BROUSSEAU, 2003) :

“ Une situation est fondamentale du premier type si elle vise à fournir un modèle qui, par le jeu de ses variables et de leurs limitations, peut convenir à n'importe quelle situation où cette notion intervient.

Exemple « Une situation statistique est une situation où un sujet essaie de reconnaître un objet dans une collection de situations fluctuantes et/ou avec un répertoire incertain ».

Une situation est fondamentale du deuxième type si elle vise à servir de référence, à représenter symboliquement au besoin, ce qui est essentiel dans les objets et dans leurs relations, de façon à pouvoir y rattacher des situations effectives par « le sens », par des « représentations ou par des transformations diverses.

Exemple : celle qui est finalement présente dans l'esprit des élèves à la fin du processus « comment se convaincre de la vérité d'un fait qui reste incertain en répétant des prises d'informations ».

Une situation est fondamentale du troisième type si elle peut engendrer un processus qui aboutit à la connaissance de la notion par le jeu des questions qu'elle conduit à se poser, et des réponses qu'elle appelle.

La situation « combien de boules blanches dans cette bouteille » appartient à ce type. Elle aboutit à la connaissance d'une situation du deuxième type. ”

Bien évidemment la situation d'enquêtes effectuées par les étudiants relève du troisième type de situations pointées par Guy Brousseau. Mais ces deux situations ne doivent pas nous faire perdre de vue que les obstacles restent nombreux dans l'apprentissage de la statistique et remarquons que parmi d'autres obstacles « ...les difficultés proviendraient des représentations (au sens de la sociologie) que se font ces institutions de la façon de pratiquer, d'apprendre ou d'enseigner ce savoir en le reliant aux mathématiques...» (BROUSSEAU, 2003). Nous reviendrons dans la partie V de ce travail sur ces fameuses représentations dont parle Guy Brousseau, mais surmonter ce type d'obstacle est un travail délicat convoquant plusieurs disciplines dans une savante catalyse qui permettra au plus grand nombre de nos étudiants un réel apprentissage de la statistique.

Dans la conclusion de cette première partie Guy Brousseau nous donne à la fois une vision générale de la statistique et un schéma de fonctionnement que voici :

« L'étude de cette expérience et de ses résultats ainsi que sa comparaison avec d'autres méthodes sont renvoyées à la fin de l'article, lorsque nous connaîtrons mieux ses motifs objet du chapitre suivant.

Il faut néanmoins retenir que la situation comporte trois systèmes d'objets

- *Ce qui a été vu devient une image du passé d'un machine de hasard, c'est à dire une statistique,*
- *Ce que l'on cherche à reproduire sont des événements, plus ou moins probables...*
- *Déterminés par la structure de la machine*
- *Les observations répondent à des hypothèses faciles à imaginer sur les relations entre ces trois types d'objets*

Elle les met en rapport :

- *Les raisonnements renvoient d'un objet à l'autre*
- *Le contenu de la bouteille ne change pas : contenu \rightarrow hypothèse (probabilité)*
- *Les observations reflètent le contenu de la bouteille : statistiques \rightarrow contenu*
- *Les observations à venir doivent refléter celles du passé puisque la machine ne change pas. Hypothèse (probabilités) \rightarrow statistiques*

Ce qui génère un processus à trois temps dont le moteur est le suivant :

- *Si ce que l'on observe donne des indications sur ce que contient la bouteille alors en reproduisant ce qu'on a fait on devrait reproduire ce qu'on a vu. (évidemment, la question implicite est : «Qu'est-ce que toutes ces expériences ont en commun ?»).*
- *Le passé \rightarrow la machine \rightarrow le futur »*

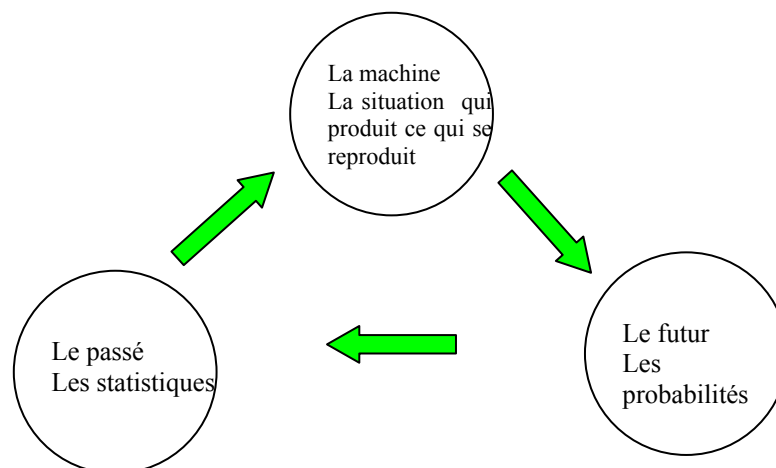


Figure 2 : Schématisation de l'activité des élèves selon Guy Brousseau
(BROUSSEAU, 2003, p.8)

2.2.2. La transposition didactique

Nous avons déjà approché le concept de transposition didactique (ORIOU, 1986) développé par Yves Chevallard afin d'examiner la nécessaire transformation du savoir savant en savoir enseigné (CHEVALLARD, 1985). On comprend bien que dans le domaine de la statistique on aura à faire des choix, un exemple parmi d'autres dans un cours de statistique traitant du théorème de De Moivre- Laplace : va-t-on exposer la démonstration de De Moivre qui se limite au cas de probabilités binomiales et utilisant la formule de Stirling, celle de Laplace, celle de Tchebycheff plus claire en montrant chaque fois quels sont les problèmes

soulevés par ces démonstrations ou bien comme cela se fait la plupart du temps directement la démonstration de Paul Lévy datant de 1919 ? La réponse dépend sans doute du temps que l'on accorde à ce passage important de la théorie de la statistique.

2.3. Enseignement de la statistique et Ingénierie didactique

Nous allons à présent décrire brièvement en termes d'ingénierie didactique l'enquête par questionnaire et la simulation dans nos dispositifs.

2.3.1. L'enquête par questionnaire

Nous avons déjà vu l'étymologie du mot enquête depuis le latin populaire « inquæsitā » et le latin classique « inquisita », avec comme substantif « inquirere » qui signifie « chercher à découvrir » et qui a donné « enquete » au début du XII^{ème}.

Nous allons essayer de donner quelques variables didactiques que nous avons rencontrées lors des enquêtes faites par nos étudiants :

- Commanditaire : extérieur, à l'intérieur de l'IUT, du cours (professeur par exemple).
- Encadrement institutionnel : une personne, une équipe.
- Nombre d'étudiants en charge de l'enquête : de deux à soixante dix.
- Questionnaire : déjà fait par le commanditaire, à modifier, à créer.
- Nombre de questionnaires à traiter : de quelques dizaines à plus d'un millier.
- Durée de l'enquête : de quelques heures à plusieurs semaines.
- Durée du traitement : de quelques heures à plusieurs mois.
- Mode d'administration du questionnaire : en face à face, distribution et récupération, envoi postal, envoi par internet, par téléphone, directement sur un serveur.
- Méthodes statistiques utilisées : échantillonnage, ACP, AFC, méthodes déjà vues en cours, méthodes nouvelles pour les étudiants (ASI par exemple).

- Déroulement du projet : rendez-vous réguliers, aléatoires, compte-rendu à mi-parcours.
- Rendu : dossier papier, exposé des résultats.
- Contrôle : notation du dossier, de l'exposé, notation collective, individuelle.

2.3.2. La simulation

Nous avons installé notre vision de la simulation en statistique sur deux idées forces :

- Ce sont des simulations construites par les étudiants : cet a priori est le résultat de notre vision des apprentissages,
- Elles ne font pas appel à des logiciels spécialisés.

2.3.3. Evaluation

Il n'est pas de dispositif d'ingénierie didactique sans dispositif de contrôle ou d'évaluation : or le mot évaluation a le vent en poupe dans l'éducation.

Si l'on se réfère aux textes gravitant autour du système éducatif français (textes officiels, commentaires, articles de la littérature pédagogique, etc.) on assiste à une récente inflation de l'usage du mot évaluation.

Avant la fin des années soixante, en effet, on construit des examens²⁷, on réfléchit à la notation en termes statistiques (depuis les années vingt Piéron a introduit la notion de docimologie) et l'évaluation est complètement absente. En revanche, à la fin des années soixante-dix, le mot « évaluation » est un des mots les plus en faveur dans le champ des sciences de l'éducation.

Il existe sans aucun doute plusieurs raisons permettant d'expliquer ce constat :

- le système éducatif a une « crise de croissance » aussi bien en termes d'effectifs qu'en termes de contenus,
- certains courants de pensée remettent en cause la « sélection par l'échec » et dénoncent publiquement les outils « examens » et « contrôle » mis à son service,

²⁷ Du latin examen, au propre «aiguille de balance», de exigere, au sens de «peser», 1339.

- le monde économique s'investit dans la formation continue et veut se donner les moyens de mesurer la rentabilité des sommes engagées.

Les grands paradigmes unifiants qui ont remis en cause le fonctionnement de la société et de l'école se sont effondrés en tant qu'instruments de rénovation laissant place aux outils du monde économique (on apprend ainsi à « gérer son temps », à « planifier » des tâches d'apprentissage, à estimer la « rentabilité » d'un système éducatif, à être un « gestionnaire des ressources humaines », etc.).

Alors « tout » devient évaluation et l'on évalue aussi bien les capacités en mathématiques d'un élève de sixième que le rendement d'un institut de formation des maîtres.

Mais cette polysémie grandissante n'est pas sans entraîner quelques difficultés et le nombre d'adjectifs va croissant : on parlera ainsi d'évaluation sommative, formative, implicite, interprétative, instituée, spontanée, diagnostique, formatrice, d'auto-évaluation, de macro-évaluation, de micro-évaluation, ...

Ayant pris place dans tous les champs du savoir, l'évaluation a acquis une position institutionnelle attestée par des domaines spécialisés, des travaux et des contenus de recherche, des définitions de postes de travail, des domaines d'étude tels que la « sociologie de l'évaluation », tandis que, sous la pression du monde économique, l'évaluation glisse, de l'évaluation des connaissances d'un individu à l'évaluation des systèmes éducatifs.

Il est difficile de trouver sa route dans un chemin aussi touffu, et, afin d'y voir plus clair, je vous propose d'appliquer le schéma suivant.

- D'un côté le contrôle²⁸, il s'applique clairement sur l'objet produit, il est externe à l'apprenant, il est nécessaire en tant que certification et garantie des apprentissages, il participe au respect de la norme.
- A l'opposé l'évaluation, interne au sujet apprenant, renseigne sur le processus d'apprentissage, sur les objectifs initiaux, les critères affichés de l'atteinte de ces objectifs, l'émergence de compétences non prévues au départ, la détermination de nouveaux objectifs, elle donne le sens des apprentissages.

²⁸ le mot vient de « contre-rôle », registre tenu en double 1367

Mais si cette dichotomie a le mérite de savoir « ce que l'on fait » elle est trop simplificatrice. En effet dans tout contrôle il y a une partie évaluation : par exemple avoir une bonne note au contrôle de mathématiques rassure le sujet.

Et réciproquement une évaluation ne peut s'envisager sans un système de référence et donc sans des procédures ressemblant à du contrôle. Nous sommes en face d'un phénomène « à curseur » :

contrôle  **évaluation**

Nous venons de le voir, nos pratiques pédagogiques sont traversées par un axe allant du contrôle à l'évaluation. Mais les apprentissages sont (ou devraient être) également structurés par la complexité des objets d'apprentissage considérés.

Et trop souvent on constate que les stratégies d'apprentissage sont une extrapolation de modèles d'apprentissages construits pour des situations simples. Ces méthodes qui sont rejetées des activités industrielles (qui irait, en effet, construire un avion d'un kilomètre de long de la même façon qu'un avion de trente mètres de long ou un avion en métal comme un avion en papier ?) restent souvent présentes dans les activités éducatives et expliquent plus d'un échec.

Il s'agit d'une non prise en compte de la complexité des processus d'apprentissage liés à des tâches complexes.

Nous ne développerons pas diverses approches de la complexité et nous resterons sur une approche peu formalisée mais assez efficace finalement : il y a complexité lorsque l'apprentissage n'est pas décomposable en une séquence d'éléments que l'on peut apprendre directement et indépendamment ; ainsi en mathématiques l'intégration est plus complexe que la dérivation. La limite entre complexe et non complexe n'est pas comme précédemment un phénomène continu mais plutôt un champ comportant des niveaux et des sauts à faire pour passer d'un niveau à l'autre.

2.3.3.1. La complexité : un élément endogène de l'obstacle épistémologique

Une telle analyse resterait dans le cabinet du chercheur et dans l'atmosphère plus ou moins feutrée des séminaires si elle n'avait pas ou peu de résonances sur le terrain. En fait il n'en est rien car elle est sans aucun doute l'un des éléments constitutifs des obstacles épistémologiques.

Survolons rapidement les classifications les plus courantes concernant les apprentissages :

- une première approche vise à distinguer l'objet d'apprentissage du processus
- une autre taxinomie sépare les apprentissages par action de ceux par instruction
- enfin on pourra distinguer les apprentissages subsymboliques des apprentissages symboliques

Sans que les catégories précédentes soient forcément exclusives les unes des autres on en retirera que l'apprentissage d'une tâche complexe demande du temps, que son contrôle est difficile ou en tout cas plus difficile, que le processus est moins « déterministe » que dans le cas d'apprentissages non complexes et enfin que la réponse de la part de l'enseignant est plus délicate.

2.3.3.2. Une fausse réponse à la complexité : la complication

On vient de l'entrevoir, il est difficile de prendre la mesure dans le déroulement didactique d'apprentissages complexes. Cette difficulté va entraîner la mise en place de dispositifs pédagogiques compliqués. Ces échafaudages ne répondent en général qu'à une partie des difficultés rencontrées par les apprenants et ils sont difficilement transposables pour d'autres équipes d'enseignants ou d'autres contextes.

Et donc ils vont aller en se compliquant pouvant doubler alors l'obstacle didactique d'obstacles pédagogiques. L'exemple de l'étude des décimaux à l'école primaire illustre ce propos.

On peut se demander pour quelles raisons la construction d'échafaudages compliqués est une réponse aux apprentissages complexes ?

Parmi les raisons premières on remarque que la complication

- a un effet de « masquage » des difficultés à venir,
- qu'elle protège le sujet par des conduites de type « maniaque » contre l'angoisse provoquée par le déséquilibre nécessaire à l'assimilation de nouveaux concepts, la phase de déconstruction précédant nécessairement la reconstruction (des concepts),
- qu'elle prend souvent sa source dans l'égalitarisme bien pensant supposé dû aux étudiants,
- et enfin qu'elle provoque de la complicité (entre enseignants, entre étudiants et enseignants),

déplaçant l'attention de la complexité vers la complication et permettant ainsi de masquer les obstacles épistémologiques.

Par conséquent on observera autour de nous que les dispositifs pédagogiques semblent attirés par des procédures compliquées.

Un point de vue sur le stage inhérent aux DUT

A partir des réflexions précédentes, observons les activités proposées aux étudiants dans le cadre du DUT. Ce dernier se caractérise par un stage effectué en entreprise par les étudiants durant le cours des études. Si les modalités exactes sont variables d'un IUT à l'autre le stage est un élément important et incontournable de la stratégie pédagogique.

Ce stage effectué par les étudiants, mais c'est aussi vrai pour les divers projets, met en jeu d'un côté des apprentissages complexes et de l'autre des comportements de l'ordre de la vie sociale. On rencontrera dans les modalités des projets tuteurés et éventuellement dans celles du stage la complication relevée précédemment.

D'autre part, et sans qu'il soit utile d'insister, on voit bien qu'il nous faudra prendre en compte une part de contrôle et une part d'évaluation.

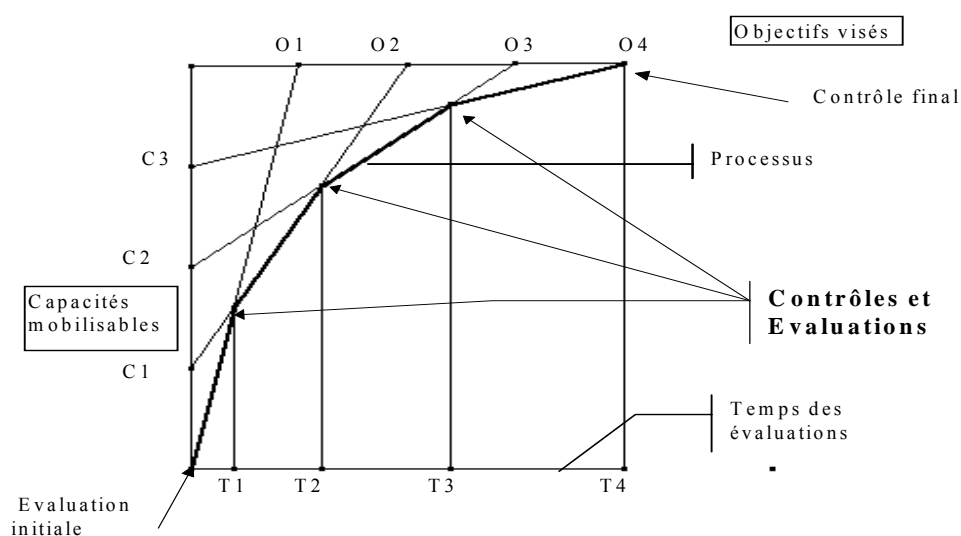
Souvent le contrôle se fait sur la production : rapport de stage, mémoire de deuxième année, etc. alors que l'évaluation prend en compte les comportements de l'étudiant sur divers plans.

Ainsi dans le livret d'évaluation de l'IUT Lumière nous avons distingué trois plans concernant le travail de l'étudiant dans l'entreprise : mise en œuvre, apprentissage, relationnel avec un système de lettres déterminant un niveau de A à E, et deux notes (de 0 à 20) l'une pour le rapport de stage (ou le mémoire en deuxième année) et l'autre concernant la soutenance du dit rapport.

Une conclusion bien provisoire

Qu'on ne se méprenne point sur ces quelques lignes : je n'appartiens pas aux encenseurs d'un norme didactique donnant leçons et récompenses, mais j'aime à faire partager mes interrogations concernant certaines pratiques pédagogiques. La complication des dispositifs d'enseignement rencontrés donne toujours matière à réflexion et souvent substance à simplification.

Un des travaux d'enquêtes de satisfaction (LENIR, 2000) a été ponctué par un dossier et un exposé donné par chaque équipe de sept étudiants chargée d'une partie de l'enquête. Afin de comptabiliser les efforts de nos étudiants nous avons donné une note dans les disciplines statistique, communication, transport de voyageurs et en cela, nous avons satisfait aux rites d'une évaluation sommative. Mais c'est lors du dialogue qui s'est installé entre les étudiants et les enseignants lors de l'exposé (dix minutes d'exposé et vingt minutes de questions/réponses), que la formation des étudiants fut la plus fructueuse ; cette évaluation formative permet au groupe d'aller plus loin dans son travail d'analyse de l'enquête. Nous sommes dans un schéma déjà décrit lors d'une communication au colloque sur l'alternance (ORIOU, 1999), et que nous rappelons :



2.4. Contexte de la recherche : l'IUT Lumière et le département STID

Les entreprises sont devant une grande quantité d'informations à extraire de bases informatiques, et cette quantité augmente au fil du temps ; les traitements statistiques de ces informations doivent permettre aux responsables de ces entreprises de prendre des décisions appropriées au développement de leurs activités.

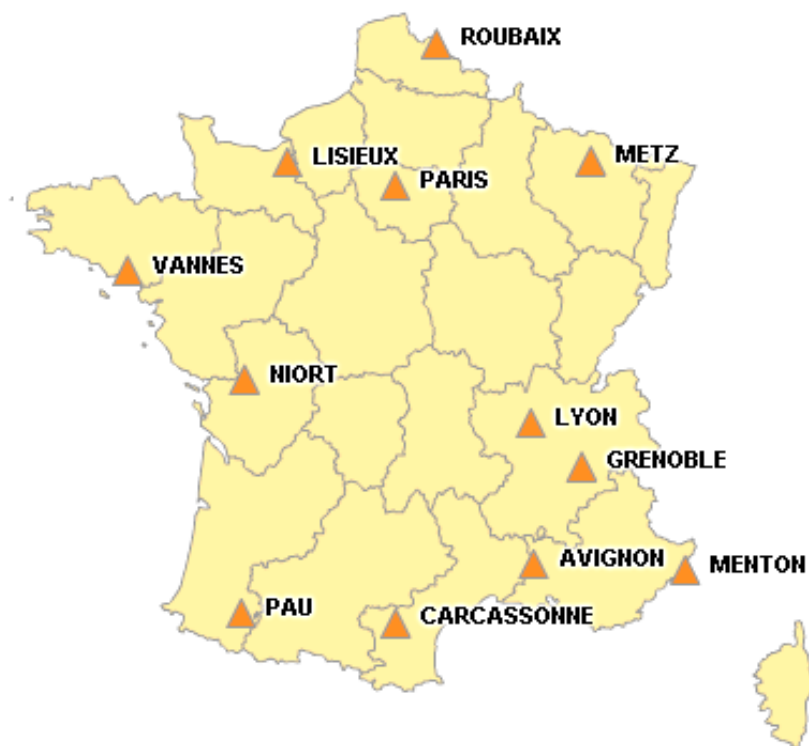
Dans ce contexte le rôle des professionnels de la statistique comporte une forte composante visant à extraire et traiter des données sous forme numérique. De tels professionnels sont de plus en plus recherchés par les responsables d'organisations et par les chefs d'entreprises.

Ainsi parmi les rôles que devrait remplir un département STID, il y a celui de former des techniciens aux métiers de la statistique, du traitement des données et de l'analyse de l'information afin d'apporter une réponse à ce besoin.

2.4.1. Le département STID de l'IUT Lumière

Le département STID de l'IUT Lumière a ouvert en septembre 1998. C'est le dixième département STID dans l'ordre chronologique parmi les 12 départements STID qui existent en France et qui sont régulièrement en contact.

En revanche, le département STID de l'IUT Lumière est le seul département STID à fonctionner complètement en alternance.



Répartition des départements STID en France

Le choix de l'alternance implique un certain nombre d'adaptations et de spécificités.

Cela a notamment un impact au niveau :

- du recrutement et profil des étudiants,
- du suivi des étudiants,
- du programme pédagogique,
- du devenir des promotions diplômées.

La préparation au DUT STID à l'IUT Lumière se déroule sur deux années selon le principe « du 1+1 » avec :

Une première année sous statut étudiant préparatoire à l'alternance salariale. Elle comporte non seulement une acquisition des techniques propres à chaque enseignement, mais également des actions spécifiques conçues pour permettre aux étudiants de réussir leur alternance salariale en 2^{ème} année :

stages intensifs d'anglais et d'informatique,

jeux d'entreprises et visites d'entreprises, interventions de professionnels,
réalisation d'enquêtes ou d'études statistiques,
réflexion sur le projet professionnel de chaque étudiant,
présentation des secteurs d'accueil et des entreprises partenaires,
accompagnement de la prospection d'entreprises faite par les étudiants,
forum de recrutement (en février) permettant aux entreprises proposant un contrat d'alternance de rencontrer les étudiants et de les sélectionner,
stage en entreprise (de 7 semaines, de début avril à mi-mai) d'essai mutuel, avec un tuteur IUT et un tuteur entreprise encadrant l'étudiant.

Une deuxième année en contrat d'apprentissage. En deuxième année, l'étudiant est salarié d'une entreprise et il suit une formation en vue de l'obtention d'un Diplôme Universitaire de Technologie. Cette deuxième année est organisée de la façon suivante :

contrat d'apprentissage de 13 mois, de début septembre de l'année N au 30 septembre de l'année N+1,

22 semaines d'enseignement universitaire selon un calendrier établi à l'avance et 35 semaines en entreprise dont les congés payés,

un rythme d'alternance de 15 jours à l'IUT et 15 jours en entreprise,

environ 750 heures de formation à l'IUT avec des cours, travaux dirigés, des travaux pratiques et des projets tuteurés.

2.4.2. L'alternance à l'IUT Lumière

Depuis sa création en 1992 l'IUT Lumière est le seul IUT à former ses étudiants complètement en alternance. Cette formation est ainsi construite : en première année les étudiants suivent un stage de sept semaines dans une entreprise pendant la période avril-mai, la deuxième année est réellement en alternance donnant lieu à la signature d'un contrat d'apprentissage (deux systèmes sont en place en GEA 2 jours + 3 jours, et pour tous les autres départements GLT STID QLIO et HSE une autre gestion du temps, à savoir 15 jours à l'IUT

et 15 jours en entreprise). Dans tous les cas la surcharge horaire nous oblige à finir l'année en septembre (jurys et délivrance des diplômes). Cette forme est désignée par "système 1+1" . On trouvera des compléments détaillés sur l'organisation des cursus dans une communication de Paul Rousset, directeur de l'IUT depuis sa création (Rousset, 1997), et plus particulièrement sur le département GEA dans un article d'Isabelle Barth (Barth, 2001) qui a dirigé le département GEA pendant plusieurs années.

2.4.3. Approche taxinomique de l'alternance

Bien entendu l'alternance est considérée a priori comme un levier qui favorise les apprentissages mais c'est également un poids ; on trouve une mise en perspective de cette pratique pédagogique chez de nombreux auteurs (Malgaive, Weber, 1982) ou (Combes, 1996). Avant d'aller plus avant il convient sans doute de faire un point sur cette notion d'alternance. Ainsi J-C Régnier (Régnier, 2000a) souligne déjà dans le titre de son article « *Des évidences de la formation en alternance qui posent questions* » l'angle sous lequel il aborde ce sujet.

L'auteur de l'article approche la notion d'alternance en formation en abordant la question du « Qu'est ce qu'apprendre ? » qu'il représente par le schéma suivant :

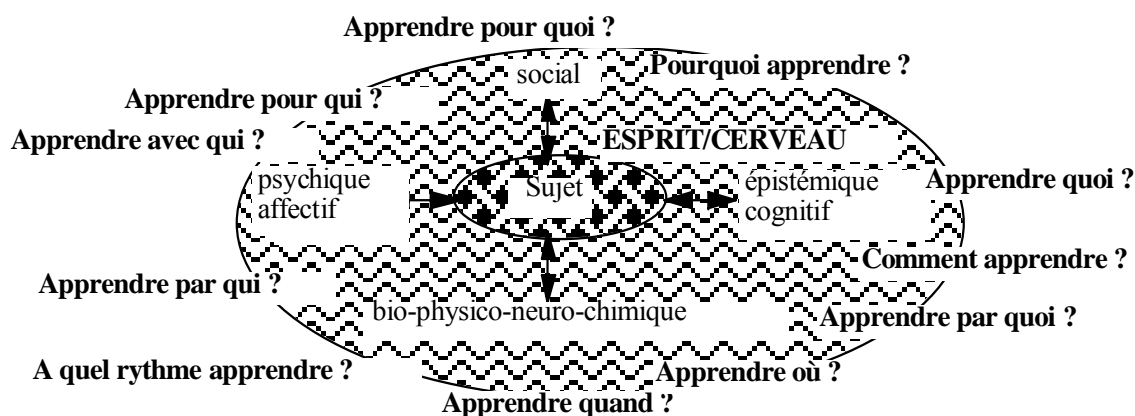


Figure 3 Qu'est ce qu'apprendre ? (Régnier, 2000a, p. 44)

Cette approche permet à l'auteur de distinguer 4 types d'alternance dans le domaine de la formation :

<i>l'alternance juxtapositive</i>	<i>l'alternance déductive</i>	<i>l'alternance inductive</i>	<i>l'alternance intégrative</i>
le sujet en formation construit ses compétences à partir des connaissances formalisées acquises à l'école parallèlement aux connaissances pratiques acquises sur le terrain professionnel	l'acquisition des connaissances pratiques est le résultat de la mise en œuvre des connaissances formalisées acquises à l'école	l'acquisition des connaissances formalisées est le résultat de la généralisation des connaissances pratiques acquises sur le terrain professionnel	le sujet en formation construit ses compétences dans un lent processus dynamique impliquant inter-activement des connaissances formalisées acquises à l'école et des connaissances pratiques acquises sur le terrain professionnel

Figure 4 : Quatre types d'alternance selon Régnier, 2000a, p. 48

D'autres taxinomies sont possibles, par exemple en se référant à une approche plus institutionnelle, Antoine, Grootaers, et Tilman (Antoine F., Grootaers D., Tilman F., 1988) distinguent quatre formes d'alternance :

l'alternance fusion : SF (situations de formation) et SP (situations de production) ont lieu sur le même site. L'école impose sa logique (de formation) et l'entreprise impose sa logique (de production) aux apprenants, le lien est uniquement géographique.

l'alternance juxtaposition : il n'y a pas de communication entre les problèmes rencontrés par l'apprenant en SF et/ou en SP.

l'alternance complémentarité : les SF et SP ont les mêmes objectifs.

l'alternance articulation : liaisons systématiques entre SF et SP, il existe un lien structurel

Ce classement est repris par Françoise Raynal et Alain Reunier dans l'article « *pédagogie de l'alternance* » (Raynal F., Reunier A., 1997, p. 226).

On le voit, en dehors des effets innovants propres à toute différenciation pédagogique, l'alternance ne peut, à elle seule porter le poids symbolique des apprentissages.

La mécanique « alternance » est plus complexe qu'il n'y paraît à la lecture d'un simple emploi du temps. Si cette complexité n'est pas prise en compte en tant que telle, elle donnera lieu à différents symptômes parmi lesquels on peut citer :

- la mise en place de procédures compliquées, cette complication donnant naissance à de la complicité entre ceux qui savent « se mouvoir » dans le système.
- la poursuite des innovations perpétuelles, ce qui peut permettre de ne pas se poser de questions.

2.4.4. La place des tuteurs

Un autre problème que l'on rencontre dans cette situation concerne la formation des tuteurs, moniteurs, maîtres de stages ou accompagnateurs, qui souvent se résume au minimum (faute de temps et/ou de moyens). Cette interrogation est aussi bien celle de JC Régnier (Régnier, 2000a) que de Antoine, Grootaers, et Tilman (Antoine F., Grootaers D., Tilman F., 1988, p. 71) ; ces derniers précisent les qualités en ces termes

« Trois critères d'égale importance devraient influencer le choix d'un tuteur :

- sa maîtrise professionnelle

- sa motivation

- ses capacités relationnelles et pédagogiques. »

Ce contexte est porteur d'intégration mais on aura aussi compris que la division du temps porte en elle des obstacles difficiles à surmonter : on a donc installé des dispositifs dits de « retour d'alternance » permettant aux étudiants de se repositionner dans leur parcours académique.

3. Partie III

Pratique d'enquête par questionnaire : un outil d'appropriation de la statistique

Nous décrirons ici plusieurs activités d'enquêtes réalisées par les étudiants dans les départements GEA, GLT et STID de l'IUT Lumière dans des contextes variés.

3.1. Enquête par questionnaires réalisée par les étudiants d'un département GEA : une ingénierie pédagogique de l'apprentissage de la statistique inférentielle

3.1.1. Présentation

L'expérience des années précédentes nous avait montré un désintérêt certain chez quelques étudiants du département GEA pour les séances consacrées à l'enseignement de la statistique inférentielle. En accord avec la direction du département nous avons décidé que les étudiants (par groupes de quatre) seraient chargés en fin de cursus de faire une enquête par questionnaires.

3.1.2. L'organisation de l'enseignement de la statistique dans le département GEA

L'alternance donne un temps passé en entreprise ayant une géométrie variable entre première et deuxième année ; il a conduit l'équipe enseignante à donner des poids différents aux cours de statistique dans chacune des années.

En première année une quarantaine d'heures est consacrée à l'étude de la statistique descriptive ; cet enseignement donne lieu à deux contrôles (devoirs individuels sur table) et à une évaluation sous la forme d'une enquête concentrée sur plusieurs disciplines (communication, marketing, statistique, etc.) et réalisée en groupes.

En deuxième année, seize heures sont consacrées à travailler sur les notions de lois de probabilités, d'échantillonnages, d'estimations, d'intervalles de confiance.

3.1.3. Mise en place du travail d'enquête par questionnaires

Par rapport aux emplois du temps (très chargés) des étudiants, nous ne pouvions envisager un travail d'enquête par questionnaires aussi lourd que dans d'autres contextes, comme celui, par exemple, de la statistique en première année GLT (Gestion Logistique et Transport) dont on pourra consulter un compte rendu dans (Le Nir, Oriol, 2000).

Nous avons donc pris le parti de consacrer, à la fin du cours, quatre heures (une matinée ou une après-midi) à cette activité d'enquête par questionnaires.

3.1.3.1. Déroulement de l'enquête

Consignes données aux étudiants (ils sont regroupés par groupes de 4 au maximum)

Phase 1

- déterminer une question (ou plusieurs),
- déterminer la population de votre étude,
- → [venir faire valider votre projet auprès de l'enseignant]

Phase 2

- Construire un questionnaire en prévoyant le traitement futur.
- → [venir faire valider votre questionnaire auprès de l'enseignant]

Phase 3

- Taper votre questionnaire sous un traitement de texte
- Déterminer votre mode de travail : combien de questionnaires à faire passer, quelles procédures, etc.
- → [venir faire valider ce travail auprès de l'enseignant]

Phase 4

- faire passer le questionnaire
- saisir les données à l'aide d'un tableur
- → [dialogues avec l'enseignant par rapport aux premières hypothèses avancées]

Phase 5 (à finir en dehors de la séance de 4 heures)

- écrire un rapport (environ dix pages) concernant cette étude et le rendre à une date précise.
- → [ce rapport donnera lieu à une évaluation et à une note]

3.1.3.2. Quelques sujets

Les sujets étaient librement choisis par les étudiants (seulement soumis à l'approbation du professeur). Nous n'avons jamais eu à refuser un sujet mais seulement à faire préciser la question que les étudiants posaient.

Parmi les nombreux sujets trouvés en voici quelques-uns :

- Quelle est la part du budget des étudiants accordée aux vacances et quelles sont leurs destinations ?
- Comportement des étudiants par rapport au sport et les influences que celui-ci exerce sur cette population.

-
- Que font les étudiants de leur temps libre ?
 - Influence des médias sur les choix politiques
 - Quelle est la distribution du budget global mensuel suivant les postes logement, transport, alimentation, sport et loisirs, habillement [de la population des étudiants de Lyon 2]
 - Origine et répartition du budget des étudiants de Lyon2.
 - Mode de transport des étudiants de Lyon2 en rapport avec diverses variables : sexe, budget, etc.
 - Influence de l'âge sur le mode d'habitation
 - Pratique d'un sport et tabagisme (étudiants de Lyon2)
 - La typologie du vacancier [étudiant de l'université Lyon2]
 - Les étudiants et le sport
 - Les étudiants et les discothèques
 - Qui sont les étudiants de l'UT Lumière ?
 - Impact de la distribution d'un quotidien gratuit
 - Etc.

3.1.3.3. Remarques immédiates :

Population étudiée : pour des raisons de commodité la population étudiée est celle de l'université Lyon2 et quelquefois celle de l'IUT Lumière.

Types de sujets : nous pouvons remarquer qu'il y a eu très peu d'enquêtes ayant la politique comme sujet et aucun sujet sur la santé.

L'anticipation concernant le traitement des données est pratiquement toujours présente.

Le questionnement sur les types de variables envisagées n'est présent qu'une fois sur deux environ.

La maîtrise de l'outil traitement de textes est largement suffisante.

La maîtrise du tableur est plutôt bonne.

Beaucoup de groupes prévoient d'interroger plus de personnes que nécessaire afin d'avoir un peu de marge.

Peu de groupes s'interrogent sur la possibilité de données aberrantes

Quelques groupes s'inquiètent de ne pas avoir dans leur échantillon 50% de garçons (en fait la population des étudiants de Lyon2 est composée d'une forte majorité d'étudiantes...)

3.1.3.4. Parcours d'apprentissage

Ces temps de dialogues entre un groupe de 4 étudiants et l'enseignant sont des moments d'évaluation qui leur permettent de se fixer de nouveaux objectifs donnant à leur parcours d'apprentissage une forme déjà explicitée (Oriol, 1999).

3.2. La loi de de Benford et la pratique d'enquêtes : un exemple de situations didactiques en statistique

3.2.1. Présentation

Nous décrivons dans ces lignes une activité dans l'enseignement de la statistique destinée à des étudiants de deuxième année d'IUT dans les départements GEA et STID autour de la loi de Benford et de la pratique d'enquêtes par questionnaire. Nous essayerons de distinguer dans ces activités ce qui relève de la théorie des situations didactiques développée par G. Brousseau.

3.2.2. La loi de Benford

3.2.2.1. Présentation de la loi de Benford²⁹

On connaît le phénomène depuis plus d'un siècle : si l'on "collecte" des nombres au hasard les premiers chiffres de ces nombres n'ont pas la même probabilité d'apparition ; dans des domaines aussi différents que les cours de la bourse, les prix présents dans des catalogues,

²⁹ F. Benford, *The law of anomalous numbers*, Proceedings of the American Philosophical Society, 78 (1938), p. 551

les constantes en physique on retrouve le même phénomène : il y a plus de nombres commençant par le chiffre 1 que par le chiffre 2, etc, et les fréquences sont en ordre décroissant jusqu'au chiffre 9 (si l'on est en base dix).

C'est dans un article de 1881 que Newcomb remarque que les premières pages des tables de logarithmes étaient les plus usées, et en se servant de raisonnements assez simples il en déduit que la probabilité $p(k)$ que k soit le premier chiffre d'un nombre pris dans un ensemble de nombres quelconques est égale à

$$p(k) = \log_{10}(1 + 1/k), \text{ avec } k = 1, 2, \dots$$

3.2.2.2. Travail avec les étudiants

Nous avons donné aux étudiants divers catalogues récoltés dans les commerces de grande distribution voisins de notre établissement.

Nous leur avons demandé ce qu'ils pensaient de la fréquence du premier chiffre des nombres contenus dans les catalogues.

Quelques uns ont dit qu'il devait y avoir équiprobabilité mais d'autres ont fait remarquer que si on leur posait la question c'est que sans doute " il y avait quelque chose à trouver ".

Et nous avons ensuite saisi (sous Excel) par groupe de 2 étudiants l'ensemble des nombres puis mis en commun l'ensemble des données.

La représentation graphique de la distribution semblait bien contredire l'hypothèse d'équiprobabilité et les étudiants ont proposé de faire un test du Khi deux qui a permis de rejeter l'hypothèse d'équiprobabilité au seuil de 0.01.

3.2.3. Faire réaliser un travail d'enquête par les étudiants

Ce point est similaire au paragraphe précédent : « Faire réaliser un travail d'enquête par questionnaires aux étudiants : un outil didactique pour la statistique inférentielle à l'Université » et ne sera pas développé dans ce paragraphe. En revanche nous manquons un peu de temps.

Afin de ne pas diminuer le nombre d'exercices étudiés par rapport aux années précédentes nous avons utilisé des documents auto-correctifs et auto-évaluatifs (Régnier, 1994) qui a renforcé l'autonomie des étudiants par rapport aux apprentissages. Il existe quatre livrets qui sont structurés de la façon suivante :

- Une présentation
- Quelques objectifs spécifiques
- Un mode d'emploi
- Quelques éléments d'histoire de la statistique
- Des propos généraux sur la notion abordée
- Plusieurs documents de travail comportant
 - une situation problème
 - un traitement (à faire)
 - une grille d'auto-évaluation
 - une feuille (à remplir) d'analyse personnelle des erreurs commises
 - un corrigé du traitement des données.

3.3. L'enquête : élément central de l'enseignement de statistique en première année du département GLT

3.3.1. Présentation

Ces quelques lignes rendent compte d'un enseignement de statistique en première année du DUT Gestion Logistique et Transport ayant comme élément central la conduite d'une véritable enquête par les étudiants.

3.3.2. La SLTC

La Loi d'Orientation sur les Transports Intérieurs (LOTI³⁰) définit entre autre les compétences en matière de transports collectifs urbains. Elle suppose en premier lieu la

³⁰ Loi n° 82-1153 du 30 décembre 1982 d'orientation des transports intérieurs, in Journal officiel, 31 décembre 1982.- pp. 4004-4011

définition d'un périmètre des transports urbains (PTU) et confie la mission à une autorité organisatrice (AO). En ce qui concerne l'agglomération lyonnaise, c'est le SYTRAL (Syndicat Mixte des Transports en Commun de l'Agglomération Lyonnaise) qui est l'autorité organisatrice en matière de déplacements urbains. Composé d'élus de la Communauté Urbaine de Lyon et du Conseil Général du Rhône, le SYTRAL élabore, programme, finance, contrôle les investissements (autobus, rames de métro...) et les infrastructures (ouverture ou prolongement de ligne de bus, de métro, tramway, création de stations et d'arrêts, etc.). Le SYTRAL a délégué, comme la loi l'y autorise, à la Société Lyonnaise des Transports en Commun, entreprise privée filiale du groupe Via GTI (Général de Transport et de l'Industrie), l'exploitation du réseau de transports. Ses missions et les objectifs qui lui sont assignés sont définis par un cahier des charges qui précise les conditions d'exploitation ainsi que les critères de qualité à respecter. Ce cahier des charges apporte notamment des précisions sur les lignes, les fréquences, les amplitudes et les modes. Six objectifs sont clairement affichés : la propreté, la ponctualité, l'information des voyageurs, la lutte contre la fraude, le respect de la production et la disponibilité des équipements. L'exploitant s'engage sur un niveau de trafic traduit en objectif de recettes, ces dernières restant propriété de l'Autorité Organisatrice, l'exploitant étant pour sa part rémunéré sur une base forfaitaire. La SLTC rend compte de l'exploitation et assume le risque de charges de l'exploitation. Outre l'exploitation du réseau, la SLTC gère le patrimoine du SYTRAL et assiste celui-ci dans son rôle de maître d'ouvrage.

Plus d'un million de voyages sont effectués quotidiennement avec le concours de sept types de modes de transport. Le réseau de transports collectifs est décomposé en douze Unités de Transports (trois unités pour le métro et neuf unités de surface) auxquelles il convient d'ajouter deux ateliers centraux et trois unités spécialisées (voies et bâtiments, électronique et électrique). Les Unités de surface correspondent à un découpage géographique du périmètre des transports urbains. Ces unités disposent d'une relative autonomie dans leur organisation en particulier en matière de politique de marketing. C'est le cas de l'Unité des Pins qui est une des unités les plus importantes du réseau. C'est en collaboration avec le service marketing de cette unité que l'enquête a été organisée.

3.3.3. Les enquêtes à la SLTC

Les enquêtes constituent pour la SLTC un outil privilégié en matière de stratégie de marketing. Les objectifs sont nombreux : fidéliser la clientèle, attirer de nouveaux clients,

développer une gamme de produits, convertir des clients occasionnels en clients réguliers, etc.. Trois types d'enquêtes sont destinés à mieux évaluer le profil et les attentes de la clientèle : les enquêtes «freins et motivations», les enquêtes «images» et les enquêtes «satisfaction». Ce sont les enquêtes de satisfaction qui ont fait l'objet d'un partenariat entre la SLTC et l'IUT Lumière.

L'origine des enquêtes de satisfaction remonte à l'année 1995. Elles constituent un instrument de sensibilisation de tout premier ordre et permettent de repérer les priorités en matière d'amélioration de l'offre de service. La difficulté est qu'une enquête nécessite la mobilisation de moyens techniques et humains importants. Ce qui explique les difficultés rencontrées pour reproduire régulièrement ce type d'opération. L'Unité des Pins a ainsi réalisé depuis 1995, six enquêtes de satisfaction. La priorité étant donnée aux lignes en cours de certification, quatre lignes sur dix avaient jusque là fait l'objet d'une enquête.

3.3.4. Vers la réalisation de dix enquêtes simultanées

Compte tenu de l'importance de ces enquêtes, les responsables marketing de l'Unité des Pins ont décidé de tenter une opération originale : la réalisation de dix enquêtes simultanées. Toutes les lignes relevant de cette unité, à l'exception des deux navettes et des services spéciaux scolaires, ont fait l'objet d'une enquête de satisfaction. L'idée consistait à ne pas s'arrêter à la production de données et l'élaboration de traitements élémentaires. Les responsables du projet souhaitaient également que des propositions soient faites en matière de communication des résultats auprès de plusieurs populations distinctes : les clients, les conducteurs, les agents de maîtrise et l'ensemble des personnels du dépôt. La commande était double et nécessitait à la fois de mobiliser un nombre d'enquêteurs suffisant mais également d'être capable de répondre aux différents besoins exprimés aussi bien d'un point de vue statistique que du point de vue de la communication.

3.3.5. Le projet SLTC comme outil pédagogique multiple

Compte tenu de ce que nous avons dit préalablement l'engagement des étudiants et des enseignants de l'IUT ne pouvait se faire sur la simple base d'une prestation de service sous la forme d'une réalisation d'enquête, d'une saisie des données et de la réalisation de traitements élémentaires des informations, mais sur la base d'un partenariat ambitieux prenant en compte les exigences légitimes des deux parties.

3.3.6. Un constat, des a priori, une stratégie

Nous avons en effet une interrogation : "Quelles statistiques nos étudiants étaient-ils capables d'utiliser dans leurs pratiques professionnelles ?"

Et, sans que cela fût une réponse absolue à la question précédente, nous avons deux indicateurs nous donnant des éléments pour nous faire une idée de la réponse, à savoir d'une part les statistiques présentes dans le rapport de stage de première année (sept semaines en entreprise) et d'autre part celles se trouvant dans le mémoire de deuxième année (trente-six semaines en entreprise). Nous avons d'autres indicateurs que ces rapports écrits et les soutenances qui s'y rapportaient. En effet, deux études de cas ponctuaient le cours de première année, et trois évaluations lors du stage en entreprise étaient mise en place lors de la deuxième année.

A la lecture de ces indicateurs nous étions un certain nombre à penser que les résultats n'étaient pas à la hauteur des efforts consentis. A priori, il nous semblait que les apprentissages de nos étudiants fonctionnaient selon le schéma bien connu de "contextualisation – décontextualisation - recontextualisation" et nous espérions que nos étudiants seraient capables de "faire des ponts" entre d'une part les diverses composantes du cours de statistique, et d'autre part l'enquête nous ramenant ainsi à l'articulation décrite par M. Develay et Ph. Mérieu : "*C'est pourquoi il est si important de pratiquer ce que les chercheurs américains appellent le "bridging" et le fait de "faire des ponts" et qui consiste à demander aux sujets apprenants après maîtrise d'une situation ou d'une procédure de chercher eux-mêmes les situations où ils peuvent la retrouver et la faire jouer*".³¹ Sans jeu de mot il nous semblait que, sans l'enquête, la recontextualisation repoussée dans le rapport de stage était "un pont trop loin". Plus généralement c'est bien entendu sur "l'effet de contexte" que notre action se portait, ou encore comme le soulignent J. Mathieu et R. Thomas sur "*La notion de cadre [qui] intervient ici pour décrire l'organisation des connaissances autour d'un même thème. Un cadre sert à expliquer les inférences qui sont faites par un sujet à partir d'un contexte, c'est à dire comment certaines informations non fournies par le texte sont quand même utilisées par le sujet*"³².

³¹ Develay M., Mérieu Ph., *Emile reviens vite...ils sont devenus fous*, ESF éditeur, 1992, p 161.

³² Mathieu J., Thomas R., *La compréhension*, in Manuel de Psychologie, Vigot, 1985, p.154

3.3.7. L'organisation du cours

L'organisation proposée ici était relativement simple : le cours magistral et les conférences de méthodes permettent aux étudiants une contextualisation et une décontextualisation des notions de base des statistiques descriptives, l'enquête leur donnant l'occasion d'une première recontextualisation, et les indicateurs d'un apprentissage réussi étant la présence de statistiques dans les rapports de stage de première année et dans les mémoires de seconde année.

3.3.8. La place de l'outil informatique

Depuis déjà quelques années, une réflexion était menée par une équipe d'enseignants de l'Université Lyon 2, quant à l'apprentissage de la statistique descriptive. Initiée à l'Institut d'Etudes Politiques de Lyon, puis engagée à l'Institut Universitaire de Technologie Lumière et enfin à la Faculté de sciences économiques, cette démarche mettait l'accent sur deux postulats de base. Il s'agissait d'intégrer le plus tôt possible l'outil informatique comme instrument naturel d'une pratique de la statistique. En matière d'outil informatique le tableur fut privilégié. Plusieurs raisons expliquent ce choix qui peut apparaître surprenant en raison de la puissance de certains logiciels de traitements statistiques. Les cours concernaient des étudiants de premier cycle. Il nous paraissait indispensable d'éviter l'effet boîte noire du logiciel.

Le tableur, en obligeant les étudiants à reconstituer les indicateurs de base de la statistique descriptive nous offrait une plus grande garantie d'éviter cet effet boîte noire. Il était en outre intéressant de faire travailler les étudiants sur l'un des outils de bureautique les plus répandus en entreprise. Enfin, le tableur présentait l'avantage d'offrir une gamme intéressante d'outils graphiques, ce que peu de logiciels de statistique proposaient. Il fallait toutefois éviter certains pièges comme par exemple l'emploi impropre dans certains logiciels du terme d'«histogramme». Au-delà de la puissance de calcul, l'outil informatique favorisait le recours plus systématique à des bases de données de taille importante, se rapprochant ainsi davantage de ce que l'on peut trouver dans la pratique professionnelle. Le recours à l'informatique permettait enfin la simulation qui dans de nombreux cas offraient la possibilité de dépasser certains concepts pouvant apparaître plus complexes aux étudiants en favorisant la multiplication des applications concrètes.

3.4. Projets d'études statistiques STID 2ème année

Les projets d'études statistiques sont des projets importants de seconde année de DUT Stid. Ils comportent les étapes suivantes :

- Répondre à la demande d'une étude statistique par un vrai commanditaire
- Lors de cette étude appliquer le cours d'analyse des données, et les cours de statistique, d'informatique, etc.
- Sous la houlette d'un enseignant responsable
- L'étude se termine par une soutenance et la remise d'un dossier

Conditions

- Par groupes d'étudiants (30 divisé par n, le nombre de projets)
- Environ 70 heures dont la moitié sont encadrées, les autres sont en autonomie
- Deux rencontres, en cours de travail à des dates à définir, seront organisées afin que chaque groupe présente l'avancement de ses travaux aux autres étudiants et enseignants.
- Projets envisagés (en 2006-2007)
 1. Chambre de commerce : faire une enquête de satisfaction concernant un concours « Grand prix du Design »
 2. Energie directe Gaz de France : étude d'impact de campagnes publicitaires
 3. ADIUT : traiter les données concernant les enquêtes des diplômés des IUT de la région Rhône Alpes
 4. Formation à distance : enquête sur l'impact d'une formation à distance
 5. Agence d'urbanisme : collecter et traiter des données concernant l'agence d'urbanisme du Rhône.

_Les deux études présentées dans la partie V de ce travail sont issues de projets d'études statistiques.

3.5. Conclusion

Favorisant la pratique de production des apprenants, ce travail d'enquête par questionnaires renforce, pour les étudiants, les signifiants, mais aussi les signifiés, à l'intérieur de même de la « situation de formation ».

Obligés de construire leurs propres énoncés, de travailler dans des situations ouvertes, les apprenants s'approprient plus aisément les schèmes qui sont convoqués dans l'apprentissage de la statistique.

Dans ces activités, ils travaillent à la fois par rapport à l'éthique du sens et à celle de la norme, organisant les concepts (statistiques) ; et l'on peut supposer que les sujets ont ainsi construit des compétences pouvant être, plus facilement et plus immédiatement, investies sur le terrain professionnel.

On peut raisonnablement penser comme G. Bachelard "*...[qu'] en établissant à propos de chaque notion une échelle de concepts, en montrant comment un concept en a produit un autre, s'est lié avec un autre. Alors il aura quelque chance de mesurer une efficacité épistémologique. Aussitôt la pensée scientifique apparaîtra comme une difficulté vaincue.*" (Bachelard, 1938, p. 18).

C'est une observation scientifique du travail d'enquête par questionnaires faits par les étudiants, qui nous donnera des indications précieuses sur les obstacles que rencontrent les étudiants dans l'apprentissage de la statistique et sur les concepts mis en œuvre ; ce travail nous permettra d'apporter une pierre à la construction de la didactique de la statistique et au passage de répondre à la question liminaire « de quoi sont capables nos étudiants tant sur le terrain de la statistique que sur celui de la statistique de terrain ? ».

4. Partie IV

Simulation en statistique et conceptualisation

4.1. Introduction

4.1.1. *Une spécificité endogène au champ de la statistique*

Notre propos s'ancre sur la prise en compte de la place de la simulation dans la pratique de l'enseignement de la statistique et dans le champ théorique de la didactique de la statistique. La simulation ressort dans nombre d'ouvrages de référence, de directives officielles, de commentaires des programmes comme un outil privilégié qui doit être mis au service des apprenants dans les situations didactiques visant des notions du champ de la statistique.

Ainsi de nombreux logiciels de simulation plus ou moins sophistiqués, élaborés à l'intention des étudiants et des enseignants sont apparus. Cette panoplie va des langages dédiés jusqu'aux outils généralistes.

À ce jour le langage Java permet de produire des Applets qui offrent des possibilités considérables. Et même si notre propos n'est pas de développer ici cette idée, nous pensons que cet intérêt pour les simulations et cette insistance à les intégrer dans des séquences didactiques concernant les apprentissages statistiques sont d'une certaine manière significatifs d'une spécificité endogène au champ de la statistique.

4.1.2. *Un illustre précurseur*

Galton, dans une lettre à son cousin Darwin, voulant illustrer certaines propriétés de la loi binomiale produit le croquis ci-dessous :

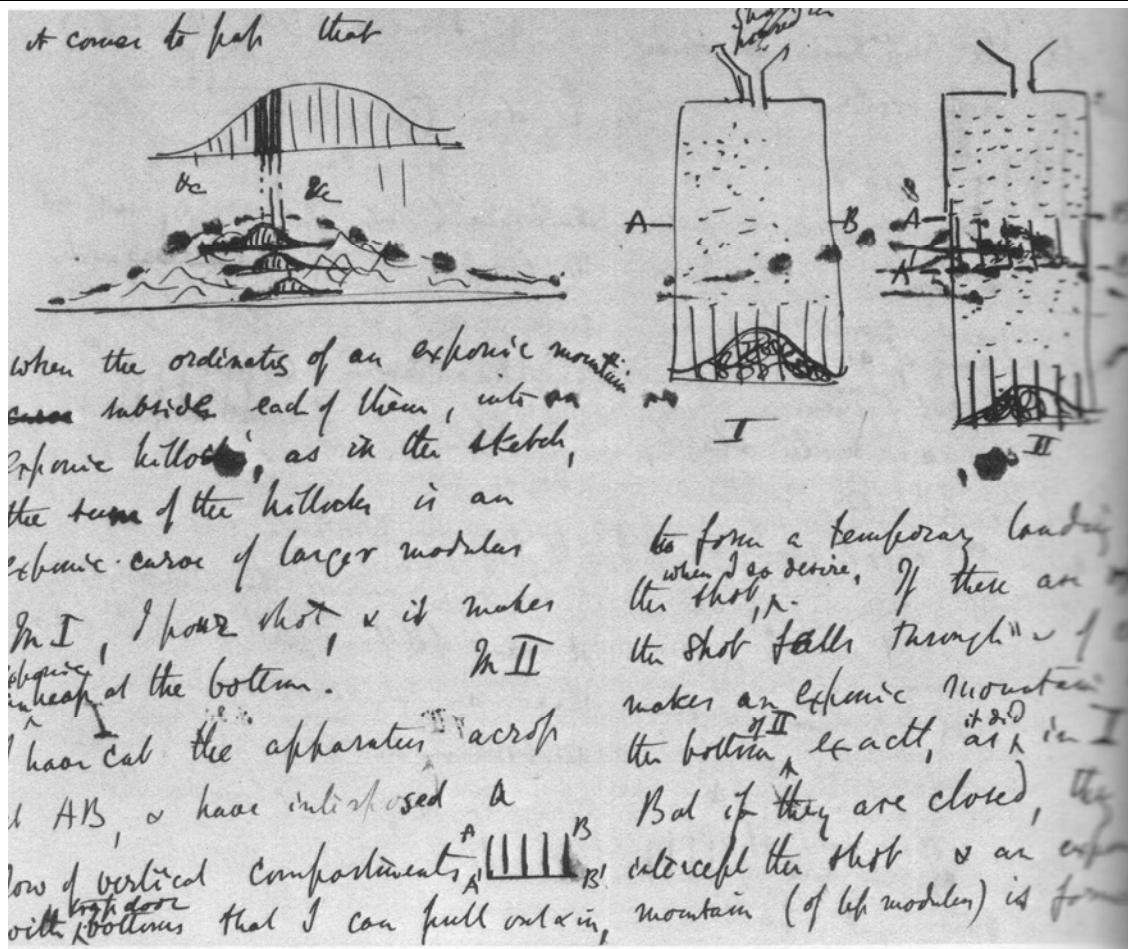


Figure 5 : Extrait d'une lettre du 12 janvier 1877 de Galton
 à son cousin G. Darwin in « The history of statistics » Stigler, 1986.

Il fait fabriquer le quinconce (dispositif connu à présent sous le nom de planche de Galton), puis le double quinconce. C'est en quelque sorte le premier à instrumentaliser la simulation pour comprendre ou faire comprendre la statistique. En effet Galilée et « La lettre au Duc de Toscane » ou les échanges entre le Chevalier de Méré et Pascal utilisent plutôt des données statistiques supposées afin d'étayer leurs théories.

4.1.3. Un usage bien répandu

Qui n'a vu dans ces grandes surfaces consacrées à l'ameublement de la maison ces machines sensées démontrer la solidité des sièges ou celle des portes de meubles ouvertes et fermées sans relâche par un dispositif mécanique ?

On imagine moins bien faire les fameux « crash tests » sur un trop grand nombre d'automobiles, et on admet avec un soulagement teinté d'inquiétude, malgré tout, que les

simulations touchant de près ou de loin le nucléaire doivent suffire pour travailler sur ce domaine.

En fait la simulation concerne actuellement un nombre sans cesse croissant d'activités industrielles : l'automobile, la conception assistée, l'industrie aéronautique, l'industrie nucléaire, les industries chimiques et pharmaceutiques, les prévisions météorologiques, l'industrie de l'espace, la micro électronique, etc.

Cette prolifération vient de deux raisons principales :

- d'une part la puissance de calcul présente dans des ordinateurs de faibles coûts,
- d'autre part une importance sans cesse accrue et une efficacité de plus en plus réelle des modèles théoriques concernant les divers domaines de la production et de l'exploitation.

De quoi parle-t-on ? On peut dire que l'on fait une simulation lorsqu'on expérimente non pas sur l'objet réel mais sur un modèle.

4.2. Les invariants de la variabilité

Yves Chevallard (Chevallard, 2003) et Floriane Wozniak (Wozniak, 2005) ont bien distingué dans les situations d'apprentissage de la statistique une notion centrale, celle de la variabilité.

Afin d'aller plus loin, remarquons que si l'on veut développer une pensée statistique, on rencontre une difficulté spécifique de cet apprentissage : comment dégager des invariants de situations dont l'essence même est la variabilité ?

Bien entendu une figure de géométrie n'est pas fixe pour un énoncé donné, mais c'est seulement dans le cas de la statistique qu'un coup de dé, un aléa, le hasard, les variations d'échantillonnage font partie de la situation.

D'un autre point de vue la multiplication des expériences rend en statistique plus de services qu'ailleurs parce que l'intuition y est également plus souvent prise en défaut.

C'est cette expérience qu'il convient de nourrir en renouvelant plusieurs fois une même situation. Comme il est très difficile de répéter un grand nombre d'expériences dans un temps raisonnable, on fait appel tout naturellement à des simulations.

Ces résultats successifs, d'une façon générale différents, permettent de se familiariser avec les invariants de la variabilité.

4.3. Quelques distinctions concernant les simulations utilisées en statistique et nos choix concernant la construction des simulations

Comme le signale Jamie D. Mills (Mills, 2002) on peut distinguer dans la littérature quatre manières d'envisager l'intégration de simulations dans l'apprentissage de la statistique : *" Four definitions of computer simulation methods were described in the literature reviewed for this article. One definition involved students writing their own programs (using SAS PROC IML®, say), setting up a model for a problem and investigating diagnostics for the model in seeking possible violations of assumptions. A second definition allowed students to experience similar advantages using a random number generator in Excel® or MINITAB®. Using Excel® or MINITAB®, the commands to generate the random samples and perform experiments on the model are mostly window-driven. Third, many instructors used some combination of the first and second definitions, by providing program templates that allowed students to change parameters during the experiments (commonly in SAS® or SPSS®). Finally, a fourth definition involved using commercial software packages designed exclusively for simulation purposes (for example, the "Samplings Distribution" program). The literature reviewed in this paper included journal articles that utilized all four operational definitions, although the majority of the authors reported using the latter three. "*

Parallèlement à cette typologie des types de simulations, nous avons fait les choix suivants en ce qui concerne les simulations construites dans le cadre de cette étude :

A. Le choix d'Excel

Notre pratique pédagogique vise à intégrer le plus tôt possible l'outil informatique comme instrument canonique de la pratique de la statistique.

Et en matière d'outil informatique, le tableur est privilégié. Plusieurs raisons expliquent ce choix surprenant en comparaison de la puissance de certains logiciels de traitements statistiques. Le tableur est dans ce cadre utilisé par différentes disciplines

(comptabilité, gestion, mathématiques, mathématiques financières) et l'investissement sur un autre logiciel serait peu rentable à ce stade. Par ailleurs nous demandons aux étudiants de ne pas utiliser Visual Basic for Applications afin de ne pas multiplier les difficultés.

B. Le choix de faire construire les simulations par les étudiants

Il nous est apparu très vite que les simulations qui apportaient le plus aux étudiants étaient celles pour lesquelles une partie du cours était appliquée dans un environnement différent, c'est la raison pour laquelle toutes les simulations sont construites par les étudiants.

4.4. Les outils de base pour construire une simulation

Pour construire une simulation nous avons besoin d'un petit nombre d'outils de base : un générateur de nombres pseudo aléatoires, être capable d'itérer c'est à dire de faire plusieurs tirages successifs, un compteur, une remise à zéro. Détaillons ces briques de base :

- générateur de nombres aléatoires : Excel comporte deux générateurs de nombres pseudo aléatoires, `alea()` qui renvoie un nombre « au hasard », c'est-à-dire avec une densité uniforme, entre 0 et 1, et qui est un « bon » générateur (il a subi avec succès l'ensemble des tests auxquels nous l'avons soumis, et `alea.entre.bornes(a;b)` qui renvoie un nombre entier avec une équiprobabilité sur les entiers compris entre a et b (a et b entiers, $a < b$), ce dernier générateur est plutôt moins « bon » que `alea()` mais il est très pratique quand on doit simuler une variable discrète.
- Pour faire des itérations sur Excel, il faut cocher la case « Itération » dans l'item « Options » du menu « Outil » ; on choisit le nombre d'itérations, ce nombre ne doit pas être supérieur à la valeur 32767 (c'est-à-dire 2 puissance 15 moins 1). Une fois la boucle finie il faudra appuyer sur la touche F9 pour lancer une nouvelle série.
- Quand la case « Itération » est cochée on peut facilement fabriquer un compteur (par exemple le contenu de la case A1 sera « =A1+1 »),
- Et également une succession d'une grande quantité de nombres pseudo-aléatoires ; par exemple le contenu de la case A2 sera « =A2+Alea()-A2 »,

- La remise à zéro se fait tout simplement à l'aide d'un SI dans les cellules que l'on veut remettre à 0, et d'une cellule qui sert à faire basculer vers la remise à zéro.

L'ensemble de ces stratagèmes va permettre de construire des feuilles de calcul simulant diverses situations et aussi d'utiliser un tableur comme aide à la décision.

Ajoutons que le modèle théorique de la notion de tableur convoqué ici est celui d'un automate cellulaire où l'état de chaque cellule est défini en fonction de ses voisins et de contraintes externes.

4.5. Premier niveau de simulations : constat de la stabilité et de la convergence

Dans ce premier niveau nous faisons construire des simulations concernant les expériences aléatoires suivantes : pile ou face, pièce truquée ou non, un dé, pipé ou non, moyenne et distance quadratique, médiane écart moyen, schéma de Bernouilli, planche de Galton. Décrivons plus en détail cette dernière activité.

Le dispositif : une bonne introduction à la loi binomiale peut s'appuyer sur le tableau ou la planche de Galton. Mais si l'expérience est aisée, elle est difficile à montrer réellement dans un cours devant une centaine d'étudiants. Ainsi beaucoup ont eu l'idée de simuler l'expérience ; et si l'on connaît diverses versions de simulation à l'aide de langage de haut niveau (Pascal, Basic, etc.) nous avons choisi de simuler la planche de Galton à l'aide d'un tableur. Reprenons sa description dans la "*Kleine Enzyklopädie der Mathematik*" page 640 : "*Il s'agit d'un tableau parsemé de tiges rigides. Ces tiges sont placées de telle façon que la distance entre deux tiges adjacentes situées dans une même rangée est partagée dans les proportions p et $(1-p)$ on peut expérimenter ainsi n'importe quelle distribution binomiale*".

Une approche cellulaire : nous utiliserons les outils fondamentaux décrits ci-dessus ; l'initialisation se fait dans la cellule O1, la valeur « vrai » dans cette cellule correspond au lâcher d'une nouvelle bille. On tire un nombre au hasard et selon le rapport entre la valeur de ce nombre et la valeur de p choisie au départ la bille "descend" à gauche ou à droite. On recommence l'opération jusqu'au bas du tableau et l'on incrémente le compteur de la cellule où la bille est finalement arrivée. On peut construire un graphique représentatif des valeurs trouvées.

Visualisation : on peut se rapprocher de l'expérience en transformant l'affichage. Les cellules contenant "Vrai" afficheront l'objet "Bille" et les cellules contenant "Faux" afficheront l'objet "Clou". Nous obtiendrons alors :

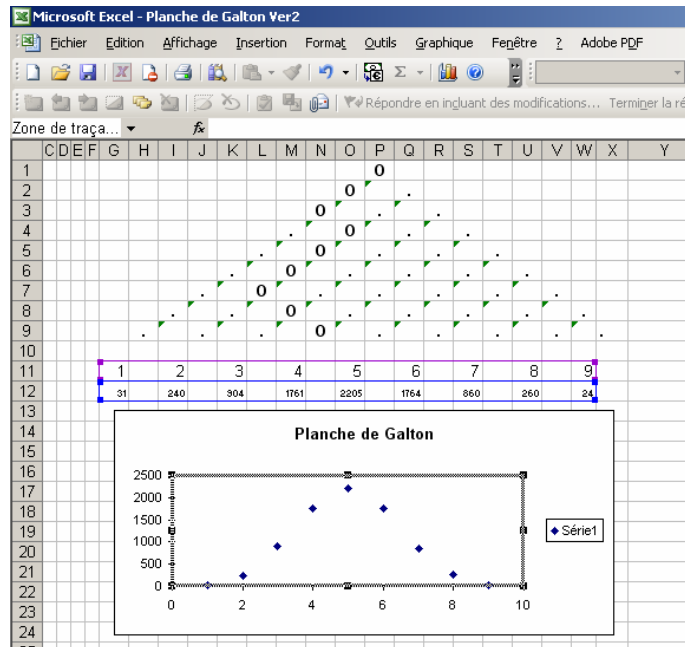


Figure 6 : Simulation de la planche de Galton

4.6. Simulation mettant en place un calcul de convergence

Une autre série de simulations prend naissance dans la pluie aléatoire de points sur un rectangle donné ; on peut avoir ainsi une approximation de π , de n'importe quelle intégrale (par exemple intégrale de 1 à e de $\ln(x)$) ou bien de l'aire sous la courbe de Gauss. En considérant la pluie aléatoire comme un phénomène binomial on calculera un intervalle de confiance de la valeur cherchée à $k\%$, de la valeur approchée de l'intégrale.

Prenons l'exemple de l'aire « sous » la courbe de la loi normale centrée réduite comprise entre les valeurs $x=-3$ et $x=3$. Projetons les points sur le rectangle défini par l'axe des x les droites $x=-3$ et $x=3$ et la droite passant par le sommet de la courbe de Gauss, ce qui pour 1000 points donne quelque chose qui ressemble à cela :

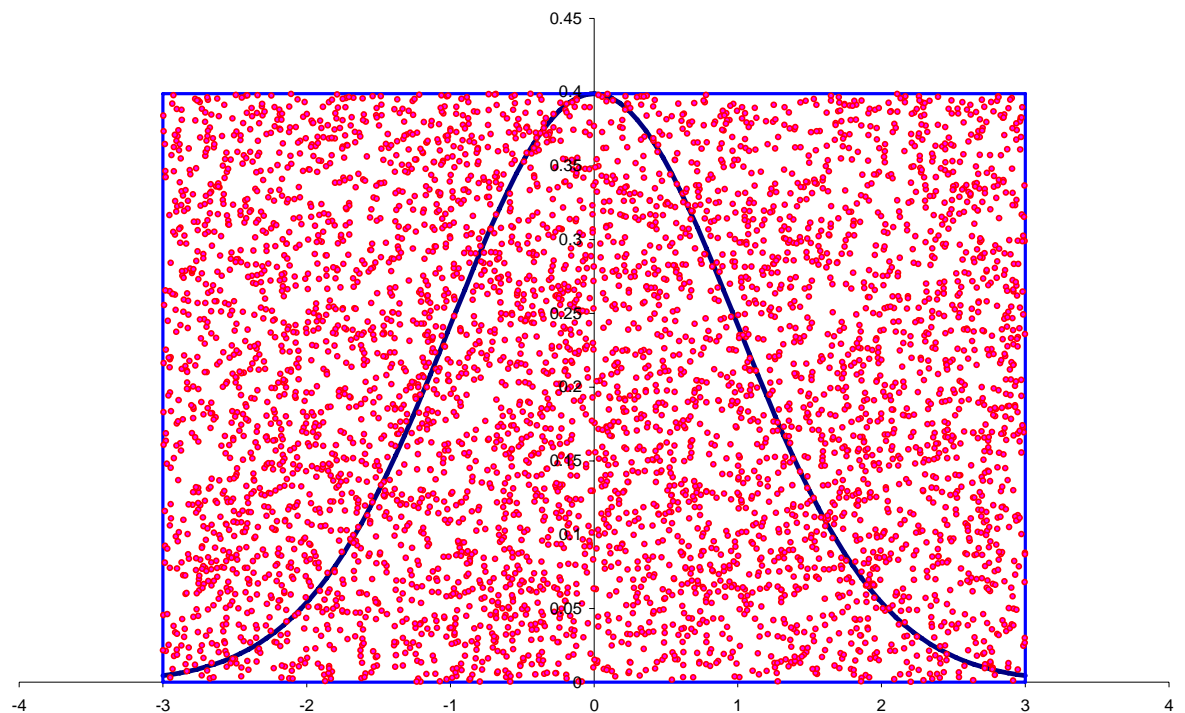


Figure 7 : Pluie aléatoire et courbe de la loi normale centrée réduite

On peut observer la fréquence des points sous la courbe par rapport au nombre total de points. Afin de calculer un intervalle de confiance des fréquences, on suppose que pour un point le fait d'être « sous » la courbe est un phénomène binomial. La probabilité p pour un point d'être sous la courbe est égale à l'aire sous la courbe (soit environ 0,997300204) divisée par l'aire totale du rectangle (c'est-à-dire $6 * 0,39894228$ soit 2,393653682) on trouve donc $p = 0,416643482$ (c'est une valeur approchée). On peut alors facilement calculer pour un tirage de 1000 points les bornes inférieures et supérieures à 90% par exemple et ensuite reporter les valeurs observées lors de 100 tirages sur un graphique comme ci-dessous :

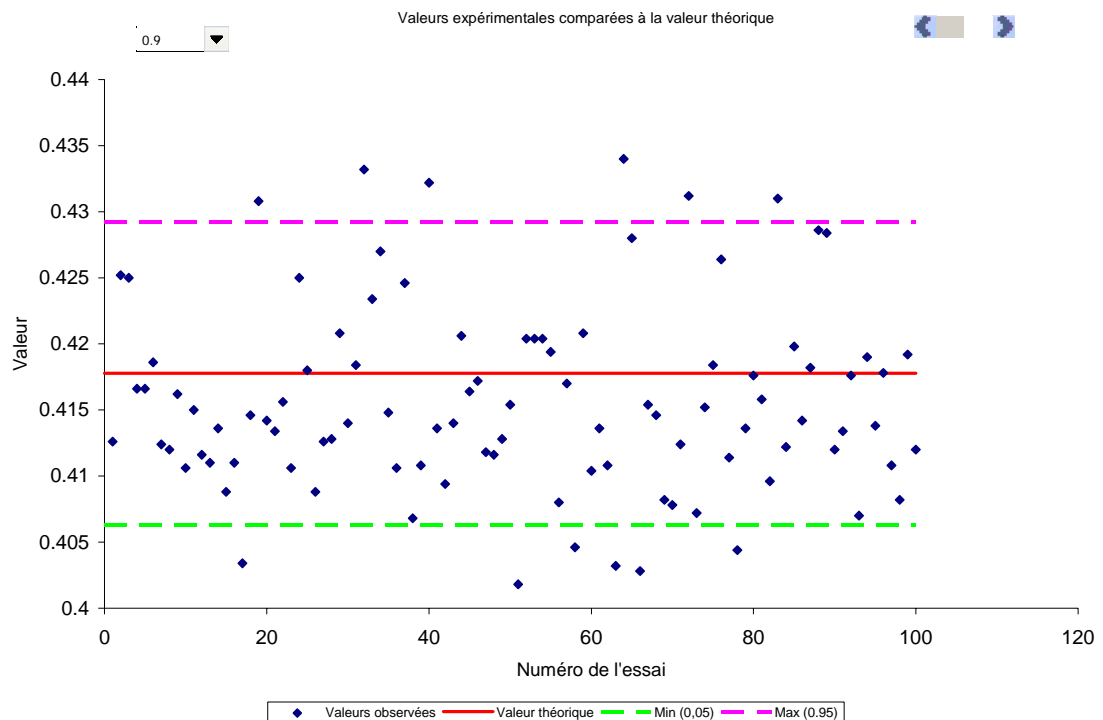


Figure 8 : Valeurs expérimentales et intervalle de confiance

4.7. Simulation concernant le concept d'intervalle de confiance d'une moyenne

4.7.1. Quelques travaux concernant les intervalles de confiance et la simulation

L'utilisation de la simulation pour l'enseignement de l'inférence statistique a fait l'objet de développements et d'expériences didactiques. Citons [Kennedy, Olinsky, and Schumacher \(1990\)](#) utilisant MINITAB®, [Marasinghe et al. \(1996\)](#) développant un module et un logiciel particulier, [Hesterberg \(1998\)](#) qui après diverses expérimentations recommande un langage interactif et plus particulièrement S-PLUS®, ou [Pedazhur \(1997\)](#) défendant le point de vue d'une compréhension plus intuitive des notions. Enfin Batanero et Godino (Batanero, Godino 2001 p.7-10) observent que " *Puesta que la media de la muestra varia de una muestra a otra, los intervalos de confianza variarán de una muestra a otra (lo mismo ocurre con la propoción). Lo que nos dice el coeficiente de confianza es que en un porcentaje dado de muestras, el verdadero valor del parámetro estará incluido en el intervalo.* "

4.7.2. La situation problème et la construction de la simulation

La séance didactique prend appui sur une bonne maîtrise du logiciel Excel par les étudiants et fait suite à une séance d'une heure sur les intervalles de confiance. Les consignes sont les suivantes : « Vous devez construire un classeur Excel, sans utiliser le langage VBA, qui contient les opérations permettant de tirer au hasard 10000 nombres compris entre 0 et 1, de calculer leur moyenne. Après quoi qui permet par tirage aléatoire, d'obtenir 10 échantillons, chacun de trente nombres, de calculer l'intervalle de confiance pour la moyenne selon un niveau choisi par menu entre 90% et 99% et d'afficher les intervalles correspondants à chaque échantillon ». Nous ajoutons que « Cet énoncé subira des modifications dans une heure avec des consignes modifiées. ». Par binôme, les étudiants produisent un classeur dont la feuille de visualisation a, aux variations individuelles près, l'aspect suivant :

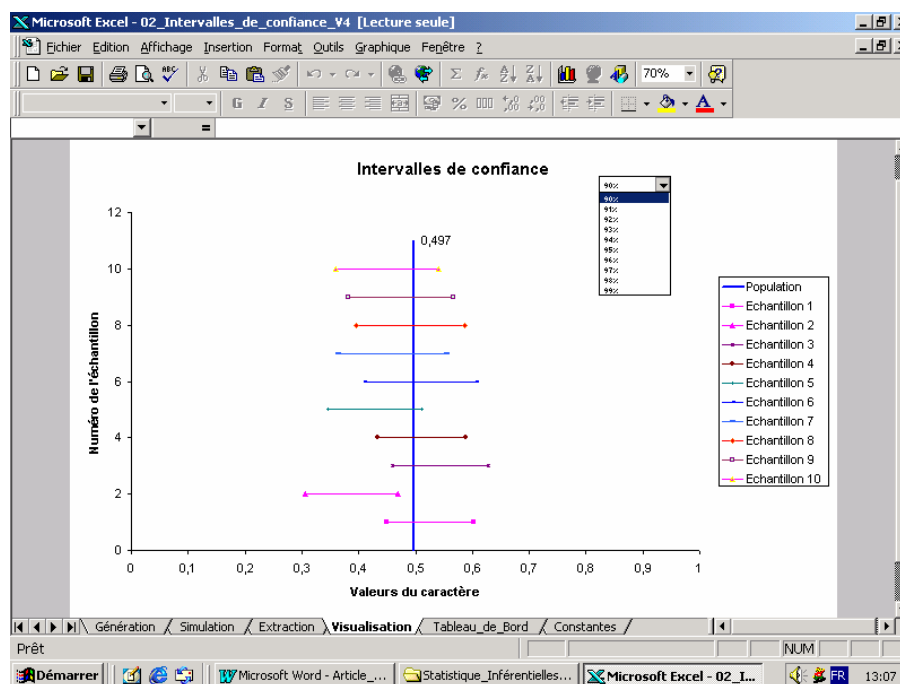


Figure 9 : Simulation d'un tirage de 10 échantillons de 30 individus sur une population de 10000

Les variations de l'énoncé concernent : le changement ou non de la population chaque fois que l'on tire 10 nouveaux échantillons, le choix de la taille de la population (<65536), le choix de la taille des échantillons, le nombre d'échantillons. Ce qui donne le " tableau de bord " suivant :

Population aléatoire à chaque nouveau tirage de 10 échantillons	Oui
Taille de la population	N = 10000
Taille de chaque échantillon	n = 30
Niveau de confiance	90%
Nombre d'échantillons	10

Figure 10 : Tableau de bord de la simulation

4.8. Simulation et concept de corrélation

4.8.1. Quelques travaux concernant la corrélation et la simulation

La plupart des travaux publiés sur ce sujet sont américains. Citons brièvement Franklin (1992) qui propose l'utilisation de Minitab®, mais également Jensen (1983), Pulley et Doolbear (1984), Olinsky et Schumacher (1990), Ferral (1995), Romeu(1995) qui parmi d'autres proposent des approches des coefficients de régression et de détermination utilisant des générateurs de nombres aléatoires. Batanero et Godino (2001 page 4-15) proposent des études de la corrélation utilisant STATGRAPHICS ; ils citent Inhelder et Piaget ces derniers relevant « que algunos sujetos analizan solamente la relacion entre los casos favorables positivos », et ils font également référence à Chapman y Chapman (1967) qui développent la notion de « contingencia empiricas ».

4.8.2. La situation problème et la construction de la simulation

La séance didactique fait suite à une séance de quatre heures sur la corrélation. Les consignes sont les suivantes : « Vous devez construire un classeur Excel, sans utiliser le langage VBA, qui contienne les opérations permettant de tirer au hasard $2N$ nombres compris

entre 0 et 1, de représenter sur un graphique N couples ayant pour coordonnées les nombres précédents, de calculer le coefficient de détermination.

Vous devez conserver les résultats dans un tableau et représenter ce tableau à l'aide d'un graphique. Vous devez construire un dispositif de remise à 0 et un autre dispositif permettant de choisir le nombre N de points. »

Par binôme, les étudiants produisent un classeur dont la feuille de visualisation a, aux variations individuelles près, l'aspect suivant :

Coordonnées des points

0,41375358	0,908240487
0,68523213	0,659984889
0,87607447	0,239837341
0,13867309	0,698268701
0,11978189	0,958255627
0,58925438	0,922923736
0,27862459	0,81292607
0,58346583	0,245340285
0,42924337	0,058812911
0,81047842	0,653848759

Représentation graphique :

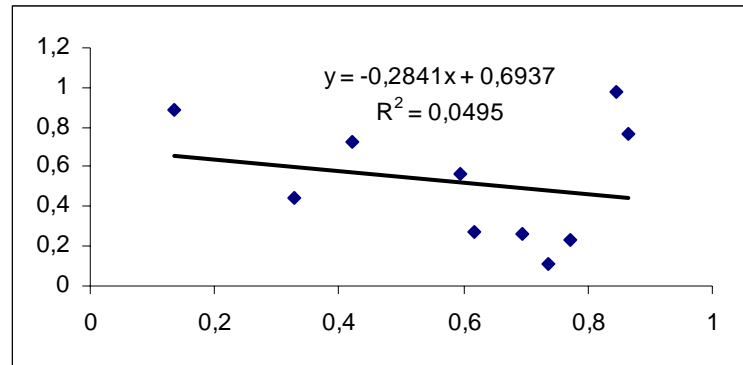


Figure 11 : Dix points au hasard ayant des coordonnées entre 0 et 1

Tableau récapitulant les résultats :

Première colonne : valeur de R^2 première ligne entre 0 et 0,1 deuxième ligne entre 0,1 et 0,2 etc, deuxième colonne : nombre de configurations, troisième colonne : pourcentage sur le total

0	3846	60,63%
1	1113	17,55%
2	573	9,03%
3	395	6,23%
4	223	3,52%
5	112	1,77%
6	51	0,80%
7	25	0,39%
8	6	0,09%
9	0	0,00%
	6344	

Et enfin un graphique représentant le tableau précédent

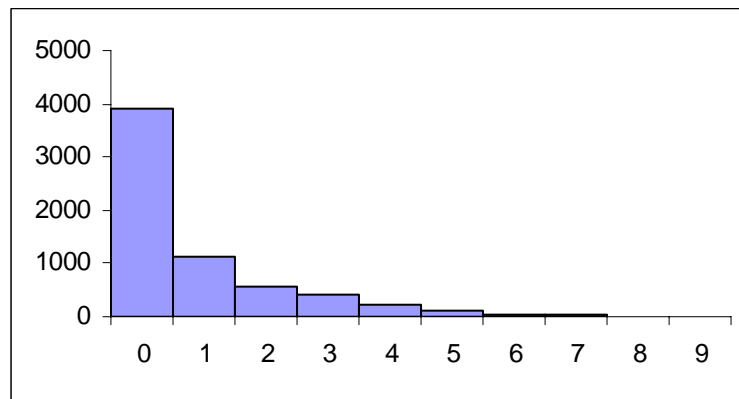


Figure 12 : Répartition de 6344 valeurs du coefficient de détermination de 10 points au hasard (coordonnées entre 0 et 1)

Voici d'autres résultats pour trois points :

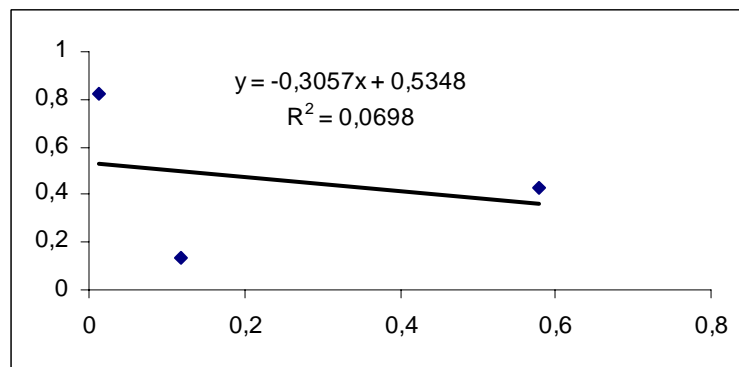


Figure 13 : Trois points au hasard ayant des coordonnées entre 0 et 1

Avec les mêmes conventions :

0	1433	59,73%
1	434	18,09%
2	236	9,84%
3	125	5,21%
4	84	3,50%
5	53	2,21%
6	26	1,08%
7	8	0,33%
8	1	0,04%
9	0	0,00%
	2400	

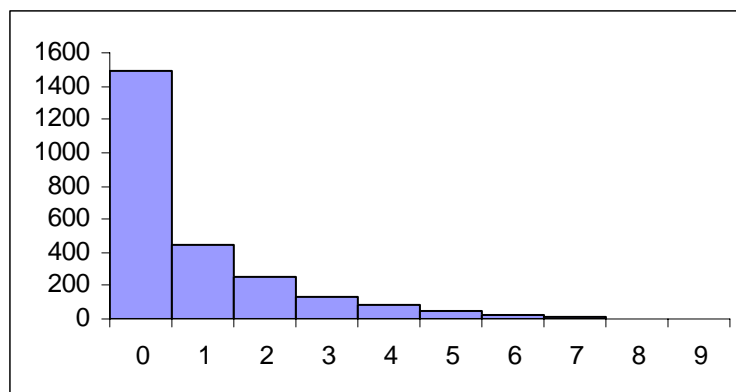


Figure 14 : Répartition de 2400 valeurs du coefficient de détermination de 3 points au hasard (coordonnées entre 0 et 1)

4.8.3. Conclusion

Nous ne serons pas ici les caudataires d'une loi supposée défendant un point de vue ou un logiciel particulier, mais plutôt dans une approche d'ingénierie didactique dans laquelle la simulation joue un rôle dans la construction des savoirs. Ici, la construction par les apprenants, eux-mêmes, de l'outil s'apparente aux méthodes d'apprentissage où les premiers travaux des apprentis consistaient à fabriquer leurs propres outils. Parmi les problèmes en suspens reste le contrôle par les étudiants du générateur de nombres aléatoires d'Excel.

4.9. Étude de la transformation de Fisher dans l'analyse du coefficient de corrélation : d'une pratique fondée sur l'argument d'autorité et la lecture d'abaques à une argumentation fondée sur la simulation.

Ce paragraphe se propose d'étudier la transposition didactique en acte. Notre propos concerne précisément l'analyse de la transformation de Fisher dans l'étude du coefficient de corrélation, rencontrée en particulier par des étudiants de deuxième année de DUT STID (ainsi que par les étudiants de licence professionnelle Chargé d'études statistiques). La particularité du travail présenté consiste dans le fait que les étudiants construisent l'outil de simulation, à partir de consignes données par l'enseignant. Nous présenterons la séquence didactique mise en place, les principaux obstacles rencontrés et la mesure des résultats obtenus dans les apprentissages par les étudiants.

4.9.1. Introduction

En nous interrogeant sur la question de la formation citoyenne en statistique, nous sommes confrontés à des questions relatives aux concepts à enseigner et aux stratégies pédagogiques fondées sur la transposition et une ingénierie didactique. Nous nous intéressons ici non pas à la démonstration « mathématique » de l'amélioration de la convergence grâce à la transformation de Fisher appliquée au coefficient de corrélation, mais aux explications données aux étudiants dans le programme universitaire du DUT STID et à leur évolution depuis les années 70. Nous constatons ainsi le passage d'arguments d'autorité couplés à l'outil « abaque », à une argumentation plus souple fondée en partie sur la simulation.

Nous décelons dans cette modification des pratiques didactiques de l'enseignement de la statistique au niveau bac +2, d'une part une évolution des pratiques sociales et d'autre part l'émergence d'une place et d'un rôle plus importants attribués aux outils informatiques dans la pédagogie universitaire de la statistique. Il s'agit de tenter de déterminer l'intérêt d'une telle démarche mais aussi d'en relever les risques de genèse d'obstacles didactiques qui pourront perturber la formation en statistique.

4.9.2. Une remarque préalable

Si l'on est attentif aux enseignements concernant le coefficient de corrélation il apparaît qu'une analyse d'un échantillon d'ouvrages fait ressortir une part non négligeable d'entre eux qui passent sous silence la question de la confiance à accorder à la réalisation du coefficient empirique obtenue sur un échantillon aléatoire extrait de population étudiée.

Comment pouvons-nous interpréter ce fait ? Peur de « rentrer » dans une question trop complexe, éloignement des justifications satisfaisantes, recommandations faites par les programmes ou leurs commentaires ?

4.9.3. Petit retour sur l'état du problème pédagogique

Nous sommes en face d'un problème d'enseignement d'une notion que nous pouvons décrire de la façon suivante :

Il est dit que nous étudions deux variables quantitatives X et Y formant un couple gaussien sur une population donnée de laquelle nous tirons un échantillon aléatoire de taille n . Sur cet échantillon nous calculons la réalisation du coefficient empirique de corrélation

linéaire R entre X et Y . La valeur r obtenue est une estimation du coefficient ρ de corrélation linéaire entre les variables X et Y sur la population.

Il s'agit alors de tester les hypothèses : $H_0 : \rho = 0$ contre $H_1 : \rho \neq 0$

Sous H_0 , on connaît une réponse à ce problème : la variable aléatoire R coefficient empirique est telle la variable aléatoire $\frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$ suit la distribution de probabilité de Student ayant $\nu = n - 2$ degrés de liberté.

Mais cette apparente simplicité est souvent mise à mal car, en particulier, la distribution de la variable R est complexe dès que ρ s'éloigne de 0.

C'est alors que l'on peut réaliser une transformation dite *transformation de Fisher* en construisant une nouvelle variable Z définie par : $Z = \frac{1}{2} \ln\left(\frac{1+R}{1-R}\right)$ ou encore la fonction $\text{Arg th}R$ (argument de tangente hyperbolique de R)

Il est dit alors que si $n > 30$, taille de l'échantillon, Z suit approximativement une loi de Laplace-Gauss de paramètres $LG(\mu, \sigma)$ avec

$$\mu = E(Z) = \frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right) + \frac{\rho}{2(n-1)} \quad \text{et} \quad \sigma^2 = V(Z) = \frac{1}{n-3}$$

De là il ressort la possibilité de construire un intervalle de confiance à niveau $1-\alpha$ par rapport à Z et donc par rapport à R moyennant la transformation réciproque :

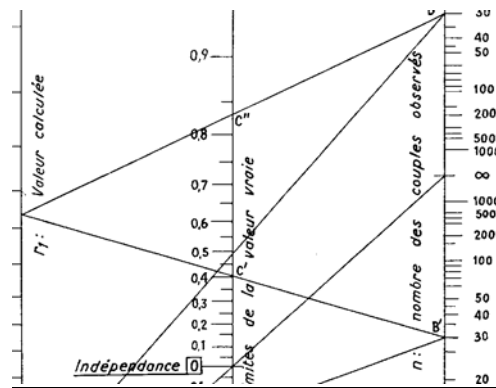
$$R = \frac{\exp(2Z) - 1}{\exp(2Z) + 1}$$

4.9.4. Quelques solutions apportées dans des ouvrages de base

Dans un certain nombre d'ouvrages, c'est une solution graphique qui est fournie. Ainsi on donne un schéma sous la forme d'abaque permettant de se dispenser de ces calculs.

Signalons deux réalisations :

Tout d'abord dans (Liorzou 1979 p.175)



Ou bien encore dans (Wonnacott Wonnacott 1972 p.545)

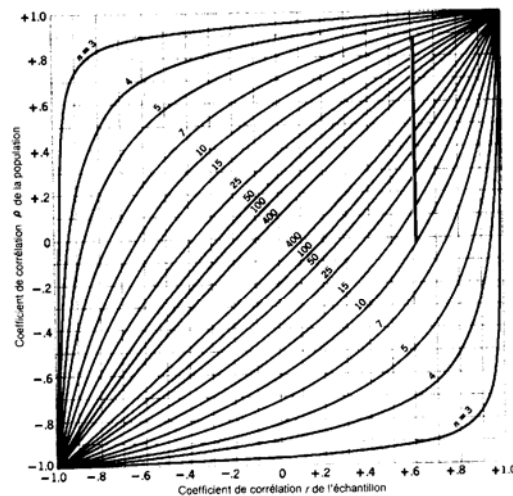


FIGURE 15-4
Bande de confiance à 95 % pour le coefficient de corrélation ρ dans le cas d'une population normale bivariee, pour différentes valeurs de la taille n de l'échantillon.

Ces deux abaques sont assez efficaces mais les étudiants qui les utilisent perdent de vue les fluctuations d'échantillonnages et ont des difficultés à faire le lien avec les éléments de théorie que nous avons rappelés ci-dessus, comme nous avons pu le constater à partir des données construites par observation dans la salle de classe. C'est pour cela que nous avons été conduit à réaliser une transposition didactique de l'approche de ce concept de corrélation linéaire fondée sur la simulation.

4.9.5. Une séquence didactique d'observation des fluctuations d'échantillonnage de R

Les étudiants ont réalisé une feuille Excel qui fournit des réalisations du coefficient empirique R à partir d'échantillons aléatoires avec remise extraits d'une population de taille $N=1000$. Les deux variables étudiées X et Y connues sur la population donnent donc la

possibilité de connaître la valeur certaine de ρ . Le graphique ci-dessus donne les résultats d'une expérience fondées sur 15 échantillons et permet de comparer les 15 réalisations notées r_1 à r_{15} à la valeur exacte de ρ .

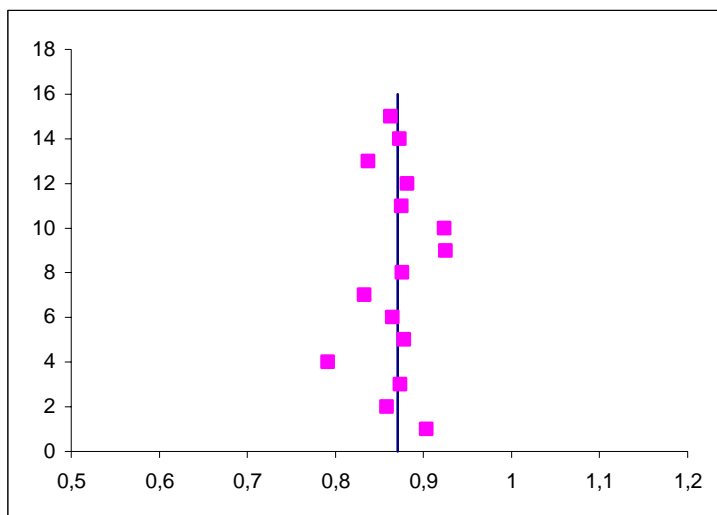


Figure 15 : Fluctuations d'échantillonnage de R coefficient de corrélation

4.9.6. L'introduction de la transformation de Fisher et de la variable Z

Sur ces quinze échantillons les étudiants ont appliqué la transformation de Fisher puis ont calculé et représenté graphiquement les intervalles de confiance. En voici un exemple :

N°échantillon	Réalisations de Z	Intervalle de confiance	
1	1,42583773	1,04864271	1,80303274
2	1,16252098	0,78532597	1,539716
3	1,51054716	1,13335214	1,88774217
4	1,40355715	1,02636214	1,78075217
5	1,45978321	1,0825882	1,83697823
6	1,38219475	1,00499973	1,75938976
7	1,26767646	0,89048145	1,64487148
8	1,374048	0,99685298	1,75124301
9	1,32848267	0,95128765	1,70567768

10	1,45518598	1,07799097	1,832381
11	1,45997334	1,08277832	1,83716835
12	1,76132777	1,38413276	2,13852279
13	1,13325308	0,75605807	1,51044809
14	1,3976812	1,02048619	1,77487621
15	1,16694815	0,78975313	1,54414316

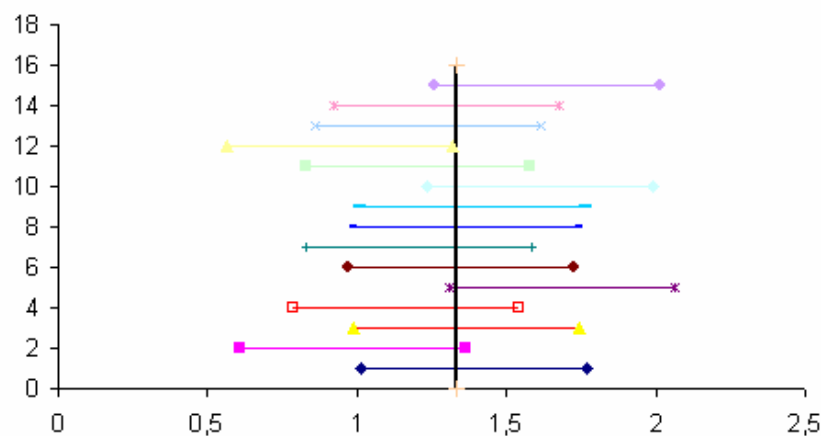


Figure 16 : Intervalles de confiance sur R après la transformation de Fischer

4.9.7. Conclusion

Nous avons réalisé dans cette activité pédagogique une transposition didactique qui permet aux étudiants une appropriation d'un savoir délicat sans perdre de vue les idées qui ont présidé à la transformation de Fisher pour étudier le coefficient de corrélation d'une population à partir de celui d'un échantillon.

4.10. Simulations concernant l'analyse statistique implicite

4.10.1. La place de la simulation

Ainsi la notion de champ conceptuel en tant qu'ensemble des situations renvoyant à l'idée de procédure, permet de situer la simulation en statistique dans sa finalité qui est de fournir à l'apprenant un éclairage sur le signifié (invariants opératoires).

Par ailleurs l'approche développée par Jean-Claude Régnier (1988) intégrant l'apprentissage fondé sur le tâtonnement expérimental de l'apprenant offre une perspective pour analyse didactique de la simulation et relation à la résolution de problème.

4.10.2.La situation proprement dite

La situation proposée est définie par le « commanditaire » et porte dans ce cas précis sur les représentations par les étudiants de première année de DUT de la statistique et de l'apprentissage. D'autres études sont menées par divers groupes pour d'autres commanditaires. Citons par exemple une Etude pour l'Agence de lutte contre l'illettrisme, une étude pour le Chambre de commerce de Lyon concernant le grand prix du design, une étude pour l'Espace Numérique des entreprises, une étude sur l'utilisation de moyens audiovisuels dans l'enseignement supérieur, etc. Les méthodes utilisées dans ces études peuvent être libres ou bien imposées par le commanditaire.

Ressources mises à disposition du groupe d'étudiants : ordinateurs avec logiciels (Excel, CHIC, etc.), en tout 50 heures de travaux tuteurés, la moitié de ces heures étant encadrées par un animateur, un questionnaire (8 pages), une explication de l'étude par le commanditaire.

Consignes : à rendre : rapport sur l'étude (environ une centaine de pages) (noté), soutenance de l'étude devant le commanditaire (noté), à mi parcours chaque groupe explique aux autres groupes l'avancement de ses travaux (non noté), diverses consignes peuvent être données selon le sujet.

Dans le cas qui nous occupe nous avons suggéré de construire une simulation avant de se lancer dans le travail sur le questionnaire et nous allons en exposer quelques lignes ci-dessous.

4.10.3.La simulation outil d'appropriation de concepts statistiques

La simulation : faire apparaître des invariants dans la variabilité : nous avons débattu de l'utilité et de l'utilisation de la simulation dans diverses situations rencontrées dans l'enseignement de la statistique entre autres concernant les intervalles de confiance (Oriol et Régnier 2003a) et le coefficient de corrélation (Oriol et Régnier 2003b).

C'est sans doute un des points spécifiques de l'enseignement de la statistique obligé de développer un enseignement s'appuyant sur les mathématiques mais hétérodoxe par rapport aux outils traditionnels construits dans l'exclusion entre le vrai et le faux. Le propre du raisonnement scientifique est que des mêmes conditions vont produire des effets identiques. En statistique il n'en est rien et c'est cela sans doute une des difficultés que rencontrent les étudiants.

La simulation permet de distinguer les invariants dans la variabilité.

Notons d'ailleurs que dans notre contexte de l'analyse statistique implicative la pensée développée est « doublement » hétérodoxe : d'une part comme toute pensée statistique et d'autre part comme s'intéressant à des énoncés $a \Rightarrow b$ « partiellement » vrais.

4.10.4. La simulation proposée

Le choix d'Excel : notre pratique pédagogique vise à intégrer le plus tôt possible l'outil informatique comme instrument canonique d'une pratique de la statistique. Et en matière d'outil informatique, le tableur est privilégié dans un premier temps. Les cours « spécialisés » à l'aide de logiciels tels que SAS, SPAD complètent la gamme des compétences de nos étudiants.

La construction de la simulation : la séance prend appui sur une bonne maîtrise du logiciel Excel par les étudiants et fait suite à une séance de deux heures sur l'approche de l'analyse statistique implicative. Les étudiants doivent construire une feuille Excel tirant 100 fois au hasard les valeurs binaires de a et b, évaluer l'indépendance des variables a et b, calculer pour ces 100 valeurs l'indice d'implication et l'intensité d'implication entre a et b.

Et recommencer...

Voici les 10 premières lignes d'un tableau (le tableau comporte en réalité 100 lignes) dans lequel les valeurs de a et de b sont 0 ou 1 obtenues aléatoirement avec le générateur de nombres aléatoires d'Excel :

	a	b					Vérif.
							100
			16	35	26	23	100
1	0	1	0	1	0	0	1
2	1	1	0	0	0	1	1
3	1	0	0	0	1	0	1
4	1	0	0	0	1	0	1

5	1	1	0	0	0	1	1
6	0	1	0	1	0	0	1
7	1	0	0	0	1	0	1
8	0	1	0	1	0	0	1
9	0	0	1	0	0	0	1
10	1	0	0	0	1	0	1
---	---	---	---	---	---	---	---

Le tableau récapitulatif correspondant obtenu :

a →	0	1	
↓b			
0	16	26	42
1	35	23	58
	51	49	100

On en déduit le tableau de contingence et le calcul de la valeur du Khi deux

Tableau d'indépendance	21,42	20,58	
	29,58	28,42	
Calcul du Khi deux	1,371	1,427	
	0,993	1,034	
Valeur du Khi2			4,826

Et également le calcul de l'indice d'implication et de l'intensité d'implication

Calcul de l'indice d'implication
 $q(a, non(b)) = 1,195$

Calcul de l'intensité d'implication
 $\Phi(a, non(b)) = 0,116$

Figure 17 : Comparaison des valeurs de q et des valeurs de Phi correspondantes

Voici une série de résultats obtenus concernant des indices d'implication et les valeurs d'intensités correspondantes

Valeurs de k	Valeurs de q	Valeurs de Phi
1	-1,74207716	0,95925257
2	-0,48564293	0,68638982
3	-0,79259392	0,78599284
4	1,46026115	0,0721092
5	-0,06428571	0,52562871
6	-0,00769231	0,50306877
7	-0,48107024	0,68476671
8	0,20225996	0,41985677
9	-0,38392627	0,6494834
10	-0,32433749	0,62715866

Et une représentation graphique :

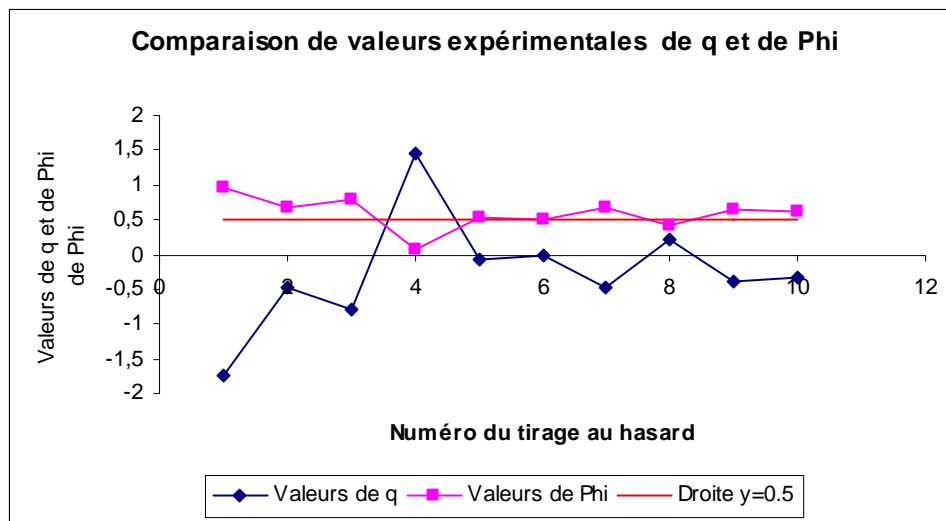


Figure 18 : Indices d'implication et intensités d'implication

Les remarques évidentes à la lecture des dizaines de graphiques obtenus, par exemple si l'indice q est inférieur à 0 alors l'intensité est supérieure à 0,5 vont permettre de passer aux démonstrations de ces propriétés puis à poursuivre et à aborder les concepts suivants du champ de l'analyse statistique implicite.

4.10.5. Recherche de bénéfices collatéraux : les valeurs trouvées du Khi deux, le test de Mc Némar

Comme nous pouvons « faire tourner » cette simulation (en appuyant sur la touche F9) nous pouvons également stocker les résultats afin de les comparer à des résultats prévisibles par la théorie statistique.

Puisque on génère des valeurs du Khi² on peut les comparer aux seuils théoriques à 5% et à 1% par exemple. Ainsi dans le tableau ci-dessous sur les 640 valeurs du Khi² obtenues nous en avons 6 au dessus du seuil de 1% et 32 au dessus du seuil de 5% ce qui correspond aux valeurs attendues.

Valeur du Khi2		2,793491	
Seuils théoriques			
A 5%	A 1 %		
3,841	6,635		
Valeurs observées			
634	6	640	0,9%
608	32	640	5,0%

Figure 19 : Comparaison des valeurs simulées et théoriques

D'une façon similaire nous avons construit le test de Mc Nemar que les étudiants venaient d'étudier dans le cours sur les méthodes consacrées aux tests non paramétriques.

4.10.6. Recherche de bénéfices directs concernant l'ASI : q est symétrique, il existe une relation entre le Khi^2 et $q(a, \text{non}(b))$

Nous avons recherché à vérifier les relations entre $q(a, \text{non}(b))$ et $q(\text{non}(b), a)$. Cela permet aux étudiants d'être plus à l'aise avec les formules

Calcul de l'indice d'implication $q(a, \text{non}(b)) = 0,909$

Calcul de l'indice d'implication $q(\text{non}(b), a) = 0,909$

Et également à vérifier la relation entre le Khi^2 et $q(a, \text{non}(b))$

$\text{Phi}(a, \text{non}(b)) =$	0,18
----------------------------------	------

$\text{Khi}^2/q(a, \text{non}(b)) =$	3,92
$n^2/nb * n\text{non}(a) =$	3,92

4.10.7. Comparaison du calcul de Phi avec la loi de Poisson et avec la loi normale

Il nous a semblé intéressant de comparer le calcul de Phi d'une part à l'aide de la loi de Poisson et d'autre part avec la loi normale

Calcul de l'intensité d'implication (loi normale) $\text{Phi}(a, \text{non}(b)) = 0,674443$	avec la loi de Poisson $\text{Lambda} = 26,3$ 0,62776283
--	---

Figure 20 : Comparaison des valeurs de l'intensité d'implication (loi de Gauss et loi de Poisson)

D'une expérience à l'autre l'écart varie peu et il est de l'ordre de 0.05.

4.10.8.Observation de 100 valeurs de q, de 500 valeurs de Phi

Nous avons ensuite représenté et observé 100 valeurs au hasard de q

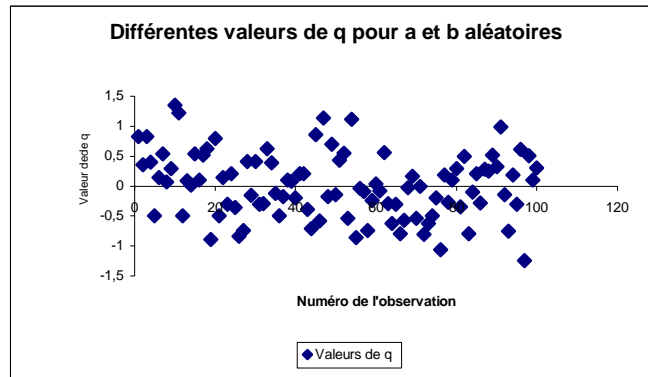


Figure 21 : 100 valeurs « au hasard » de l'indice d'implication

Puis 500 valeurs de l'intensité d'implication :

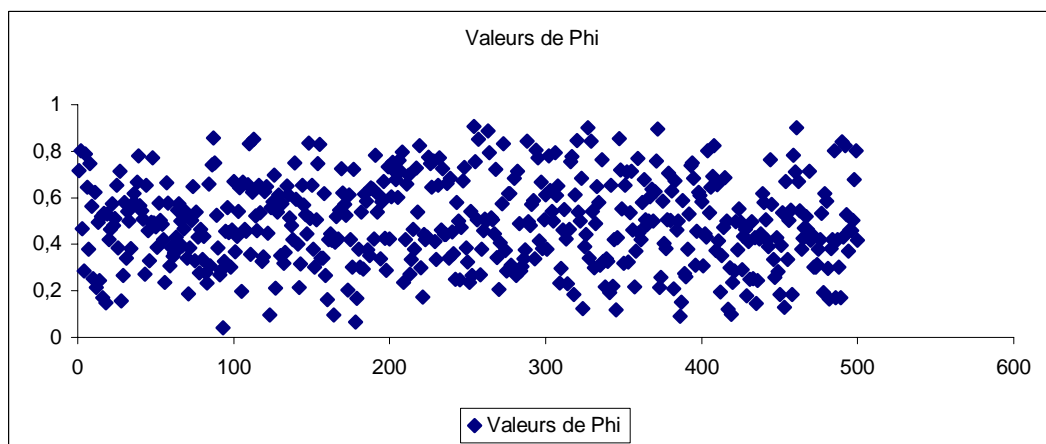
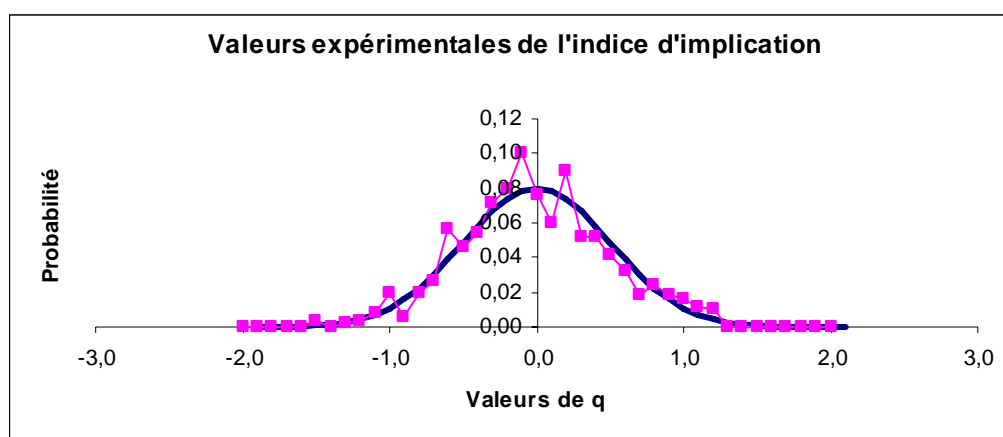


Figure 22 : 500 valeurs de l'intensité d'implication

Regroupement des observations de 100 valeurs de q, de 500 valeurs de Phi

Dans notre dispositif nous avons regroupé les valeurs précédentes et généré les courbes suivantes :



Nous avons fait une opération similaire sur l'intensité d'implication

4.10.9. Conclusion sur les simulations concernant l'ASI

Il nous apparaît que l'organisation du couple situation schème permet une meilleure appropriation des concepts statistiques par les étudiants.

Ici l'exigence d'utiliser une théorie nouvelle et un outil inconnu oblige les étudiants à un aller retour entre le réel et ses représentations, entre signifié et signifiant, à construire des invariants opératoires, à construire le sens des situations, bref à conceptualiser l'analyse statistique implicite.

Dans le cas présenté la construction par les apprenants de l'outil de simulation leur permet de dégager des invariants de la variabilité de phénomènes non déterministes. Nous exposerons dans la partie V les résultats de cette activité de simulation qui trouve une contextualisation dans la pratique d'enquêtes réalisées par les étudiants dans la suite de ces travaux tuteurés.

4.11. Conclusion de la partie IV

Les simulations «prêtes à être admirées» emmènent le plus souvent les étudiants à une «bovarysation» des notions étudiées.

Ici, en revanche, la construction par les apprenants, eux-mêmes, de l'outil s'apparente aux méthodes d'apprentissage où les premiers travaux des apprentis consistaient à fabriquer leurs propres outils (l'équerre pour un ouvrier par exemple, ou encore une brouette pour un menuisier) et à la mise en scène du tâtonnement expérimental producteur d'apprentissages.

La théorie des champs conceptuels et le modèle d'apprentissage fondé sur le tâtonnement expérimental permettent de situer ces activités de simulation dans l'apprentissage de la statistique, lieu où elles ont une place, mais bien entendu pas toute la place. Elles permettent comme nous l'avons dit de dégager les invariants de la variabilité des situations statistiques.

5. Partie V : Etude des impacts de la pratique d'enquête sur les représentations de la statistique chez les étudiants de DUT

Présentation

Notre allons présenter deux enquêtes effectuées par les étudiants de STID deuxième année ayant pour objet les représentations de la statistique chez les étudiants de DUT, études utilisant l'analyse statistique implicite. La première enquête vise à comparer les représentations de la statistique chez les étudiants de première année présents à l'IUT Lumière en 2005-2006 dans le DUT STID et le DUT QLIO. La deuxième enquête qui voulait comparer les représentations de la statistique entre deux populations d'étudiants STID deuxième année à Grenoble (sans alternance) et à Lyon (en alternance) a dévié de son projet initial en raison du faible taux de retour des étudiants de STID-Grenoble et du biais que cela aurait entraîné ; elle s'est alors concentrée sur les représentations de la statistique des étudiants STID-Lyon et l'implication des variables entre elles

5.1. Quelques brefs rappels théoriques concernant l'analyse statistique implicite.

Quand a implique b mais pas toujours

L'analyse implicite se propose de donner un sens statistique à des expressions du type $a \Rightarrow b$ dans le cas où l'implication est non stricte, c'est-à-dire que lorsque la variable a est vraie, b a peu de chances d'être fausse.

Selon les valeurs logiques prises par a et b, la proposition $(a \Rightarrow b)$ est vraie ou fausse.

- si (a est faux et b est faux) alors $a \Rightarrow b$ est vrai.
- si (a est faux et b est vrai) alors $a \Rightarrow b$ est vrai.
- si (a est vrai et b est vrai) alors $a \Rightarrow b$ est vrai.
- si (a est vrai et b est faux) alors $a \Rightarrow b$ est faux.

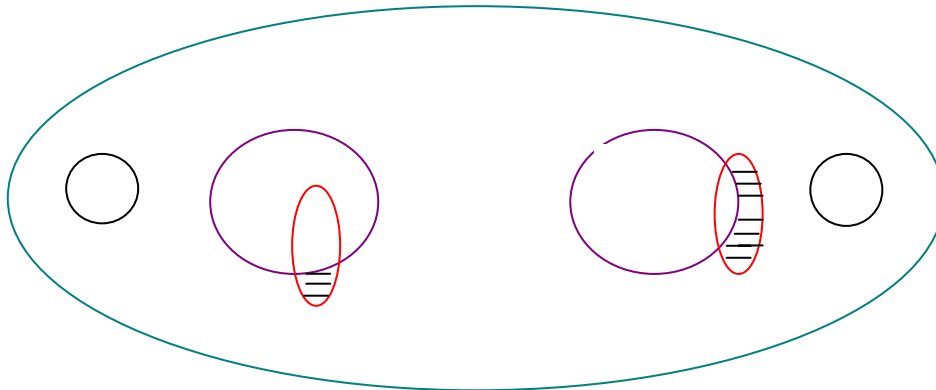
Nous pouvons résumer ces cas dans les tableaux suivants :

a	b	$a \Rightarrow b$
Faux	Faux	Vrai
Faux	Vrai	Vrai
Vrai	Faux	Faux
Vrai	Vrai	Vrai

		a	
		1	0
b	1	1	1
	0	0	1

Notons que 1 correspond à la modalité « vrai » et 0 à la modalité « faux ».

Partons de l'exemple donné par Régis Gras (GRAS, 1996, p.29) que l'on ait une population E , avec $\text{Card}(E) = 100$ (sachant que Card correspond au nombre d'évènements réalisés dans E). Soit deux variables A et B (binaires), avec $\text{Card}(A) = 36$ et $\text{Card}(B) = 50$.



Dans le cas 1, $\text{Card}(A \cap \bar{B}) = 3$. Comparée à la valeur des cardinaux de A et B , cette valeur est très faible : A est presque contenu dans B . On a donc une implication statistique entre les deux variables ($a \Rightarrow b$). Dans le cas 2, $\text{Card}(A \cap B)$ est très faible ce qui nous permet de déduire une absence de lien entre a et b . Ci-dessous un exemple (GRAS, 1996, p.27) d'implication entre deux variables binaires :

Sujets	V	
	a	b
1	0	0
2	0	1
3	1	1
4	1	0
5	0	0
6	1	1
7	1	1
8	0	0
9	0	1
10	0	0
Total	4	5

A partir de ce tableau, on calcule la somme des couples définis précédemment ce qui donne le tableau ci-dessous.

	a	1	0	Marge
b		3	2	5
	1	3	2	5
	0	1	4	5
Marge		4	6	10

Ce tableau est utilisé pour représenter la situation d'implication statistique. En effet, le nombre 3 du tableau de contingence signifie qu'il y a 3 sujets qui ont la réponse vrai (=1) pour la variable a et vrai (=1) pour la variable b (sujets 3, 6 et 7 du tableau précédent).

Ci-dessous un tableau qui permet de comprendre comment le précédent est construit.

Tableau des notations des effectifs conjoints et marginaux :

	a	1	0	Total
b				
1		$n_{a \cap b}$	$n_{\bar{a} \cap b}$	n_b
0		$n_{a \cap \bar{b}}$	$n_{\bar{a} \cap \bar{b}}$	$n_{\bar{b}}$
Total		n_a	$n_{\bar{a}}$	n

De ce tableau, nous pouvons tirer la formule d'indépendance ci-dessous :

$$n_{a \cap \bar{b}} = \frac{n_a * n_{\bar{b}}}{n}$$

Dans notre cas, si a et b étaient indépendants, le nombre de réponses lorsque a est vrai et b faux) serait égal à $5*4/10$ ce qui n'est pas le cas.

La quantité $n_{a \cap \bar{b}}$ est la réalisation de la variable aléatoire $Card(X \cap \bar{Y})$ qui suit sous certaines conditions (Gras, 1996, p.30) la loi de Poisson de paramètre $n * p(a) * p(\bar{b})$.

A l'origine, pour ce genre de modélisation, la loi pressentie était la loi binomiale. La méthode de maximum de vraisemblance consiste à proposer la valeur pour laquelle la probabilité de l'observation dans le modèle est la plus forte. En appliquant cette méthode, on a pu approximer cette loi par une loi de Poisson.

De l'indice d'implication à l'intensité d'implication

On peut alors construire (Gras, 1996, p.32) un indice d'implication et une intensité d'implication de la façon suivante :

$$Q(a, \bar{b}) = \frac{\text{Card}(X \cap \bar{Y}) - np(a)p(\bar{b})}{\sqrt{np(a)p(\bar{b})}} = \frac{\text{Card}(X \cap \bar{Y}) - \frac{na\bar{b}}{n}}{\sqrt{\frac{na\bar{b}}{n}}}$$

La variable aléatoire $Q(a, \bar{b})$ suit une loi normale centrée réduite. Elle permet de définir l'intensité d'implication $\varphi(a, \bar{b})$ qui correspond à la qualité de l'admissibilité de $a \Rightarrow b$.

$$\varphi(a, \bar{b}) = 1 - P[Q(a, \bar{b}) \leq q(a, \bar{b})] = \frac{1}{\sqrt{2\pi}} \int_{q(a, \bar{b})}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$$

NB : Q est la variable aléatoire de l'indice d'implication. q est sa réalisation.

A ne pas confondre avec ϕ qui correspond à l'intensité d'implication.

Le graphe implicatif, l'arbre hiérarchique ou cohésitif

De ce concept statistique dérivent deux représentations graphiques :

- le graphe implicatif
- l'arbre hiérarchique dit cohésitif

Tout d'abord, le graphe implicatif permet de voir les liens d'implication entre plusieurs variables. En effet, le principe du graphe est de calculer les indices d'implication. Puis on représente les variables dont les liens sont les plus forts.

Ci-dessous un exemple de graphe implicatif à partir de données prises au hasard. Les flèches rouges (ex : entre V109 et V117) représentent un lien à 99%, les bleues (ex : entre V75 et V83) un lien à 97% et les vertes (ex : entre V48 et V01) un lien à 95%.

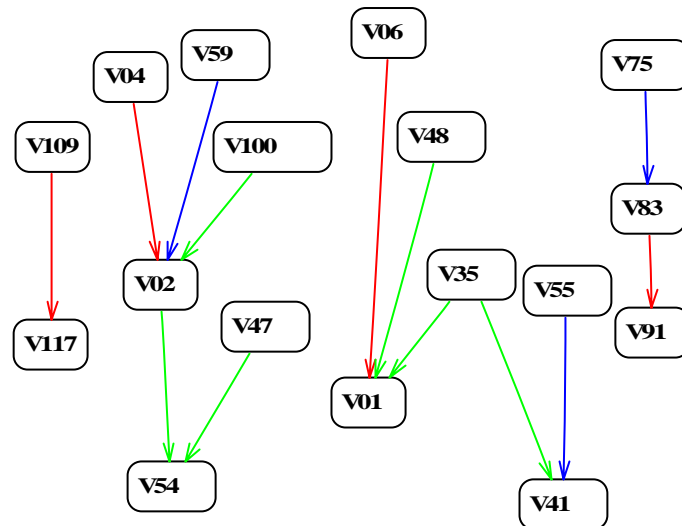


Figure 23 : Exemple de graphe implicatif

Par exemple, si V04 est vraie alors V02 est vraie dans 99% des cas. Lorsque V59 est réalisée, cela implique que V02 est également réalisée à 97%. Quand V100 est égale à 1, on a 95% de chances que V02 soit égale à 1.

Lorsque V02 est vraie, il y a une probabilité de 95% que V54 soit vraie. Par conséquent, si nous reprenons le raisonnement précédent, nous pouvons en déduire que lorsque au moins une des trois variables (V04, V59 et V100) est vraie, V02 se réalise, ce qui entraîne la réalisation de V54.

La deuxième représentation illustrant le concept d'implication statistique est l'arbre hiérarchique. Cette figure représente les variables avec leur significativité présentée sous forme de niveaux, à partir de données tirées au hasard.

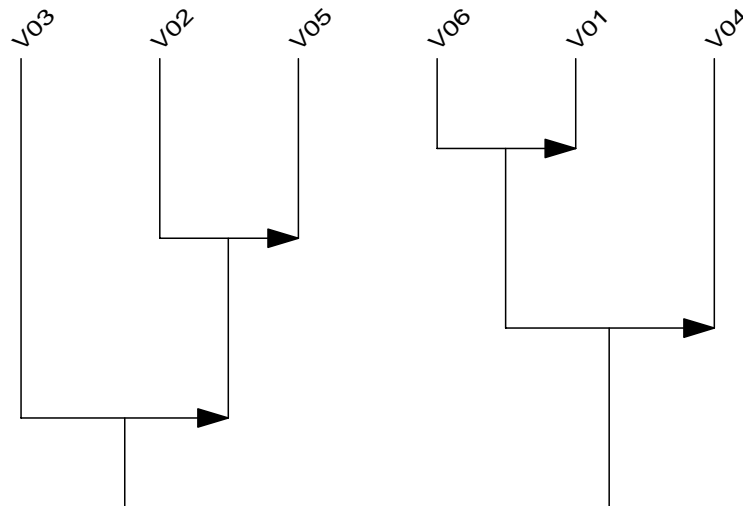


Figure 24 : Exemple d'un arbre cohésitif

Par ce graphique, nous pouvons en conclure qu'il y a une implication entre la variable V02 et la V05 (de même pour les variables V06 et V01). De plus, la variable V03 implique le couple de variables constitué de V02 et V05. L'implication est contraire pour ce qui concerne le deuxième cas. En effet, c'est le couple de variables V06/V01 qui contribue à la réalisation de la variable V04.

L'analyse des similarités

Il existe un deuxième concept nécessaire pour comprendre le fonctionnement de la classification ascendante hiérarchique : l'analyse des similarités. Selon I.C. LERMAN, cette analyse consiste à étudier des classes de variables semblables, grâce à la création d'une typologie. Cette typologie est créée de la manière suivante : à partir de l'ensemble des variables, le logiciel regroupe celles-ci en classes de taille de plus en plus grande, dont l'effectif est de moins en moins important.

Plus le lien entre deux variables est fort, plus celles-ci sont semblables. En effet, la similarité se définit de la sorte : soit deux variables quelconques A et B, plus la valeur de $A \cap B$ est élevée, plus ces deux variables sont similaires. L'indice de similarité est calculé par la probabilité que la valeur de $A \cap B$ soit supérieure à un nombre au hasard.

Nous allons vous exposer brièvement l'arbre des similarités. Ce graphique se construit de la manière suivante : on réunit en une classe, au premier niveau, les deux variables qui se ressemblent le plus au sens de l'indice de similarité ; puis on fait de même pour deux autres variables ou une variable et une classe déjà formée et ainsi de suite.

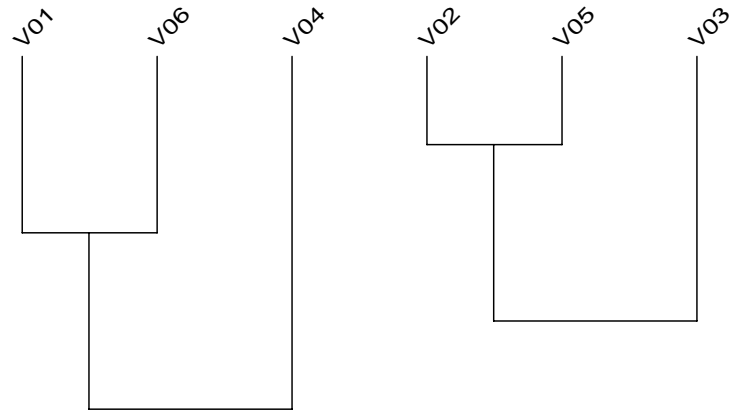


Figure 25 : Exemple d'un arbre des similarités

Plus le niveau de significativité est haut (exemple : V02/V05), plus on a une forte corrélation entre les variables.

« La distance implicative d'un individu x à la classe C est le nombre :

$$d(x, C) = \left[\frac{1}{n} \sum_{i=1}^{i=n} \frac{(\varphi_i - \varphi_{x,i})^2}{1 - \varphi_i} \right]^{\frac{1}{2}} \gg^{33}$$

Rappelons que φ correspond à l'intensité d'implication définie précédemment.

Ce calcul permet de comprendre comment le niveau de significativité est déterminé. Plus cette distance est grande, moins les individus sont liés.

Il est possible d'effectuer le même type de traitement sous le logiciel StatBox que sous CHIC. Voici ce que l'on obtient :

³³ L'implication statistique – Nouvelle méthode exploratoire de données, Régis GRAS, La pensée sauvage Editions, 1996, p.299

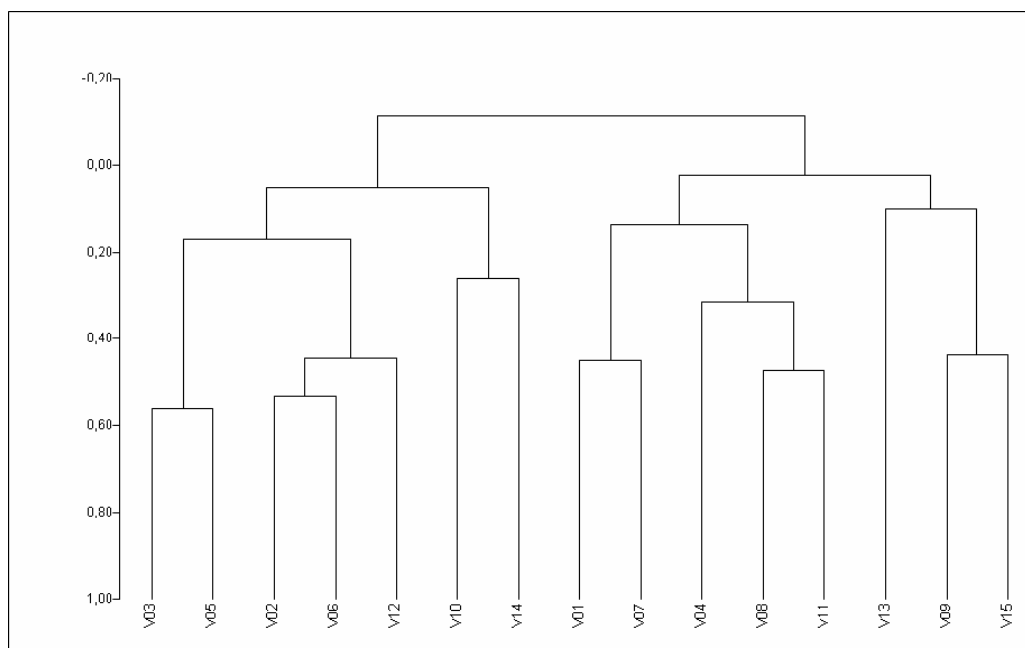


Figure 26 : Exemple d'un dendrogramme obtenu à l'aide du logiciel StatBox

Contrairement au logiciel CHIC, StatBox propose des dendrogrammes où la racine se situe en haut. En effet, l'axe des ordonnées est inversé : les valeurs négatives sont situées en haut du graphique et les valeurs positives en bas. Ainsi, les nœuds représentant les similarités les plus fortes sont positionnés dans le bas de la représentation. Lorsqu'on tronque le dendrogramme à un certain niveau de ressemblance, cela crée une partition regroupant plusieurs classes de modalités dont la similarité est proche.

Citons pour mémoire d'autres exemples d'approches de l'implication statistique cités par R. GRAS (Gras, 1996, p.34-36) :

J. LOEVINGER (1947) : indice définissant la quasi-implication de a sur b

$$H(a,b) = 1 - \frac{nn_{a \cap \bar{b}}}{na n \bar{b}} \quad H(a,b) \text{ appartient à l'intervalle }]-\infty; 1].$$

« Si $H(a,b)$ est « proche » de 1, l'implication est « presque » satisfaite. Mais cet indice présente l'inconvénient, ne se référant pas à une échelle de probabilité, de ne pas fournir de seuil de vraisemblance et d'être invariant dans toute dilatation de E , A , B et $A \cap \bar{B}$. »

La citation précédente signifie que lorsque $H(a,b)$ est proche de 1, l'implication est forte. Cependant, cet indice a l'inconvénient de porter sur un intervalle large ($]-\infty; 1]$), ce qui rend difficile le calcul d'un seuil de vraisemblance qui se situe entre 0 et 1 (échelle de probabilité).

J. PEARL (1988), S. ACID (1991), A. GAMMERMAN (1991), Z. LUO (1991)

$$\left| P(A \cap \bar{B}) - P(A) * P(\bar{B}) \right|$$

Pour ces derniers, cette expression correspond à l'implication entre a et b.

H. RALAMBRODRAINNY (1991) avec le système GENRED

Ce système part du principe que la règle d'inférence est pertinente lorsque, pour deux seuils a et b donnés par l'utilisateur, les deux équations suivantes se vérifient :

$$\text{Card}(A \cap \bar{B}) \leq \alpha$$

$$\text{Card}(A \cap B) \geq \beta$$

❖ M. SEBAG (1991) et M. SCHOENAUER (1991)

Leur système, comparable au précédent, retient la règle d'inférence au seuil α , si :

$$\frac{\text{Card}(A \cap B)}{\text{Card}(A \cap \bar{B})} \geq \alpha$$

ou cette relation équivalente :

$$\frac{\text{Card}(A \cap B) - \text{Card}(A \cap \bar{B})}{\text{Card}(A \cap B)} \geq \alpha$$

Y. SCHEKTMAN (1992), J. TREJOS (1992), N. TROUPE (1992), E. DIDAY (1991), M. O. MENESSIER (1991)

Pour juger de l'implication statistique de a sur b, ces chercheurs calculent la probabilité conditionnelle suivante : P [B/A].

❖ J. G. GANASCIA (1991)

$$2 * P(B / A) - 1$$

Cet indice permet de calculer l'incertitude de l'implication $a \Rightarrow b$.

5.2. Comparaison des représentations de la statistique chez deux populations d'étudiants

Cette étude a été menée, en 2005-2006, sous ma responsabilité par un groupe d'étudiants de deuxième année du DUT Stid Lyon de l'IUT Lumière. Il s'agit de Delphine BERQUIER, Renaud BONNOT, Marie-Hélène CHANG, Alexina PYE, Amandine RYAN et

Nathalie TOURNOUX. Sans ces étudiants nous aurions eu du mal à venir à bout de cette enquête et je tiens à les remercier pour le sérieux qu'ils ont apporté à ce travail.

Population de l'étude : les étudiants du DUT STID (30) et du DUT QLIO (30) de première année (2005-2006) de l'IUT Lumière

5.2.1. Introduction

Le travail à faire se décompose comme suit :

- administrer le questionnaire à une nouvelle population
- saisir les questionnaires retournés
- procéder au nettoyage de la base et au recodage nécessaire avant l'importation des données
- traiter les données à l'aide du logiciel CHIC
- analyser les résultats grâce à la méthode de classification hiérarchique des données

Dans le but d'appliquer en détail la théorie de l'implication statistique à un exemple concret, et d'en tester la pertinence et la sensibilité, nous avons, dans le cadre des projets longs de statistique de deuxième année de STID, été amenés à étudier les représentations de la statistique que se font les étudiants de première année d'IUT, à l'aide du logiciel CHIC. C'est ce que nous vous présentons dans la deuxième partie de ce travail.

5.2.2. Présentation de l'étude

5.2.2.1. Présentation du logiciel CHIC

Le sigle CHIC signifie : Classification Hiérarchique Implicative et Cohésitive. C'est un logiciel d'analyse de données, initialement conçu par Régis Gras.

Ce logiciel traite des données numériques sous forme binaires ou non pour des valeurs comprises entre 0 et 1 se présentant en un tableau à 2 entrées, (par exemple les sujets en ligne et les variables en colonne). Ces données doivent être au préalable saisies sous Excel. Il permet de fournir des statistiques brutes (ex : moyennes, écart-types, coefficients de corrélation).

De plus, il est possible d'établir une classification hiérarchique des similarités (valeurs des similarités, arbre hiérarchique avec son processus et ses niveaux significatifs) et une analyse implicative, tout en donnant la possibilité de conjointre, disjoindre, supprimer, ajouter des variables.

Ce logiciel fournit les valeurs d'intensité d'implication, ainsi que le graphe implicatif selon différents seuils d'intensité, les sujets et les catégories de sujets contribuant aux chemins du graphe, les chemins significatifs, etc ...

Il permet d'obtenir une classification cohésitive en un arbre, avec ses niveaux significatifs, la contribution de sujets et catégories de sujets à ces niveaux ainsi qu'un graphe inclusif.

5.2.2.2. Méthodologie d'enquête

Le questionnaire (en annexe) interroge les étudiants sur :

- leur parcours de formation à l'entrée à l'IUT,
- l'évocation de la statistique en général,
- l'approche pédagogique adoptée dans le dispositif de formation à l'IUT

Nous avons été chargés d'administrer une soixantaine de questionnaires, répartis de la façon suivante :

- 30 questionnaires aux étudiants STID de 1^{ère} année
- 30 questionnaires aux étudiants QLIO de 1^{ère} année

La deuxième étape a été la saisie des réponses. Cette étape est essentielle pour la suite du travail (cf. partie « Présentation du logiciel CHIC »).

Nous nous sommes heurtés à plusieurs difficultés : certaines réponses étaient manquantes, d'autres n'avaient aucun sens ; mais le plus difficile a été de résumer les réponses par un mot afin de les regrouper en différents thèmes.

Ensuite, il a fallu recoder le nom des différents thèmes pour une meilleure lisibilité dans les sorties des graphiques.

Le fichier devait être épuré au maximum afin de le rendre plus compréhensible.

Les résultats de l'étude

A partir du fichier obtenu, nous avons effectué différents regroupements de questions ce qui nous a permis de conclure sur les caractéristiques qui ressortent sur la représentation de la statistique pour les deux profils d'étudiants interrogés.

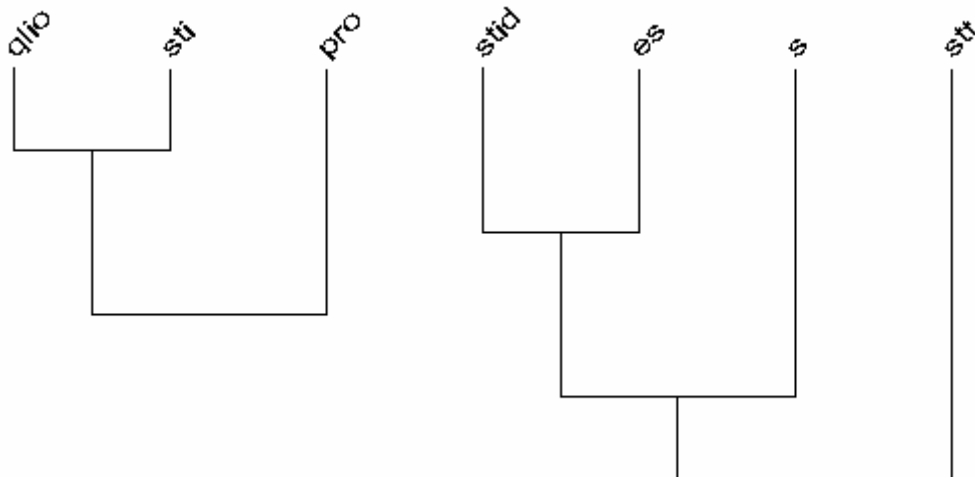
Pour ce faire, nous avons construit, pour chacun des groupes de questions, un arbre de similarités et un arbre cohésitif. Nous avons également élaboré un graphe implicatif qui reprend l'ensemble des questions.

Pour certains graphiques, la signification des modalités ne sera pas donnée à cause de la taille de ceux-ci. Pour avoir la signification des modalités, il faudra vous référer au dictionnaire des données situé en annexe : cette notification sera signalée par la phrase suivante : « **En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.** »

5.2.2.3. D'où proviennent les étudiants STID et QLIO ?

V2. Département de l'IUT ?
 V3. Baccalauréat obtenu ?

- *Arbre de similarité*



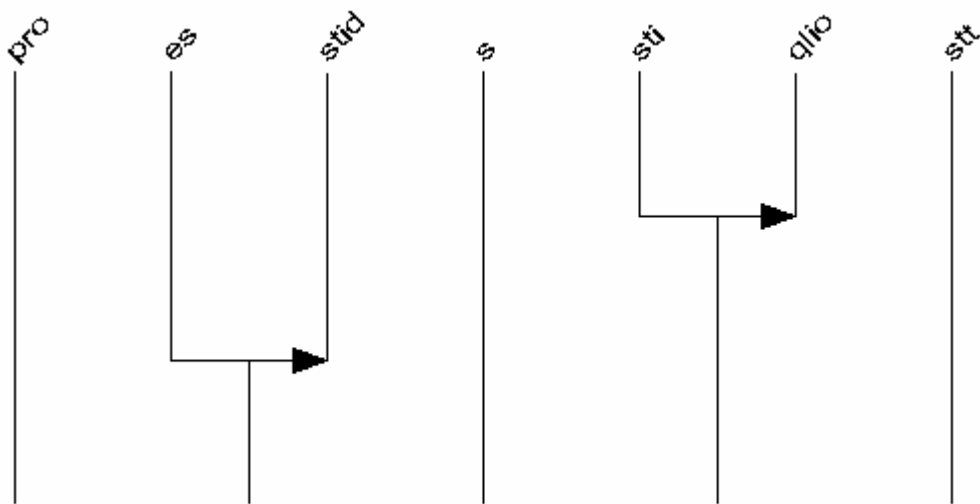
stid	STID
qlio	QLIO
pro	Bac professionnel
s	S
es	ES
stt	STT
sti	STI

tel-00191166, version 1 - 27 Nov 2007

Ici, nous pouvons d'abord voir que la modalité « QLIO » est proche des modalités « stid » et « pro ». On peut en déduire que les QLIO proviennent en général des baccalauréats STI et PRO.

Quand au STID, on constate que ceux-ci ont un lien avec les baccalauréats «ES », « S », et « STT ». Les STID contiennent donc en général ces trois types de baccalauréats.

- *Arbre cohésitif*



stid	STID
qlio	QLIO
pro	bac professionnel
s	S
es	ES
stt	STT
sti	STI

Niveau 1 « stid » implique « qlio » : ceux qui viennent de la filière STI sont les étudiants QLIO

Niveau 2 « es » implique « stid » : ceux qui viennent de la filière ES sont les étudiants STID

Dans ce graphique, on remarque une implication entre la modalité « ES » et la modalité « STID ». Autrement dit, si l'on provient d'un bac ES, alors on est en STID.

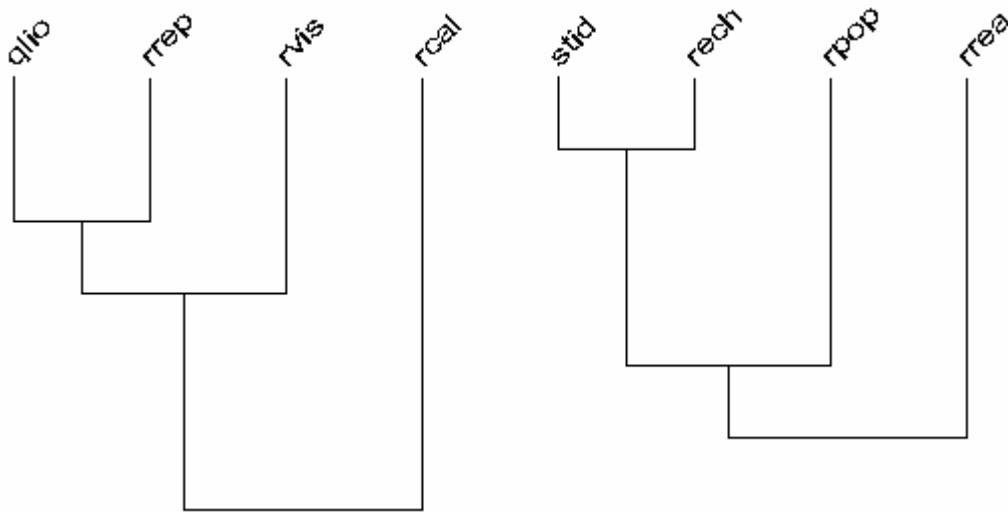
Ensuite, on constate une implication encore plus forte entre la modalité « stid » et la modalité « QLIO ». Donc, si l'on provient d'un bac STI, alors on est forcément en QLIO.

5.2.2.4. La signification de certains mots statistiques selon le bac et le DUT

V9.A ce jour que signifient pour vous les termes suivants ?

Nous avons demandé aux interrogés quelles significations ils donnent à certains mots statistiques. Nous avons les mots : représentatif, significatif, estimation, test d'hypothèse, probabilité et sondage.

- **Analyse du mot « représentatif »**
 - *Arbre de similarité*



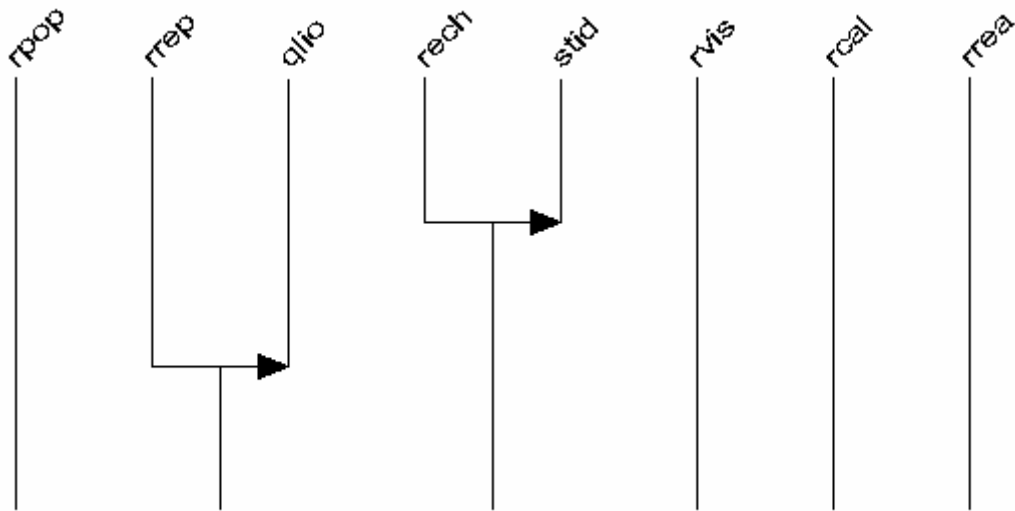
rpop	Population
rrep	Représentatif
rech	Echantillon
rvis	Visualiser
rvid	Vide
rcal	Calcul
rrea	Réalité

Ici, on voit que si on est « QLIO », alors on pense que le mot « représentatif » signifie « représentatif » (modalité « rrep ») de façon significative. On pense aussi que ce mot signifie « visualiser » (modalité « rvis ») et « calcul » (modalité « rcal »).

En ce qui concerne les « STID », ils pensent à un niveau très significatif que le mot « représentatif » signifie « échantillon » (modalité « rech »). Les mots qui ressortent pour

définir ce mot selon eux sont aussi « population » (modalité « rpop ») et « réalité » (modalité « rrea »).

o *Arbre cohésitif*



rpop	Population
rrep	Représentatif
rech	Echantillon
rvis	Visualiser
rvid	Vide
rcal	Calcul
rrea	Réalité

Niveau 1 « rech » implique « stid » : ceux qui pensent que représentatif signifie échantillon sont en STID

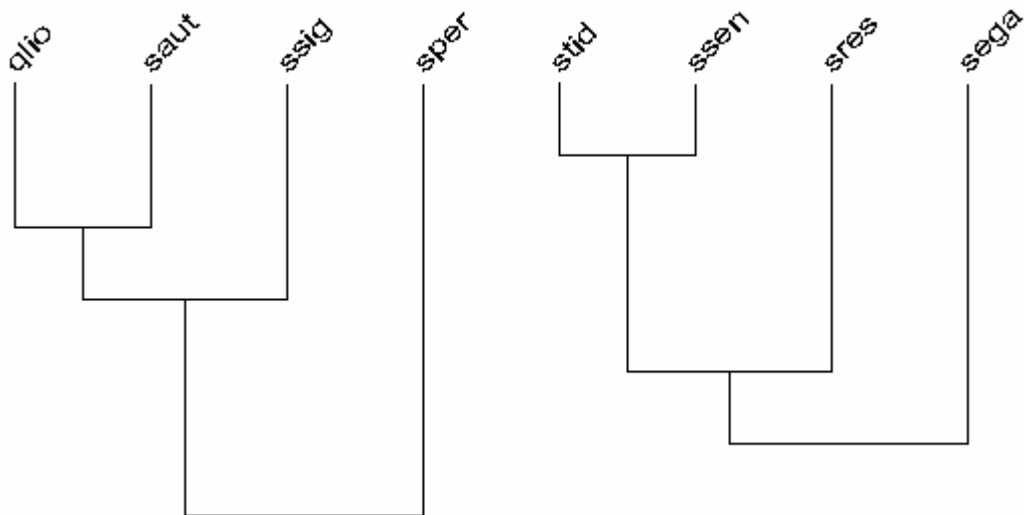
Niveau 2 « rrep » implique « qlio » : ceux qui pensent que représentatif signifie représentatif sont en QLIO

Sur ce graphique, on remarque que donner « représentatif » (modalité « rrep ») comme signification du mot « représentatif » implique d’être « QLIO ».

Aussi, si l’on défini ce mot comme « échantillon » (modalité « rech »), alors cela implique fortement d’être en « STID ».

- *Analyse du mot « significatif »*
 - o *Arbre de similarité*

tel-00191166, version 1 - 27 Nov 2007

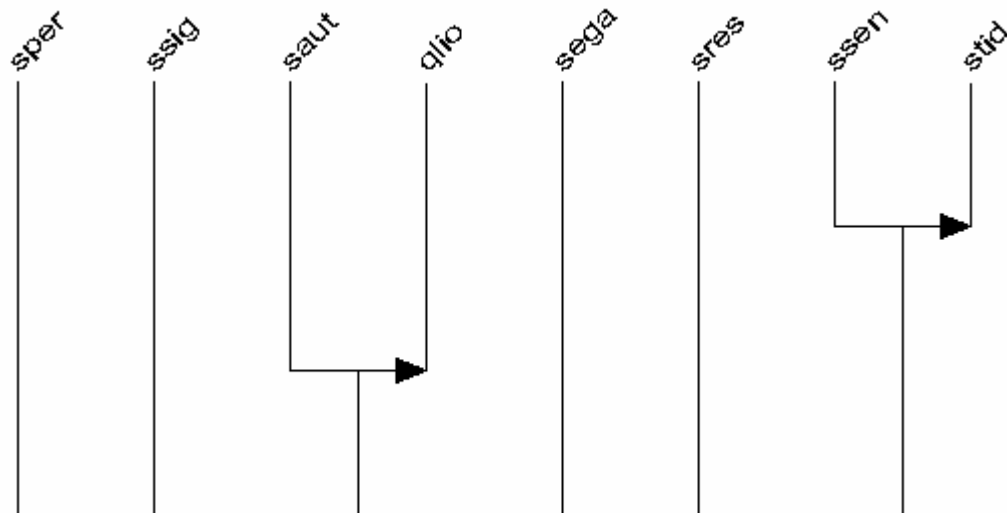


sper	Pertinence
ssig	Significatif
svid	Vide
saut	Autre
sega	Egal
sres	Résultat
ssen	Sens

Nous voyons sur le graphique que la modalité « QLIO » est très proche de la modalité « saut » et de façon moins significative des modalités « ssig » et « sper ». Autrement dit, si l'on est en « QLIO », alors on donne plutôt comme définition de « significatif » les mots « significatif », « pertinence », mais surtout d'autres mots.

Quand aux « STID », on voit que le mot « significatif » signifie pour eux « sens », « résultat », « égal ».

- *Arbre cohésitif*



sper	Pertinence
ssig	Significatif
svid	Vide
saut	Autre
sega	Egal
sres	Résultat
ssen	Sens

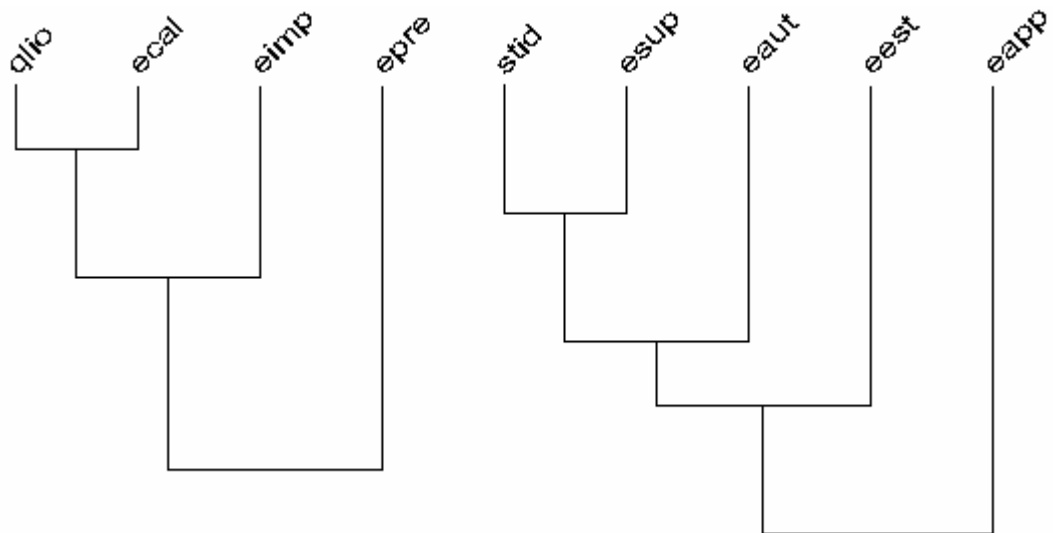
Niveau 1 « ssen » implique « stid » : ceux qui pensent que significatif veut dire sens sont en STID

Niveau 2 « saut » implique « qlio » : ceux qui pensent que significatif signifie autre sont en QLIO

On remarque sur ce graphique que si on pense que « significatif » n’a pas de signification particulière (modalité « saut »), alors on a des chances d’être en « QLIO ». Par contre, si l’on pense que ce mot veut dire « sens », alors on a de fortes chances d’être en « STID ».

Analyse du mot « estimation »

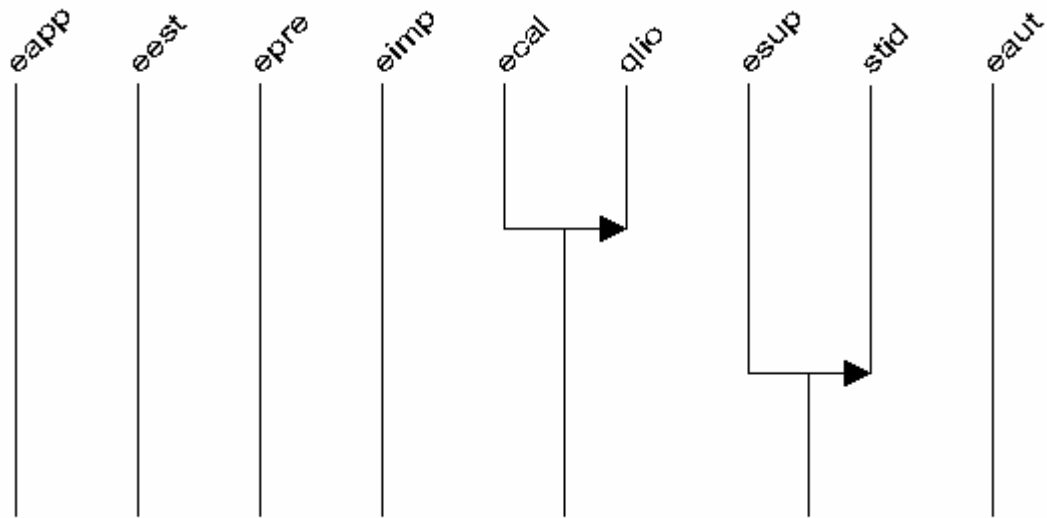
Arbre de similarité



eapp	Approximation
eest	Estimation
epre	Prévision
evide	Vide
eimp	Imprécis
ecal	Calcul
esup	Supposition
eaut	Autre

Ici, on constate d'abord des similarités entre la modalité « QLIO » et la modalité « ecal », mais aussi dans une moindre mesure les modalités « eimp » et « epre ». Si on traduit, Les QLIO pensent que ce mot signifie « calcul », « imprécis », « précision ».

Les « STID » donne comme signification les mots « supposition » (modalité « esup »), « estimation » (modalité « eest »), « approximation » (modalité « eapp ») et d'autres mots (modalité « eaut »).

Arbre cohésitif

eapp	Approximation
eest	Estimation
epre	Prévision
evide	Vide
eimp	Imprécis
ecal	Calcul
esup	Supposition
eaut	Autre

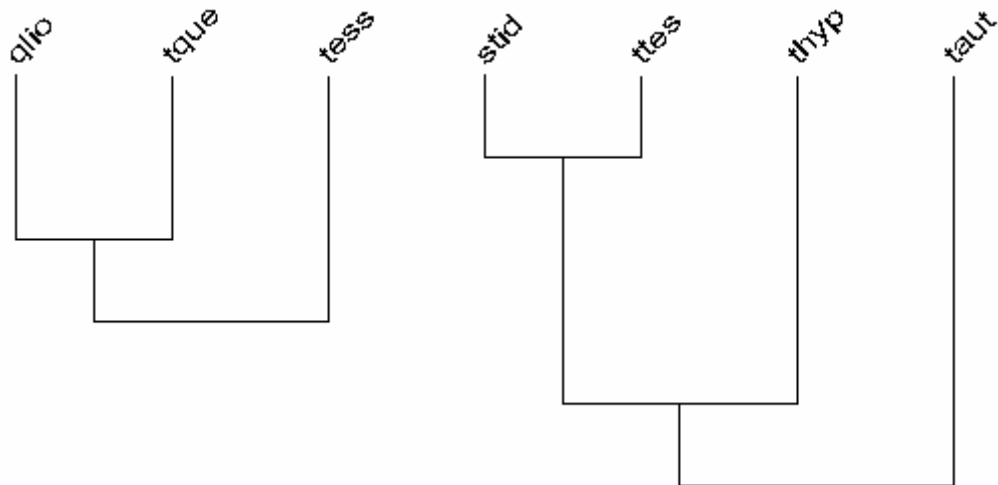
Niveau 1 « ecal » implique « qlio » : ceux qui pensent que estimation veut dire calcul sont en QLIO

Niveau 2 « esup » implique « stid » : ceux qui pensent que estimation signifie supposition sont en QLIO

Ici, si l'on donne comme définition le mot « calcul » (modalité « ecal »), alors cela implique d'être en « QLIO ». Si l'on donne comme définition le mot « supposition » (modalité « esup »), alors cela implique d'être « STID ».

- *Analyse du mot « test d'hypothèse »*

- *Arbre de similarité*

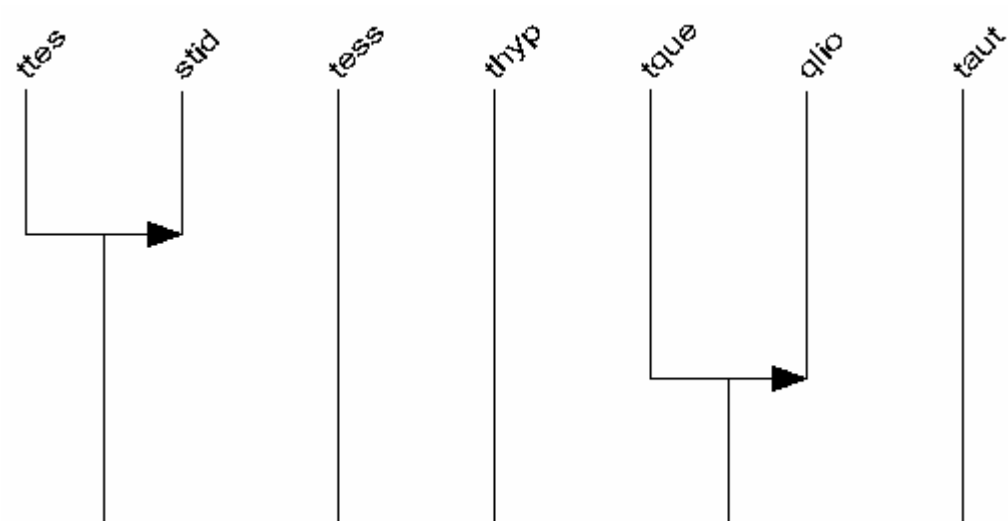


ttes	Tester
tess	Essai
tvid	Vide
thyp	Hypothèse
tque	Question
taut	Autre

En ce qui concerne ce graphique, on constate en premier que le fait d'être en « QLIO » nous amène à une liaison avec les mots « question » (modalité « tque ») et « essai » (modalité « tess ») comme définition de « test d'hypothèse ».

Si on regarde les « STID », on voit qu'ils donnent plutôt comme définition « estimation » (modalité « ttes »), « hypothèse » (modalité « thyp ») et d'autres mots (modalité « taut »).

○ *Arbre cohésitif*



ttes	Tester
tess	Essai
tvid	Vide
thyp	Hypothèse
tque	Question
taut	Autre

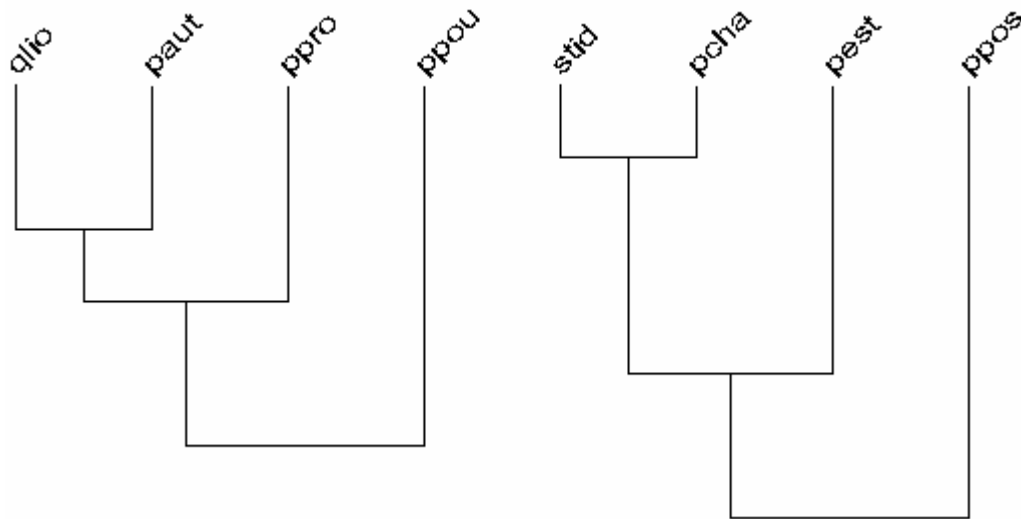
Niveau 1 « ttes » implique « stid » : ceux qui pensent que test d'hypothèse veut dire tester sont en STID

Niveau 2 « tque » implique « qlio » : ceux qui pensent que test d'hypothèse signifie question sont en QLIO

Ici, on voit très bien que le mot « tester » (modalité « ttes ») comme définition du terme implique beaucoup d'être en « STID ». Si l'on pense que « test d'hypothèse » signifie plutôt « question » (modalité « tque »), alors on a des chances d'être en « QLIO ».

- *Analyse du mot « probabilité »*

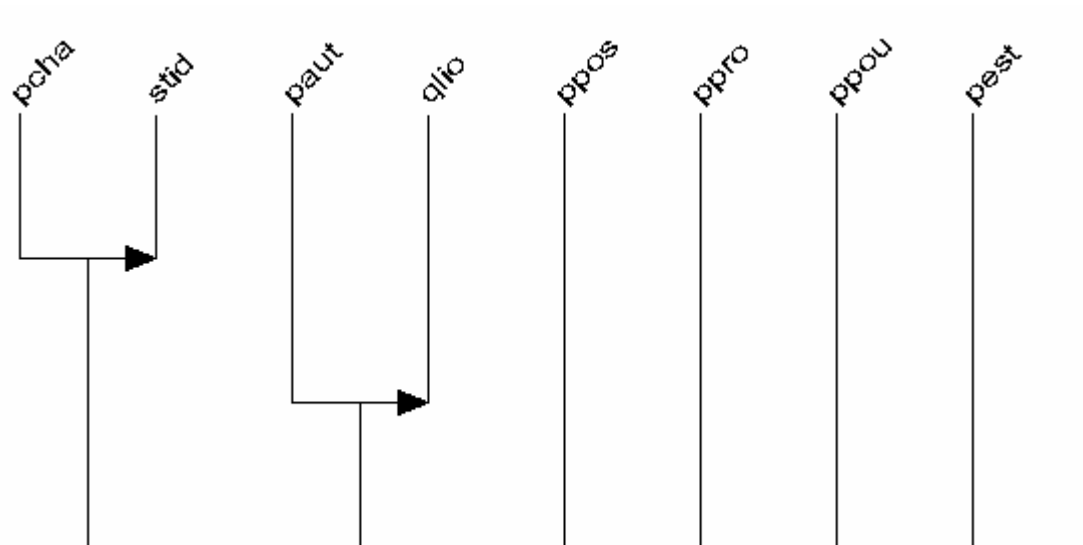
- *Arbre de similarité*



pcha	Chance
pvid	Vide
paut	Autre
ppos	Possibilité
ppro	Probabilité
ppou	Pourcentage
pest	Estimation

D'après ce graphique, on voit une ressemblance entre le fait d'être en « QLIO » et penser que le terme « probabilité » n'a pas de signification particulière (modalité « paut »). Ceux qui se lancent pensent que cela signifie « probabilité » (modalité « ppro ») donc ne savent pas ce que c'est. D'autre évoque le mot « pourcentage » (modalité « ppou »). Les « STID », quand à eux, pensent que ce terme est défini par le mot « chance » de façon très significative ou encore le mot « estimation » et « possibilité ».

o *Arbre de similarité*



pcha	Chance
pvid	Vide
paut	Autre
ppos	Possibilité
ppro	Probabilité
ppou	Pourcentage
pest	Estimation

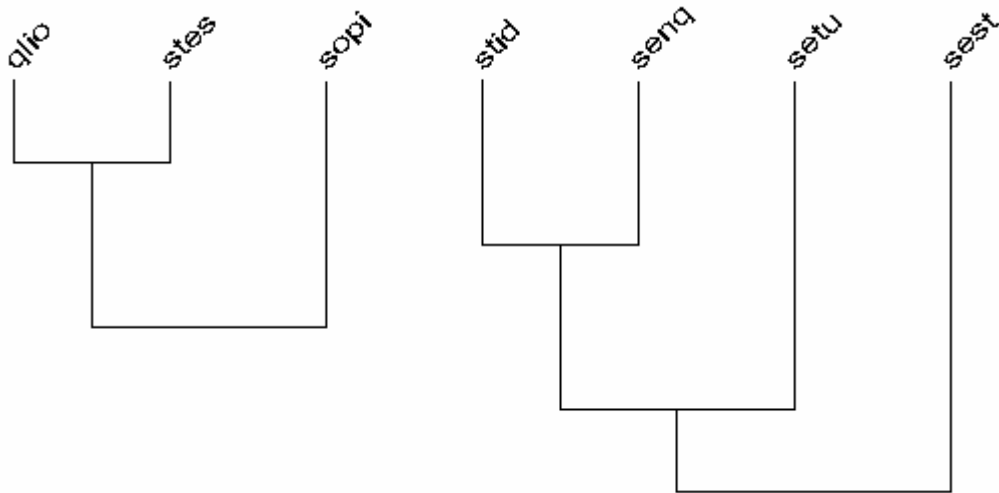
Niveau 1 « pcha » implique « stid » : ceux qui pensent que « probabilité » veut dire « chance » sont en STID

Niveau 2 « paut » implique « qlio » : ceux qui pensent que « probabilité » signifie « autre » (pas de signification) sont en QLIO

On voit bien sur ce graphique que la variable « pcha », c'est-à-dire penser que « probabilité » signifie « chance » implique beaucoup d'être en « STID ». Si l'on n'a pas de signification particulière, alors on a des chances d'être en « QLIO ».

- *Analyse du mot « sondage »*

- *Arbre de similarité*

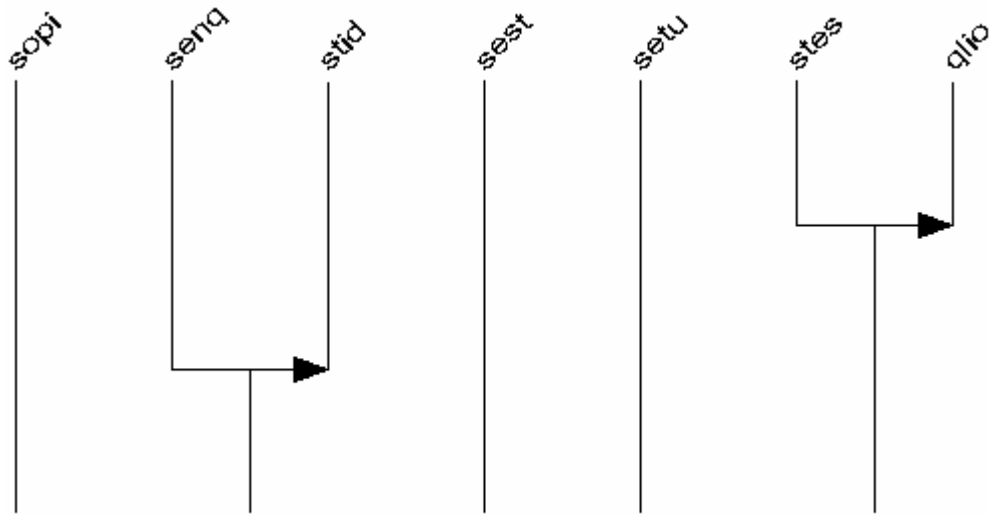


sopi	Opinion
senq	Enquête
sest	Estimation
svid	Vide
setu	Etude
stes	Test

Dans cet arbre, on peut noter que le fait d’être en « QLIO » est similaire au fait de penser que le terme « sondage » signifie « test » (modalité « stes ») et « opinion » (modalité « sopi »).

Les « STID », eux, ont affirmé que le terme leur évoquait plutôt « enquête », « étude » ou « estimation ».

○ *Arbre cohésitif*



sopi	Opinion
senq	Enquête
sest	Estimation
svid	Vide
setu	Etude
stes	Test

Niveau 1 « stes » implique « qlio » : ceux qui pensent que « sondage » veut dire « estimation » sont en QLIO

Niveau 2 « senq » implique « stid » : ceux qui pensent que « sondage » signifie « enquête » sont en STID

On voit que si le terme « probabilité » évoque l'enquête (modalité « senq »), alors on affirme que l'on est en « STID ».

Par contre, si les personnes interrogées pensent que c'est plus un test (modalité « stes »), alors nous avons des chances que ce soit des « STID ».

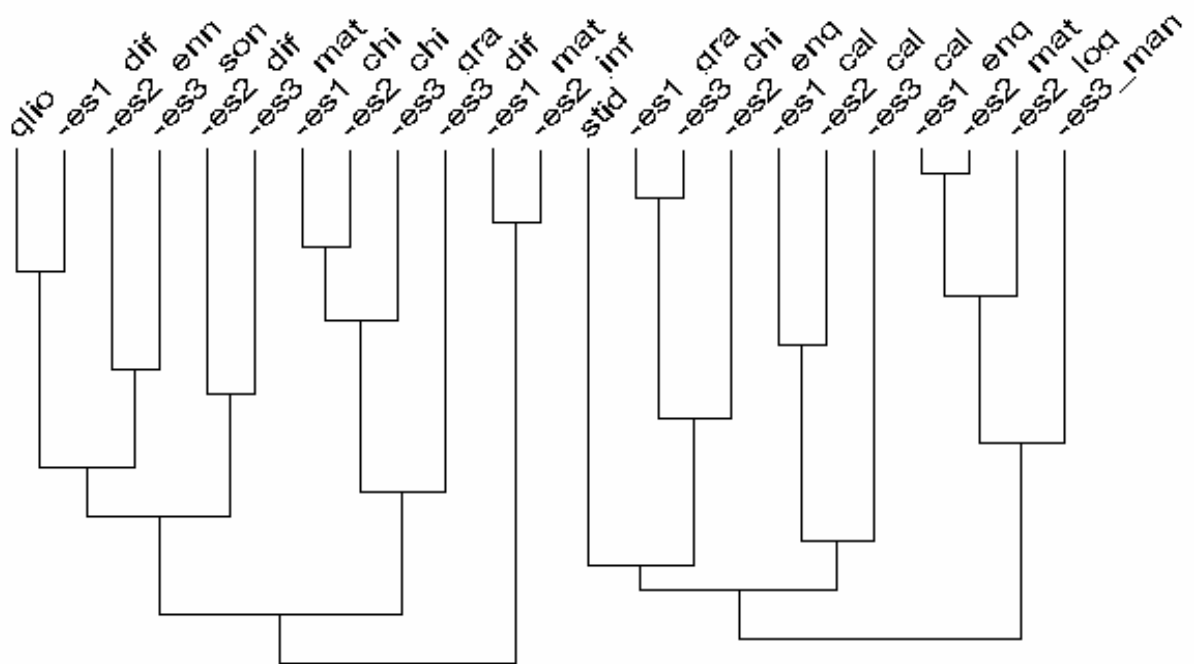
5.2.2.5. Les étudiants et l'évocation de la statistique à l'entrée de leur formation à l'IUT

V7.A l'entrée dans cette formation en IUT, qu'évoquait pour vous la statistique ?

A cette question, les étudiants devaient répondre par trois mots de la façon suivante :

(Donnez trois mots)

- *Arbre de similarité*



En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.

D'abord, on peut voir que la modalité « QLIO » est proche de la modalité « -es1_dif ». Si on traduit, les QLIO pensaient à un niveau de significativité élevé qu'ils trouvent que les statistiques étaient difficiles. Ensuite, on constate que la modalité « -es1__mat » ressemble à la modalité « -es2_inf ». En clair, quand on pensait que statistiques voulaient dire les

tel-00191166, version 1 - 27 Nov 2007

mathématiques, alors on pensait que cela voulait dire informatique à un niveau de significativité encore plus fort. C'est valable plutôt chez les QLIO.

La modalité « -es1_chi » est proche de la modalité « -es2_chi » puis proche de « -es3_gra » et « -es3_dif ». Donc, quand on pensait « chiffre » en évocation de la statistique, alors on pense « graphique » en général et de manière moins significative au mot « difficile », toujours chez les QLIO.

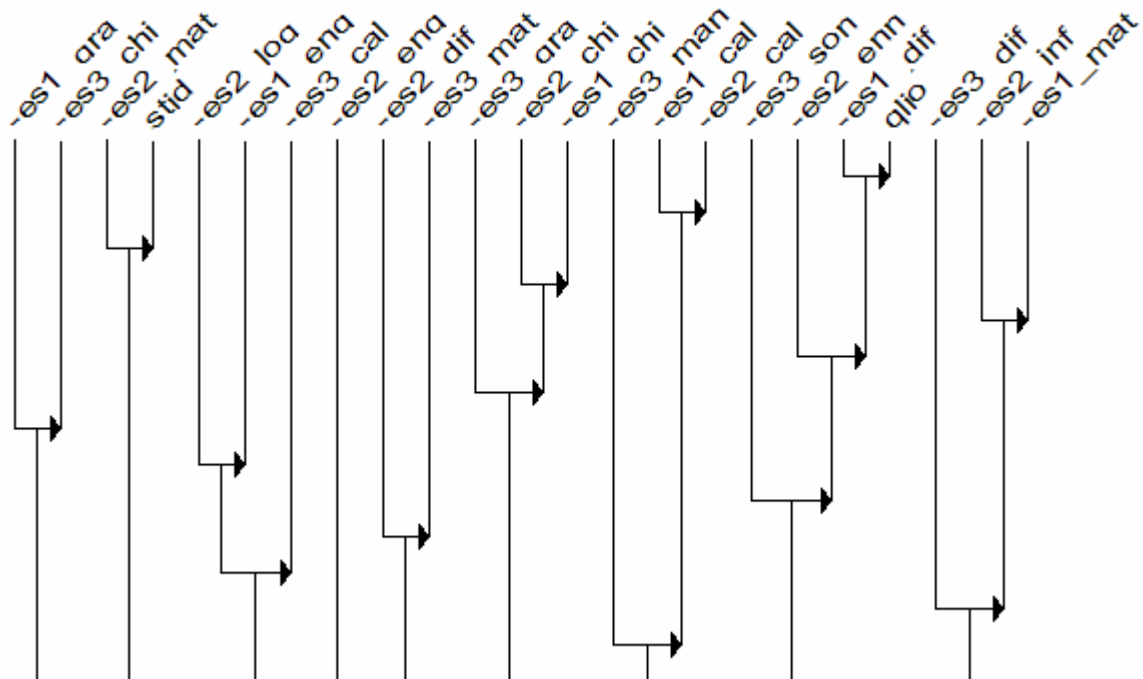
On peut remarquer aussi que la modalité « -es2_enq » ressemble à la modalité « -es3_son ». Effectivement, quand on pensait que les statistiques voulaient dire « enquête », alors on pensait à « sondage » de façon très significative. Ceci a des chances de se produire plutôt chez les étudiants STID.

La modalité « -es2_dif » ressemble à la modalité « -es3_mat ». Quand la statistique nous évoquait la difficulté, alors on pensait au mot « mathématique ».

La modalité « -es1_gra » est proche de la modalité « -es3_chi » puis proche de « -es2_enq ». En effet, ceux qui pensaient que les statistiques leur évoquaient le mot « graphique », alors on pensait aussi « chiffre » et « enquête ».

Enfin, « -es1_enq » proche de « -es2_mat » puis proche « -es2_log » et de façon beaucoup moins significative proche de « -es3_man ». Si on traduit, quand on évoquait le mot « enquête », alors on pensait « mathématiques », « logique » et de manière moins significative à « manipulation », toujours chez les étudiants QLIO.

- *Arbre cohésitif*



En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.

Interprétation de quelques niveaux :

- Niveau 1** « -es1_dif » implique « qlio » : ceux qui pensaient que la statistique était difficile sont les étudiants QLIO
- Niveau 2** « -es1_cal » implique « -es2_cal » : les statistiques étaient associés aux calculs
- Niveau 3** « -es2_mat » implique « std » : ceux qui pensaient que la statistiques étaient des mathématiques sont les étudiants STID
- Niveau 4** « -es2_chi » implique « -es1_chi » : la statistique évoquait les chiffres
- Niveau 5** « -es2_inf » implique « -es1_mat » : l'inférence fait appel aux mathématiques

D'abord, on voit que les individus qui pensaient, en rentrant à l'IUT, que les statistiques étaient une discipline difficile (« -es1_dif »), implique fortement des étudiants provenant de la promo QLIO. Cette appréhension se retrouve essentiellement chez les étudiants pour qui la statistique présente un aspect ennuyeux (« -es2_enn »).

Ensuite, les étudiants pour qui statistique rimait avec mathématiques (« -es2_mat ») implique de manière importante ceux de la promo STID.

Les étudiants pour qui la statistique évoquaient en deuxième choix l'informatique (« -es2_inf »), implique ceux qui ont évoqué en premier lieu les mathématiques (« -es1_mat »). Le fait d'évoquer le mot « difficile » (« -es3_dif ») implique de façon beaucoup moins significative le fait de penser que la statistique évoque la difficulté et les mathématiques.

Les personnes pour lesquelles la statistique évoquait avant le DUT les graphiques (« -es3_gra ») à l'entrée de leur formation à l'IUT implique également ceux pour qui cela évoque les chiffres (« -es2_chi » et « -es1_chi »), ce qui est logique puisqu'un graphique s'établit à partir de données chiffrées. Cela se retrouve dans une autre relation à gauche du graphique.

On peut voir aussi que le mot « manipulation » (« -es3_man ») implique le mot « calcul » (« -es1_cal » et « -es2_cal »).

La modalité « -es2_dif » implique la modalité « -es3_mat ». Donc le fait évoquer la difficulté avant l'entrée en DUT en statistique, implique le fait de penser au mot « mathématique ».

Enfin, on voit que le mot « logique » (« -es2_log ») implique le mot « enquête » (« -es1_enq ») et « calcul » (« -es3_cal »).

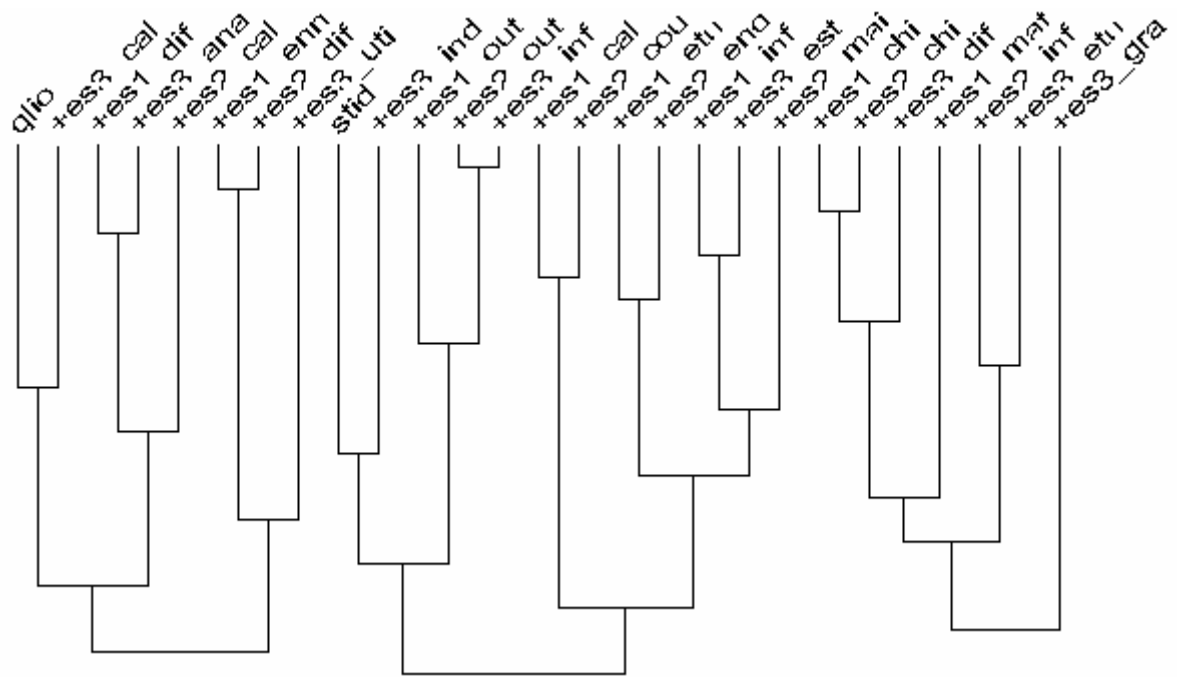
5.2.2.6. Les étudiants et l'évocation de la statistique à ce jour

V8.A ce jour qu'évoque pour vous la statistique ?

A cette question, les étudiants devaient répondre par trois mots de la façon suivante :

(Donnez trois mots)

- *Arbre de similarité*



En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.

Sur cet arbre de similarité, nous constatons trois groupes de modalités.

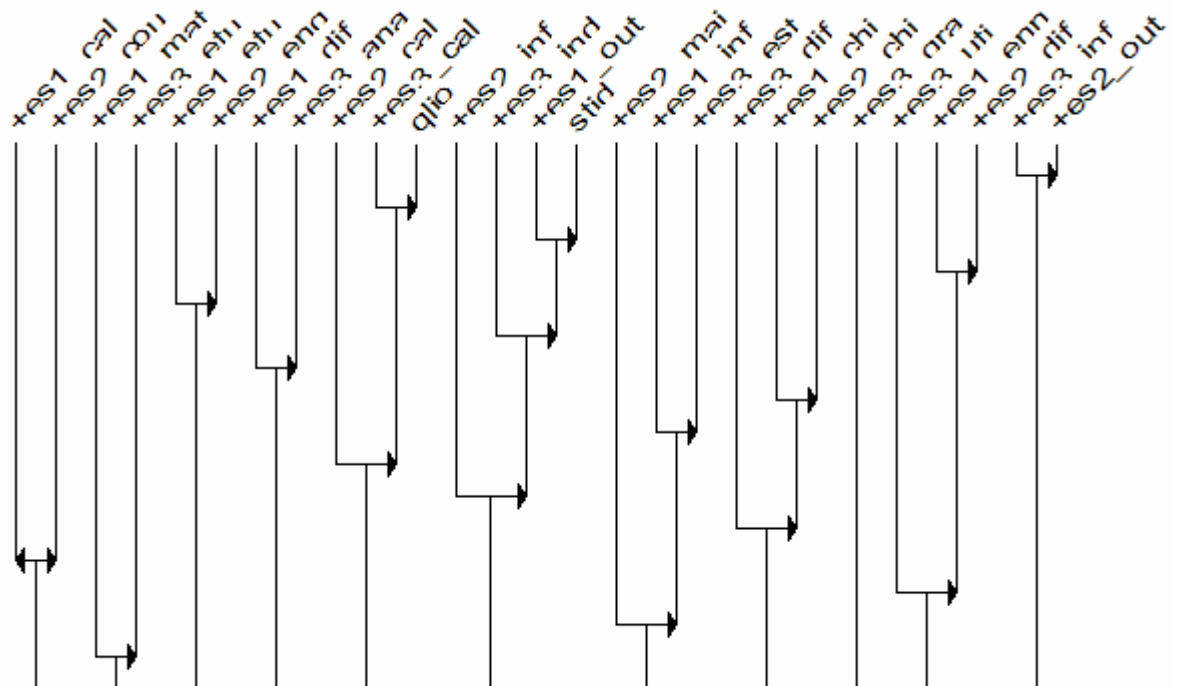
Le premier concerne les étudiants QLIO. On peut voir que la modalité « qlio » est proche de la modalité « +es3_cal », ce qui signifie que les étudiants QLIO associent la statistique aux calculs. Ce groupe de modalités est également proche des modalités « +es1_dif » et « +es3_ana » et de façon moins significative de la modalité « +es2_cal » : certains étudiants QLIO pensent que la statistique est difficile et demande de l'analyse et pour d'autres, la statistique sont des calculs. Pour d'autres encore, mais de façon moins marquée, la statistique évoque l'ennui (+es1_enn) et la difficulté (+es2_dif) mais que c'est utile (+es3_uti).

Le deuxième groupe concerne les étudiants STID. Ici, la modalité « stid » est proche de la modalité « +es3_ind » : les étudiants STID ont une vision plus précise ; en effet, ils associent la statistique aux calculs d'indicateurs (moyenne, variance, écart-type,...). Ce groupe de modalités est aussi proche des modalités « +es3_inf » : ils associent également la statistique à un outil pour l'inférence. Pour une autre partie des étudiants, de façon plus ou

moins significative, la statistique est également des calculs, des cours, des études, des enquêtes, de l'estimation, demande de la maîtrise.

Enfin, le troisième groupe ne fait pas de distinction entre les deux promotions : il concerne les étudiants en général. De façon la plus significative à la moins significative, la statistique leur évoque les chiffres, la difficulté, les mathématiques, l'inférence, les études puis les graphiques.

- *Arbre cohésitif*



En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.

Interprétation de quelques niveaux:

- Niveau 1** « +es3_inf » implique « +es2_ou » : l'inférence est un outil statistique
- Niveau 2** « +es3_cal » implique « qlio » : ceux qui pensent que la statistique est associée aux calculs sont les étudiants QLIO
- Niveau 3** « +es1_out » implique « stid » : ceux qui pensent que la statistique est un outil sont les étudiants STID
- Niveau 4** « +es1_enn » implique « +es2_dif » : l'ennui implique la difficulté

Niveau 5 « +es1_etu » implique « +es2_enq » : la statistique évoque la réalisation d'étude pour faire des enquêtes

En premier lieu, on remarque qu'il existe une forte implication entre l'inférence (« +es3_inf ») et les outils (« +es2_ou ») : en effet, avoir répondu que la statistique évoque à ce jour l'inférence, implique d'avoir également répondu qu'elle évoque des outils. Ces 2 notions sont en effet liées puisque la statistique inférentielle peut être assimilée à un outil.

A ce jour, les étudiants pour lesquels la statistique évoque des outils sont ceux issus de la formation STID. En effet, au cours de cette formation, les étudiants font l'apprentissage de différentes méthodes de travail et d'une variété d'outils utiles pour faire des statistiques. Les personnes pour lesquelles la statistique évoque des indicateurs sont celles citées précédemment. Cela semble logique puisque ces personnes apprennent également l'utilisation et le calcul de différents indicateurs lors de leur formation STID. Enfin, ces étudiants sont aussi ceux pour lesquels la statistique évoque l'informatique, ce qui conforte ce que l'on a dit précédemment : l'informatique fait aussi partie des enseignements fondamentaux de la formation STID.

On remarque également qu'avoir répondu « des outils » à la question « qu'évoque à ce jour pour vous la statistique ? » implique fortement d'être issu de la formation QLIO. Cela peut s'expliquer par le fait que ces étudiants ne suivent pas une formation axée sur la statistique, et s'en font donc une représentation plus basique, plus mathématique et peut-être un peu moins précise et technique que les étudiants STID, ce qui semble logique.

De plus, on constate que les étudiants pour lesquels la statistique évoque une certaine difficulté sont également ceux pour qui elle évoque l'ennui... Cela paraît logique, puisque ceux qui trouvent le cours ennuyeux ne le suivent donc pas, mais éprouvent par la suite des difficultés.

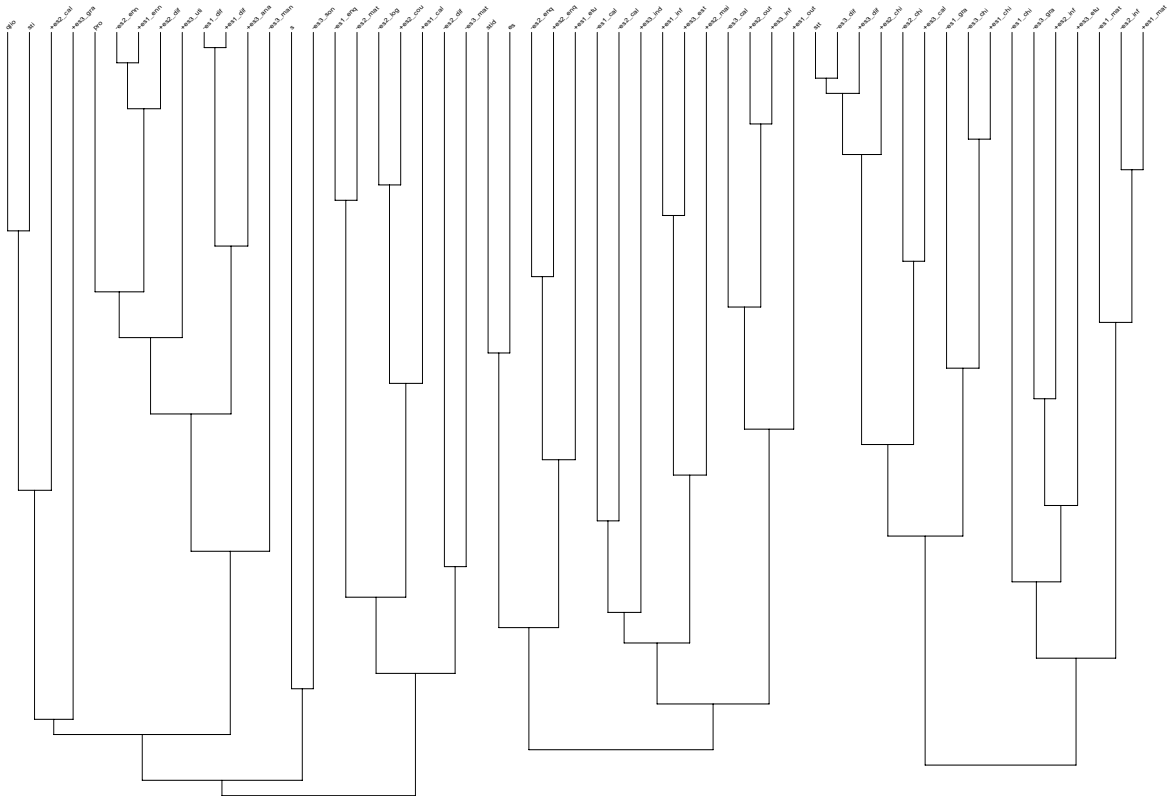
Enfin, on observe que le fait que la statistique évoque la difficulté implique le fait qu'elle évoque également les chiffres. Cela signifie que les personnes ont des difficultés en statistique plus spécifiquement concernant l'aspect mathématique, les calculs que les théories de méthodologie des enquêtes par exemple.

5.2.2.7. Les étudiants et ce qu'évoquait la statistique avant le DUT puis à ce jour selon le bac et le DUT ?

V7.A l'entrée dans cette formation en IUT, qu'évoquait pour vous la statistique ?

V8.A ce jour qu'évoque pour vous la statistique ?

- *Arbre de similarité*



Arbre des similarités : X:\projet\stat\Clpse de fichier à partir d'un clic\Doc

En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.

Voici une analyse des principales similitudes :

Nous pouvons d'abord constater d'après ce graphique que quand on pense avant le DUT que la statistique est difficile, à ce jour on pense la même chose en général. Ceci est valable plus pour les « QLIO » et les « S » que pour les autres car ils appartiennent au même groupe de nœuds. En effet, on voit à un niveau de significativité très élevé que les modalités « -es1_dif » et « +es1_dif » sont très similaires. On pense aussi dans une moindre mesure qu'il y a un lien entre cette relation avec le fait de penser pendant le DUT que les statistiques

signifient l'analyse « +es3_ana ». Ensuite, on peut remarquer que les interrogés provenant d'un bac Pro pensent avant le DUT et actuellement que les statistiques signifient l'ennui et la difficulté. Effectivement, les modalités « -es2_enn », « +es2_enn » et « +es2_dif » sont très liés significativement et de manière plus lointaine avec la modalité « pro ». On peut signaler également que les anciens « stt » pensait que les statistiques étaient difficiles avant le DUT (« -es3_dif »). Actuellement ils ont le même état d'esprit et évoque aussi les chiffres (« +es2_chi ») quand on leur parle de Statistique. Ces anciens « stt » ne changent pas d'avis avant et après le DUT quand les statistiques leurs évoque les chiffres. Enfin, pour cette catégorie, quand les statistiques leurs évoquait l'informatique (« -es2_inf ») avant le DUT, cela leur évoque maintenant plutôt les mathématiques (« +es1_mat »).

En ce qui concerne le groupe de nœuds rattachant plutôt les « STID » et « ES », quand ils pensent actuellement que la statistique est un outil (« +es2_out »), ils pensent aussi à un niveau de significativité élevé que c'est de l'informatique (« +es3_inf ») et également du calcul (« +es3_cal ») mais de façon plus atténué. Cette catégorie pensent aussi actuellement que cette discipline leur évoque l'informatique et l'estimation. Ces modalités (« +es1_inf » et « +es3_est ») sont similaires. Enfin, quand ces types de personnes évoquaient l'enquête (« -es2_enq ») avant, ils pensent la même chose maintenant (« +es2_enq »).

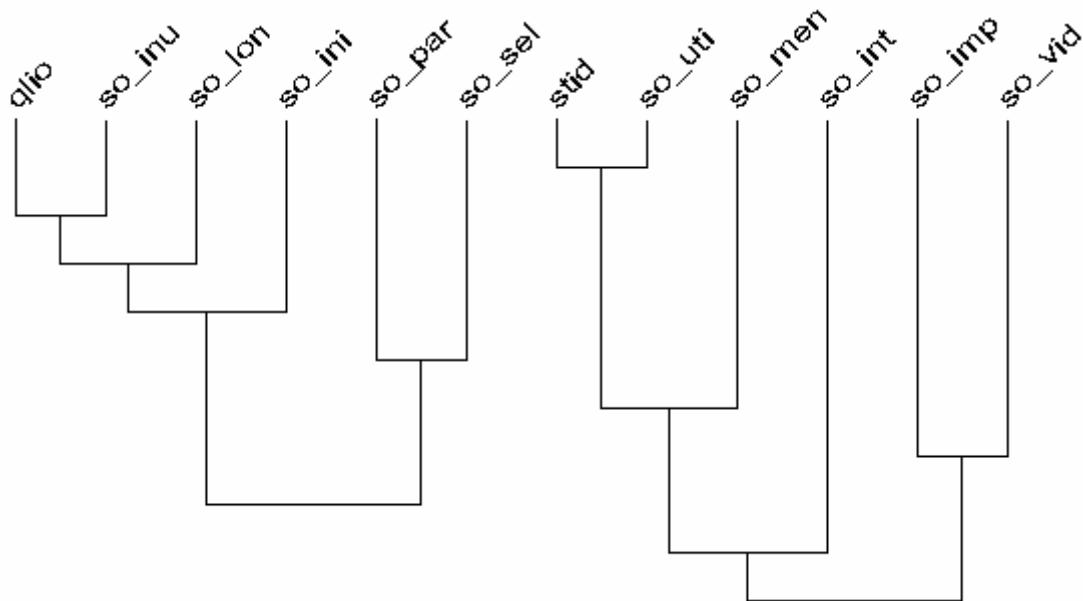
- *Arbre cohésitif*

avant le DUT que la statistique était ennuyeuse (« -es2_enn »), alors on a de fortes chances de penser maintenant que c'est toujours ennuyeux (« +es1_enn ») et même que c'est difficile (« +es2_dif »). Enfin, le fait d'être ancien « stt » implique le fait d'évoquer la difficulté actuellement (« +es3_dif ») et avant le DUT (« -es3_dif ») et d'une façon plus significative les chiffres actuellement (« +es2_chi »).

5.2.2.8. Les étudiants et les sondages d'opinion

V10. A ce jour que pensez-vous des sondages d'opinion ?

- *Arbre de similarités*



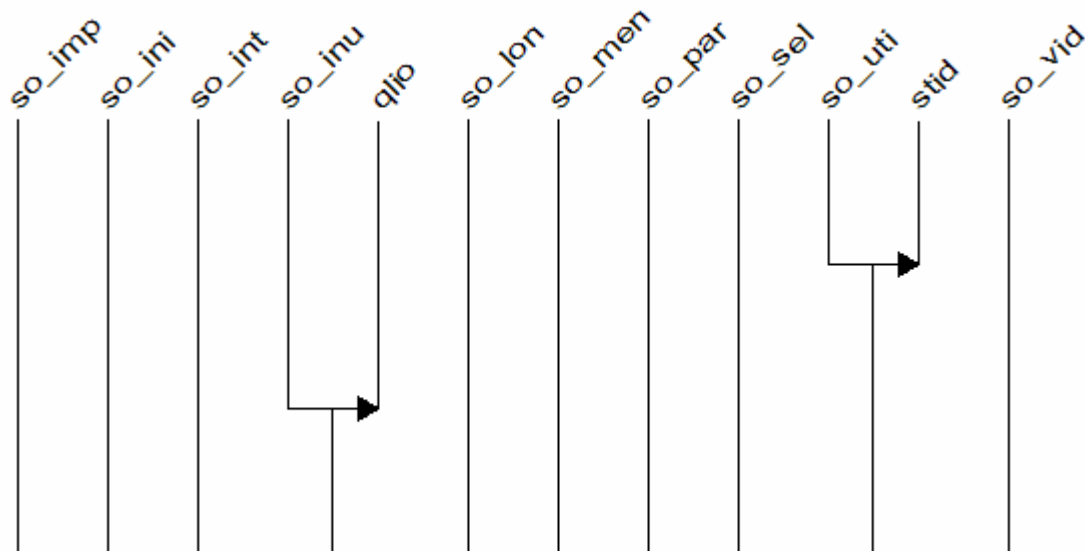
so_imp	Imprécis
so_ini	Inintéressant
so_int	Intéressant
so_inu	Inutile
so_lon	Long
so_men	Mensonger
so_par	Participatif
so_sel	Sélectif
so_uti	Utile
so_vid	Vide

tel-00191166, version 1 - 27 Nov 2007

Nous pouvons tout d'abord remarquer qu'il existe un lien entre le fait d'avoir suivi la formation QLIO et de penser que les sondages d'opinion sont inutiles. Ce sont ces mêmes étudiants qui caractérisent également les sondages de longs et inintéressants.

A l'inverse, il existe une corrélation entre être un étudiant STID et juger les sondages d'opinion utiles. Ce sont ces mêmes individus pour qui les sondages d'opinion apparaissent intéressants. Cela semble logique puisqu'ils suivent une formation en statistique et donc en rapport avec les sondages d'opinion qu'ils auront peut-être plus tard à analyser. Cependant, on observe également qu'ils les considèrent parfois comme mensonger, ce qui traduit qu'ils savent rester lucides.

- *Arbre cohésitif*



so_imp	Imprécis
so_ini	Inintéressant
so_int	Intéressant
so_inu	Inutile
so_lon	Long
so_men	Mensonger
so_par	Participatif
so_sel	Sélectif
so_uti	Utile
so_vid	Vide

Niveau 1 « so_uti » implique « stid » : les sondages d'opinion sont utiles selon les étudiants STID

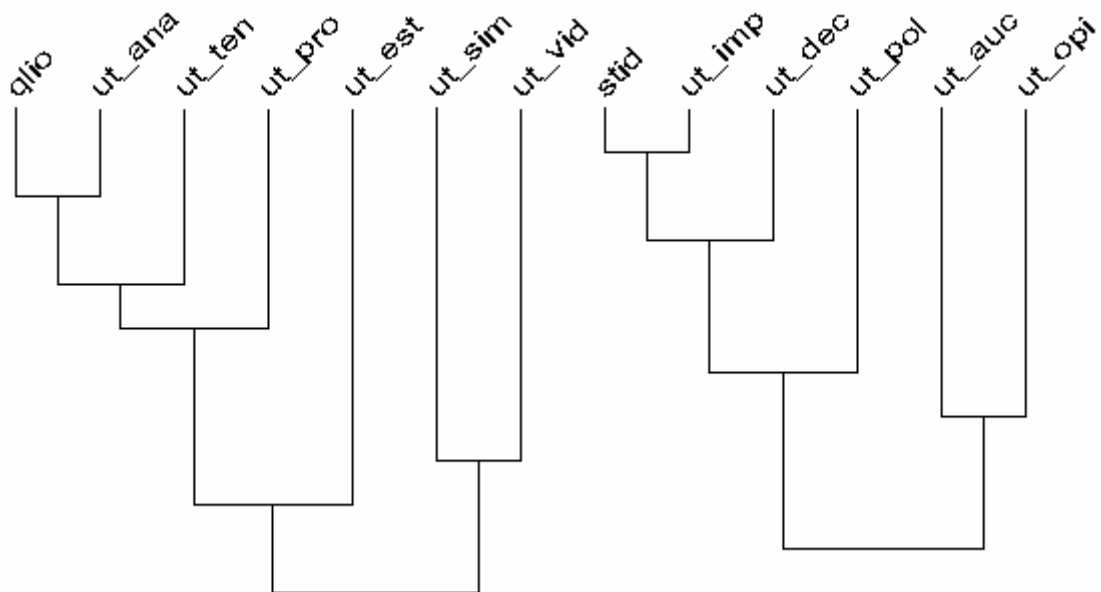
Niveau 2 « so_inu » implique « qlio » : les sondages d'opinion sont inutiles selon les étudiants QLIO

Une idée apparaît relativement frappante, concernant la position des étudiants par rapport aux sondages d'opinion. En effet, nous pouvons constater que les personnes qui trouvent les sondages utiles sont les étudiants STID, tandis que celles qui trouvent ceux-ci inutiles sont les étudiants de la promo QLIO. Cela confirme une idée préconçue : un des apprentissages de la formation STID consiste à maîtriser la réalisation de sondages. Il serait contradictoire que les étudiants qui suivent cette formation, trouvent les sondages d'opinion inutiles.

5.2.2.9. Les étudiants et l'utilité de la statistique

V11. Dans la vie sociale actuelle, quelle utilité attribuez-vous à la statistique ?

- *Arbre de similarités*



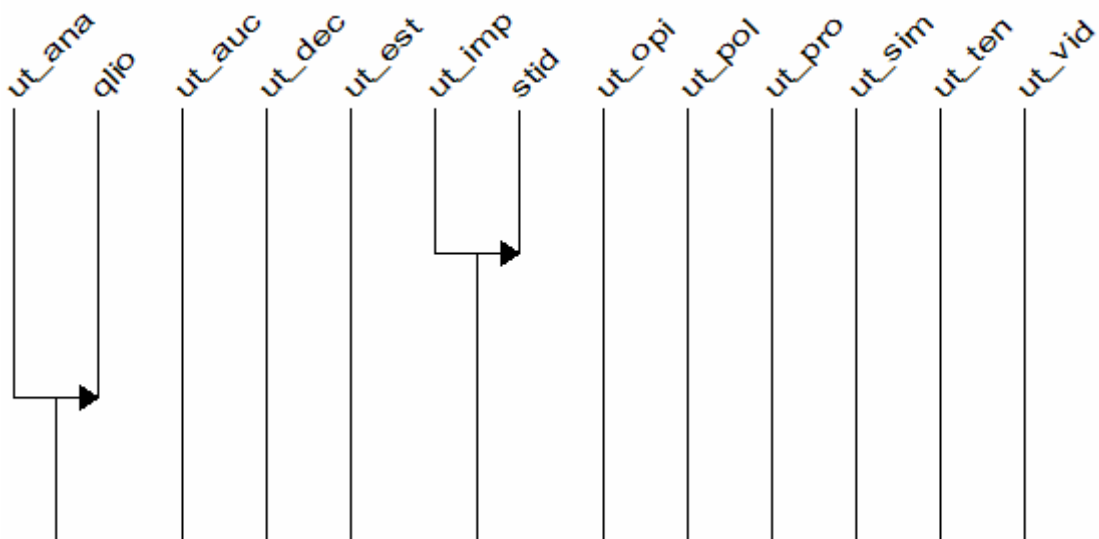
ut_ana	Analyse
ut_auc	Aucune
ut_dec	Décisionnel
ut_est	Estimation
ut_imp	Importante
ut_opi	Opinion

ut_pol	Politique
ut_pro	Probabilité
ut_sim	Simplification
ut_ten	Tendance

En premier lieu, nous pouvons constater qu’il existe une liaison entre le fait d’avoir suivi ou de suivre la formation QLIO et de trouver la statistique utile pour l’analyse. Ces étudiants trouvent aussi que la statistique permet de calculer des probabilités et de faire des estimations. Ils semblent avoir une vision assez scolaire de la statistique.

Nous pouvons également remarquer que la modalité « STID » est très proche de la modalité « importante » : cela signifie que ces étudiants trouvent que la statistique occupe une place primordiale. En effet, d’après ces étudiants, la statistique fournit une aide à la décision et est utile au niveau de la politique.

- *Arbre cohésitif*



ut_ana	Analyse
ut_auc	Aucune
ut_dec	Décisionnel
ut_est	Estimation
ut_imp	Importante
ut_opi	Opinion
ut_pol	Politique
ut_pro	Probabilité

ut_sim	Simplification
ut_ten	Tendance

Niveau 1 « ut_est » implique « stid » : la statistique est utile pour l'estimation selon les étudiants STID

Niveau 2 « ut_ana » implique « qlio » : la statistique est utile pour l'analyse selon les étudiants QLIO

Dans la vie sociale actuelle, les personnes attribuant à la statistique une utilité importante sont bien sûr les STID. En effet cette discipline est le cœur de leur formation et de leur savoir faire : cela paraît évident qu'ils ne peuvent pas trouver les statistiques inutiles.

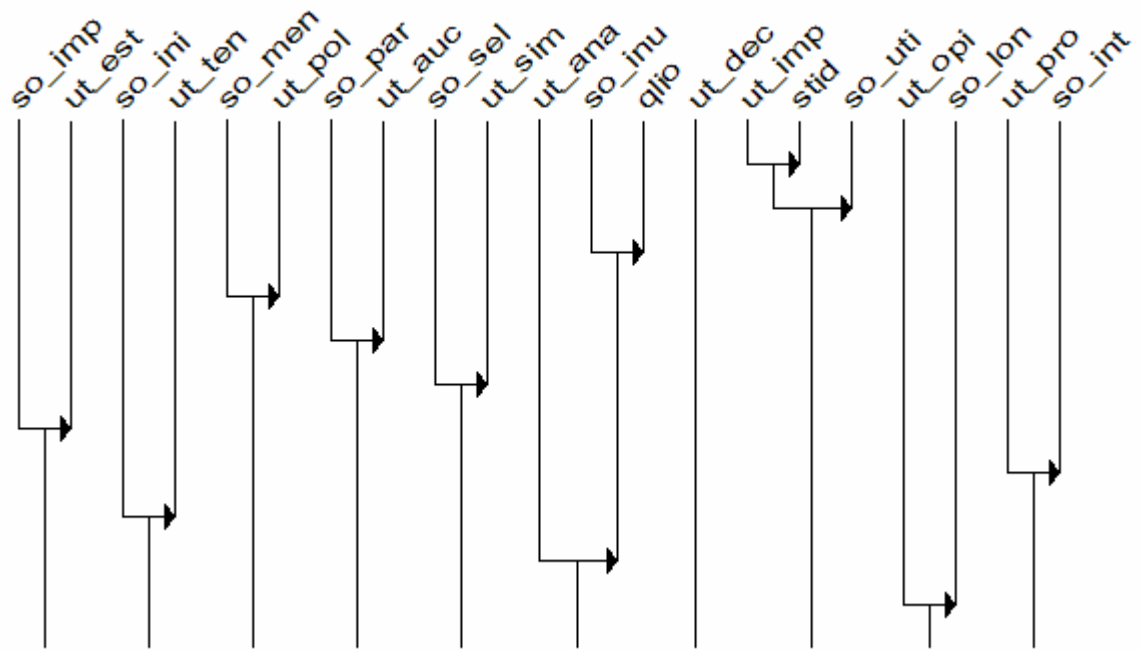
Les personnes pour qui l'utilité principale de la statistique est de fournir une analyse sont les étudiants issus de la formation QLIO. Ces derniers semblent associer très clairement la statistique à l'analyse.

5.2.2.10. L'avis des étudiants sur les sondages d'opinion et l'utilité de la statistique.

V10. A ce jour que pensez-vous des sondages d'opinion ?

V11. Dans la vie sociale actuelle, quelle utilité attribuez-vous à la statistique ?

- *Arbre cohésitif*



so_imp	Imprécis
so_ini	Inintéressant
so_int	Intéressant
so_inu	Inutile
so_lon	Long
so_men	Mensonger
so_par	Participatif
so_sel	Sélectif
so_uti	Utile
so_vid	Vide

ut_ana	Analyse
ut_auc	Aucune
ut_dec	Décisionnel
ut_est	Estimation
ut_imp	Importante
ut_opi	Opinion
ut_pol	Politique
ut_pro	Probabilité
ut_sim	Simplification
ut_ten	Tendance

Interprétation de quelques niveaux :

Niveau 1 « ut_imp » et « stid » impliquent « so_uti » : la statistique est important dans la vie sociale pour les étudiants STID ce qui implique que les sondages d’opinion sont utiles

Niveau 4 « so_men » implique « ut_pol » : les sondages d’opinion mensongers entraînent une utilité politique

- Niveau 5** « so_par » implique « ut_auc » : les sondages d'opinion participatifs montrent que la statistique n'a aucune utilité
- Niveau 6** « so_sel » implique « ut_sim » : le fait que les sondages d'opinion soient sélectifs implique une utilité simplifiée de la statistique
- Niveau 10** « ut_ana » implique « so_inu » et « qlio » : l'analyse comme utilité de la statistique implique que les sondages d'opinion sont inutiles pour les étudiants QLIO

Certains étudiants, dans la vie sociale actuelle, estiment que la place de la statistique est d'une utilité importante : ces personnes sont issues de la promo STID, ce qui s'explique par le choix qu'ils ont entrepris pour rentrer dans cette filière. Ces étudiants, de formation STID et considérant que la statistique est utile, pensent de même que les sondages d'opinion sont importants à ce jour.

Par opposition, les personnes trouvant que les sondages d'opinion sont d'une inutilité dans la vie sociale d'aujourd'hui, sont des étudiants QLIO. Leur formation n'étant pas axée sur les sondages, cela peut justifier leur désintéressement de la matière.

De plus, on peut remarquer, grâce à ce graphique, que les sondages d'opinion considérée comme mensongers de la part des sondés sont en lien avec la science politique. C'est-à-dire que les personnes considérant que les sondages sont mensongers, sont des personnes qui associent l'utilité de la statistique à la politique.

On constate aussi, que les individus pensant à ce jour que les sondages sont sélectifs sont des personnes qui trouvent que l'utilité de la statistique dans la vie sociale actuelle est la simplification.

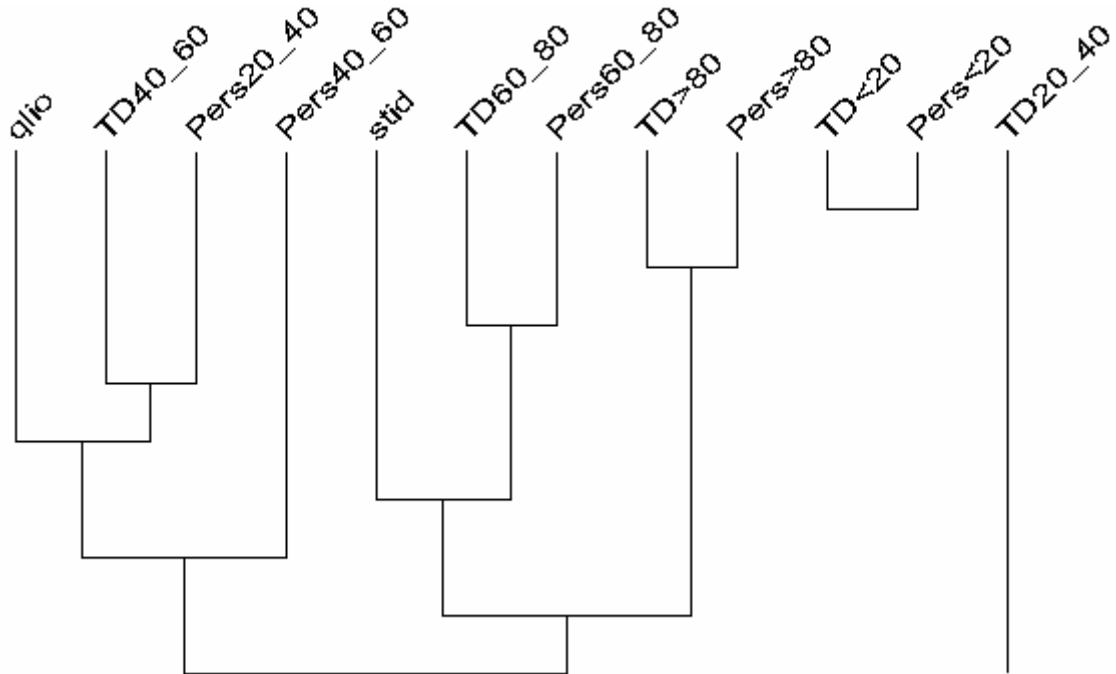
Enfin une dernière remarque intéressante, que l'on peut apporter d'après ce graphique : les gens qui jugent les sondages d'opinion imprécis, sont des personnes qui pensent que l'utilité de la statistique dans la vie sociale actuelle est d'estimer.

5.2.2.11. Nombre d'heures de cours suivies et nombre d'heures de travail personnel à ce jour pour les interrogés.

V12. A ce jour, combien d'heures de cours et de TD de statistique avez-vous suivies ?

V21. A ce jour, durée totale estimée (h) de travail personnel pour étudier la statistique.

- *Arbre de similarité*

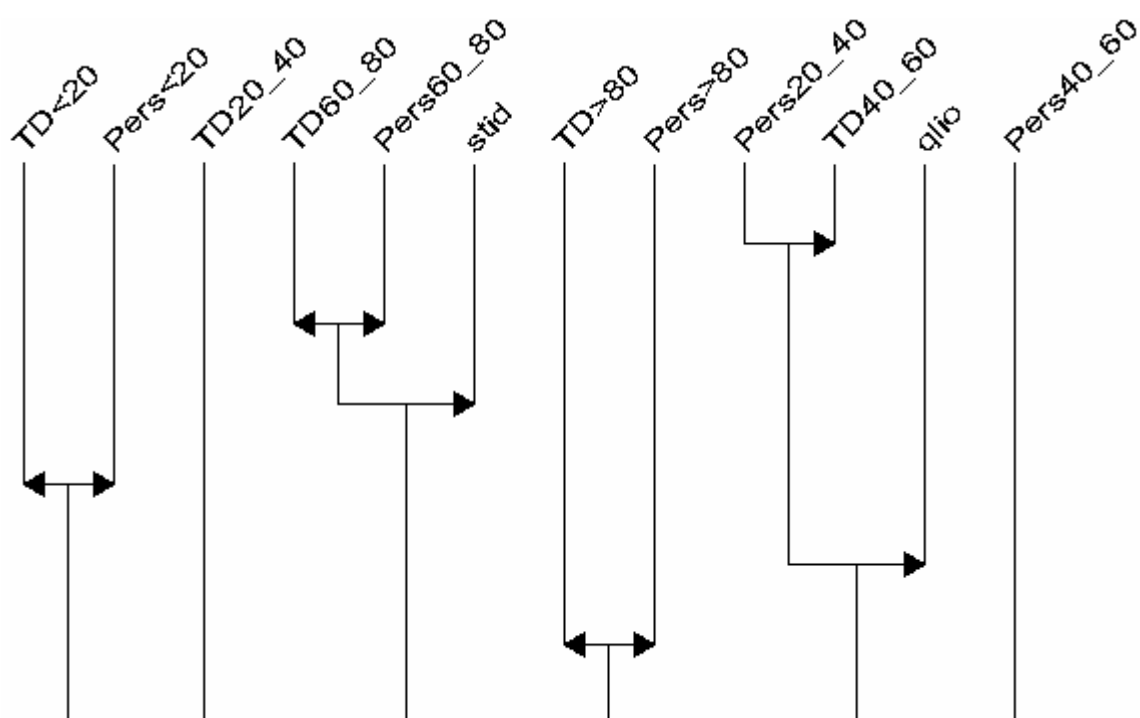


TD<20	Moins de 20
TD20_40	De 20 à 40
TD40_60	De 40 à 60
TD60_80	De 60 à 80
TD>80	Plus de 80
Pers<20	Moins de 20 H
Pers20_40	Entre 20H et 40H
Pers40_60	Entre 40H et 60H
Pers60_80	Entre 60H et 80H
Pers>80	Plus de 80 H

Dans ce graphique, on peut noter que la modalité « QLIO » et les modalités « TD40_60 », « Pers20_40 », « Pers40_60 » sont assez similaires. Si l'on traduit, Les QLIO ont suivi entre 40 et 60 heures de cours et TD et ont consacré en général entre 20 et 60 heures de travail personnel. Ensuite, on voit que la modalité « Pers60_80 » et « TD60_80 » sont très proche et rejoignent à un niveau de significativité moins élevé la modalité « STID ». Donc,

les STID ont suivent entre 60 et 80 heures de cours et TD et consacrent autant de temps pour le travail personnel. On peut aussi ajouter dans une moindre mesure que certains STID estiment avoir suivi plus de 80 heures de cours et TD (modalité « TD>80 ») et ont consacré autant d'heures à la maison pour le travail personnel (modalité « Pers>80 »). Enfin, on peut dire qu'il y a une liaison entre estimer avoir suivi moins de 20 heures de cours et TD (modalité « TD<20 ») et consacrer moins de 20 dans son travail personnel (modalité « Pers<20 »).

- *Arbre cohésitif*



TD<20	Moins de 20
TD20_40	De 20 à 40
TD40_60	De 40 à 60
TD60_80	De 60 à 80
TD>80	Plus de 80
Pers<20	Moins de 20 H
Pers20_40	Entre 20H et 40H
Pers40_60	Entre 40H et 60H
Pers60_80	Entre 60H et 80H
Pers>80	Plus de 80 H

Interprétation de quelques niveaux :

Niveau 1 « pers20_40 » et « TD40_60 » impliquent « qlio » : un travail personnel entre d'environ 30 heures implique entre environ 50 heures de cours et de TD de statistique pour les étudiants QLIO

Niveau 2 « TD60_80 » et « pers60_80 » impliquent « stid » : environ 70 heures de cours et de TD de statistique implique un travail personnel d'environ 70 heures pour les étudiants QLIO

Ici, on voit au niveau de significativité le plus élevé le fait de consacrer entre 20 et 40 heures de travail personnel (modalité « Pers20_40 ») impliquent fortement le fait de dire que l'on a suivi entre 40 et 60 heures de cours et TD (modalité « TD40_60 »). Tout ceci implique le fait d'être en QLIO. On constate aussi que le fait de dire que l'on a suivi entre 60 et 80 heures de TD (modalité « TD60_80 ») impliquent beaucoup le fait de consacrer autant d'heures dans le travail personnel (modalité « Pers60_80 ») et vice-versa. Ceci a des chances de se produire chez les STID. On peut remarquer aussi une implication réciproque entre avoir moins de 20 heures de TD (modalité « TD<20 ») et moins de 20 heures de travail personnel (modalité « Pers<20 »). Il y a également une implication réciproque pour les modalités « TD>80 » et « Pers>80 ». Autrement dit, si on estime que l'on a fait plus de 80 heures de cours et TD, alors on consacre autant d'heures pour le travail personnel.

5.2.2.12. Les étudiants et les notions assez bien comprises

V13. Parmi les notions visées par le cours de statistique, quelles sont celles que vous pensez avoir plutôt bien comprises ?
--

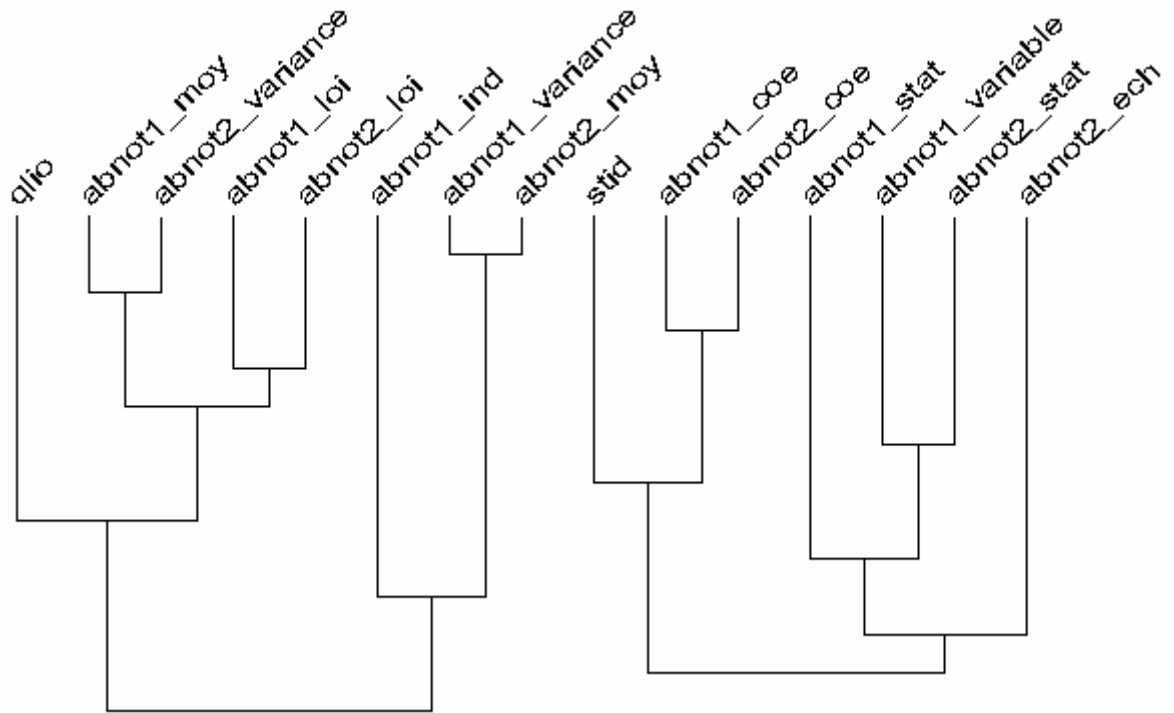
A cette question, les étudiants devaient répondre de la façon suivante :

...assez bien comprises

... plutôt mal comprises

Pour traiter cette question, nous avons choisi de ne prendre que les trois premières réponses.

- *Arbre de similarités*



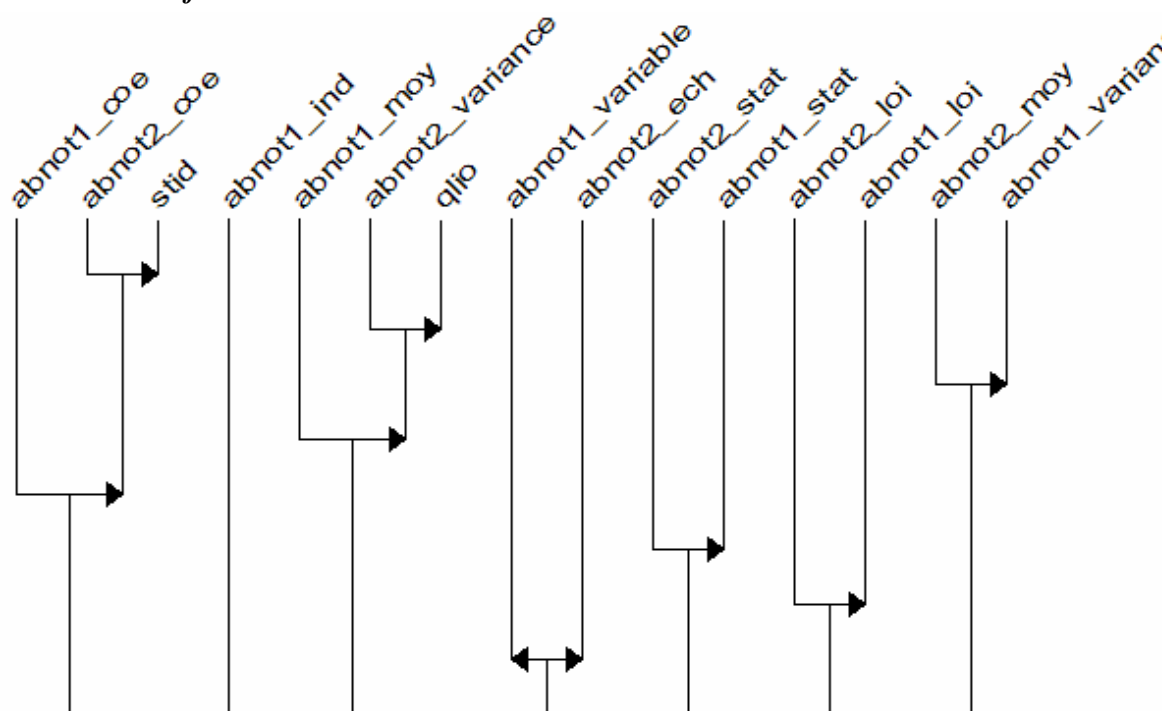
abnot1_coe	Notion1	Coefficients
abnot1_ind		Indicateurs
abnot1_stat		Statistique descriptive
abnot1_moy		Moyenne
abnot1_variance		Variance
abnot1_loi		Lois
abnot1_variable		Variables
abnot1_vid		Vide
abnot2_stat	Notion2	Statistique descriptive
abnot2_loi		Lois
abnot2_variance		Variance
abnot2_moy		Moyenne
abnot2_ech		Echantillon

abnot2_coe		Coefficients
abnot2_vid		Vide

Tout d'abord, nous pouvons observer que la moyenne et la variance sont des notions qui ont été bien assimilées par les étudiants : en effet, ce sont des notions liées constituant les bases des statistiques. Ces étudiants ont aussi compris les notions de loi et sont des individus issus de la formation QLIO.

En parallèle on remarque que les personnes qui ont mis comme notions comprises les coefficients et qui sont des étudiants STID, ont assimilé aussi toutes les notions de la statistique descriptive.

- *Arbre cohésitif*



abnot1_coe	Notion1	Coefficients
abnot1_ind		Indicateurs
abnot1_stat		Statistique descriptive
abnot1_moy		Moyenne
abnot1_variance		Variance
abnot1_loi		Lois
abnot1_variable		Variables
abnot1_vid		Vide

tel-00191166, version 1 - 27 Nov 2007

abnot2_stat	Notion2	Statistique descriptive
abnot2_loi		Lois
abnot2_variance		Variance
abnot2_moy		Moyenne
abnot2_ech		Echantillon
abnot2_coe		Coefficients
abnot2_vid		Vide

Interprétation de quelques niveaux :

Niveau 1 « abnot2_coe » implique « stid » : les coefficients de Cramer et de corrélation sont les notions assez bien comprises par les étudiants STID

Niveau 2 « abnot2_variance » implique « qlio » : la notion de variance est assez bien comprise des étudiants QLIO

Niveau 3 « abnot2_moy » implique « abnot1_variance » : la compréhension de la moyenne implique la compréhension de la variance

Ce graphique concerne les notions visées par le cours de statistique que l'étudiant considère avoir comprises. On constate que ceux qui ont acquis les coefficients, sont des étudiants STID.

Concernant ceux qui ont choisi de mettre en première notion acquise la moyenne, on constate que ce sont des étudiants qui ont mis en second la variance ; ces personnes sont des QLIO. On constate par rapport à la remarque précédente que les QLIO ont acquis des notions de base (moyenne, variance,...) alors que pour les STID ce sont des notions plus spécifiques qui sont divers coefficients comme par exemple Cramer.

Dans ce graphique on remarque enfin une double implication concernant la notion 1 la plus comprise par les étudiants qui touche les variables et la notion 2 qui concerne les échantillons. Ces deux notions sont liées : ce sont les bases sur lesquelles on s'appuie pour effectuer une enquête. En effet, nous avons besoin d'un échantillon d'individus et des variables sur lesquelles interroger les individus, ce qui explique que ces notions ne vont pas l'une sans l'autre.

5.2.2.13. Les étudiants et les notions plutôt mal comprises

V13. Parmi les notions visées par le cours de statistique, quelles sont celles que vous pensez avoir plutôt mal comprises ?

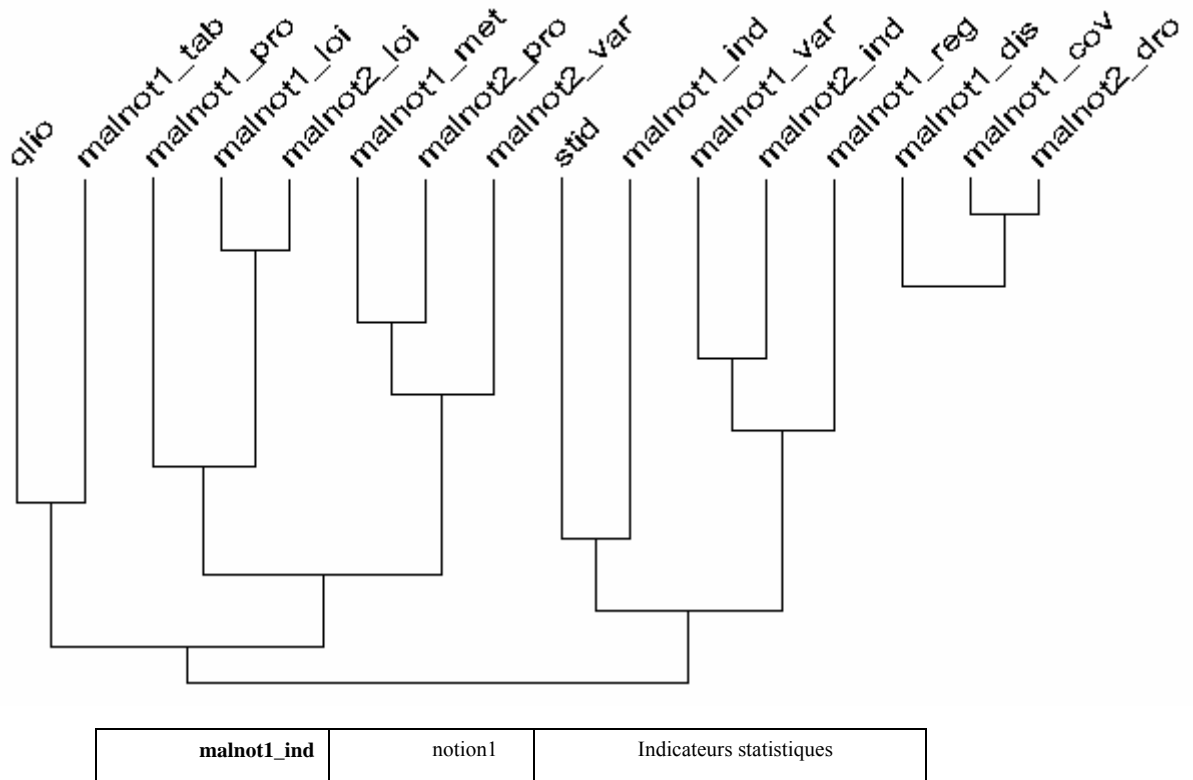
A cette question, les étudiants devaient répondre de la façon suivante :

...assez bien comprises

... plutôt mal comprises

Pour traiter cette question, nous avons choisi de ne prendre que les trois premières réponses.

- *Arbre de similarités*

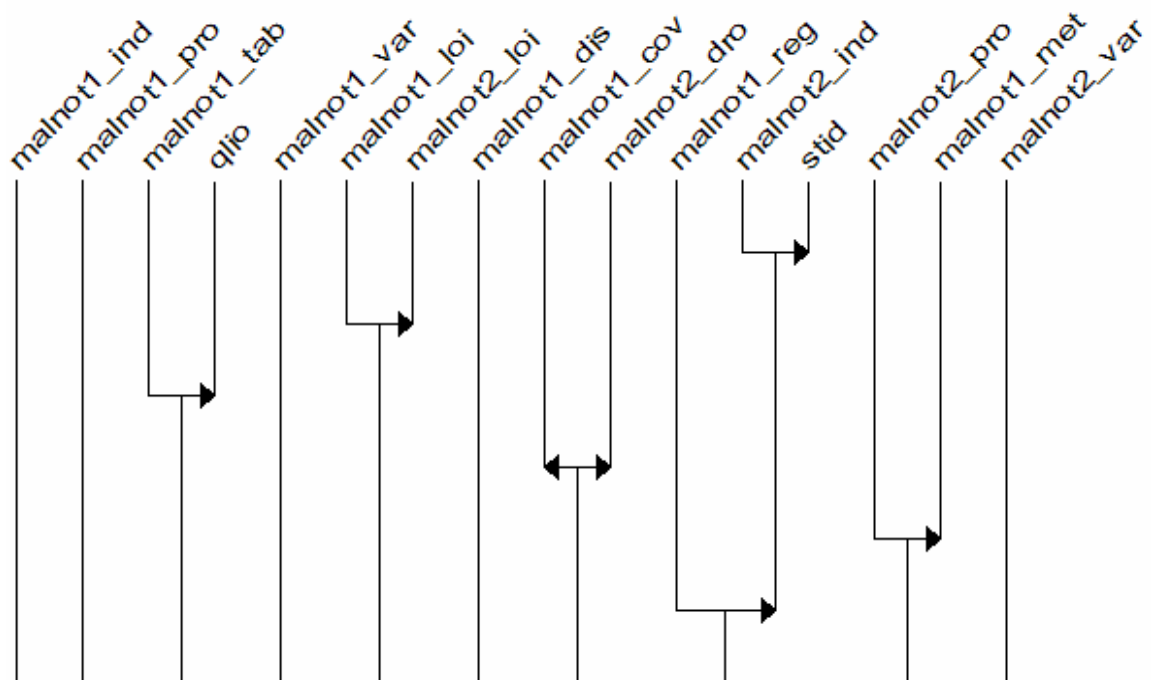


malnot1_pro		Probabilité
malnot1_tab		Tableaux croisés
malnot1_var		Variance
malnot1_met		Méthodes statistiques
malnot1_loi		Lois
malnot1_dis		Dispersion
malnot1_cov		Covariance
malnot1_reg		Régression
malnot1_vid		Vide
malnot2_loi	notion2	Lois
malnot2_pro		Probabilités
malnot2_var		Variance
malnot2_dro		Droite de régression
malnot2_ind		Indicateurs statistiques
malnot2_vid		Vide

On remarque qu'il existe des similarités, dans les notions mal comprises, entre dispersion, covariance et droite de régression. On observe aussi qu'il existe un lien entre avoir des difficultés à comprendre la notion de loi, et rencontrer des problèmes pour assimiler les probabilités. Ces deux notions sont ainsi fortement liées.

Les étudiants issus de la formation QLIO sont ceux qui rencontrent des difficultés afin d'assimiler les tableaux croisés. Les STID, quant à eux, ont du mal à assimiler la notion d'indicateurs statistiques.

- *Arbre cohésitif*



malnot1_ind	notion1	Indicateurs statistiques
malnot1_pro		Probabilité
malnot1_tab		Tableaux croisés
malnot1_var		Variance
malnot1_met		Méthodes statistiques
malnot1_loi		Lois
malnot1_dis		Dispersion
malnot1_cov		Covariance
malnot1_reg		Régression
malnot1_vid		Vide
malnot2_loi	notion2	Lois
malnot2_pro		Probabilités
malnot2_var		Variance
malnot2_dro		Droite de régression
malnot2_ind		Indicateurs statistiques
malnot2_vid		Vide

- Niveau 1** « malnot2_ind » implique « stid » : les indicateurs statistiques sont mal compris par les étudiants STID
- Niveau 3** « malnot1_tab » implique « qlio » : les tableaux croisés sont les notions mal comprises par les étudiants QLIO
- Niveau 4** « malnot1_cov » implique « malnot2_dro » et inversement : la non compréhension de la covariance entraîne la non compréhension de la droite de régression et inversement
- Niveau 5** « malnot2_pro » implique « malnot2_var » : le fait que les probabilités soient mal comprises impliquent le fait que la variance soit mal comprise
- Niveau 6** « malnot1_reg » implique le niveau 1 : la régression comme notion mal comprise implique le fait que les indicateurs statistiques soient mal compris par les étudiants STID

Par opposition au graphique précédent, celui-ci nous montre les notions visées par le cours de statistique que l'étudiant estime avoir, cette fois-ci, mal comprises.

On remarque tout d'abord que les étudiants qui pensent avoir des difficultés dans l'établissement de certains indicateurs statistiques sont des étudiants provenant de la promotion STID. Les gens qui pensent avoir des difficultés pour la régression sont les étudiants que l'on vient de citer, c'est-à-dire des STID dérangés par le calcul d'indicateurs statistiques.

En ce qui concerne les individus qui ont des problèmes avec les tableaux croisés sont des étudiants QLIO. Cela rejoint la remarque que nous avons faite précédemment sur les notions bien comprises visées par le cour de statistique : on peut discerner une différence entre les personnes issues de la formation STID et celles issues de QLIO. En effet, on constate que les QLIO ont des difficultés sur des notions de base que les STID ont sûrement acquis, ces derniers ayant des difficultés sur des éléments plus précis.

5.2.2.14. Les étudiants et les difficultés majeures (pour alternance)

V14. Quelles sont les trois difficultés majeures que vous pensez devoir affronter pour répondre aux exigences du dispositif pédagogique fondé sur l'alternance ?

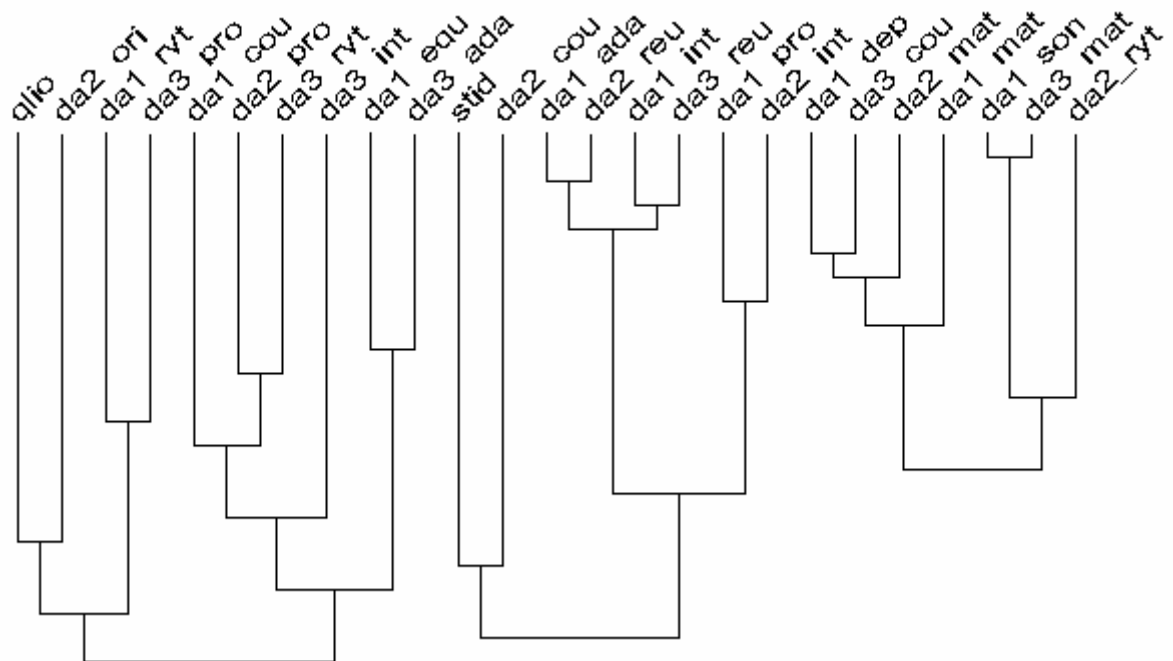
A cette question, les étudiants devaient répondre de la façon suivante :

Diff_1 :

Diff_2 :

Diff_3 :

- *Arbre de similarité*



En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.

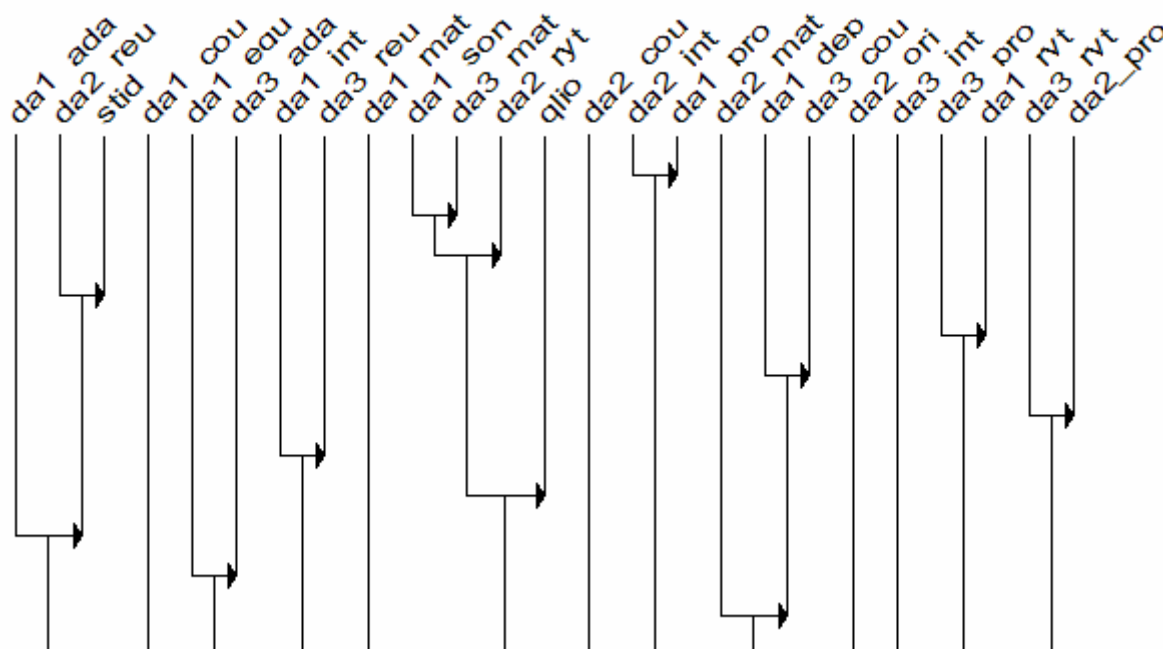
Tout d'abord, on remarque que la modalité « QLIO » est très proche de la modalité « da2_ori ». En fait, l'orientation des QLIO est leur principale préoccupation. Celle-ci se rapproche beaucoup à des difficultés dans le rythme d'alternance et la prospection téléphonique. Cela se voit dans le fait que le nœud entre la relation précédente nous amène à une relation entre les modalités « da1_rvt » et « da3_pro ». On peut noter aussi que les personnes, ayant senti toutes les difficultés énoncées, ont du mal à s'intégrer, à s'adapter, à

suivre les cours et à travailler en équipe. En effet, le nœud regroupement toutes les modalités précédentes se rattachent à un autre nœud regroupant les modalités « da1_cou », « da3_int », « da1_equ » « da3_ada ». Les modalités « da1_rvt » et « da3_pro » sont une fois de plus représentées.

Si l'on regarde les réponses des STID, on remarque que leur principale difficulté est le fait de suivre en cours. On voit effectivement que la modalité « STID » est similaire à la modalité « da1_ada ». Cette dernière implique d'ailleurs le fait d'avoir du mal à s'adapter, réussir, s'intégrer et prospecter. Ceci se voit dans le fait que le nœud regroupant les modalités « STID » et « da1_ada » se rattache à un autre nœud possédant les quatre difficultés citées précédemment.

On peut voir enfin une classe de difficultés qui ne sont pas forcément rattaché aux départements mais qui permettent de montrer des liens. En effet, on voit que les difficultés de se déplacer, de suivre les cours, d'être mature, et de suivre le rythme sont assez liées. Ceci est matérialisé par le dernier groupe à droite du graphique.

- *Arbre cohésitif*



En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.

Interprétation de quelques niveaux :

- Niveau 1** « da2_int » implique « da1_pro » : l'intégration est une difficulté majeure pour la prospection
- Niveau 4** « da2_reu » implique « stid » : la réussite est une difficulté pour les étudiants STID
- Niveau 5** « da3_pro » implique « da1_ryt » : la prospection est une difficulté pour suivre le rythme
- Niveau 9** les niveaux 2 et 3 implique « qlio » : le sondage, la maturité et le rythme sont des difficultés pour les étudiants QLIO

Ce graphique traite des difficultés majeures que les étudiants, STID et QLIO pensent devoir affronter pour répondre aux exigences du dispositif pédagogique fondé sur l'alternance. Avant tout commentaire, rappelons que les étudiants interrogés sont en première année et qu'ils n'ont pas encore goûté au plaisir d'être en alternance.

Tout d'abord on constate que les individus qui ont mis en deuxième difficulté l'intégration dans l'entreprise sont des personnes qui ont mis en premier lieu la prospection. En effet ces personnes là comme nous l'avons dit ci-dessus sont des étudiants de première année et auront un stage deux mois après suivant le questionnaire que nous leur avons fait passer. Ce qui signifie que leur préoccupation actuelle était la recherche d'entreprise.

On remarque que les enquêtes sont jugées comme une difficulté majeure pour l'alternance de la part de certaines personnes et que ces derniers sont des personnes qui estiment que la maturité est une difficulté capitale pour affronter l'alternance. Ces personnes là apeurées par les enquêtes et la maturité qu'ils nécessitent pour réussir au mieux l'alternance, sont inquiets aussi à propos du rythme de l'alternance auquel il faudra effectivement s'habituer. Enfin ces individus préoccupés par les 3 difficultés majeures que

l'on vient de citer (enquêtes, maturité et rythme) sont des étudiants issus de la formation QLIO.

Ensuite, on observe que les personnes s'inquiétant de la réussite pour l'alternance sont les STID. Et les individus trouvant que l'adaptation est une des difficultés essentielles à affronter l'alternance sont les personnes que l'on vient de citer ci-dessus, c'est-à-dire les STID s'inquiétant de leur réussite.

On peut constater sur la partie droite du graphique deux implications qui se ressemblent. En effet les gens qui ont mis en troisième difficulté la prospection sont des personnes s'inquiétant en premier lieu du rythme de l'alternance. Et l'autre implication montre que les étudiants s'inquiétant en troisième choix du rythme soutenu de l'alternance sont des personnes s'inquiétant en deuxième lieu de la prospection. On pourrait presque sous entendre une double implication pour ces deux termes bien qu'ils n'aient pas la même place dans l'ordre des difficultés à affronter.

5.2.2.15. Les étudiants et les difficultés majeures dans formation en statistique

V15. Quelles sont les trois difficultés majeures que vous pensez devoir affronter pour répondre aux exigences dans la formation en statistique dans le cadre du dispositif pédagogique fondé sur l'alternance ?

A cette question, les étudiants devaient répondre de la façon suivante :

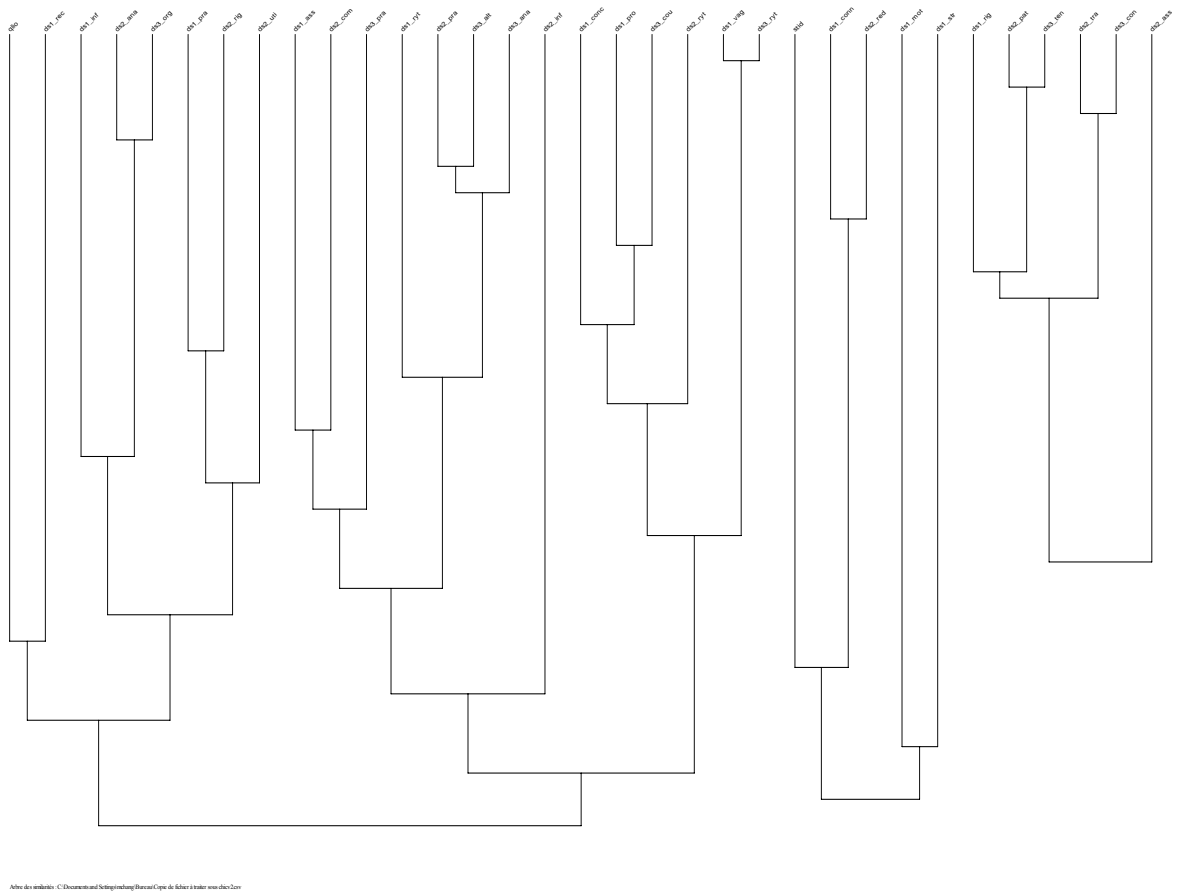
Diff_4 :

Diff_5 :

Diff_6 :

Nous avons remplacé « diff_4 », « diff_5 » et « diff_6 » par, respectivement, « diff_1 », « diff_2 » et « diff_3 »

- *Arbre de similarité*



En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.

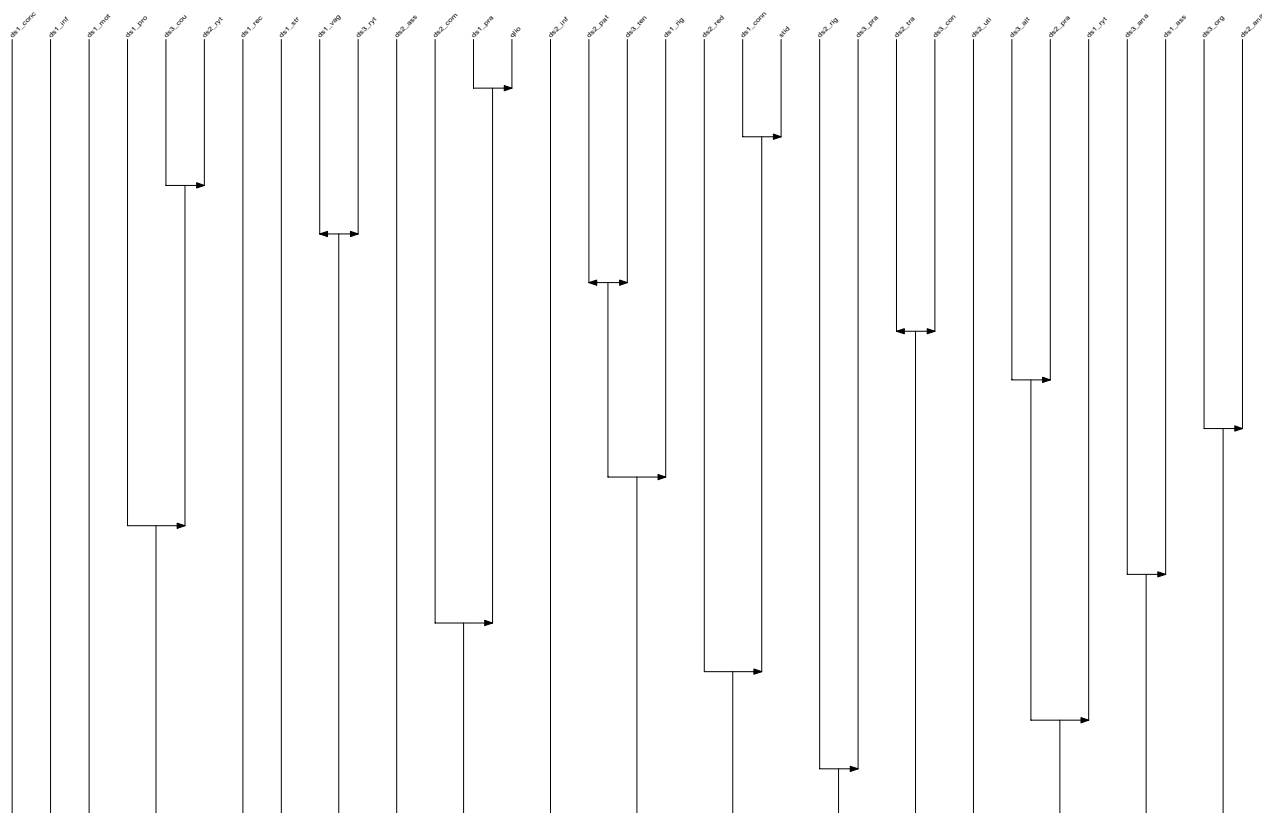
Dans ce graphique, beaucoup de relations sont présentes. D'abord, la modalité « QLIO » est proche de la modalité « ds1_rec ». Les QLIO ont donc du mal à prendre du recul en statistique. Si l'on regarde ensuite par niveau de significativité au niveau du groupe des QLIO, on constate d'abord une forte ressemblance entre le fait de trouver les statistiques vagues (modalité « ds1_vag ») et le fait de trouver le rythme des cours de statistique trop rapide (modalité « ds3_ryt »). La modalité « ds2_ana » et « ds3_org » sont liées très significativement également. Si l'on traduit, avoir des difficultés d'analyse est similaire au fait d'avoir des difficultés d'organiser. On voit dans une moindre mesure que les modalités « ds2_pra », « ds3_alt » et « ds3_ana » semblent similaires. Cela veut dire que les QLIO ont des difficultés de pratiquer, mettre en relation les statistiques et l'alternance, analyser cette matière, et que ces difficultés sont liées. Ceci est similaire aussi au fait d'avoir du mal à se

concentrer (modalité « ds1_conc »). Autre remarque, les modalités « ds1_conc », « ds1_pro » et « ds3_cou » font aussi partie d'un groupe. Autrement dit, le département « QLIO » éprouve des difficultés liées soient la concentration, la prospection et les cours. Ceci est aussi lié à la difficulté de suivre le rythme (modalité « ds2_ryt »).

En ce qui concerne les STID, on voit d'abord une liaison entre le fait d'avoir des difficultés de connaissances (« ds1_conn ») et le fait d'avoir du mal à rédiger (« ds2_red »), tout ceci en statistique toujours. Cette relation est liée au fait d'être STID.

On remarque enfin, un groupe qui ne concerne aucun département. On voit un nœud qui lie « ds2_pat » et « ds3_ten ». Ne pas avoir de patience est similaire à avoir du mal à être tenace. Ce nœud est lié à la modalité « ds1_rig » c'est-à-dire à la difficulté de rigueur. L'autre nœud plus loin nous informe que la difficulté de travailler est liée à la difficulté de concentration. Pour finir avec les relations les plus importantes du graphique, les trois relations précédentes sont liées avec la difficulté d'assimilation.

Arbre cohésitif



Atlas cobaltif: G:\Nouveaux dossiers\CHC\travaux\11_1120

En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.

Tout d'abord, nous pouvons remarquer que de considérer la pratique comme une difficulté majeure de la formation en statistique en alternance implique fortement d'être issu de la formation QLIO. Les étudiants QLIO semblent ainsi éprouver des difficultés à mettre en pratique les connaissances acquises à l'IUT afin de pouvoir les utiliser en entreprise.

Ensuite, nous constatons que les étudiants qui ont des difficultés au niveau de leurs connaissances sont les étudiants STID. Ces derniers, tout comme les QLIO, ne savent peut-être pas trop comment mettre en œuvre en entreprise les connaissances acquises à l'IUT. Ils ont aussi peut-être peur de ne pas savoir assez de choses afin de pouvoir s'en sortir en entreprise...

Les étudiants pour lesquels les cours en général constituent une difficulté sont également pour qui le rythme de travail est un problème. Cela signifie que la raison principale

des difficultés de suivi en cours des étudiants provient du rythme imposé, trop soutenu. Ce sont ces mêmes étudiants qui rencontrent des difficultés pour la prospection d'entreprises en vue du contrat en alternance.

De plus, on remarque une double implication : les étudiants qui trouvent le rythme soutenu et qui considèrent cela comme une difficulté trouvent également les cours trop vagues, et vice-versa. Cela peut s'expliquer par le fait qu'ils aient l'impression de ne pas approfondir suffisamment les cours, qu'ils restent dans le flou, à cause d'un rythme trop rapide.

On observe de la même façon une implication à double sens concernant le travail et la concentration perçus comme des difficultés. En effet, cela va de paire, il est nécessaire d'être concentré pour effectuer du bon travail.

On observe par la suite que les étudiants pour lesquels la patience constitue une difficulté dans la formation en statistique en alternance sont les mêmes que ceux pour qui la ténacité constitue un problème, et vice-versa. Cela est logique puisque les deux termes signifient globalement la même. Ce sont ces mêmes étudiants pour qui la rigueur est une difficulté.

On constate aussi qu'avoir des difficultés à s'organiser implique d'avoir des problèmes pour analyser. En effet, l'organisation est nécessaire avant de procéder à l'analyse : un manque de rigueur peut nuire à cette dernière.

Enfin, on peut remarquer que les étudiants qui rencontrent des problèmes en ce qui concerne l'alternance, considèrent également la pratique comme une difficulté. Cela paraît logique puisque l'alternance est en fait une mise en pratique pour l'entreprise de la théorie acquise à l'IUT. Ces étudiants considèrent également le rythme comme une difficulté.

5.2.2.16. Les étudiants et les apports majeurs dans l'alternance.

V16. Quels sont les trois apports majeurs que vous attendez de ce dispositif de pédagogie en alternance quant à la formation en statistique ?

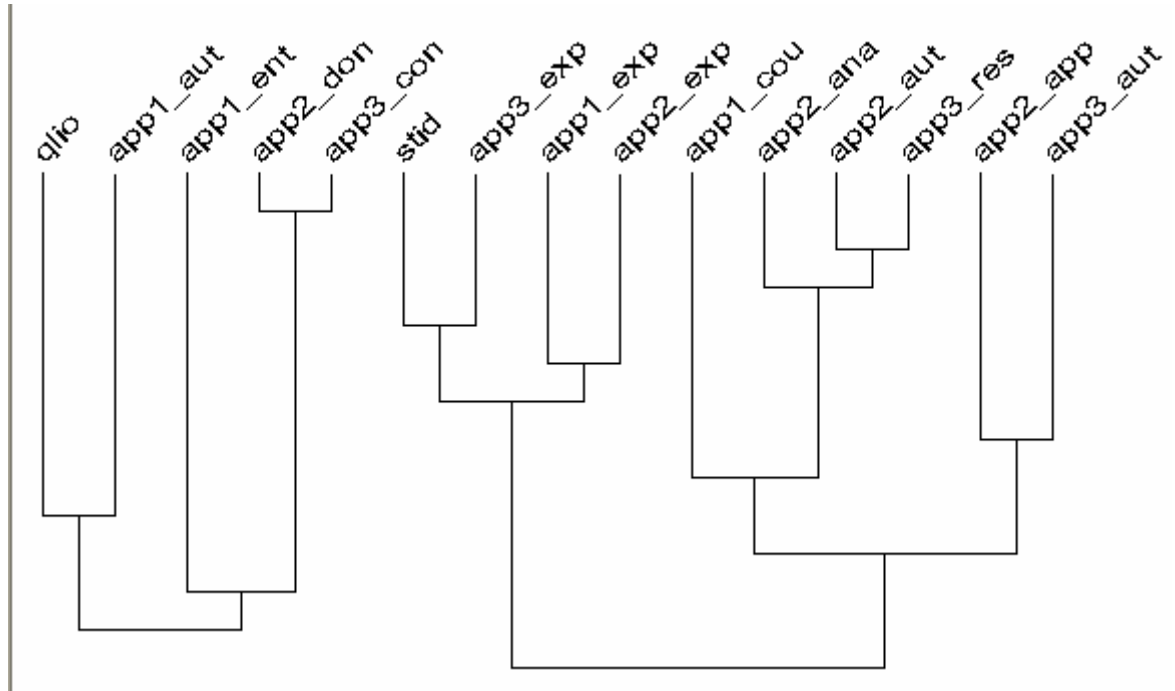
A cette question, les étudiants devaient répondre de la façon suivante :

App 1 :

App 2 :

App 3 :

- *Arbre de similarité*

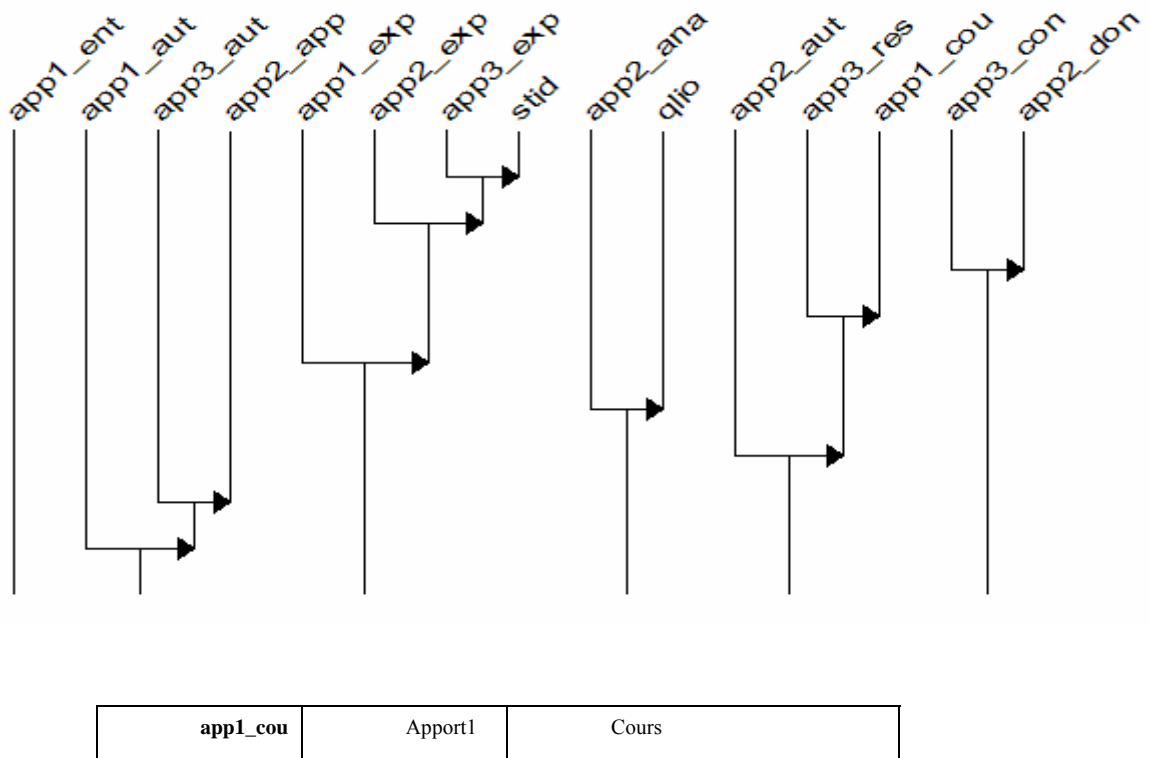


app1_cou	Apport1	Cours
app1_ent		Entreprise
app1_aut		Autre
app1_exp		Expérience
app2_app	apport2	Application
app2_ana		Analyser
app2_aut		Autre
app2_don		Donner
app2_exp		Expérience entreprise
app3_aut	apport3	Autre
app3_res		Résultats
app3_con		Concret
app3_exp		Expérience

Nous remarquons d'abord des liens entre le fait d'être en QLIO et le fait de ne pas attendre d'apports bien précis. En effet, la modalité « QLIO » est similaire à la modalité « app1_ent ». Ce nœud est similaire à un autre groupe de 3 modalités. On voit que la modalité « app2_don » et « app3_con » sont très liées et que la modalité « app1_ent » se rapproche à ces deux modalités. On peut donc dire que le département « QLIO » attend que l'on s'occupe plus de lui et qu'on lui donne plus de concret au niveau statistique. Ceci est aussi lié au fait qu'il attend plus de rapport avec l'entreprise.

Chez les STID, nous constatons qu'ils attendent principalement plus d'expériences grâce aux statistiques. La modalité « STID » est effectivement très proche de la modalité « app3_exp », « app1_exp » et « app2_exp ». Nous pouvons voir enfin un dernier groupe. Dans celui-ci, nous constatons à des niveaux de significativité élevé que « app2_ana » ressemble à « app2_aut » qui elle est liée à « app3_res ». Ceci est aussi attaché à la modalité « app1_cou » mais moins significativement. Autrement dit, les STID veulent avoir de meilleurs résultats, plus d'analyse en cours et même plus de cours mais de façon moins importante. A cela s'ajoutent d'autres attentes diverses.

- **Arbre cohésitif**



tel-00191166, version 1 - 27 Nov 2007

app1_ent		Entreprise
app1_aut		Autre
app1_exp		Expérience
app2_app	apport2	Application
app2_ana		Analyser
app2_aut		Autre
app2_don		Donner
app2_exp		Expérience entreprise
app3_aut	apport3	Autre
app3_res		Résultats
app3_con		Concret
app3_exp		Expérience

Interprétation de quelques niveaux :

Niveau 1 « app3_exp » implique « stid » : l'expérience est un apport majeur pour les étudiants STID

Niveau 2 « app2_exp » implique le niveau 1 : l'expérience entreprise est un apport majeur pour les étudiants STID

Niveau 4 « app3_res » implique « app1_cou » : apport de résultats dans les cours

Niveau 6 « app2_ana » implique « qlio » : l'analyse est un apport majeur pour les étudiants QLIO

Les étudiants qui pensent que l'expérience est un des apports majeurs fournis par la formation en alternance sont les STID. En effet, ces derniers peuvent, grâce à l'alternance, appliquer leurs connaissances statistiques théoriques acquises à l'IUT et ainsi acquérir un maximum d'expérience dans ce domaine.

Le fait d'avoir cité comme apport majeur l'apprentissage de l'analyse implique que l'on ait suivi une formation QLIO. Ceci peut s'expliquer par le fait que dans la formation QLIO la partie analyse semble occuper une place très importante.

5.2.2.17. Les étudiants et les suggestions majeures

V17. Quelles sont les trois suggestions majeures que vous feriez a priori pour faciliter l'apprentissage de la statistique dans ce dispositif pédagogique en alternance ?

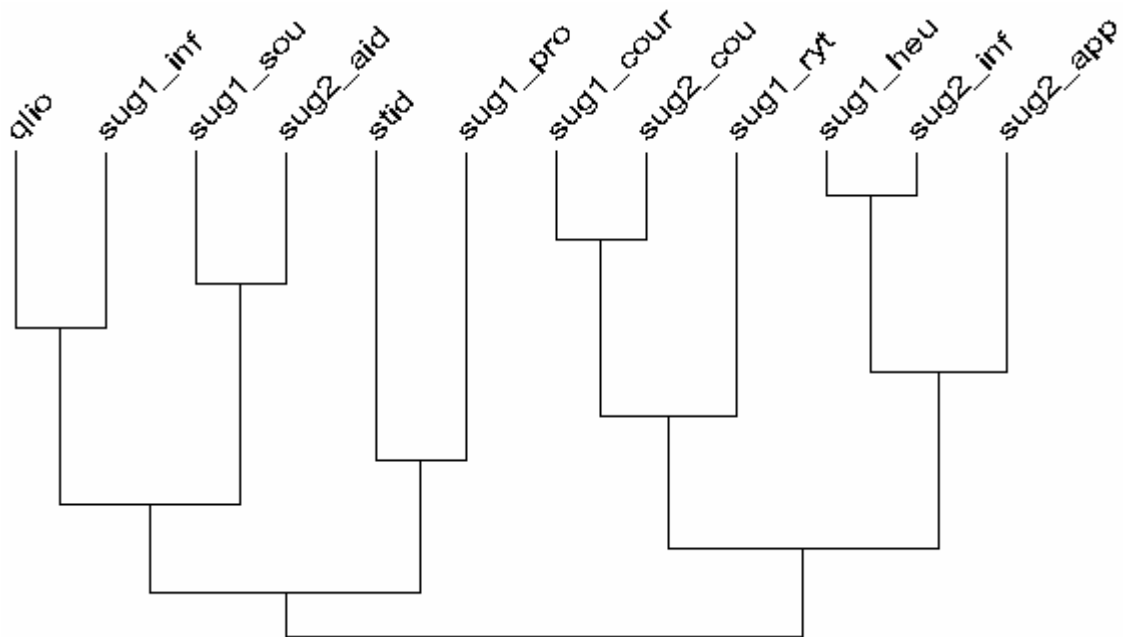
A cette question, les étudiants devaient répondre de la façon suivante :

Sugg_1 :

Sugg_2 :

Sugg_3 :

• *Arbre de similarités*



sug1_cour	Suggestion1	Plus de cours en groupes
sug1_ryt		Ralentir le rythme
sug1_pro		Plus de projets
sug1_inf		Plus d'informatique
sug1_sou		Cours de soutien
sug1_heu		Plus d'heures de cours

sug2_app	Suggestion2	Application concrète
sug2_cou		Cours plus précis
sug2_inf		Informatique
sug2_aid		Plus d'aide, plus de TD

D'après cet arbre de similarités, on constate un grand groupe de modalités. En effet, on ne distingue pas une séparation QLIO/STID.

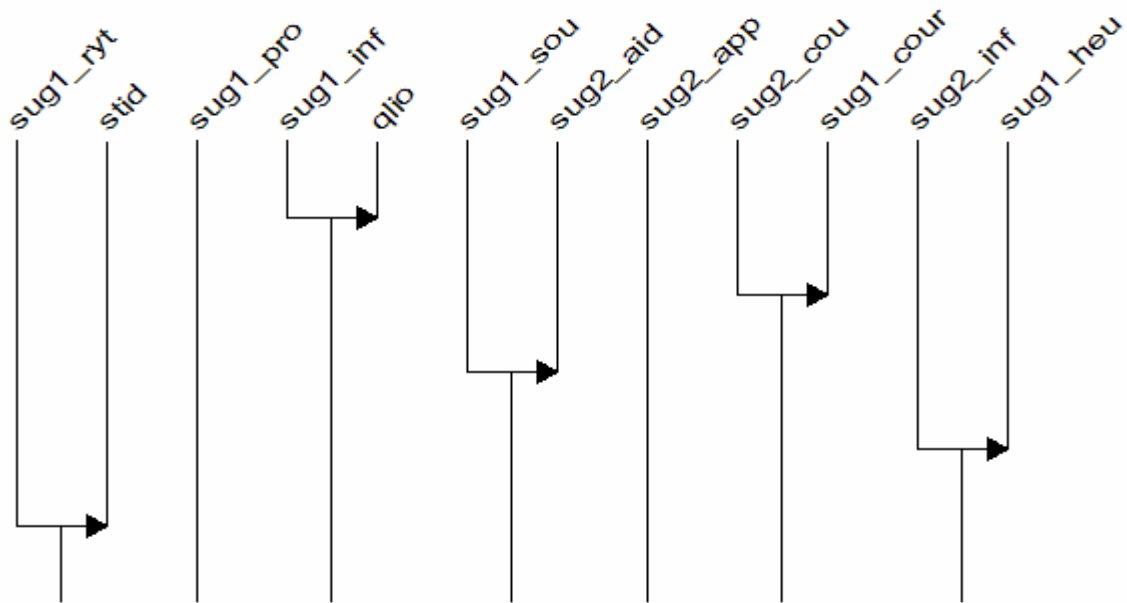
Pour les étudiants QLIO, la suggestion majeure pour faciliter l'apprentissage de la statistique dans le dispositif de l'alternance serait d'avoir plus d'informatique (la modalité « qlio » est très similaire à « sug1_inf »). Mais de façon moins significative, ils suggéreraient également des cours de soutien et plus d'aide et de TD (les modalités « sug1_sou » et « sug2_aid » sont associées).

Pour les étudiants STID, la suggestion majeure serait de réaliser plus de projets (en effet la modalité « stid » est très proche de « sug1_pro »). Cependant, on peut voir que leurs autres suggestions rejoignent ceux des étudiants QLIO.

On peut remarquer un groupe composé uniquement de modalités où n'intervient pas les modalités « qlio » et « stid » : ces suggestions sont ensuite reliées à l'autre groupe qui contiennent les modalités « qlio » et « stid ».

De manière générale, pour les étudiants des deux promos, le fait d'avoir plus de cours en groupes (« sug1_cour ») et des cours plus précis (« sug2_cou ») faciliteraient l'apprentissage de la statistique. Egalement, il faudrait ralentir le rythme (« sug1_ryt »). De plus, il faudrait plus d'heures de cours (« sug1_heu ») afin d'utiliser plus l'informatique (« sug2_app ») pour des applications concrètes (« sug2_app »).

- *Arbre cohésitif*



sug1_cour	Suggestion1	Plus de cours en groupes
sug1_ryt		Ralentir le rythme
sug1_pro		Plus de projets
sug1_inf		Plus d'informatique
sug1_sou		Cours de soutien
sug1_heu		Plus d'heures de cours
sug2_app	Suggestion2	Application concrète
sug2_cou		Cours plus précis
sug2_inf		Informatique
sug2_aid		Plus d'aide, plus de TD

Niveau 1 « sug1_pro » implique « qlio » : plus de projets pour faciliter l'apprentissage de la statistique sont suggérés par les étudiants QLIO

Niveau 2 « sug2_cou » implique « sug1_cour » : des cours plus précis impliquent plus de cours en groupes

Niveau 3 « sug1_sou » implique « sug2_aid » : des cours de soutien entraîne plus d'aide, plus de TD

Niveau 4 « sug2_inf » implique « sug1_heu » : l'informatique pour faciliter l'apprentissage de la statistique demande plus d'heures de cours

Niveau 5 « sug1_ryt » implique « stid » : ralentir le rythme est suggéré par les étudiants STID

Les étudiants qui suggèrent une formation davantage axée sur l'informatique sont des étudiants QLIO. En effet, la formation QLIO comptabilise moins d'heures d'informatique que la formation STID.

De plus, on remarque que les étudiants qui souhaiteraient avoir des cours de soutien sont également ceux qui suggèrent un nombre d'heures de TD plus important. Cela est cohérent : ces personnes souhaitent être plus encadrés et faire ainsi plus d'exercices d'application.

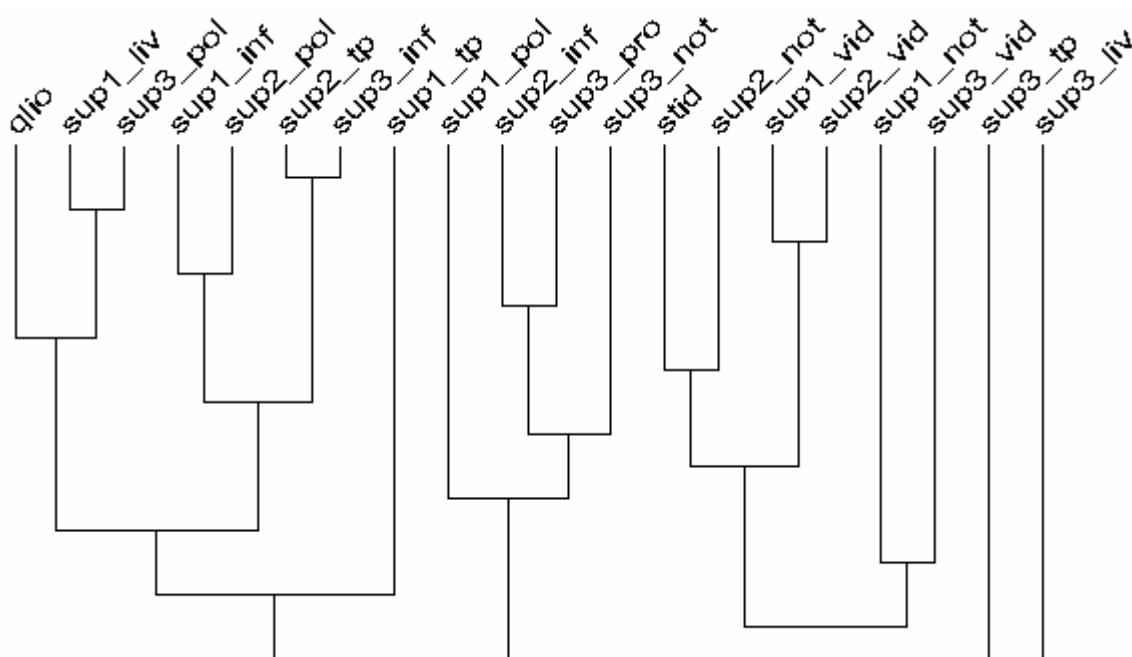
Enfin ceux qui suggèrent de ralentir le rythme sont les étudiants STID : cela laisse penser que les STID suivent une formation plus soutenue.

5.2.2.18. Les étudiants et les supports de cours

V18. De quels supports de ce cours disposez-vous ?

Pour traiter cette question, nous avons pris les trois premiers mots qui revenaient souvent.

- *Arbre de similarités*



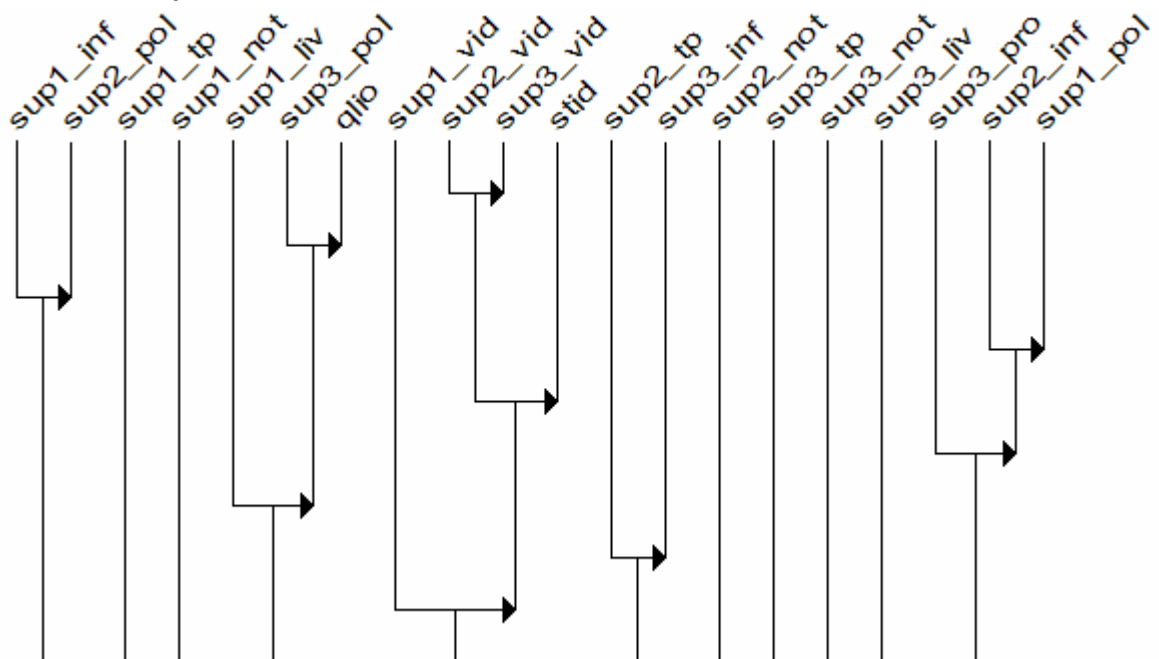
sup1_pol	support 1	Polycopiés
sup1_inf		Informatique
sup1_tp		TP
sup1_not		Notes personnelles
sup1_liv		Livre
sup1_vid		Vide
sup2_pol	support 2	Polycopiés
sup2_inf		Informatique
sup2_tp		TP
sup2_not		Notes personnelles
sup2_vid		Vide
sup3_pol	support 3	Polycopiés
sup3_inf		Informatique
sup3_tp		TP
sup3_not		Notes personnelles
sup3_liv		Livre
sup3_pro		Prof
sup3_vid		Vide

D'après cet arbre de similarité, les supports de cours cités par les étudiants QLIO sont : « livre », « poly », « informatique », « tps », « notes personnelles » et « prof ». Tandis que les supports de cours cités par les étudiants STID sont : « notes personnelles », « TPs », et « livres ». A noter qu'ils sont moins nombreux que ceux cités par les étudiants QLIO : on peut alors déjà dire que les étudiants STID utilisent moins de supports de cours que les étudiants QLIO.

De façon plus détaillé, on constate que la modalité « sup1_liv » est similaire à « sup3_pol », qui eux-mêmes sont similaires à « QLIO » : les étudiants QLIO utilisent plus le livres et le photocopié comme supports de cours.

Pour les étudiants STID, on constate que seulement « sup2_not » est similaire à « STID » ce qui veut dire que seuls les notes personnelles sont utilisées comme supports de cours.

- *Arbre cohésitif*



sup1_pol	support 1	Photocopiés
sup1_inf		Informatique
sup1_tp		TP
sup1_not		Notes personnelles
sup1_liv		Livre

sup1_vid		Vide
sup2_pol	support 2	Polycopiés
sup2_inf		Informatique
sup2_tp		TP
sup2_not		Notes personnelles
sup2_vid		Vide
sup3_pol	support 3	Polycopiés
sup3_inf		Informatique
sup3_tp		TP
sup3_not		Notes personnelles
sup3_liv		Livre
sup3_pro		Prof
sup3_vid		Vide

Interprétation de quelques niveaux :

Niveau 2 « sup3_pol » implique « qlio » : ceux qui disposent de polycopiés comme support sont les étudiants QLIO.

Niveau 3 « sup1_inf » implique « sup2_pol » : le support informatique (1^{er} support) entraîne l'utilisation des polycopiés comme second support.

Niveau 4 « sup2_inf » implique « sup1_pol » : le support informatique (2^{ème} support) entraîne l'utilisation des polycopiés comme second support.

Niveau 5 le niveau 1 implique « stid » : ceux qui n'ont pas répondu à la question sont les étudiants STID.

On remarque, en premier lieu, que les étudiants n'utilisant pas de supports de cours sont issus de la formation STID. On peut alors supposer que leurs cours sont suffisamment complets et qu'ils s'en contentent.

De plus, utiliser des livres ou des polycopiés comme supports de cours implique de suivre la formation QLIO. Ces derniers passent ainsi plus de temps à compléter leurs cours par d'autres sources.

Enfin, on observe que les étudiants qui utilisent des supports informatiques sont également ceux qui se servent des photocopies.

5.2.2.19. Les étudiants et les ressources en plus

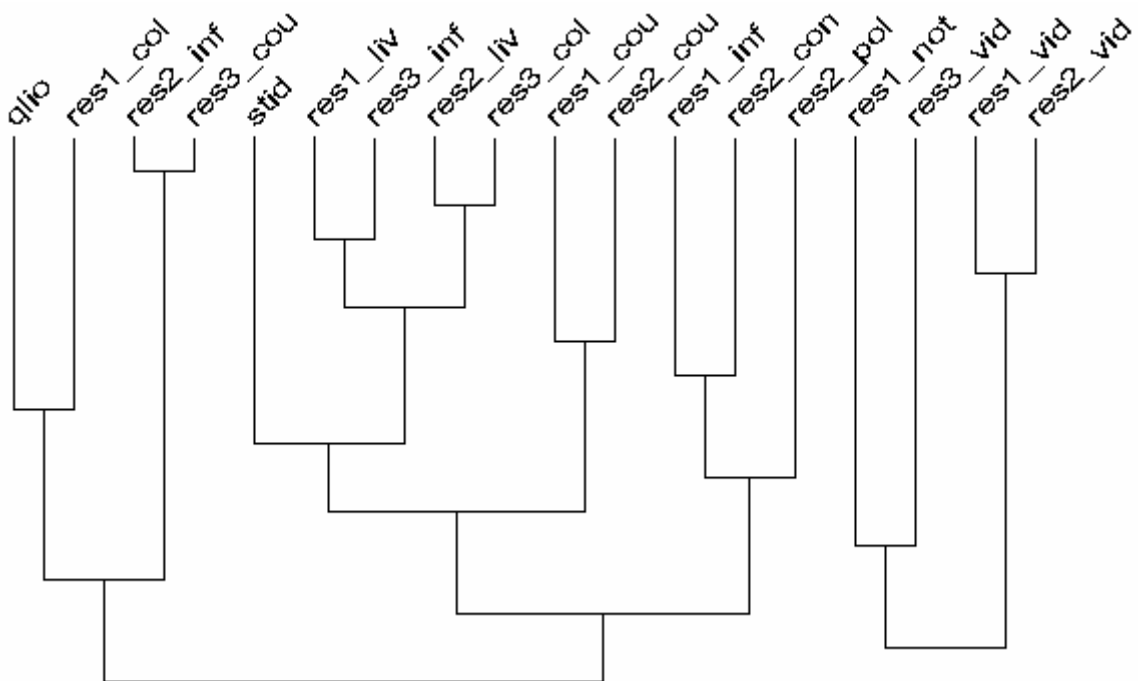
V22. Quelles ressources avez-vous utilisées en plus des documents de cours distribués ?

A cette question, les étudiants devaient répondre de la façon suivante :

(Citer au plus 4)

Pour traiter cette question, nous avons choisi de ne prendre que les trois premières réponses.

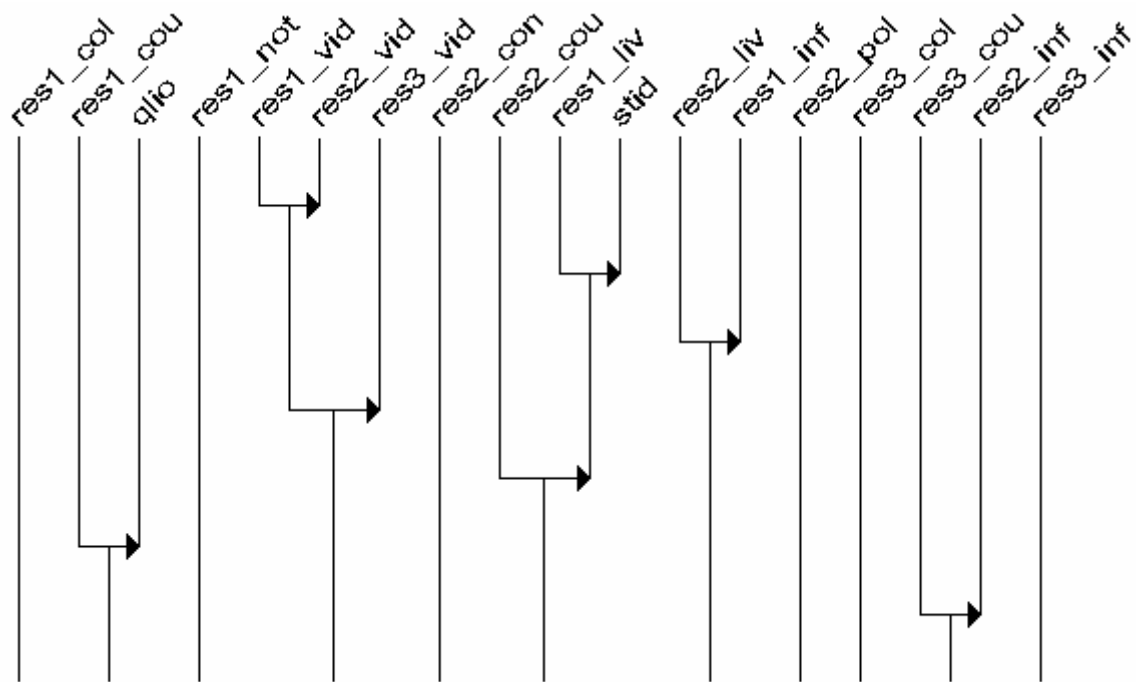
- *Arbre de similarités*



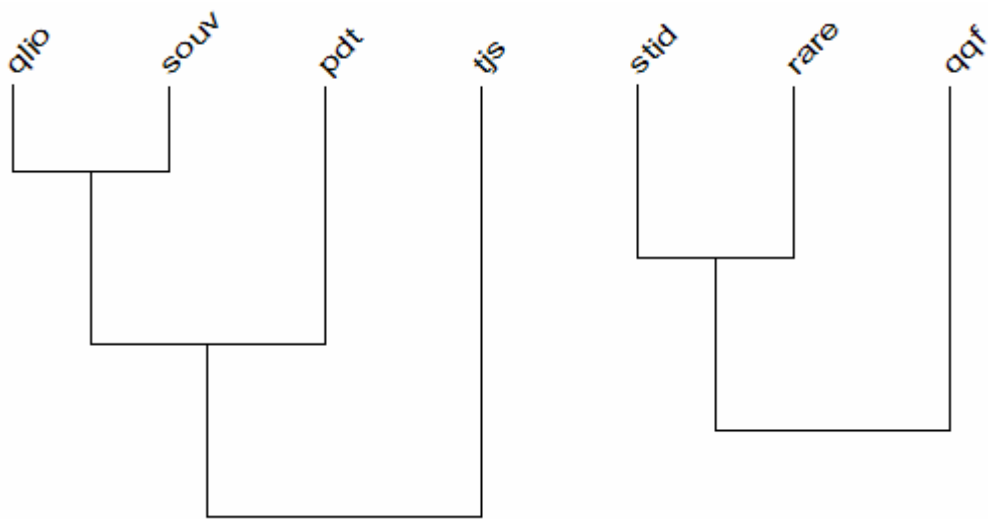
	ressource	
res1_col	1	Collègues
res1_cou		Cours
res1_inf		Informatique
res1_liv		Livres
res1_not		Notes personnelles
res1_vid		Vide
	ressource	
res2_con	2	Connaissances personnelles
res2_cou		Cours
res2_inf		Informatique
res2_liv		Livres
res2_pol		Poly
res2_vid		Vide
	ressource	
res3_col	3	Collègues
res3_cou		Cours
res3_inf		Informatique
res3_vid		Vide

On constate que la modalité « qlio » est très proche de « res1_col » ce qui signifie que les étudiants QLIO, en plus des documents de cours demandent de l'aide à leur collègue. De plus, de façon moins significative, ils utilisent l'outil informatique (« res2_inf ») et des cours (« res3_cou »). Au contraire, en plus de leurs cours, les étudiants STID utilisent plusieurs ressources principalement les livres (« res1_liv »), l'informatique (« res3_inf ») ou demandent l'aide à leur collègue (« res3_col »). Mais aussi d'autres cours (manuels scolaires), leur photocopié (« res2_pol ») ou les connaissances personnelles (« res2_con »).

- *Arbre cohésitif*



res1_col	1	ressource	Collègues
res1_cou			Cours
res1_inf			Informatique
res1_liv			Livres
res1_not			Notes personnelles
res1_vid			Vide
res2_con	2	ressource	Connaissances personnelles
res2_cou			Cours
res2_inf			Informatique
res2_liv			Livres
res2_pol			Poly
res2_vid			Vide
res3_col	3	ressource	Collègues
res3_cou			Cours
res3_inf			Informatique
res3_vid			Vide

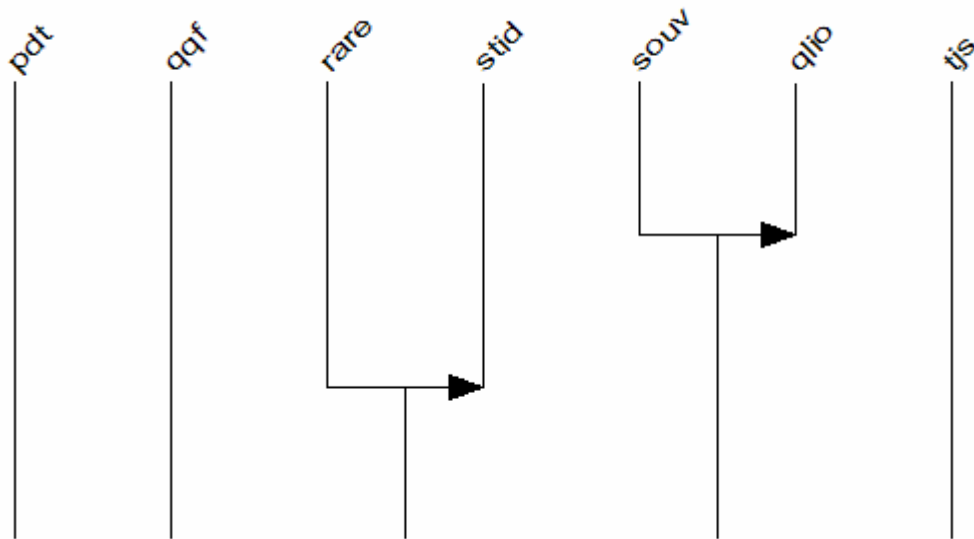


Pdt	Pas du tout
Qqf	Quelquefois
Rare	Rarement
Souv	Souvent
Tjs	Toujours

Pour les deux promotions, on remarque que la nécessité peut être très variable : en effet, pour les étudiants QLIO, ceux-ci vont souvent chercher d'autres ressources (« qlio » similaire à « souv »). Cependant, parfois ils n'ont pas besoin d'aller en chercher (« pdt ») où vont tout le temps en chercher (« tjs »).

Par contre, pour les étudiants STID, il est moins utile d'aller chercher d'autres ressources : le plus souvent, ils vont rarement en chercher mais vont en chercher quelquefois.

- *Arbre cohésitif*



Pdt	Pas du tout
Qqf	Quelquefois
Rare	Rarement
Souv	Souvent
Tjs	Toujours

Niveau 1 « rare » implique « stid » : ceux qui vont rarement chercher d'autres ressources sont les étudiants STID

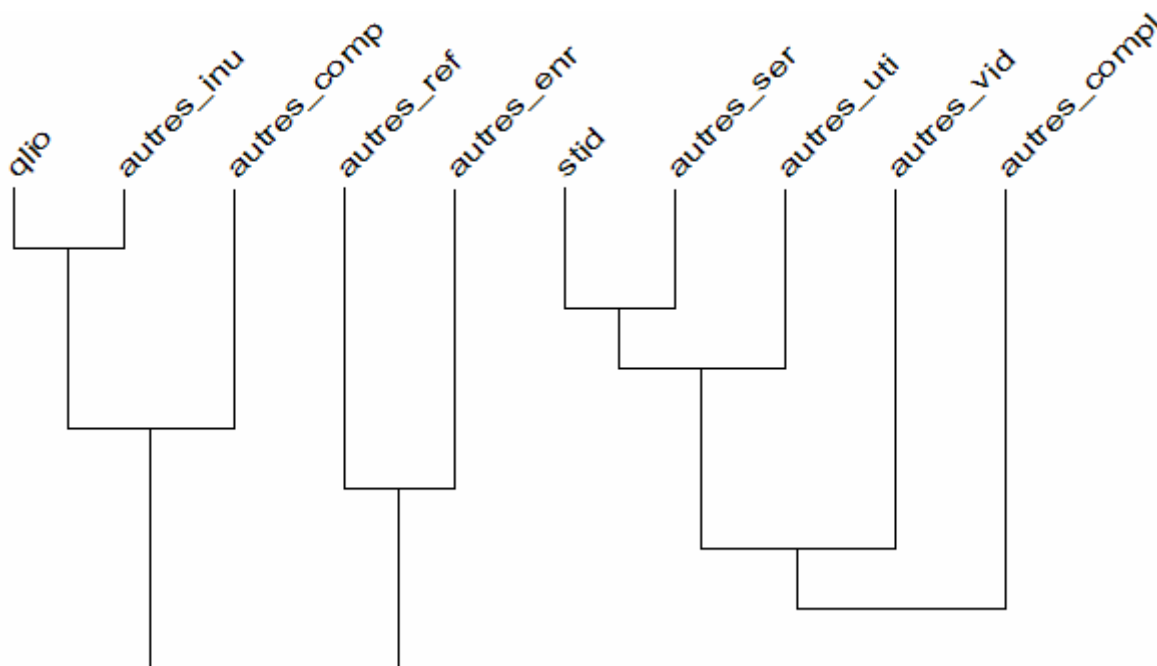
Niveau 2 « souv » implique « qlio » : ceux qui vont souvent chercher d'autres ressources sont les étudiants QLIO.

Concernant la nécessité d'aller chercher d'autres ressources que les documents de cours distribués, on constate que les étudiants QLIO vont souvent en chercher et que les étudiants STID vont rarement en chercher. Cela peut s'expliquer par le nombre d'heures de statistiques qui diffère selon la promotion. En effet, les étudiants STID ayant le plus d'heures de statistiques, ceux-ci n'ont pas vraiment besoin d'aller en chercher.

5.2.2.21. Les étudiants et les autres ressources

V24. De manière générale pour le cours de statistique, que pensez-vous de la nécessité de compléter les cours par d'autres ressources ?

- *Arbre de similarité*

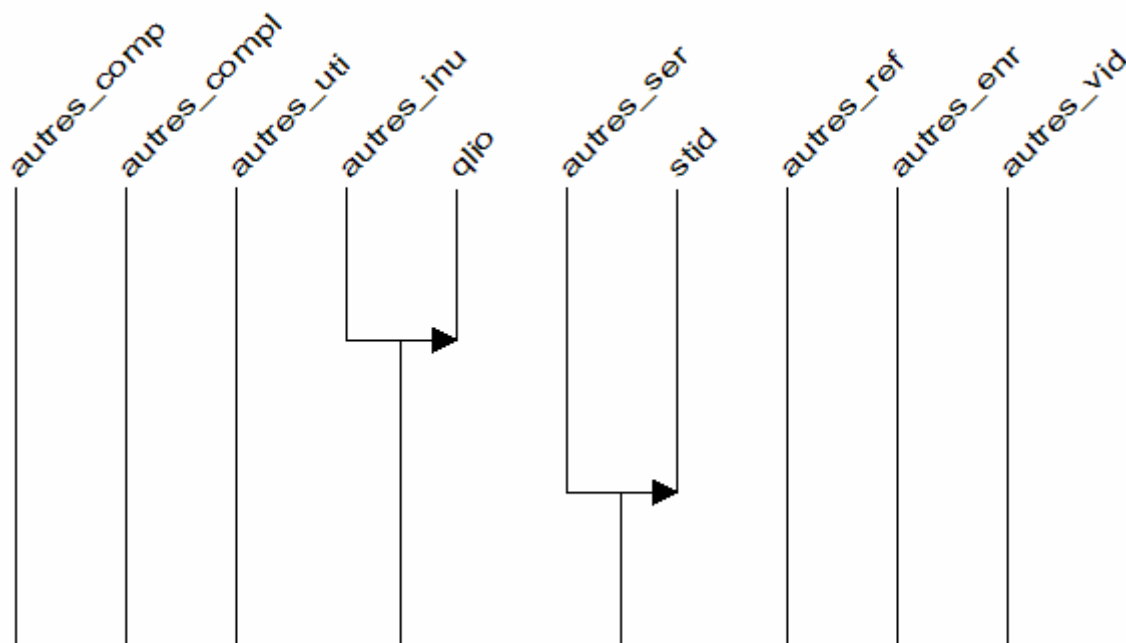


autres_comp	Compréhension
autres_compl	Complément
autres_uti	Utile
autres_inu	Inutile
autres_ser	Peut servir
autres_ref	Reformulation
autres_enr	Enrichissement
autres_vid	Vide

Dans ces des groupes de modalités, on remarque que les avis sont très partagés. En ce qui concerne les étudiants QLIO, ceux-ci pensent principalement qu'il est inutile de compléter les cours par d'autres ressources (« qlio » proche de « autres_inu »). D'autres pensent qu'il est utile pour la compréhension (« autres_comp ») mais aussi pour une reformulation des cours et pour être enrichi.

Les étudiants STID pensent que cela peut servir (« autres_ser ») (mais pas forcément nécessaire). D'autres pensent que c'est utile (« autres_uti ») et que cela peut être un complément (« autres_compl »).

- *Arbre cohésitif*



autres_comp	Compréhension
autres_compl	Complément
autres_uti	Utile
autres_inu	Inutile
autres_ser	Peut servir
autres_ref	Reformulation
autres_enr	Enrichissement
autres_vid	Vide

Niveau 1 « autres_inu » implique « qlio » : il est inutile pour les étudiants QLIO de compléter les cours par d'autres ressources

Niveau 2 « autres_ser » implique « stid » : compléter les cours peut être utile pour les étudiants STID

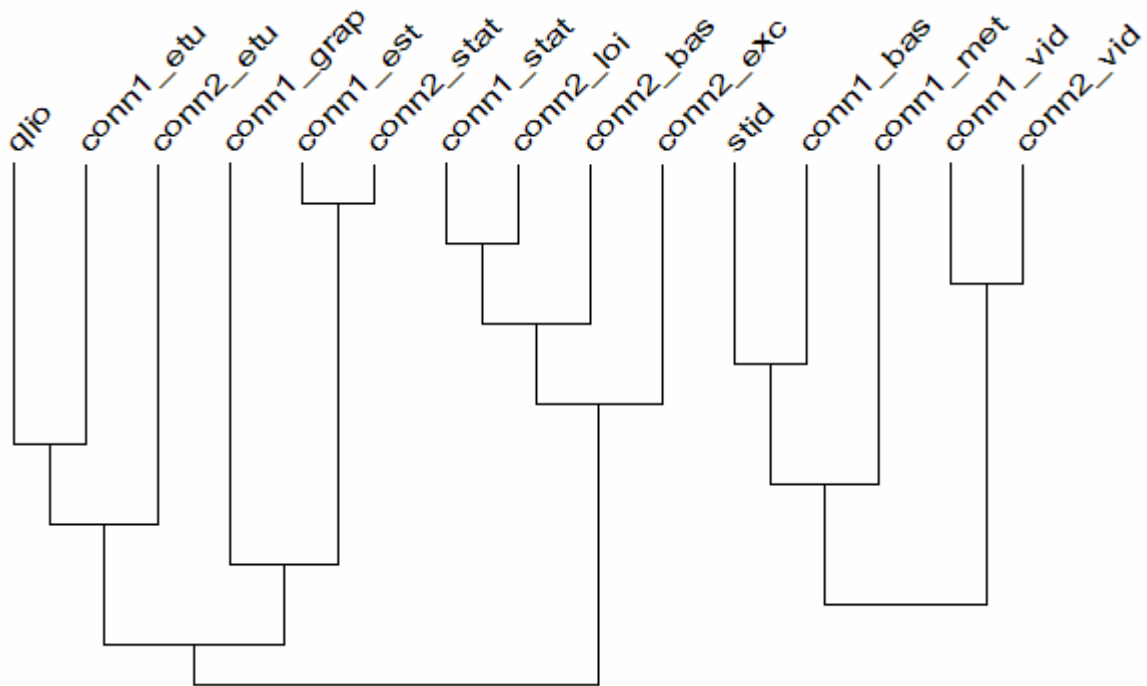
Concernant la nécessité de compléter les cours, les étudiants QLIO pensent qu'il est inutile tandis que les étudiants STID pensent que cela peut servir.

5.2.2.22. Les étudiants et la formation en statistique

V25.A ce jour que pensez-vous avoir appris en ce qui concerne le domaine de la statistique ?

Pour traiter cette question, nous avons pris les deux premières réponses qui revenaient le plus souvent.

- *Arbre de similarité*



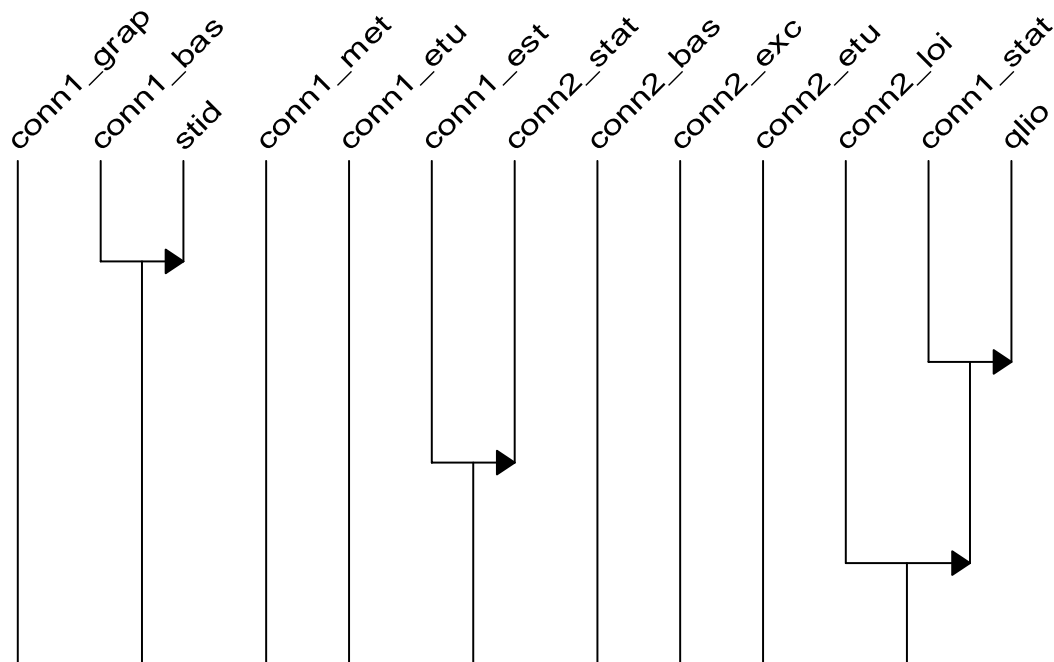
conn1_stat	connaissance 1	Statistique ciblée
conn1_grap		Graphiques
conn1_bas		Bases
conn1_met		Ensemble de méthodes et d'outils
conn1_etu		Etudes
conn1_est		Estimation
conn1_vid		Vide
conn2_bas	connaissance 2	Bases
conn2_exc		Excel
conn2_etu		Etudes

conn2_stat		Statistique ciblée
conn2_loi		Lois
conn2_vid		Vide

En ce qui concerne le domaine de la statistique, les étudiants QLIO ont appris que cela peut servir pour réaliser des études (« conn1_etu »). Egalement que l'on pouvait illustrer les résultats à l'aide de graphiques, faire de l'estimation, que l'on pouvait faire des statistiques ciblées, connaître des lois. Faire de la statistique en utilisant Excel.

Les étudiants STID, quant à eux, ont appris les bases de la statistique et les méthodes et outils nécessaires.

- *Arbre cohésitif*



conn1_stat	connaissance 1	Statistique ciblée
conn1_grap		Graphiques
conn1_bas		Bases
conn1_met		Ensemble de méthodes et d'outils
conn1_etu		Etudes
conn1_est		Estimation
conn1_vid		Vide
conn2_bas	connaissance 2	Bases
conn2_exc		Excel
conn2_etu		Etudes
conn2_stat		Statistique ciblée
conn2_loi		Lois
conn2_vid		Vide

Niveau 1 « conn1_bas » implique « stid » : ceux qui ont appris les bases sont les étudiants STID

Niveau 2 « conn1_stat » implique « qlio » : ceux qui ont appris une statistique ciblée sont les étudiants QLIO

Niveau 3 « conn1_est » implique « conn2_bas » : l'estimation implique une statistique ciblée

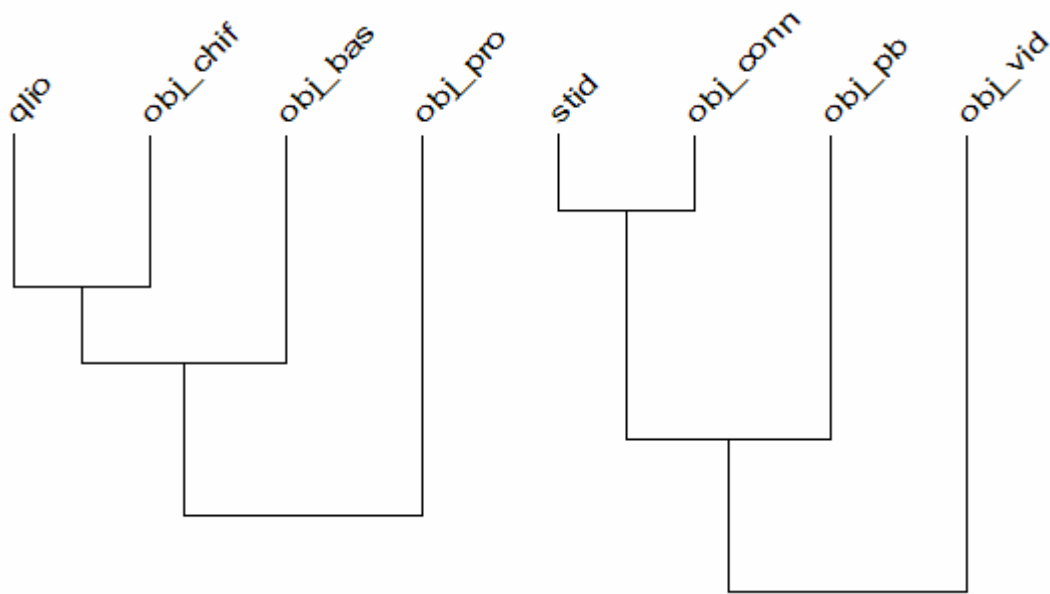
Niveau 4 « conn2_loi » implique niveau 2 : la connaissance de lois entraîne que ce soit une statistique ciblée utilisée par les étudiants QLIO.

A la question « à ce jour que pensez-vous avoir appris en ce qui concerne le domaine de la statistique ? », ceux qui ont répondu avoir appris les bases sont des STID (ce qui est normal en 1^{ère} année). De plus, ceux qui répondent avoir appris « une statistique ciblée » sont les étudiants QLIO.

5.2.2.23. Les étudiants et leur objectif

V27. Quel objectif principal formuleriez-vous pour une formation en statistique à l'IUT ?

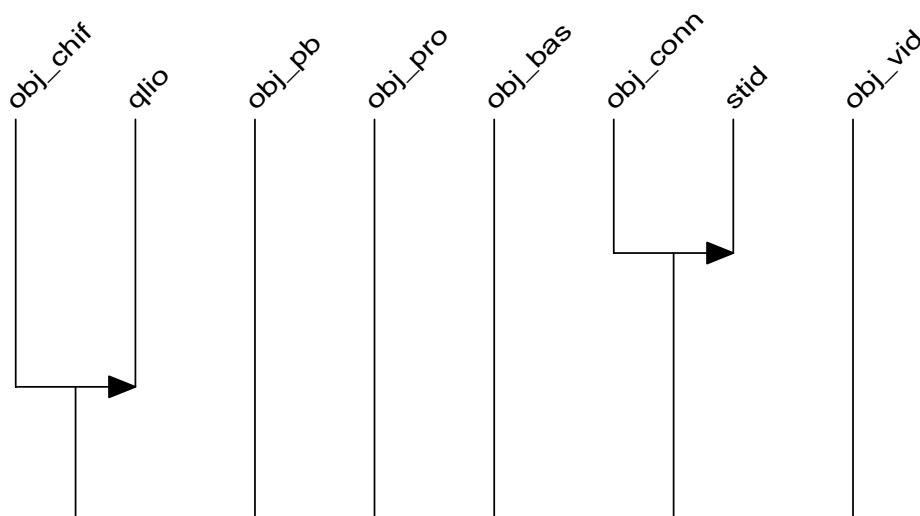
- *Arbre de similarité*



obj_chif	Utilisation des chiffres
obj_pb	Réponse à une problématique
obj_pro	Nécessité monde professionnel
obj_bas	Connaître les bases
obj_conn	Maximum de connaissances statistiques
obj_vid	Vide

On constate deux classes de modalités bien distinctes :

- « qlio » est associé à « obj_chif », « obj_bas » et « obj_pro » : ce qui montre que les étudiants QLIO ont d'abord pour objectif l'utilisation des chiffres, puis connaître les bases et enfin la statistique est une nécessité pour le monde professionnel.
- « stid » est associé à « obj_conn », « obj_pb » et « obj_vid » : ceci signifie que les étudiants STID ont pour principal objectif d'acquérir un maximum de connaissances en statistiques puis en second une formation en statistique a pour objectif de répondre à une problématique.
- **Arbre cohésitif**



obj_chif	Utilisation des chiffres
obj_pb	Réponse à une problématique
obj_pro	Nécessité monde professionnel
obj_bas	Connaître les bases
obj_conn	Maximum de connaissances statistiques
obj_vid	Vide

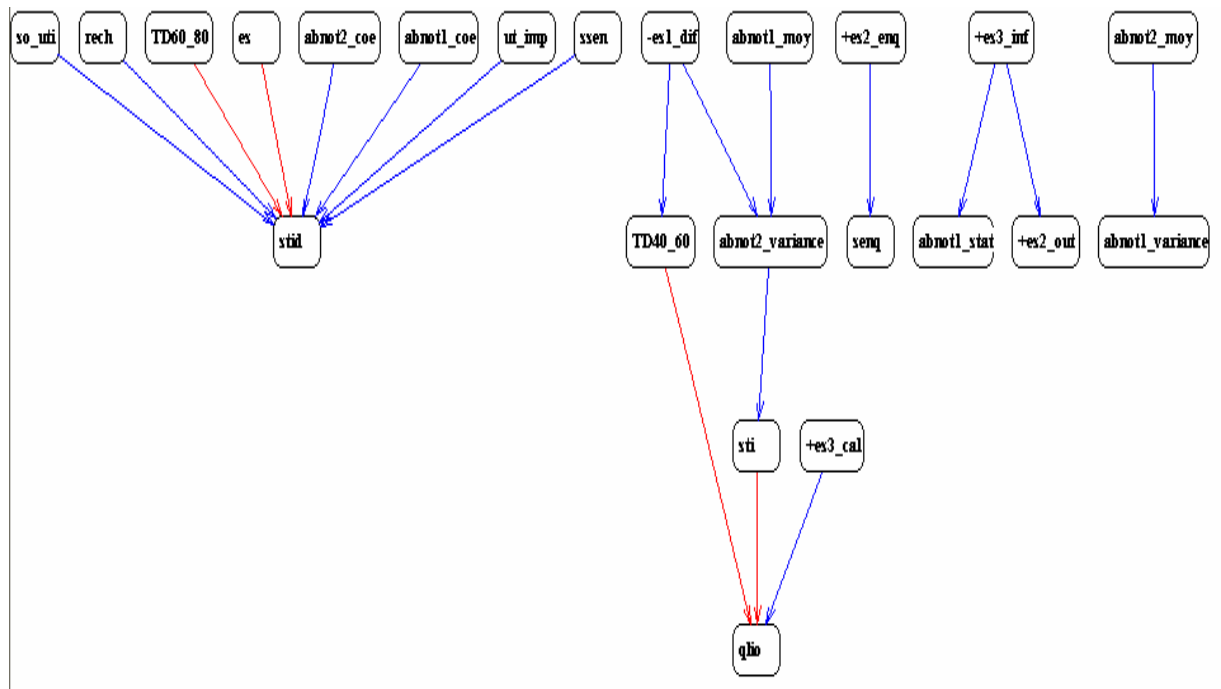
Niveau 1 « obj_conn » implique « stid » : acquérir un maximum de connaissance en statistiques est l'objectif des étudiants STID

Niveau 2 « obj_chif » implique « qlio » : l'utilisation des chiffres est l'objectif des étudiants QLIO.

Pour une formation en statistique à l'IUT, l'objectif pour le département :

- QLIO : utilisation des chiffres afin d'être opérationnel pour des missions ponctuels en entreprise qui leur demanderaient de calculer par exemple quelques indicateurs statistiques.
- STID : privilégier un maximum de connaissances en statistiques afin de bien connaître toutes les techniques pour être un bon statisticien.

5.2.2.24. Graphe implicatif – partie 1



En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.

On peut remarquer tout d'abord que la modalité « TD60_80 » et la modalité « es » implique à 99% la modalité « STID ». Donc le fait d'avoir entre 60 et 80 heures de cours et TD et d'être issu d'un bac ES se réalise à 99% chez les étudiants STID.

De même, nous remarquons que le fait d'avoir entre 40 et 60 heures de cours et TD (modalité « TD40_60 ») et d'être issu d'un bac STI implique à 99% d'être en QLIO.

On peut aussi voir que les modalités « so_util », « rech », « abnot2_coef », « abnot1_coe », « ut_imp », « ssen » se réalisent à 95% sur la modalité « STID ». En effet, le fait de trouver les sondages d'opinion utiles, de penser que le mot « Représentatif » signifie « échantillon », d'avoir bien compris les coefficients, de penser que la statistique est importante, et de penser que le mot « significatif » signifie « sens » implique à 95% le fait d'être un étudiant STID.

On peut noter également le fait d'avoir bien compris la moyenne implique à 95% d'avoir bien compris la variance (dernière implication à droite). Quand la statistique nous évoque l'enquête (modalité « +es3_enq ») implique à 95% de penser que les sondages signifient l'enquête.

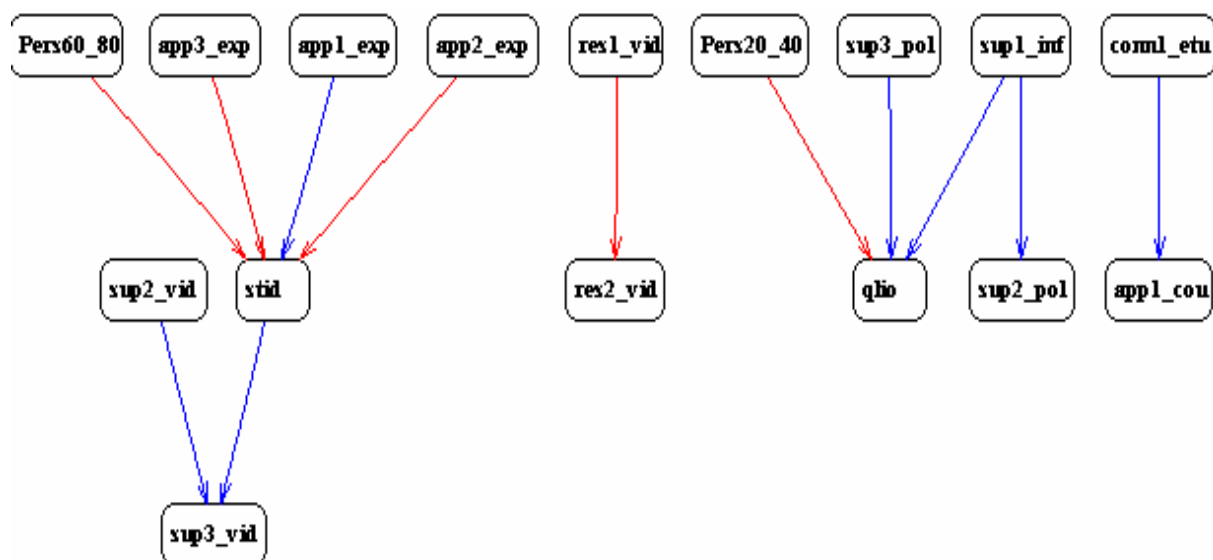
Si la statistique nous évoque à ce jour l'informatique (modalité « +es3_inf »), alors à 95% on a bien compris l'informatique et on considère la statistique comme un outil.

Si on pensait avant le DUT que la statistique est difficile (modalité « -es1_dif »), alors à 95% d'implication, on suit entre 40 et 60 heures de cours et de TD et on a bien compris le terme « variance ».

Le fait d'avoir bien compris le terme « moyenne » (modalité « abnot1_moy ») nous entraîne une bonne compréhension de la variance à 95%.

Enfin, on s'aperçoit que si la statistique nous évoque à ce jour le calcul, alors on fait partie du département QLIO à 95% de chance.

5.2.2.25. Graphe implicatif – partie 2



En cas de difficulté pour trouver la signification des modalités, veuillez vous reporter au dictionnaire des données disponible en annexe.

Dans ce graphique, nous pouvons en premier apercevoir que si on consacre entre 60 et 80 heures de travail personnel (modalité « Pers60_80 ») implique à 99% d'être en STID. Si on attend plus d'expérience en statistique (modalité « app1_exp », « app2_exp », « app3_exp ») implique aussi 95 et 99% d'être en STID. Cette même modalité implique de ne pas avoir de support particulier de cours à 95% (modalité « sup3_vid »).

Si la durée de travail personnel est entre 20 et 40 heures (modalité « Pers20_40 ») se réalisent à 99% chez les étudiants QLIO. De même si on affirme que l'on a des photocopies (modalité « sup3_pol ») et l'informatique comme support de cours (modalité « sup1_inf »), alors on est à 95% de chance en QLIO. Cette même modalité « sup1_inf » implique à 95% la modalité « sup2_pol ». Si on a l'informatique comme support de cours, alors on a à 95% de chance des photocopies.

On remarque enfin que le fait d'avoir appris à ce jour comment faire des études statistiques nous amène à 95% le fait d'attendre plus de cours dans la formation (dernière relation à droite).

5.2.3. Synthèse des résultats

En premier lieu, nous avons remarqué que les étudiants issus de la formation STID possèdent un bac ES, S ou STT et ceux de la formation QLIO ont un bac STI ou Professionnel.

Ensuite, nous avons constaté qu'avant leur entrée à l'IUT, les étudiants QLIO considéraient la statistique comme une discipline difficile. Pour les STID, la statistique correspondait à des enquêtes et aux mathématiques. Actuellement, la statistique évoque des chiffres et des calculs pour les étudiants QLIO. Les STID, quant à eux, considèrent maintenant la statistique comme un outil en utilisant l'informatique.

Concernant les sondages d'opinion, il existe une grande différence de perception suivant les formations : les QLIO les jugent inutiles, longs et inintéressants tandis que les STID les qualifient d'utiles, intéressants mais parfois mensongers (notamment concernant la politique).

Nous avons ensuite observé que les individus issus de la formation QLIO considèrent la statistique utile surtout pour l'analyse. Ils ont une vision plutôt scolaire de la statistique. Les STID pensent que la statistique occupe une place primordiale dans la vie sociale actuelle.

Nous avons également pu constater que les étudiants QLIO ont acquis des notions de base en statistique alors que pour les STID, ce sont des notions plus spécifiques. Ainsi, les difficultés auxquelles les étudiants sont susceptibles d'être confrontés concernent ces points respectifs.

Concernant les difficultés majeures rencontrées en alternance, on peut remarquer que les étudiants STID s'inquiètent pour leur réussite et leur capacité à s'adapter. Les QLIO ont plutôt des difficultés à suivre le rythme et à acquérir une certaine maturité. De plus, concernant plus précisément la statistique, les individus issus aussi bien de la formation STID que QLIO semblent éprouver des difficultés à mettre en pratique les connaissances acquises à l'IUT.

Nous avons aussi observé que les étudiants STID s'accordent à penser que l'expérience est un des apports majeurs fournis par l'alternance : ils peuvent en effet appliquer les connaissances acquises à l'IUT. Les QLIO, quant à eux, ont cité comme apport majeur l'apprentissage de l'analyse : cette dernière occupe ainsi une place importante dans la formation QLIO.

Par ailleurs, les étudiants QLIO suggèrent une formation davantage axée sur l'informatique tandis que les STID souhaiteraient avoir des cours de soutien, un nombre d'heures de TD plus important, ainsi qu'un rythme moins soutenu.

Les étudiants QLIO font souvent appel à des ressources supplémentaires, alors que les STID, ayant un nombre plus important d'heures, n'en ressentent pas forcément le besoin. En effet, les QLIO complètent leurs cours par des photocopiés et des livres tandis que les STID n'utilisent pas de supports.

Pour une formation en statistique à l'IUT, les QLIO ont pour objectif d'utiliser des chiffres afin d'être opérationnel pour des missions ponctuelles en entreprise. Les STID privilégient l'acquisition d'un maximum de connaissances en statistique afin de connaître un large éventail de techniques pour être un bon statisticien

5.2.4. Annexes : Le questionnaire

Année universitaire 2005-2006

Université Lyon 2

Étude des difficultés et des facilités d'apprentissage de la statistique au travers des méthodes quantitatives et qualitatives à l'entrée à l'Université : cas de l'IUT Lumière Lyon 2, formation en alternance.

L'objet de cette enquête par questionnaire est de construire un corpus de données à partir duquel, d'une part nous étudierons la question de la formation à la statistique, et d'autre part nous chercherons à dégager des éléments permettant une régulation de l'approche pédagogique proposée dans le dispositif de la formation en alternance.

Ces informations relèvent d'un travail de recherche et les réponses fournies ne peuvent en aucun cas intervenir dans l'évaluation.

Nous vous remercions de votre coopération en tant qu'informateur, sans laquelle les informations nécessaires ne peuvent être obtenues.

Le questionnaire est à remettre à Jean-Claude Oriol.

Jean-Claude Oriol & Jean-Claude Régnier

Questionnaire auprès des étudiants

Vous et votre parcours de formation antérieur à l'entrée à l'IUT :

Nom patronymique :	Prénom :	[V1] Date de naissance :
Nom marital :	N° étudiant	<i>Codage pour anonymat</i> :

[V2] Département de l'IUT :

[V3] Baccalauréat : Série..... [V4] Année d'obtention :

[V5] Vos diplômes universitaires **préparés** avant l'entrée à l'IUT:

	<i>Nom</i>	Année	Obtenu = O Non obtenu = N
Diplômes comportant de la statistique sous forme de cours, d'usage, etc.	-	-	-
	-	-	-
Diplômes ne comportant aucun cours de statistique ni aucune référence au domaine de la statistique.	-	-	-
	-	-	-
	-	-	-

[V6] Vos diplômes professionnels dont la **préparation** a requis des connaissances du domaine de la statistique :

<i>Nom</i>	Année	Obtenu = O Non obtenu = N

--	--	--

Vous et la statistique

[V7] A l'entrée dans cette formation en IUT, qu'évoquait pour vous la statistique ?

(Donnez trois mots)

--	--	--

(Complétez par une phrase)

--

[V8] A ce jour qu'évoque pour vous la statistique ?

(Donnez trois mots)

--	--	--

(Complétez par une phrase)

--

[V9] A ce jour que signifient pour vous les termes suivants ?

Représentatif	
Significatif	
Estimation	
Test d'hypothèses	
Probabilité	
Sondage	

[V10] A ce jour que pensez-vous des sondages d'opinion ?

[V11] Dans la vie sociale actuelle, quelle utilité attribuez-vous à la statistique ?

Vous et l'approche pédagogique adoptée dans le dispositif de formation à l'IUT

[V12] A ce jour, combien d'heures de cours et de TD de statistique avez-vous suivies ? —
 — h

[V13] Parmi les notions visées par le cours de statistique, quelles sont celles que vous pensez avoir (*citer au plus six*)

...assez bien comprises	... plutôt mal comprises

Formation en alternance et apprentissage de la statistique

[V14] Quelles sont les trois **difficultés majeures** que vous pensez devoir affronter pour répondre aux exigences du dispositif pédagogique fondé sur l'alternance ?

Diff_1 :	Diff_2 :	Diff_3 :
----------	----------	----------

[V15] Quelles sont les trois **difficultés majeures** que vous pensez devoir affronter pour répondre aux exigences dans la formation en statistique dans le cadre du dispositif pédagogique fondé sur l'alternance ?

Diff_4 :	Diff_5 :	Diff_6 :
----------	----------	----------

[V16] Quels sont les trois **apports majeurs** que vous attendez de ce dispositif de pédagogie en alternance quant à la formation en statistique ?

App_1 :	App_2 :	App_3 :
---------	---------	---------

[V17] Quelles sont les trois **suggestions majeures** que vous feriez *a priori* pour faciliter l'apprentissage de la statistique dans ce dispositif pédagogique en alternance ?

Sugg_1 :	Sugg_2 :	Sugg_3 :
----------	----------	----------

Supports du cours de statistique

[V18] De quels supports de cours disposez-vous ?

--

[V18] Quel niveau d'efficacité attribuez-vous à l'ensemble des documents distribués ?

1	2	3	4	5	6	7	8	9	10
le plus faible									le plus fort

[V19] Quel niveau d'intérêt attribuez-vous à l'ensemble des documents distribués ?

1	2	3	4	5	6	7	8	9	10
le plus faible									le plus fort

[V20] A ce jour, durée totale estimée (h) de travail personnel pour étudier la statistique

:

entre ___ et ___ h

Usage des ressources diverses pour l'étude des méthodes ?

[V21] Quelles ressources avez-vous utilisées en plus des documents de cours distribués ? (*citer au plus 4*)

[V22] Pensez-vous qu'il est nécessaire d'aller chercher d'autres ressources que les documents de cours distribués ?

1_Pas du tout	2_Rarement	3_Quelquefois	4_Souvent	5_Très souvent	6_Toujours
---------------	------------	---------------	-----------	----------------	------------

[V23] De manière générale pour le cours de statistique, que pensez-vous de la nécessité de compléter les cours par d'autres ressources ?

--

Formation en statistique

[V24] A ce jour que pensez-vous avoir appris en ce qui concerne le domaine de la statistique ?

--

[V25] Voici 10 objectifs énoncés concernant la formation en statistique :

[V25.1] Rangez-les par ordre décroissant d'importance [1 = le plus important à 10 = le moins important] en reportant le rang dans la colonne « **Rang attribué** »

[V25.2] Sur chaque ligne, indiquez, en cochant la case appropriée, si vous êtes

- TOUT A FAIT D'ACCORD = case de la colonne A
- PLUTÔT D'ACCORD = case de la colonne B
- PLUTÔT EN DESACCORD = case de la colonne C
- TOUT A FAIT EN DESACCORD = case de la colonne D

...avec l'énoncé correspondant

<i>Énoncés des objectifs</i>	<i>ang attribué</i>				
1) expliciter les questions d'une problématique dont les réponses relèvent d'une approche statistique,					
2) décrire, traiter, analyser des données de manière pertinente dans le cadre d'une étude en particulier dans le domaine professionnel choisi,					
3) faire le lien entre la réflexion analytique sur des questions relevant du champ professionnel, leur formalisation et leur traitement quantitatif,					
4) lire avec un regard critique et distancié, les conclusions de diverses études statistiques apparaissant dans des rapports de recherche de votre domaine professionnel,					
5) poursuivre de façon autonome et personnalisée un apprentissage en statistique afin d'enrichir ses acquis,					
6) poser un regard distancié à l'égard du domaine de la statistique, largement exploité dans les <i>media</i> , dans le sens de ne pas considérer les résultats dans l'ordre du tout ou rien mais en les replaçant judicieusement dans leur domaine de validité,					
7) exploiter des notions et des démarches mathématiques à des fins d'outils, et de ce fait, modifier dans un sens positif le rapport souvent négatif que certains entretiennent avec la statistique,					
8) mettre en œuvre un raisonnement intégrant l'idée de risque d'erreur dans l'énoncé de ses conclusions,					
9) mettre en œuvre l'interprétation de phénomènes sur la base de données statistiques sur des faits et sur des relations entre ces faits,					
10) communiquer des résultats d'analyses de données en distinguant clairement le modèle utilisé de la réalité qu'il est supposé représenter, et en séparant bien les traitements menés à l'intérieur du modèle des interprétations reformulées dans le contexte du problème.					

[V26] Quel objectif principal formuleriez-vous pour une formation en statistique à l'IUT ?

--

Vos suggestions

[V27] Parmi les 26 questions précédentes quelles sont pour vous les quatre questions les plus importantes :

N° des questions			

[V28] Autres suggestions

5.2.4.1. Le dictionnaire des données

	Questions			réponse	code
V2	département de l'IUT			STID	stid
				QLIO	qlio
V3	baccalauréat			bac pro	pro
				S	S
				ES	Es
				STT	Stt
				STI	sti
V4	année d'obtention			1988	Anbac88
				2001	Anbac01
				2003	Anbac03
				2004	Anbac04
				2005	Anbac05
V5	vos diplômes universitaires préparés avant l'entrée à l'IUT			CPGE éco commerce	Cpge
				DPECF	Dpecf
				Licence géographie	Geo
				Pas diplôme stat	Pdstat
		Année obtention		2004	Anuniv04
				2005	Anuniv05
				Pas répondu	Anuniv_0
V6	diplômes professionnels dont la préparation a requis des connaissances du domaine de la statistique			BEP	Pro_BEP
				Bac pro	Pro_bac_pro
				STI	Pro_STI
				Aucun	Pro_aucun
		Année obtention		2001	Pro_2001
				2003	Pro_2003
				2005	Pro_2005
				aucune	Pro_aucune
V7	A l'entrée dans cette formation en IUT, qu'évoquait pour vous la statistique ?	Mot1		Chiffres	-es1_chi
				Maths	-es1_mat
				Calculs	-es1_cal
				Difficile	-es1_dif
				Graphique	-es1_gra
				Enquête	-es1_enq
				Vide	-es1_vid
		Mot2		Calculs	-es2_cal
				Chiffres	-es2_chi
				Maths	-es2_mat
				Informatique	-es2_inf
				Logique	-es2_log
				Ennuyeux	-es2_enn
				Enquête	-es2_enq
				Difficile	-es2_dif
				Vide	-es2_vid

		Mot3	Graphique	-es3_gra
			Chiffres	-es3_chi
			Maths	-es3_mat
			Calculs	-es3_cal
			Manipulation	-es3_man
			Sondage	-es3_son
			Difficile	-es3_dif
			Vide	-es3_vid
		phrase	Graphiques	-esp_gra
			Enquête	-esp_enq
			Manipulation	-esp_man
			Calculs	-esp_cal
			Chiffres	-esp_chi
			Difficile	-esp_dif
			Science	-esp_sci
			Ennuyeux	-esp_enn
			Vide	-esp_vid
V8	A ce jour qu'évoque pour vous la statistique ?	Mot1	Calculs	+es1_cal
			Chiffres	+es1_chi
			Maths	+es1_mat
			Outils	+es1_out
			Etude	+es1_etu
			Informatique	+es1_inf
			Ennuyeux	+es1_enn
			Difficile	+es1_dif
			Vide	+es1_vid
		Mot2	Calculs	+es2_cal
			Chiffres	+es2_chi
			Enquête	+es2_enq
			Outils	+es2_out
			Inférentielle	+es2_inf
			Cours	+es2_cou
			Difficile	+es2_dif
			Maîtrise	+es2_mai
			Vide	+es2_vid
		Mot3	Analyse	+es3_ana
			Calculs	+es3_cal
			Difficile	+es3_dif
			Estimation	+es3_est
			Graphique	+es3_gra
			Indicateurs	+es3_ind
			Utile	+es3_uti
			Etude	+es3_etu
			Inférentielle	+es3_inf
			Vide	+es3_vid
		phrase	Calculs	+esp_cal
			Chiffres	+esp_chi
			Utile	+esp_uti
			Difficile	+esp_dif
			Enquête	+esp_enq
			Ennuyeux	+esp_enn
			Méthodes	+esp_met
			Science	+esp_sci
			Vide	+esp_vid
V9	A ce jour que signifient pour vous les termes suivants ?	Représentatif	Population	rpop
			Représentatif	rrep
			Echantillon	rech
			Visualiser	rvis
			Vide	rvid

				Calcul	rcal
				Réalité	rrea
		significatif		Pertinence	sper
				Significatif	ssig
				vide	svid
				Autre	saut
				Egal	sega
				Résultat	sres
				Sens	ssen
		Estimation		Approximation	eapp
				Estimation	eest
				Provision	epre
				Vide	evide
				Imprécis	eimp
				Calcul	ecal
				Supposition	esup
				Autre	eaut
		Test d'hypothèse		Tester	ttes
				Essai	tess
				Vide	tvide
				Hypothèse	thyp
				Question	tque
				Autre	taut
		probabilité		Chance	pcha
				Vide	pvide
				Autre	paut
				Possibilité	ppos
				Probabilité	ppro
				Pourcentage	ppou
				Estimation	pest
		Sondage		Opinion	sopi
				Enquête	senq
				Estimation	sest
				Vide	svid
				Etude	setu
				Test	stes
V10	A ce jour que pensez-vous des sondages d'opinion ?			Imprécis	so_imp
				Inintéressant	so_ini
				Intéressant	so_int
				Inutile	so_inu
				Long	so_lon
				Mensonger	so_men
				Participatif	so_par
				Sélectif	so_sel
				Utile	so_uti
				vide	so_vid
V11	Dans la vie sociale actuelle, quelle utilité attribuez-vous à la statistique ?			Analyse	ut_ana
				Aucune	ut_auc
				Décisionnel	ut_dec
				Estimation	ut_est
				Importante	ut_imp
				Opinion	ut_opi
				Politique	ut_pol
				Probabilité	ut_pro
				Simplification	ut_sim
				Tendance	ut_ten
				Vide	ut_vid

V12	A ce jour, combien d'heures de cours et de TD de statistique avez-vous suivies ?			moins de 20	td<20
				de 20 à 40	td20_40
				de 40 à 60	td40_60
				de 60 à 80	td60_80
				plus de 80	td>80
V13	Parmi les notions visées par le cours de statistique, quelles sont celles que vous pensez avoir...	...assez bien comprises	Notion1	Coefficients	abnot1_coe
				Indicateurs	abnot1_ind
				Statistique descriptive	abnot1_stat
				Moyenne	abnot1_moy
				Variance	abnot1_variance
				Lois	abnot1_loi
				Variables	abnot1_variable
				Vide	abnot1_vid
			Notion2	Statistique descriptive	abnot2_stat
				Lois	abnot2_loi
				Variance	abnot2_variance
				Moyenne	abnot2_moy
				Echantillon	abnot2_ech
				Coefficients	abnot2_coe
				Vide	abnot2_vid
			Notion3	Coefficients	abnot3_coe
				variance	abnot3_variance
				Statistique descriptive	abnot3_stat
				Moyenne	abnot3_moy
				Vide	abnot3_vid
			notion4	coefficients	abnot4_coe
				Lois	abnot4_loi
				Variance	abnot4_variance
				Moyenne	abnot4_moy
				Echantillon	abnot4_ech
				Statistique descriptive	abnot4_stat
				vide	abnot4_vid
			Notion5	coefficients	abnot5_coe
				lois	abnot5_loi
				moyenne	abnot5_moy
				statistique descriptive	abnot5_stat
				variance	abnot5_var
				vide	abnot5_vid
			notion6	médiane	abnot6_med
				courbe	abnot6_cou
				quali quanti	abnot6_qq
				binomiale	abnot6_bin
				vide	abnot6_vid
		...plutôt mal comprises	notion1	indicateurs stats	malnot1_ind
				probabilité	malnot1_pro
				tableaux croisés	malnot1_tab
				variance	malnot1_var
				méthodes stats	malnot1_met
				lois	malnot1_loi
				dispersion	malnot1_dis
				covariance	malnot1_cov
				régression	malnot1_reg
				vide	malnot1_vid
			notion2	lois	malnot2_loi
				probabilités	malnot2_pro
				variance	malnot2_var
				droite régression	malnot2_dro
				indicateurs stats	malnot2_ind

				vide	malnot2_vid
			notion3	lois	malnot3_loi
				tableaux croisés	malnot3_tab
				vide	malnot3_vid
V14	Quelles sont les trois difficultés majeures que vous pensez devoir affronter pour répondre aux exigences du dispositif pédagogique fondé sur l'alternance ?	difficulté 1		adaptation	da1_ada
				cours	da1_cou
				déplacements	da1_dep
				équipe	da1_equ
				intégration	da1_int
				maturité	da1_mat
				prospection	da1_pro
				rythme	da1_ryt
				sondage	da1_son
				vide	da1_vid
		difficulté 2		cours	da2_cou
				intégration	da2_int
				maturité	da2_mat
				orientation	da2_ori
				prospection	da2_pro
				réussite	da2_reu
				rythme	da2_ryt
				vide	da2_vid
		difficulté 3		adaptation	da3_ada
				Cours	da3_cou
				Intégration	da3_int
				Maturité	da3_mat
				Prospection	da3_pro
				Réussite	da3_reu
				Rythme	da3_ryt
				vide	da3_vid
V15	Quelles sont les trois difficultés majeures que vous pensez devoir affronter pour répondre aux exigences dans la formation en statistique dans le cadre du dispositif pédagogique fondé sur l'alternance ?	difficulté 1		assimilation	ds1_ass
				Concentration	ds1_conc
				Connaissances	ds1_conn
				Informatique	ds1_inf
				Motivation	ds1_mot
				Pratique	ds1_pra
				Prospection	ds1_pro
				Recul	ds1_rec
				Rigueur	ds1_rig
				Rythme	ds1_ryt
				Stress	ds1_str
				Vague	ds1_vag
				vide	ds1_vid
		difficulté 2		Analyse	ds2_ana
				Assimilation	ds2_ass
				Compréhension	ds2_com
				Informatique	ds2_inf
				Patience	ds2_pat
				Pratique	ds2_pra
				Rédaction	ds2_red
				Rigueur	ds2_rig
				Rythme	ds2_ryt
				Travail	ds2_tra
				Utile	ds2_uti

				Vide	ds2_vid
		difficulté 3		Alternance	ds3_alt
				Analyse	ds3_ana
				Concentration	ds3_con
				Cours	ds3_cou
				Organiser	ds3_org
				Pratique	ds3_pra
				Rythme	ds3_ryt
				Tenacité	ds3_ten
				Vide	ds3_vid
V16	Quels sont les trois apports majeurs que vous attendez de ce dispositif de pédagogie en alternance quant à la formation en statistique ?	Apport1		Vide	app1_vid
				Cours	app1_cou
				Entreprise	app1_ent
				Autre	app1_aut
				Expérience	app1_exp
		apport2		vide	app2_vid
				Application	app2_app
				Analyser	app2_ana
				Autre	app2_aut
				Donner	app2_don
				Expérience entreprise	app2_exp
		apport3		vide	app3_vid
				Autre	app3_aut
				résultats	app3_res
				Concret	app3_con
				expérience	app3_exp
V17	Quelles sont les trois suggestions majeures que vous feriez a priori pour faciliter l'apprentissage de la statistique dans ce dispositif pédagogique en alternance ?	suggestion1		plus de cours en groupes	sug1_cour
				ralentir le rythme	sug1_ryt
				plus de projets	sug1_pro
				plus d'informatique	sug1_inf
				cours de soutien	sug1_sou
				plus d'heures de cours	sug1_heu
				Vide	sug1_vid
		suggestion2		application concrète	sug2_app
				cours plus précis	sug2_cou
				informatique	sug2_inf
				plus d'aide, plus de TD	sug2_aid
				Vide	sug2_vid
		Suggestion3		vide	sug3_vid
V18	De quels supports de ce cours disposez-vous ?	support 1		Polycopiés	sup1_pol
				Informatique	sup1_inf
				TP	sup1_tp
				Notes personnelles	sup1_not
				Livres	sup1_liv
				Vide	sup1_vid
		support 2		Polycopiés	sup2_pol
				Informatique	sup2_inf
				TP	sup2_tp
				Notes personnelles	sup2_not
				Vide	sup2_vid
		support 3		Polycopiés	sup3_pol
				informatique	sup3_inf
				TP	sup3_tp
				Notes personnelles	sup3_not

				Livre	sup3_liv
				Prof	sup3_pro
				Vide	sup3_vid
V19	quel niveau d'efficacité attribuez-vous à l'ensemble des documents distribués ?				niv_eff
V20	quel niveau d'intérêt attribuez-vous à l'ensemble des documents distribués ?				niv_int
V21	A ce jour, durée totale estimée (h) de travail personnel pour étudier la statistique			moins de 20 H	pers<20
				entre 20H et 40H	pers20_40
				entre 40H et 60H	pers40_60
				entre 60H et 80H	pers60_80
				plus de 80 H	pers>80
V22	Quelles ressources avez-vous utilisées en plus des documents de cours distribués ?	ressource 1		collègues	res1_col
				Cours	res1_cou
				Informatique	res1_inf
				Livres	res1_liv
				Notes personnelles	res1_not
				Vide	res1_vid
		ressource 2		Connaissances personnelles	res2_con
				Cours	res2_cou
				Informatique	res2_inf
				Livres	res2_liv
				Polycopié	res2_pol
				Vide	res2_vid
		ressource 3		Collègues	res3_col
				Cours	res3_cou
				Informatique	res3_inf
				Vide	res3_vid
V23	Pensez-vous qu'il est nécessaire d'aller chercher d'autres ressources que les documents de cours distribués ?			Pas du tout	Pdt
				Quelquefois	Qqf
				Rarement	Rare
				Souvent	Svt
				Toujours	tjs
V24	De manière générale pour le cours de statistique, que pensez-vous de la nécessité de compléter les cours par d'autres ressources ?			Compréhension	autres_comp
				Complément	autres_compl
				Utile	autres_uti
				Inutile	autres_inu
				Peut servir	autres_ser
				Reformulation	autres_ref
				Enrichissement	autres_enr
				vide	autres_vid
V25	A ce jour que pensez-vous avoir appris en ce qui concerne le domaine de la statistique ?	connaissance 1		Statistique ciblée	conn1_stat
				Graphiques	conn1_grap

				Bases	conn1_bas
				Ensemble de méthodes et d'outils	conn1_met
				Etudes	conn1_etu
				Estimation	conn1_est
				Vide	conn1_vid
		connaissance 2		Bases	conn2_bas
				Excel	conn2_exc
				Etudes	conn2_etu
				Statistique ciblée	Conn2_stat
				Lois	conn2_loi
				Vide	conn2_vid
V27	Quel objectif principal formuleriez-vous pour une formation en statistique à l'IUT ?			utilisation des chiffres	obj_chif
				réponse à une problématique	obj_pb
				nécessité monde professionnel	obj_pro
				connaître les bases	obj_bas
				maximum de connaissances statistiques	obj_conn
				vide	obj_vid

tel-00191166, version 1 - 27 Nov 2007

5.3. Etude des représentations de la statistique chez les étudiants STID en alternance de Lyon

5.3.1. Présentation

Cette deuxième enquête sur les représentations de la statistique a été conduite en 2006-2007 sous ma responsabilité par un groupe d'étudiants de deuxième année du DUT Stid Lyon ; il s'agit de Carole BELLO, Marion BRETHER, Anthony DUSSURGEY, Laure GIRAUDET, Stéphanie REGAIRAZ et Myriam SOULIÉ que nous remercions pour le sérieux qu'ils ont apporté à leur travail. Nous tenons également à remercier les étudiants de 2^{ème} année STID de Lyon et de Grenoble, pour avoir répondu aux questionnaires ainsi que Gérard GRÉGOIRE, chef du département Stid Grenoble pour avoir facilité cette enquête auprès des étudiants du département Stid qu'il dirige.

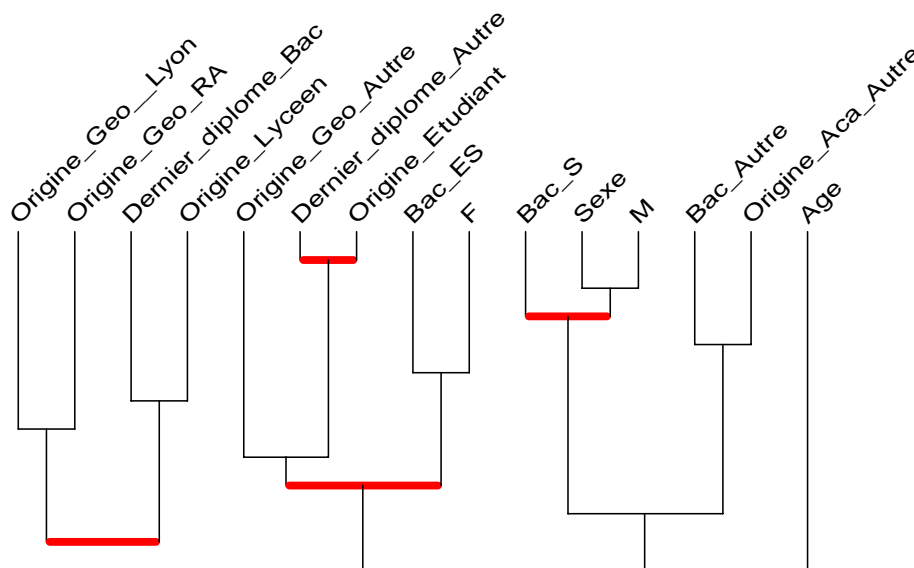
En ce qui concerne ce groupe, l'étude portait sur les représentations de la statistique. Ce projet était commandité par l'équipe de recherche ADATIC de l'Université Lyon 2. La commande était de voir s'il existait une différence entre les représentations de la statistique chez des étudiants de deuxième année STID, certains étant en alternance (à Lyon) et d'autres suivant un cycle sans alternance (à Grenoble). En fait ce projet initial a été perturbé par le stage que suivaient les étudiants du département STID de Grenoble. Le nombre de questionnaires retournés par ces étudiants, du fait de leur présence en entreprise durant le déroulement de l'enquête a été trop faible. Cela nous a amené à traiter les questionnaires de STID Lyon et à analyser uniquement les représentations de la statistique chez les étudiants de STID Lyon avant l'entrée dans le département et en fin de deuxième année. Il s'agit de connaissances déclaratives. Le questionnaire des années précédentes a été réajusté. Pour cela, les questions ont été modifiées par les étudiants afin qu'elles répondent plus précisément à notre nouvelle problématique : la commande était de voir s'il existait une différence entre les représentations de la statistique chez des étudiants de deuxième année STID, certains étant en alternance et d'autres suivant un cycle sans alternance. En fait ce projet initial a été perturbé par le stage que suivaient les étudiants du département STID de Grenoble. Le nombre de questionnaires retournés par ces étudiants, du fait de leur présence en entreprise durant le déroulement de l'enquête. Cela nous a amené à traiter les questionnaires de STID Lyon et à

analyser uniquement les représentations de la statistique chez les étudiants de STID Lyon avant l'entrée dans le département et en fin de deuxième année.

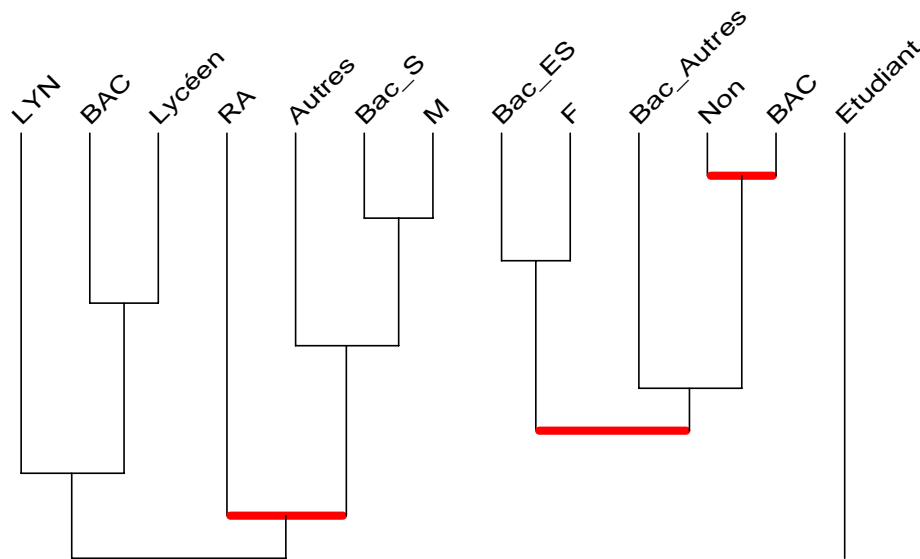
En termes d'ingénierie pédagogique le travail des étudiants s'est déroulé sur une durée de 70 heures, la moitié environ étant animée. A mi parcours un « séminaire » a amené les équipes à exposer l'avancée de leurs travaux et à répondre aux questions de l'ensemble des étudiants et animateurs de projets. Le travail de ce groupe a consisté à : comprendre les bases de l'analyse statistique implicite (lectures de documents, d'articles, construction de simulations), apprendre à utiliser le logiciel CHIC en prenant un fichier d'essai ayant un faible nombre d'individus et de variables, construire le questionnaire, administrer ce questionnaire à la population prévue, saisir les données des questionnaires retournés, nettoyer la base des données et recoder certaines variables avant la transformation de cette base afin qu'elle puisse être traitée par le logiciel CHIC et enfin analyser les résultats en s'appuyant sur l'analyse statistique implicite.

5.3.2. Travail sur un premier fichier d'essai

Les étudiants ont travaillé sur un fichier d'essai de 24 individus. Cela a permis de rencontrer différentes confusions en particulier la variable Sexe qui devait pour devenir pertinente être transformée en deux variables M et F ainsi qu'un problème concernant la variable Origine géographique Lyon et la variable Origine géographique Rhône Alpes.



Arbre des similarités : H\ASI_Articles_Etudiants\Etude_apres_lecture.csv



Arbre des similarités : C:\Documents and Settings\Oriol\Bureau\ASI Articles Etudiants\STID1 2006 2007.csv

5.3.3. Présentation de l'étude

5.3.3.1. Le questionnaire.

Le questionnaire (*voir Annexe*) interroge les étudiants sur :

- ✓ Leur cursus scolaire à savoir le Baccalauréat obtenu
- ✓ Leur vision de la statistique avant l'entrée dans le département STID, en quelques mots, et leur vision à l'heure actuelle, après l'apprentissage.
- ✓ Leur compréhension des méthodes enseignées, les notions de cours plus ou moins bien comprises, les difficultés rencontrées, les compétences qu'ils pensent acquérir au cours de leur formation STID.
- ✓ Des questions à thème plus « professionnel » afin de comparer les populations en alternance ou en formation initiale : quelles difficultés pensent-ils rencontrer dans le monde du travail, quelles méthodes pourraient-ils appliquer en entreprise...
- ✓ Leurs suggestions pour améliorer : l'apprentissage de la statistique, les supports de cours utilisés.

Nous avons été chargés d'administrer des questionnaires, répartis de la façon suivante :

- ✓ 25 questionnaires aux étudiants STID de 2nde année de Lyon 2

-
- ✓ 55 questionnaires aux étudiants STID de 2nde année de Grenoble

Face au taux de retour très faible de la part des étudiants de Grenoble (14,5%) et donc à un biais très important, nous avons décidé de mener notre étude uniquement sur les étudiants de 2nde année de Lyon 2. Nous avons décidé d'axer plus précisément notre recherche sur les différences possibles entre une population masculine et une population féminine.

5.3.3.2. La saisie et la codification.

Dès les premiers retours, nous avons saisi, sous Excel, les données obtenues dans les questionnaires. Pour cela, nous avons, au préalable, défini le masque de saisie pour codifier les données, utile pour le traitement des données sous le logiciel CHIC.

Nous avons rencontré plusieurs difficultés : certaines réponses étaient manquantes, d'autres n'avaient aucun sens ; mais le plus difficile a été de résumer les réponses par un mot afin de les regrouper en différents thèmes.

Ensuite, il a fallu recoder le nom de ces thèmes pour une meilleure lisibilité dans les sorties des graphiques. Les codifications seront détaillées pour une meilleure compréhension des résultats, (Voir également la feuille récapitulative des codifications en annexe).

Le fichier devait être épuré au maximum afin de le rendre plus compréhensible. Suite au nombre trop insuffisant de questionnaires retournés par les étudiants grenoblois, nous avons éliminé ces données de notre analyse, elles devenaient peu pertinentes. Ainsi, les questions 18, 19 et 20 concernant les supports de cours n'ont pas été traitées car les étudiants ont les mêmes supports de cours étant donné qu'ils proviennent de la même promotion. Les questions 12 et 16 ont été également écartées, car l'implication statistique n'était plus possible sans une comparaison entre deux populations.

5.3.4. Résultats

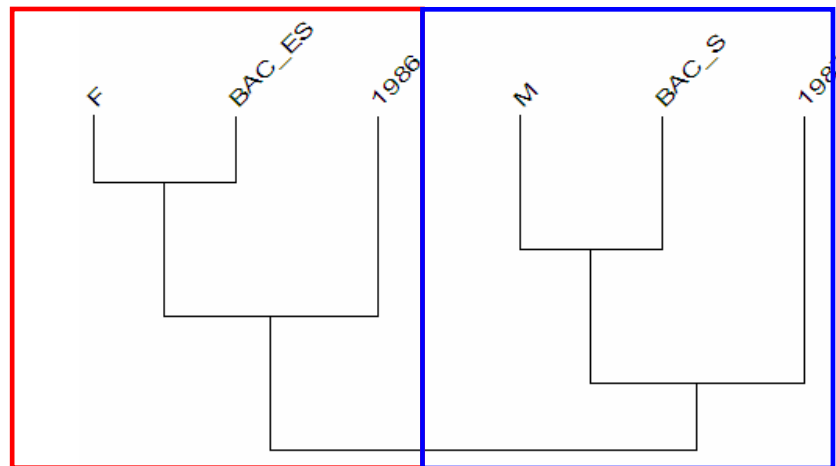
Nous allons maintenant vous présenter les résultats de l'étude. Nous analyserons chaque question en respectant l'ordre du questionnaire.

5.3.5. Le cursus scolaire des interrogés (questions 1 à 8)

Nous avons tout d'abord étudié l'implication des cursus scolaires (variables du bac obtenu : ES et S et année d'obtention de celui-ci) sur nos individus à savoir les étudiants STID de 2^{de} année de l'IUT Lumière (variables année de naissance et sexe : F et M).

5.3.5.1. Etude des similarités :

5.3.5.1.1. Arbre des similarités



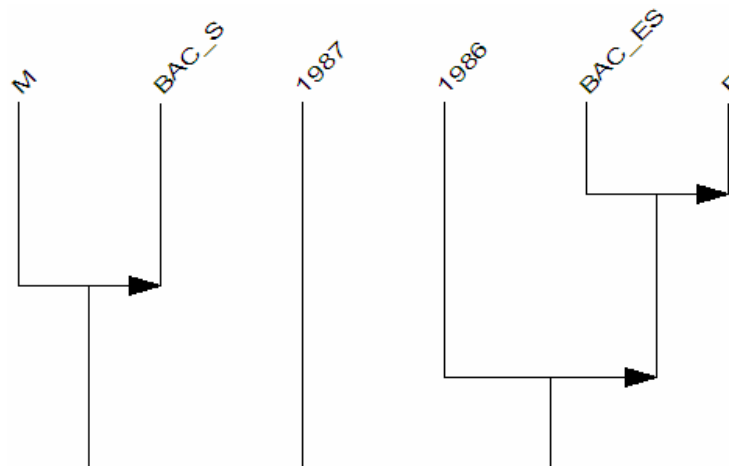
5.3.5.1.2. Analyse

L'arbre des similarités ci-dessus nous montre que les variables « BAC_ES » et sexe « F » (féminin) sont très similaires (0,956) rejoint par l'année « 1986 » (0,604) ; ce qui forme un premier groupe. Nous remarquons aussi que le « BAC_S » et le sexe « M » (masculin) sont aussi très similaires (0,952) rejoint cette fois par l'année « 1987 » (0,384) ; ce qui nous donne le deuxième groupe.

5.3.5.2. Etude de la cohésion

5.3.5.2.1. Arbre cohésitif

Si l'on s'intéresse à la cohésion, nous pouvons observer le sens des relations :

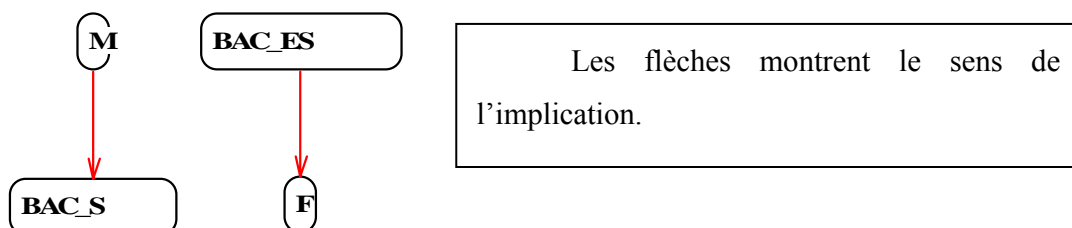


5.3.5.2.2. Analyse

Ici, nous voyons que le sens de l'implication est différent selon les 2 groupes que nous avons mis en évidence plus haut. Ainsi, le sexe « M » : masculin influence le bac choisi (« BAC_S ») à 0,961 tandis que le « BAC_ES » influence le sexe « F » (féminin) à 0,995. Nous pouvons également remarquer que l'année de naissance « 1986 » implique que l'individu soit probablement une fille ayant un bac ES. C'est-à-dire que toutes les filles nées en 1986 ont fait un bac ES. « 1987 » n'a pas réellement d'influence puisque cette variable n'est pas reliée à une ou plusieurs autres.

5.3.5.3. Etude de l'implication

De manière générale, un individu de sexe masculin implique qu'il a obtenu un bac S et un individu ayant un bac ES implique que c'est une fille. C'est ce que nous confirme le graphe implicatif :



5.3.6. La vision de la statistique avant et en fin de DUT

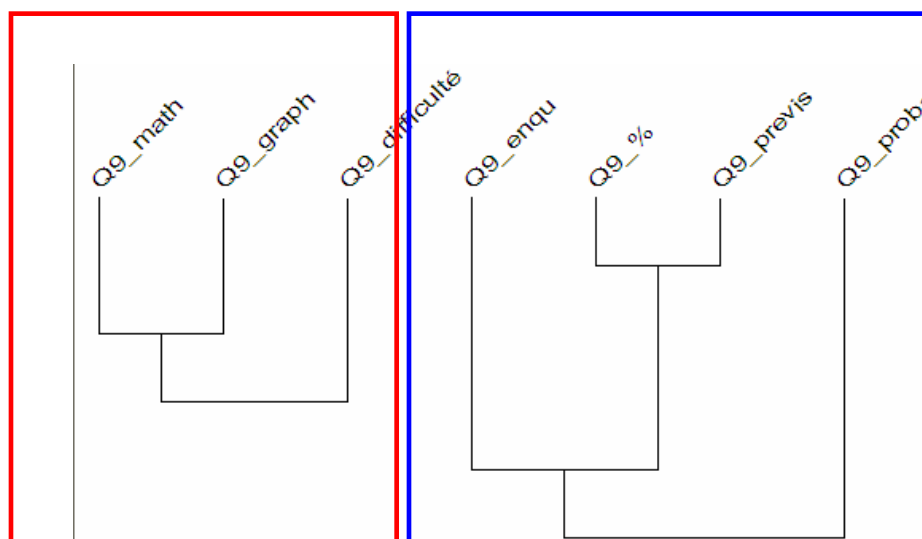
5.3.6.1. Introduction

Nous nous intéressons à présent à la vision que les individus ont de la statistique et ce, à 2 moments distincts : avant l'entrée en DUT et au cours de la seconde année de DUT. Nous utiliserons 2 questions de notre questionnaire pour analyser cette vision de la statistique dans le temps ; la question 9 : *A l'entrée dans cette formation en IUT, qu'évoquait pour vous la statistique ?* Et la question 10 : *A ce jour, qu'évoque pour vous la statistique ?*

5.3.6.2. Avant le DUT :

5.3.6.2.1. Etude des similarités

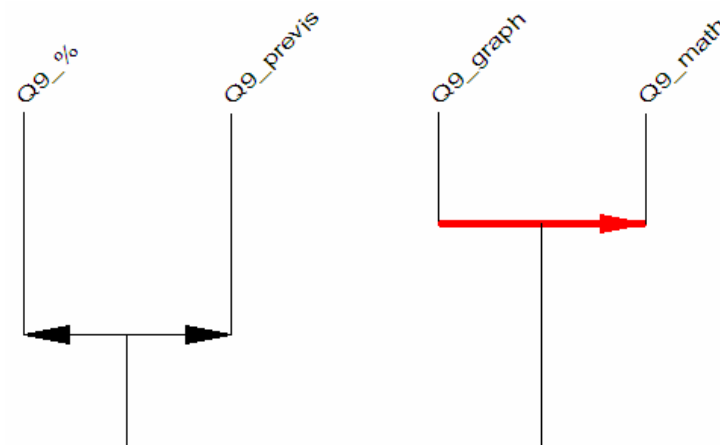
Avant l'entrée en DUT (question 9), les individus qualifiaient la statistique de mathématiques et de graphiques (0,748) puis de difficultés (0,462) pour les uns, et la considéraient comme des pourcentages et prévisions (0,955) et enquêtes (0,462) puis probabilités pour les autres :



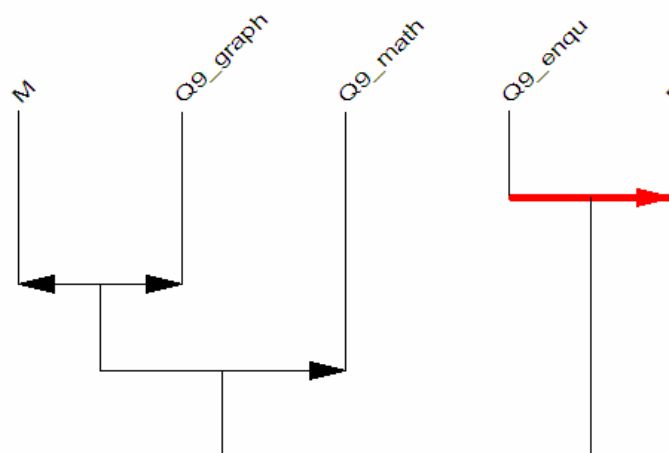
5.3.6.2.2. Etude de la cohésion

Plus précisément, nous pouvons remarquer, sur l'arbre cohésitif ci-après (page suivante), que les individus ayant répondu « Graphique » ont aussi répondu « Mathématique »

car l'un implique aussi l'autre (0,626). « Pourcentage » et « Prévision » s'impliquent l'un l'autre (0,363) ce qui signifie que si un individu a répondu l'un, il a probablement répondu l'autre et inversement.



Remarque : nous pouvons établir un lien entre la vision de la statistique avant l'entrée en DUT et le sexe :

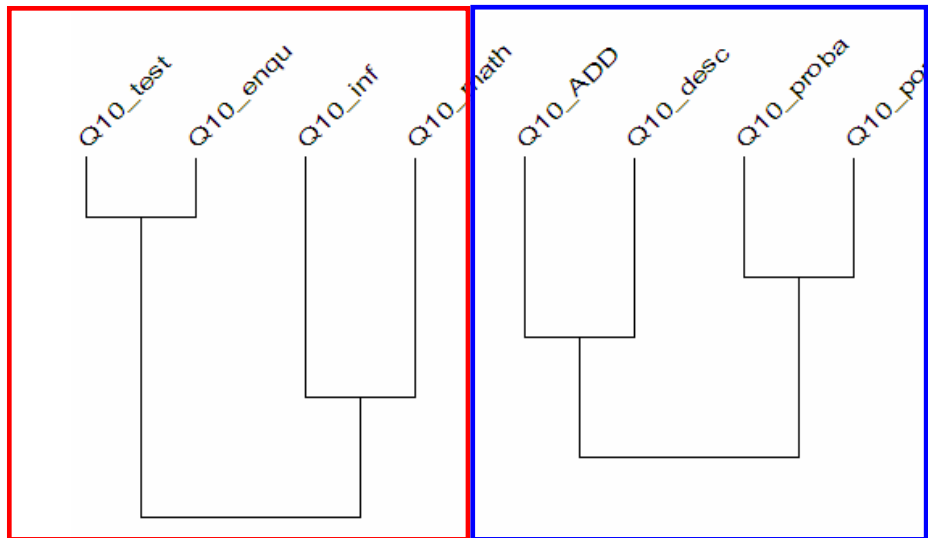


Les garçons, plus scientifiques (bac S, cf. plus haut), ont répondu « Graphique » et « Mathématique », mots liés et appartenant au même groupe que nous avons établi plus haut ; tandis que les filles ont répondu « Enquêtes », mot appartenant au second groupe.

5.3.7. Pendant la seconde année de DUT :

Après l'entrée en DUT (question 10), on retrouve certains termes d'avant mais ajoutés à des termes de vocabulaire plus statistique.

5.3.7.1. Etude des similarités

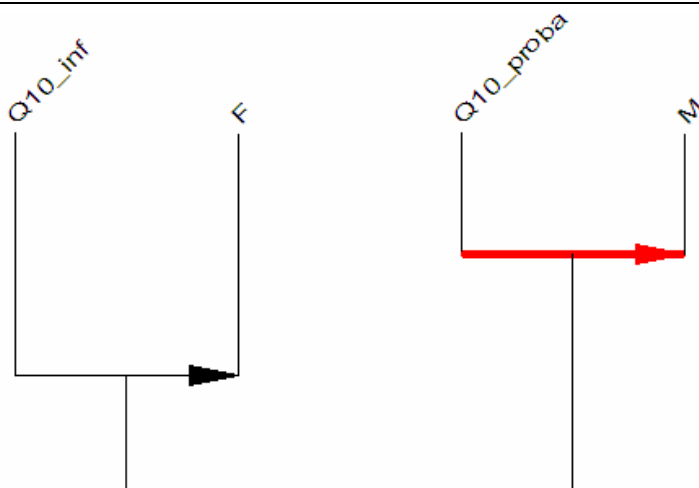


On distingue encore 2 groupes : « Test » et « Enquête » sont très similaires (0,745) ajoutés à « Inférence » et « Mathématique » (0,455) forment le premier groupe. « Population » et « Probabilité » vont de pair (0,674), ainsi que « ADD » et « Statistiques descriptives » (0,119) qui forment le second groupe.

En ce qui concerne le sens des liaisons, seul le mot « Enquête » implique « Test ». L'étudiant STID de Lyon, pensant que « Enquête » représente la statistique, le pense aussi des « Tests ».

5.3.7.2. Remarque

En mettant en évidence le sexe « M » et « F », on observe que certains termes sont liés à un type d'individus particulier. Ainsi, la variable « Inférence » implique que l'individu est une fille tandis que la variable « Probabilité » implique fortement que l'individu est un garçon.



5.3.8. Conclusion

Vision de la statistique avant le DUT STID

Les garçons, plus scientifiques, qualifient la statistique avec les mots « Graphiques » et donc « Mathématiques » selon le lien cohésitif que nous avons remarqué plus haut. Ce qui est bien représentatif de leur bac S.

Les filles, qui ont un côté plus marketing et sondage, voient la statistique comme des pourcentages et de la prévision (qualification aussi avec les mots « Enquête » et « Probabilité » dans une moindre mesure).

Avant d'entrée à l'IUT, les élèves interrogés avaient une vision très peu claire de la statistique. Pour les garçons scientifiques, Statistique signifiait *Mathématique*. Et pour les filles, plus marketing, Statistique voulait dire *Enquête* ou *Prévision*. Les mots choisis sont en lien très fort avec le contexte de l'élève juste après le bac, ce sont des mots encore très

Vision de la statistique pendant la 2nd année de DUT STID

A l'inverse, après coup, ce sont les garçons qui voient la statistique comme des « Probabilités » en lui associant le mot « Population », terme plus statistique. De plus, pour eux, ce sont les cours « d'Analyse de données » et de « Statistique descriptive » qui les ont le plus marqué.

Tandis que pour les filles, on retrouve le mot « Enquête » qu'elles ont associé cette fois au terme « Test ». Ce sont donc les cours de « Statistique inférentielle » qui les ont le plus marqué. Après coup, elles constatent que la statistique s'apparente aux « Mathématiques ».

généraux et qui donnent une impression de « *préjugés* » de la part de ces futurs STID de Lyon. A la fin du DUT, on constate un enrichissement du vocabulaire (exemple : le mot « Population ») et donc un éclaircissement de la signification du mot Statistique. Les garçons ont retenu une perception plus théorique avec de grands concepts : les mots employés se rapportent souvent à un intitulé de matière étudiée pendant l'année telle que « Analyse des données », « Statistique descriptive » ou encore « Probabilité ». Les filles ont compris que la statistique s'apparentait à une matière scientifique, et la qualifient avec des mots qui sont toujours des intitulés de matières : « Test », « Inférence » et on retrouve même le terme « Mathématique ».

5.3.9. Question 11 : termes « représentatifs » de la statistique

Pensez-vous que les termes suivants sont : très représentatif, assez représentatif, peu représentatif ou pas du tout représentatif de la statistique ? (cochez une case par ligne)

5.3.9.1. Le recodage des variables

5.3.9.1.1. Présentation de la question

Termes	Très	Assez	Peu	Pas du tout
Contrôle qualité				
Enquête				
Estimation				
Probabilité				
Sondage				
Tests d'hypothèse				
Traitement des données				

5.3.9.1.2. Recodage des modalités

Termes	Très	Assez	Peu	Pas du tout
Contrôle qualité	CQT	CQA	CQP	CQPT
Enquête	ENT	ENQA	ENQP	ENQPT
Estimation	ESTT	ESTA	ESTP	ESTPT
Probabilité	PRT	PRA	PRP	PRPT
<u>Sondage</u>	<u>SOT</u>	<u>SOA</u>	<u>SOP</u>	<u>SOPT</u>

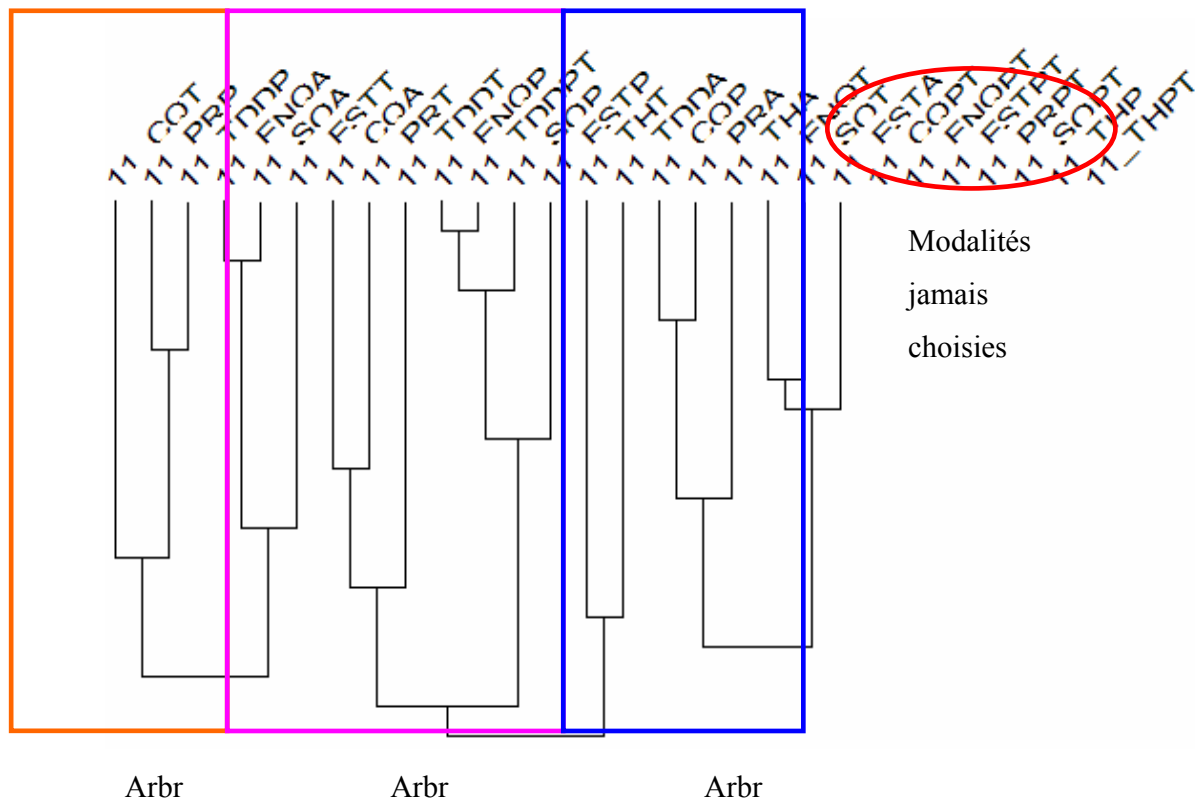
<u>Tests d'hypothèse</u>	<u>THT</u>	<u>THA</u>	<u>THP</u>	<u>THPT</u>
<u>Traitement des données</u>	<u>TDDT</u>	<u>TDDA</u>	<u>TDDP</u>	<u>TDDPT</u>

Nous obtenons à la suite du recodage 28 variables binaires. Avant ce recodage, on avait 7 variables comportant 4 modalités chacune.

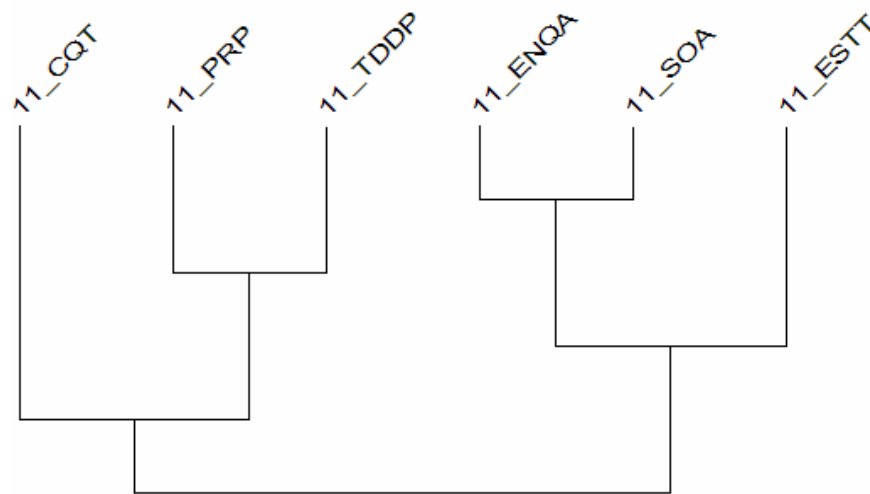
5.3.10. Etude de la question en général

Ce graphique semble compliqué à expliquer on peut voir qu'un certains nombres de modalités n'ont jamais été choisis par les interrogés, on ne les prendra pas en considération pour l'analyse de cette question. Les variables non utilisées sont les variables « CQPT », « ENQPT », « ESTPT », « PRPT », « SOPT », « THPT » et « TDDPT » : pour les interrogés aucun des termes proposés ne représente « pas du tout » la statistique.

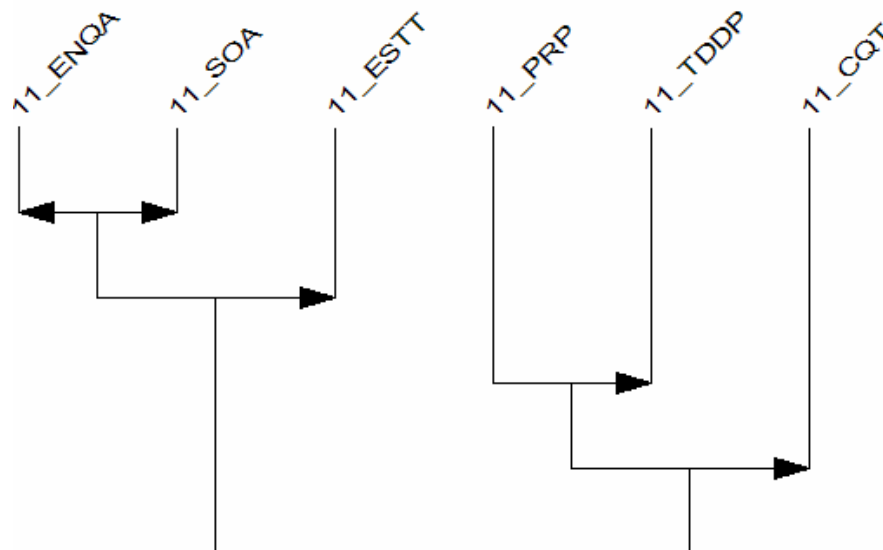
On représente trois groupes différents, chaque groupe forme un arbre indépendant.



5.3.11. Etude de la première partie de l'arbre des similarités (arbre A)



5.3.11.1. Etude de la cohésion

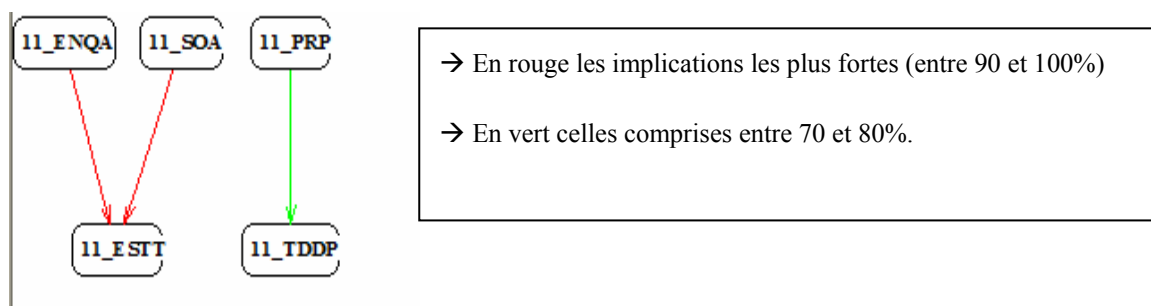


Notre étude montre qu'il y a équivalence (0,997) entre les réponses « Enquête, assez » et « Sondage, assez » c'est-à-dire que les personnes qui pensent que le mot « Enquête » représente assez bien la statistique, pensent également que le mot « Sondage » représente assez bien la statistique, et vice versa (d'après l'arbre cohésitif). On remarque également que lorsque l'on trouve que la statistique est assez bien représentée par les mots « Sondage » et « Enquête » alors on pense que « Estimation » est très représentatif de statistique (0,807).

On peut remarquer aussi que la modalité « Probabilité peu » ressemble à la modalité « Traitement des données peu » (0,985). Quand les étudiants STID2 de Lyon interrogés pensent que la statistique est peu représentée par « Probabilité », alors ils pensent que « Traitement des données » représente peu la statistique. Ces deux modalités sont également proches.

5.3.11.2. Etude de l'implication

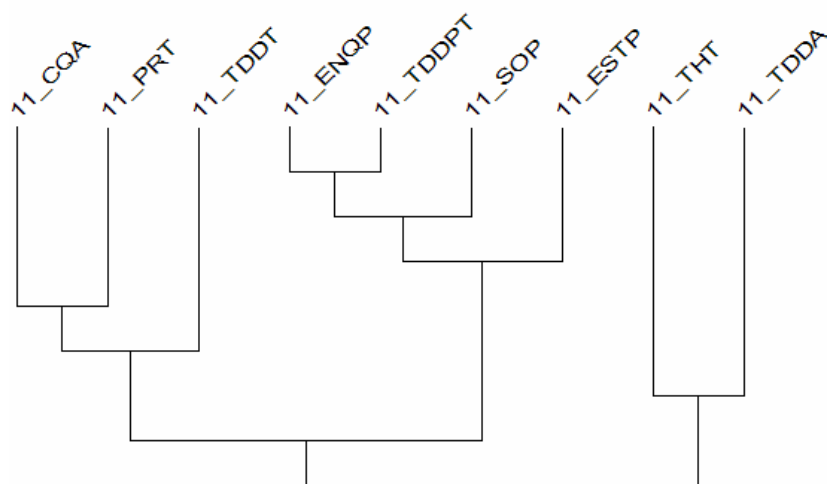
Regardons si le graphe implicatif complète ce que l'on vient d'observer :



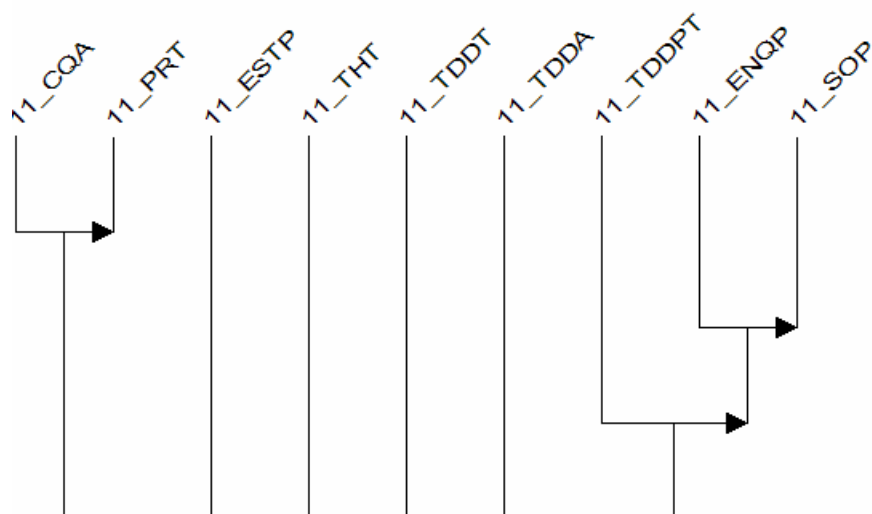
Nous retrouvons les mêmes conclusions que précédemment. Si une personne interrogée répond que le mot « Enquête » et le mot « Sondage » représentent assez bien la statistique alors cela implique fortement qu'elle ait répondu que « Estimation » représentait très bien la statistique.

Si les 2^{de} année STID de l'IUT Lumière ont répondu que le mot « Probabilité » représentait peu la statistique alors ils ont également répondu que le « Traitement de donnée » représentait peu la statistique. Cette implication est comprise entre 70% et 80%.

5.3.12. Etude de la deuxième partie de l'arbre de similarité (arbre B)



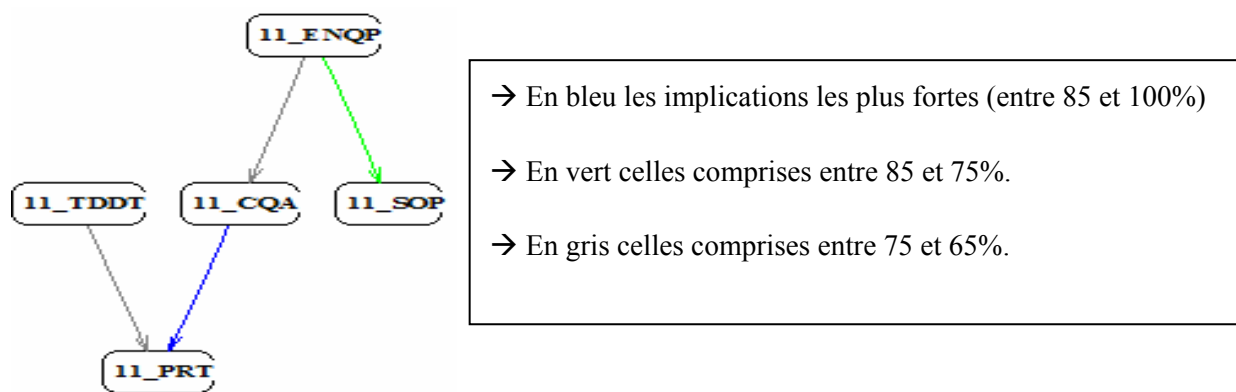
5.3.12.1. Etude de la cohésion



Au premier niveau, il y a un lien entre les variables « Contrôle qualité assez » et « Probabilité très » (0,873). Cela signifie que le fait de répondre que « Contrôle qualité » représente assez bien la statistique implique que la personne interrogée ait également répondu que le mot « Probabilité » est très représentatif de la statistique.

Au second niveau on voit que le fait d'avoir répondu que « Enquête » représente peu la statistique implique que l'on a répondu que « Sondage » représente aussi peu la statistique.

5.3.12.2. Etude de l'implication

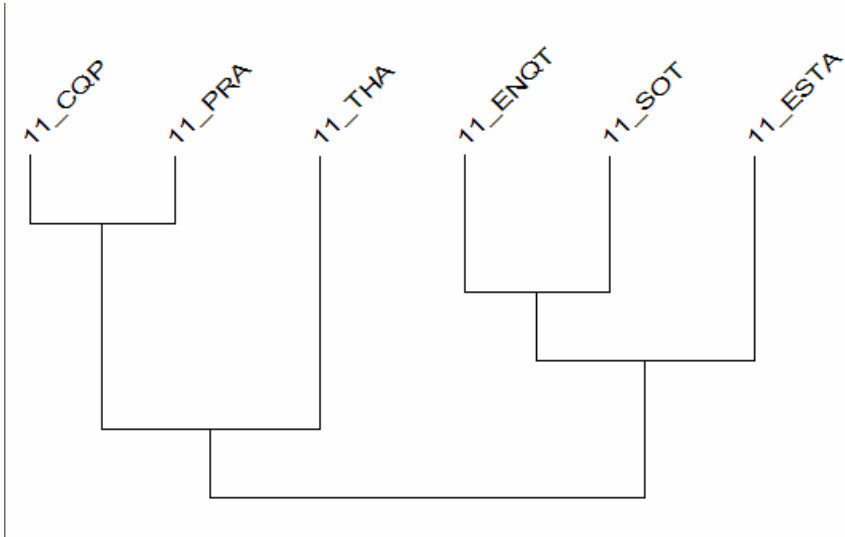


L'implication la plus forte est entre les variables « CQA » et « PRT », c'est-à-dire que si une personne répond que « Contrôle qualité » représente assez bien la statistique alors dans au moins 85 % des cas elle répond également que « Probabilité » représente très bien la statistique.

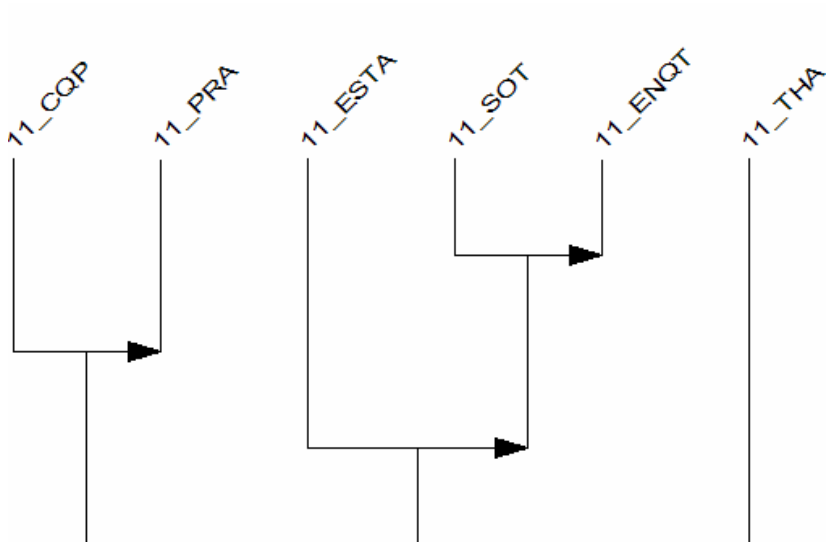
La deuxième implication est que si une personne interrogée trouve que « Enquête » représente peu la statistique alors elle trouve aussi que « Sondage » représente peu la statistique. Il se peut également, avec une implication plus faible, que cette personne trouve que « Contrôle qualité » représente assez bien la statistique.

5.3.13. Etude de la dernière partie de l'arbre des similarités (arbre C)

5.3.13.1. Arbre des similarités

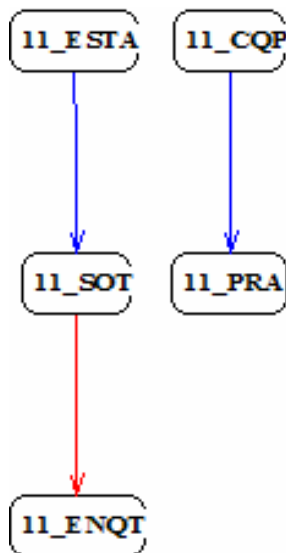


5.3.13.2. Arbre cohésitif :



tel-00191166, version 1 - 27 Nov 2007

5.3.13.3. Graphe implicatif :



→ En rouge les implications les plus fortes (entre 95 et 100%)

→ En bleu celles comprises entre 95 et 75%.

5.3.13.4. Analyse

D'après ces trois graphiques, on peut dire que les variables « Contrôle qualité, peu » et « Probabilité, assez » sont liées (0,989).

Effectivement les étudiants de 2^{nde} année STID de Lyon, trouvant que le terme « Contrôle qualité » est peu représentatif de la statistique, trouvent alors que le mot « Probabilité » est assez représentatif de celle-ci. On peut voir aussi que la réponse « Sondage, très » implique fortement la réponse « Enquête très » (0,963). Sur le graphe implicatif on peut voir que les personnes répondant que le mot « Estimation » est assez représentatif de la statistique répondent alors que « Sondage » est très représentatif.

5.3.14. Graphe implicatif pour toutes les modalités de la question 11

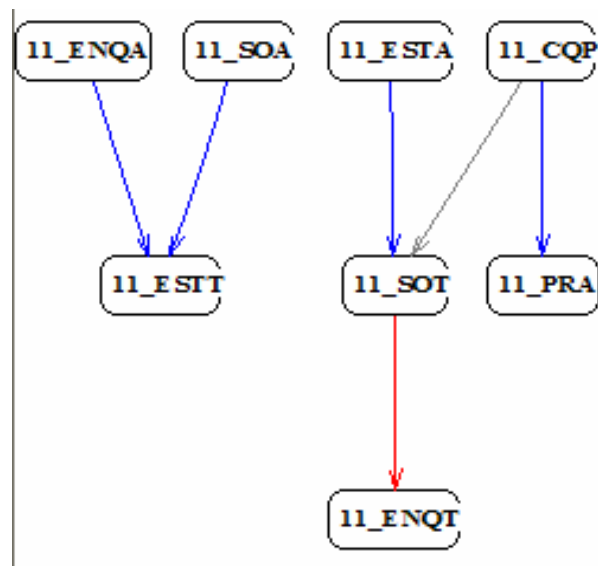
→ En rouge les implications les plus fortes (entre 95 et 100%)

→ En bleu celles comprises entre 95 et 90%.

→ En vert celles comprises entre 90 et 85%.

→ En gris celles comprises entre 85 et 80%

Remarque : Nous n'avons gardé que les liens supérieurs à 80%



5.3.15. Conclusion générale concernant le traitement de la question 11

L'implication la plus forte s'effectue donc entre « Sondage, très » et « Enquête, très ». Ces 2 variables sont donc très proches. A partir de cette observation on peut dire que les étudiants STID de 2^{de} année pensent que les mots « Sondage » et « Enquête » sont très représentatifs de la statistique. Ensuite on peut voir 4 implications comprises entre 90 et 95 %. « Enquête, assez » et « Sondage, assez » impliquent « Estimation, très ». C'est-à-dire que lorsque les interrogés répondent à la fois que « Enquête » et « Sondage » sont assez représentatif de la statistique alors ils pensent que « Estimation » représente très bien la statistique. « Estimation, assez » implique « Sondage, très » lorsque les étudiants STID de 2^{de} année pensent que « Estimation » est assez représentatif de la statistique alors il y a des chances pour qu'ils pensent que « Sondage » soit très représentatif de la statistique.

5.3.16.Question 13 : notions plus ou moins bien comprises dans le cours de statistique

Parmi les notions visées par le cours de statistique, quelles sont celles que vous pensez avoir : assez bien comprises ? Plutôt mal comprises ?

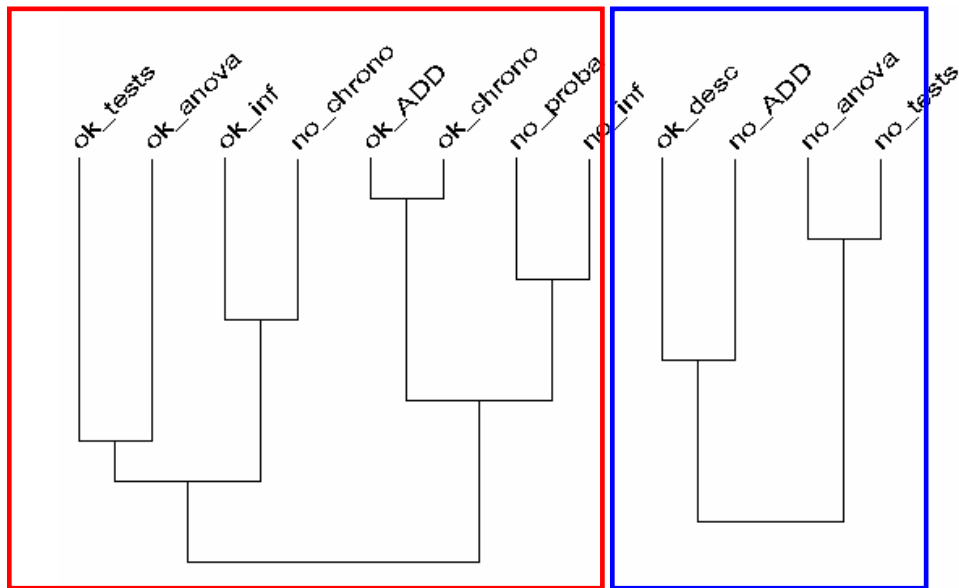
5.3.16.1.Recodification

Nous avons recodé les différentes réponses en 12 variables binaires :

	recodifications	Notions complètes
Notions assez bien comprises	Ok_tests	tests
	Ok_anova	ANOVA
	Ok_inf	Statistique inférentielle
	Ok_ADD	Analyse Des Données
	Ok_chrono	Séries chronologiques
	Ok_desc	Statistique descriptive
Notions plutôt mal comprises	No_chrono	Séries chronologiques
	No_proba	probabilité
	No_inf	Statistique inférentielle
	No_ADD	Analyse Des Données
	No_anova	ANOVA
	No_tests	tests

5.3.16.2.Etude de la similarité :

5.3.16.2.1.L'arbre des similarités



Arbre de similarité de la question 13

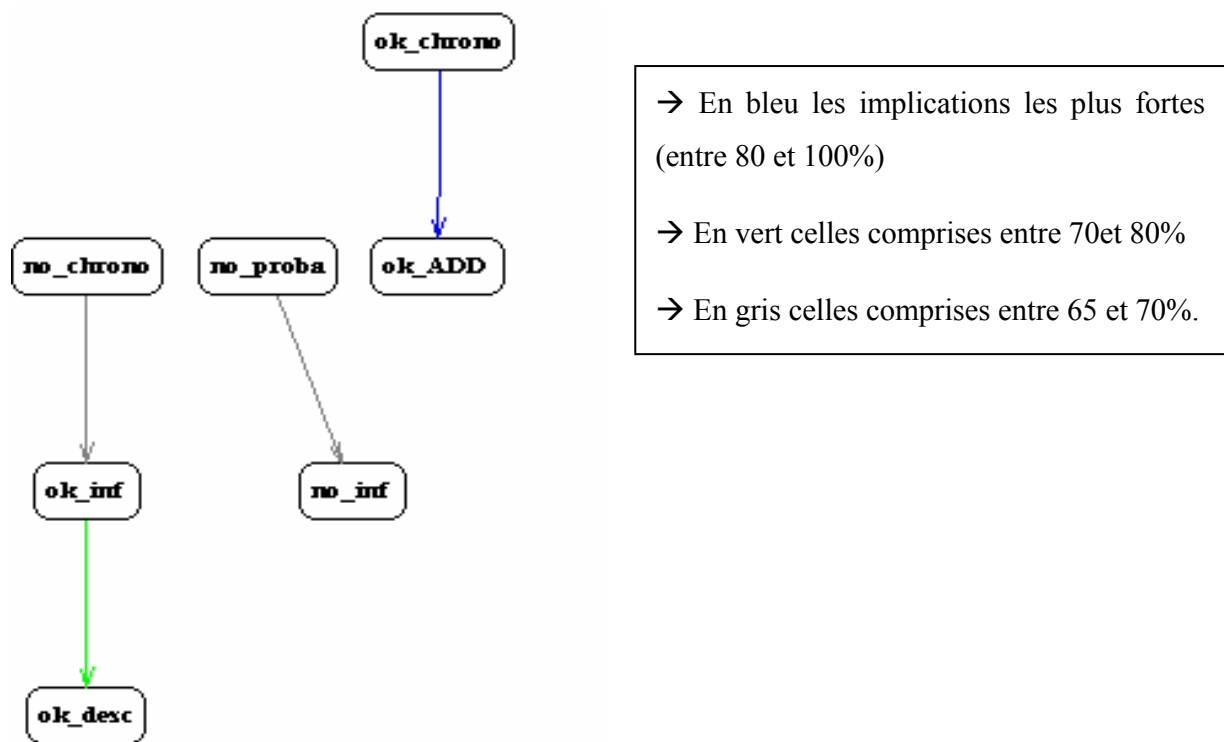
5.3.16.2.2. Analyse

On peut voir sur cet arbre des similarités, que les personnes qui ont compris les cours « d'ADD » ont également compris les cours de « Séries chronologiques », à un niveau de significativité élevé (0,998). De même, ceux qui n'ont pas compris les cours de « Probabilité » n'ont pas compris les cours de « Statistique inférentielle » (0,934).

On remarque également que ceux qui n'ont pas bien assimilés les « Tests », n'ont pas bien assimilés « l'ANOVA » (0,939). En revanche, les personnes ayant bien compris « l'ANOVA », ont bien compris les notions visées en cours de « Tests » mais à un niveau de significativité beaucoup moins élevé (0,527).

5.3.16.3. Etude de l'implication

5.3.16.3.1. Le graphe implicatif



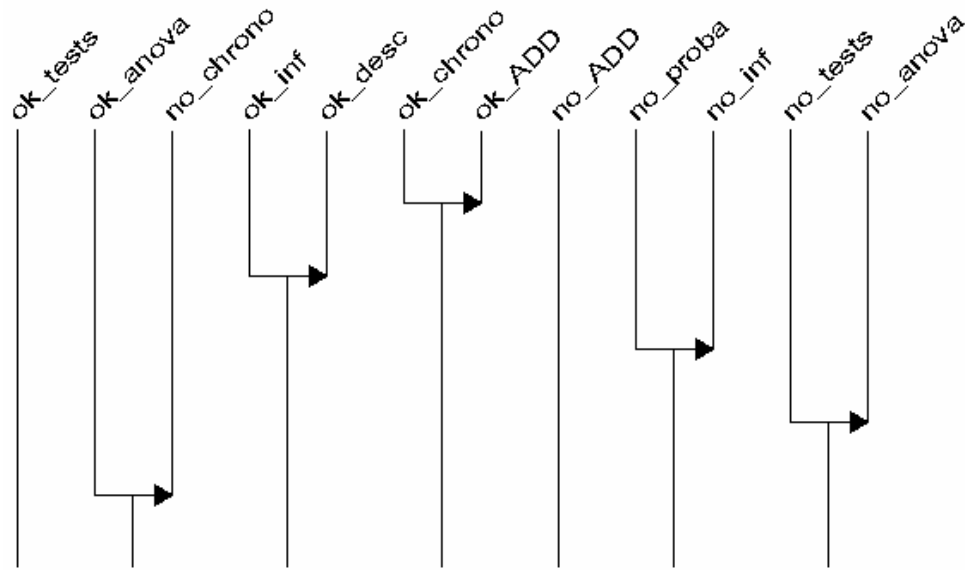
Graphe implicatif de la question 13

5.3.16.3.2. Analyse

Sur ce graphique nous pouvons voir que si une personne interrogée a compris les cours de « Séries chronologiques » cela implique fortement qu'elle ait compris les cours « d'ADD ». De même, si une personne a compris les cours de « Statistique inférentielle », cela implique assez fortement qu'elle ait également compris les cours de « Statistique descriptive ». Les autres implications ne sont pas assez significatives (le pourcentage d'implication est trop faible) pour être interprétées.

5.3.16.4. Etude de la cohésion

5.3.16.4.1. Arbre



Arbre cohésitif de la question 13

5.3.16.4.2. Analyse

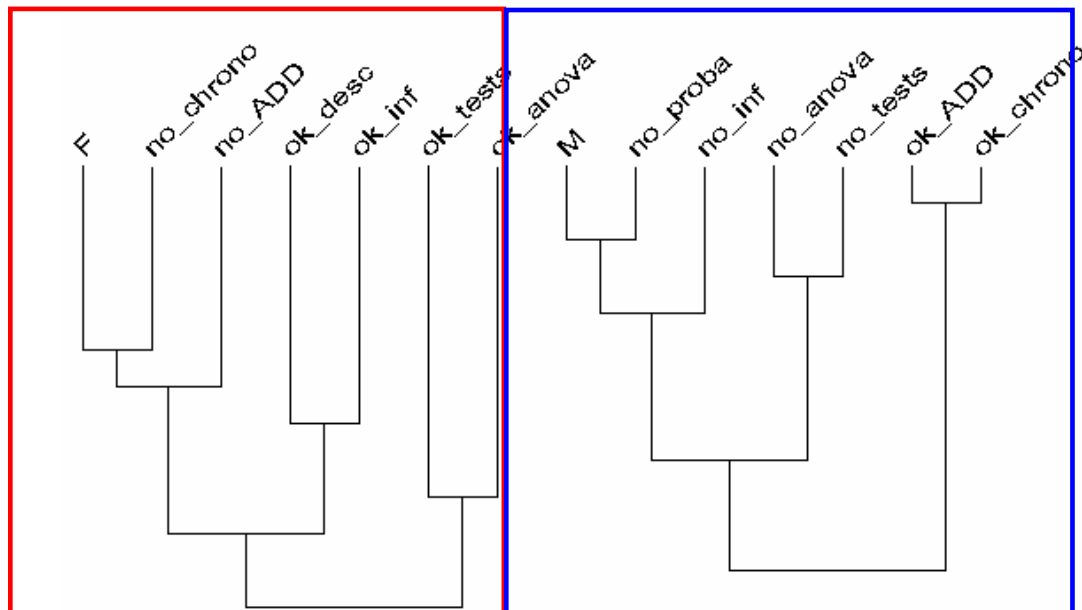
Sur ce graphique, on peut observer que c'est le fait d'avoir bien compris les cours de « Séries chronologiques » qui implique que l'on a bien compris les cours « d'ADD » (à un niveau de significativité de 0,72). Si l'on a bien compris les cours de « Statistique inférentielle », on a des chances d'avoir compris les cours « Statistique descriptive » (0,522).

5.3.17. Croisement avec les variables du sexe M (masculin) et F (féminin)

5.3.17.1. Introduction

Nous allons croiser les variables de la question 13 avec les variables de M et F. Nous espérons de cette manière pouvoir mettre en avant des différences au niveau des compréhensions de cours entre les filles et les garçons.

5.3.17.2. Arbre des similarités



Arbre de similarité

5.3.17.3. Analyse

Ce graphique nous montre les grandes tendances que l'on peut caricaturer. Ainsi : les filles n'ont pas compris les cours de « Séries chronologiques » et « d'ADD ». Mais elles ont compris les cours de « Statistique descriptive », de « Statistique inférentielle », de « Tests » et « d'ANOVA ». En revanche les garçons, n'ont pas compris les cours de « Probabilité », de « Statistique inférentielle », « d'ANOVA » et de « Tests ». Mais ils ont compris les cours « d'ADD » et de « Séries chronologiques ». On peut voir ici une opposition nette entre filles et garçons.

5.3.17.4. Conclusion de la question 13

In fine, sur cette question on ne remarque pas d'implication très forte (plus de 80%), cependant on peut remarquer que les filles et les garçons n'ont pas assimilés les mêmes cours. En effet, les filles semblent avoir mieux assimilés les cours de statistiques que les garçons. Nous pouvons supposer que les garçons ont peut être plus assimilé les cours d'informatique. On retrouve une implication qui paraît être assez logique : si on a compris les cours de

« Statistique inférentielle » c'est que l'on a, au préalable, compris ceux de « Statistique descriptive ».

5.3.18.Question 14 : difficultés majeures

5.3.18.1.Recodification

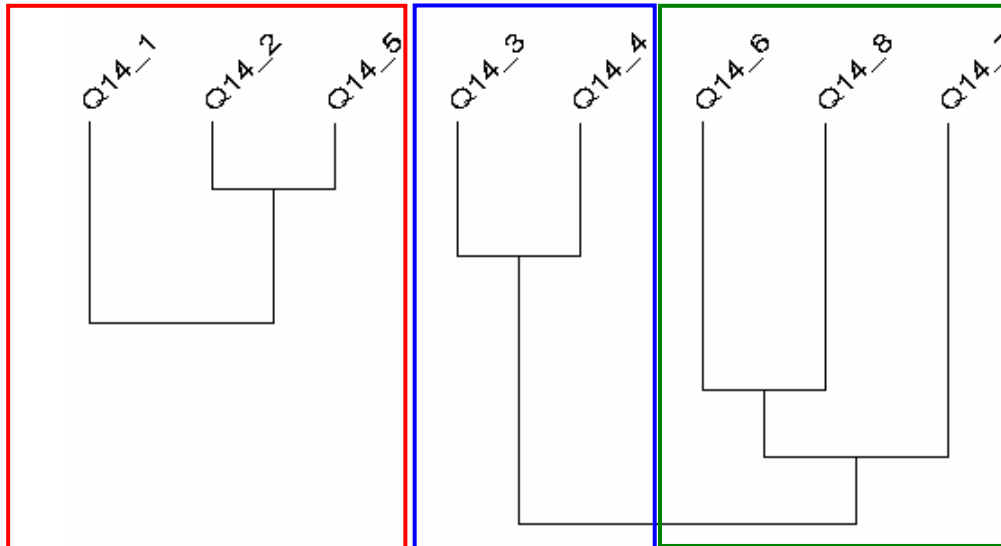
Quelles sont les trois difficultés majeures que vous pensez devoir affronter dans le monde du travail en statistique ?

- Q14_1 Réaliser seul une étude complète
- Q14_2 Appliquer les acquis théoriques
- Q14_3 Avoir la capacité d'augmenter ses connaissances
- Q14_4 Choisir la bonne méthode statistique
- Q14_5 Faire de l'inférence
- Q14_6 S'adapter aux logiciels inconnus
- Q14_7 Manquer de crédibilité face aux manques d'expériences
- Q14_8 Autre

Sur le questionnaire d'origine, nous avons une question avec 8 modalités et maintenant, nous avons 8 variables binaires, afin d'effectuer le traitement sous CHIC

5.3.18.2. Etude de la similarité

5.3.18.2.1. Arbre de similarité :



Arbre de similarité de la question

5.3.18.2.2. Analyse

On constate 3 classes de modalités distinctes :

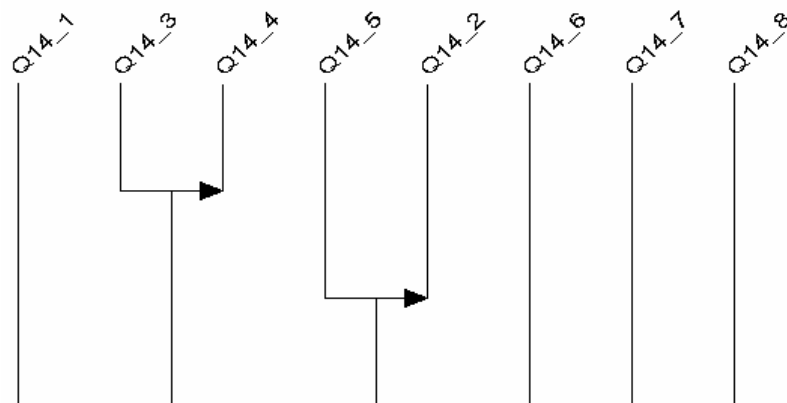
- Une première classe qui regroupe les personnes interrogées qui pensent avoir des difficultés pour « Réaliser seule une enquête ». Ces personnes ont également répondu penser avoir des difficultés pour « Appliquer les acquis théoriques » et « Faire de l'inférence ». Ces 3 items représentent d'ailleurs la capacité à appliquer dans le monde professionnel des connaissances purement théoriques.
- Un autre groupe se détache : les personnes interrogées qui ont répondu qu'elles pensaient avoir des difficultés pour « Avoir la capacité d'augmenter leurs connaissances » ont répondu qu'elles pourraient avoir des difficultés pour « Choisir la bonne méthode statistique ».
- Enfin, un 3ème groupe apparaît : les personnes qui pensent avoir des difficultés pour « S'adapter aux logiciels inconnus » sont aussi les personnes ayant la

crainte de « Manquer de crédibilité face aux manques d'expériences ». Nous retrouvons ici la crainte de rentrer dans ce monde professionnel nouveau en étant débutant, d'être trop inexpérimentés pour être pris au sérieux. Nous sommes en présence de valeurs plus professionnelles.

Nous pouvons remarquer que les classes 2 et 3 sont liées, mais l'implication est plus faible.

5.3.18.3. Etude de la cohésion

5.3.18.3.1. L'arbre cohésitif :



Arbre cohésitif de la question 14

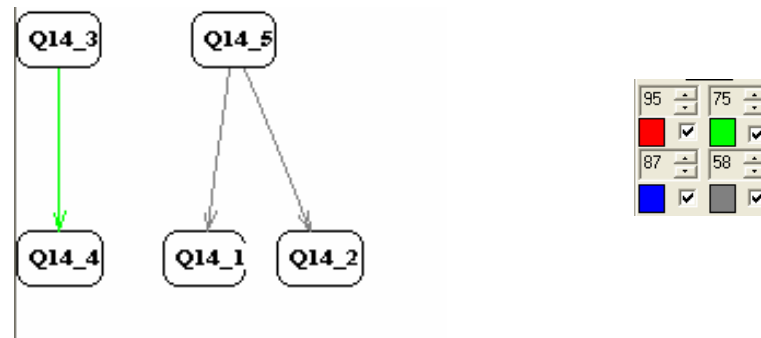
5.3.18.3.2. Analyse

Cet arbre nous montre que les étudiants de 2nde année STID ayant répondu penser avoir des difficultés pour « Avoir la capacité d'augmenter ses connaissances » sont aussi ceux ayant répondu penser avoir des difficultés pour « Choisir la bonne méthode statistique ». L'indice d'implication entre ces deux variables binaires est de 0,613.

Une autre implication, plus faible (0,302), nous montre que les personnes ayant la crainte de « Faire de l'inférence » sont aussi celles ayant peur « D'appliquer leurs acquis théoriques ».

5.3.18.4. Etude de l'implication

5.3.18.4.1. Graphe implicatif :



Graphe implicatif de la question 14

5.3.18.4.2. Analyse

Nous remarquons que l'item Q14_3 (« Avoir la capacité d'augmenter ses connaissances ») est lié à 75 % à l'item Q14_4 (« Choisir la bonne méthode statistique »). Les personnes ayant la crainte de ne pas pouvoir élargir leurs connaissances dans leur vie active sont à 75 % des personnes qui ont peur de ne pas choisir la bonne méthode statistique.

Nous n'expliquons pas les implications inférieures à 75 % qui ne sont pas pertinentes.

5.3.18.5. Conclusion de la question 14

Nous avons remarqué une réelle séparation des étudiants de 2ème année STID entre ceux qui ont des craintes par rapport aux notions théoriques déjà acquises (ceux qui pensent avoir des difficultés à augmenter leur connaissances une fois entrés dans le monde professionnel, ceux qui ont peur de ne pas appliquer la bonne méthode statistique...), et ceux qui ont des craintes plus en référence au monde du travail, à l'inconnu professionnel (la peur de manque d'expériences, la crainte de ne pas savoir utiliser des logiciels inconnus...).

5.3.19. Question 15 : les apports majeurs que les étudiants attendent d'une formation STID

5.3.19.1. Introduction

La question administrée aux étudiants STID de Lyon de seconde année était la suivante : Quels sont les trois apports majeurs que vous attendez de la formation STID ? Cette

question étant ouverte, nous avons regroupé les réponses qui se ressemblaient. Nous obtenons donc 8 variables binaires énumérées ci-dessous.

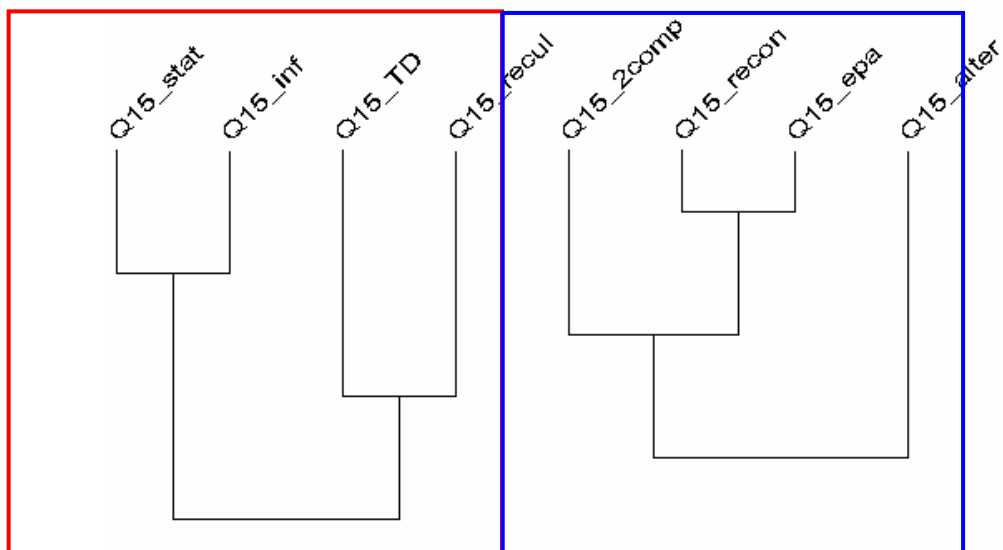
5.3.19.2. Recodification des variables

Nous obtenons donc les variables suivantes :

stat	Connaissance en stat
2comp	Double compétence
Inf	Connaissance en info
Recon	Diplôme reconnu
Alter	Insertion + alternance
TD	Méthode
recul	Prise de recul
Epa	Epanouissement

5.3.19.3. Etude de la similarité

5.3.19.3.1. Arbre des similarités et indice de similarité



Arbre de similarité de la question 15

Classification au niveau : 1 : (Q15_recon Q15_epa) similarité : 0.885334

Classification au niveau : 2 : (Q15_stat Q15_inf) similarité : 0.858234

Classification au niveau : 3 : (Q15_2comp (Q15_recon Q15_epa)) similarité : 0.781007

Classification au niveau : 4 : (Q15_TD Q15_recul) similarité : 0.516282

5.3.19.3.2.Analyse :

On observe deux groupes distincts que nous allons définir ci-dessous :

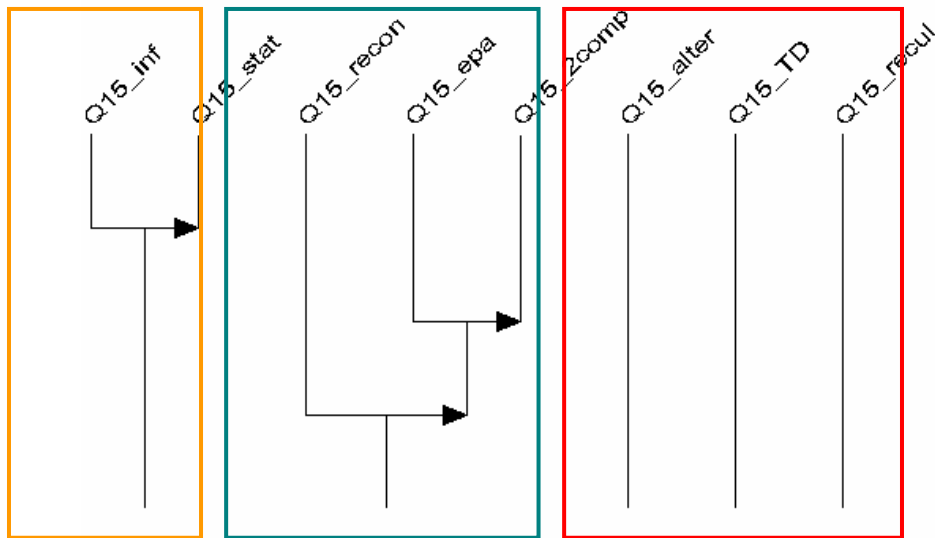
Il y aurait une similarité des variables « Epanouissement » et « Reconnaissances du diplôme » (0,885). La variable « Double compétence » s'ajoute à cette similarité pour former un premier groupe (0,781).

Il y aurait également un lien logique : un étudiant STID attend autant de connaissances en informatiques qu'en statistiques (0,858). Les variables « Méthode » et « Prise de recul » sont également mais faiblement liées (0,516).

Les étudiants attendent de la formation STID des connaissances en informatique ainsi que des connaissances en statistique. Ceux qui espèrent avoir une reconnaissance de leur diplôme espèrent aussi s'épanouir.

5.3.19.4.Etude de la cohésion

5.3.19.4.1.Arbre cohésitif et indices de cohésion



Arbre cohésitif de la question 15

Classification au niveau : 1 : (Q15_inf Q15_stat) cohésion : 0.703

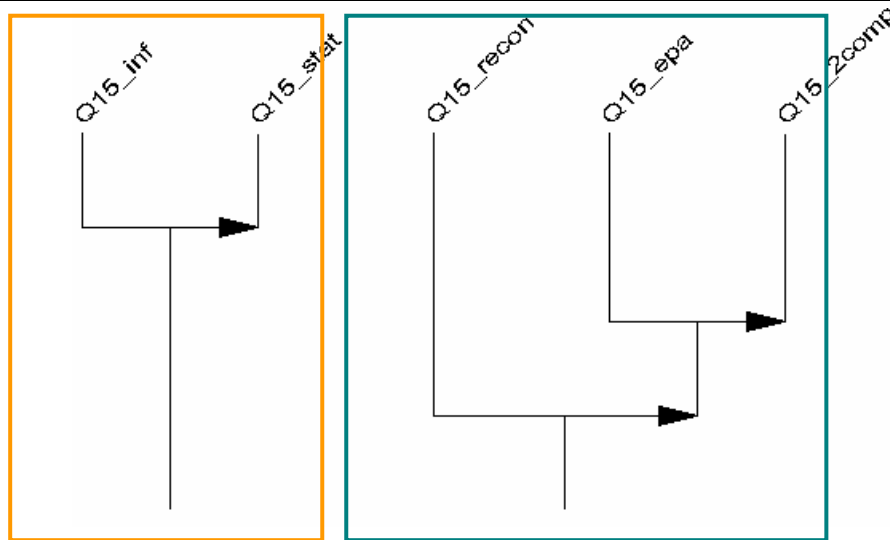
Classification au niveau : 2 : (Q15_epa Q15_2comp) cohésion : 0.412

Classification au niveau : 3 : (Q15_recon (Q15_epa Q15_2comp)) cohésion : 0.084

Les variables « Q15_alter », « Q15_TD » et « Q15_recul » n'apportent pas suffisamment d'informations pour être traitées. On génère un second arbre cohésitif en excluant les variables suivantes : « l'insertion professionnelle et l'alternance », « la méthodologie » et le « recul ».

5.3.19.4.2. Analyse

On obtient donc ce nouvel arbre ci-après :



Arbre cohésitif de la question 15 sans les variables ne donnant pas assez d'informations

Cet arbre cohésitif peut se diviser en deux sous-arbres distincts. Cependant, un seul donne réellement de l'information.

Pour les individus de la population interrogée, « Attente des connaissances en informatique » impliquent qu'ils attendent des « Connaissances en statistique » (niveau de cohésion de 0,703)

Les étudiants ayant répondu qu'ils attendaient de « l'épanouissement » ont également répondu qu'ils attendaient « Double compétence » (0,412).

5.3.20.Question 17

Quelles sont les trois suggestions majeures que vous feriez a priori pour faciliter l'apprentissage de la statistique dans cette formation STID ?

5.3.20.1.Recodification

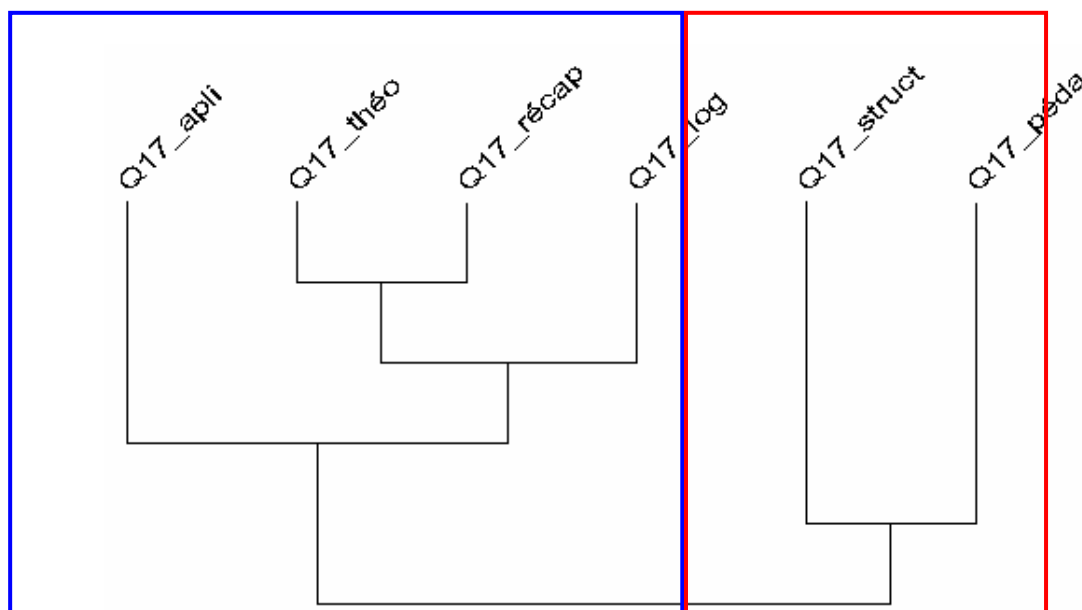
Nous avons du recoder cette question ouverte, selon les réponses retournées. Les différentes recodifications sont les suivantes :

Apli	Plus d'applications pratiques et concrètes
Struct	Cours plus structurés
Theo	Avoir davantage de théories

Recap	Avoir des cours récapitulatifs
Log	Utiliser plus régulièrement des logiciels de statistique
Péda	Faire plus de sorties pédagogiques

5.3.20.2. Etude de la similarité

5.3.20.2.1. Arbre de similarité :



Arbre de similarité de la question 17

5.3.20.2.2. Analyse

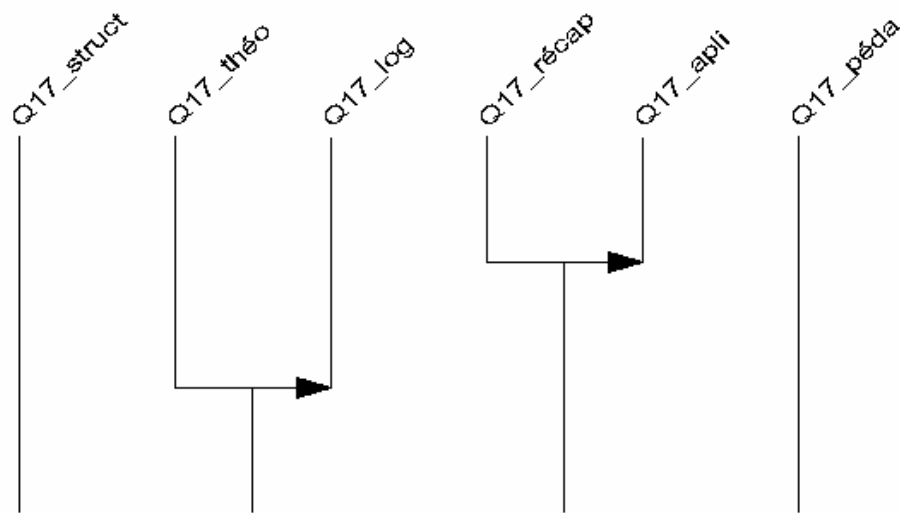
Deux classes se détachent :

Les étudiants de 2^{nde} année STID demandeurs de « Plus de théories » sont aussi ceux qui demandent plus de « Fiches récapitulatives de cours ». Ce groupe d'étudiants souhaite également avoir « Plus d'applications sur les logiciels, d'exercices appliqués sur logiciels ».

Une autre classe montre que des étudiants désirant des « Cours plus structurés » sont aussi les étudiants souhaitant « Plus de pédagogie » de la part des professeurs.

5.3.20.3. Etude de la cohésion

5.3.20.3.1. L'arbre cohésitif :



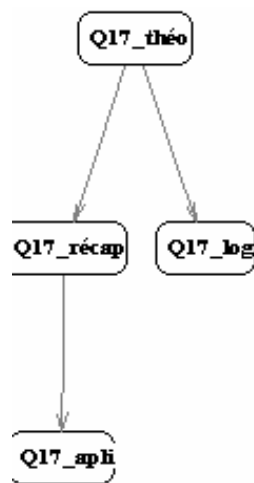
Arbre cohésitif de la question 17

5.3.20.4. Analyse

Les étudiants demandant « Plus de théories » demandent « Plus d'applications concrètes sur logiciels ». L'intensité d'implication est de 0,0604, donc on peut dire que l'implication entre ces deux variables est très faible.

Les étudiants désirant « Plus de fiches récapitulatives de cours » sont aussi demandeurs de « Plus d'applications pratiques et concrètes ». L'implication entre les deux variables est de 0,089, donc une implication légèrement plus forte.

5.3.20.5. Etude de l'implication



Graphe implicatif de la question 17

Les implications sont trop faibles, car aucune n'est supérieure à 50%. Nous ne les interpréterons pas.

5.3.20.6. Conclusion de la question 17

Il y a deux types de demandes, qui sont contradictoires : des étudiants qui demandent plus d'exercices pratiques, plus d'applications sur logiciels (variable liée à une demande de plus de théories d'ailleurs). Et d'autres qui demandent plus de démonstrations mathématiques, des connaissances plus théoriques sur les statistiques.

5.3.21. Question 21

Si vous aviez, demain, un poste en statistique, dans une entreprise, pensez-vous être capable de :

5.3.21.1. Rappel de la question

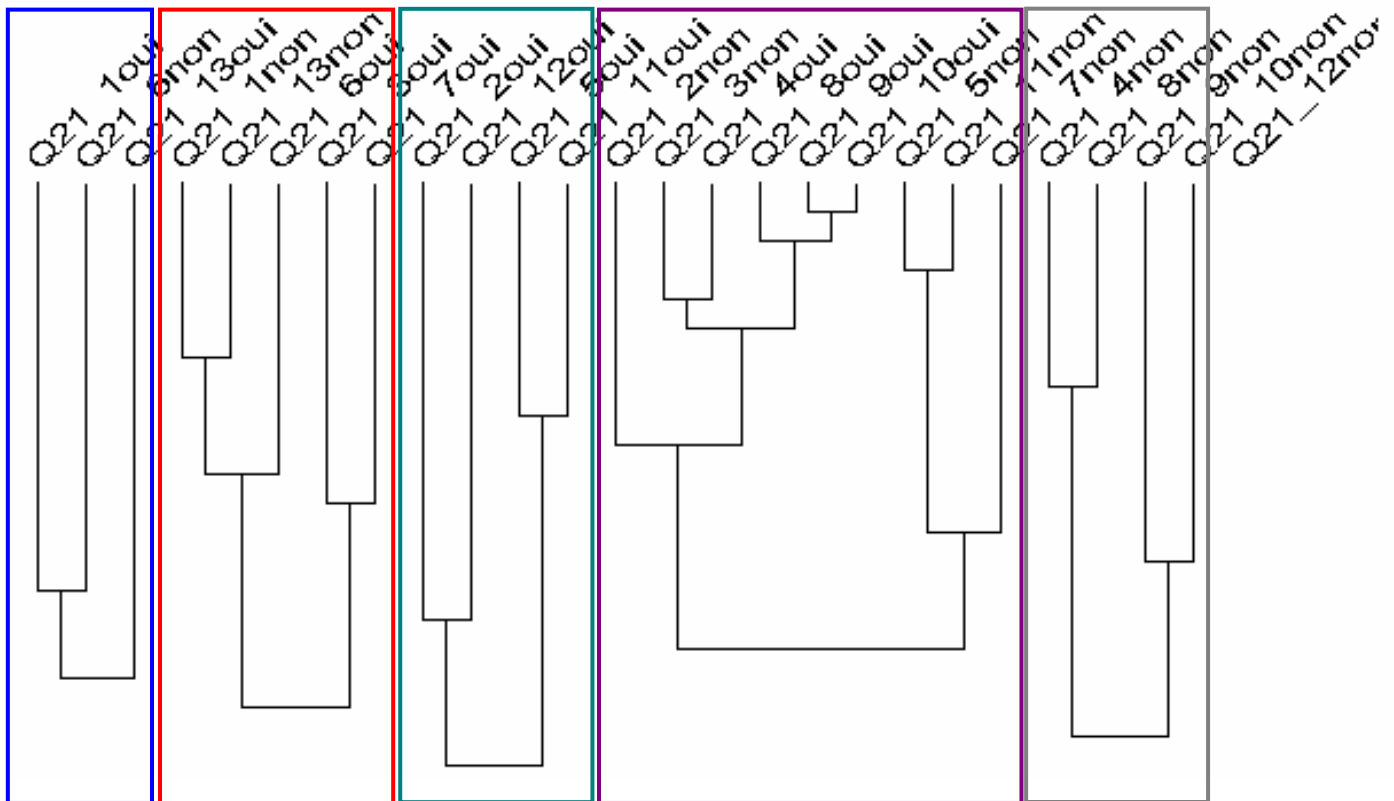
-
- 1 Appliquer des méthodes inférentielles
 - 2 Créer des indicateurs
 - 3 Créer des tableaux de bord
 - 4 Effectuer un contrôle qualité
 - 5 Elaborer un questionnaire

- 6 Faire de la cartographie
- 7 Faire du datamining
- 8 Mettre en place des cartes de contrôle
- 9 Mettre en place un plan d'expériences
- 10 Mettre en place un SPC
- 11 Réaliser seul une enquête
- 12 Traiter un questionnaire
- 13 Utiliser un logiciel de statistique

Les interrogés devaient, pour chacun des 13 intitulés de mission ci-dessus, répondre s'il serait capable ou non d'effectuer cette mission en entreprise. Nous avons donc, pour cette question, créé 26 modalités de variables distinctes (2 pour chaque missions : capable ou pas capable). Avant le recodage sous CHIC, nous avons 13 variables binaires et après le recodage, nous avons 26 variables binaires. Chaque variable est nommé de la façon suivante : Q21_ + numéro de la mission + (oui pour capable ou non pour incapable).

5.3.21.2. Etude des similarités de la question 21

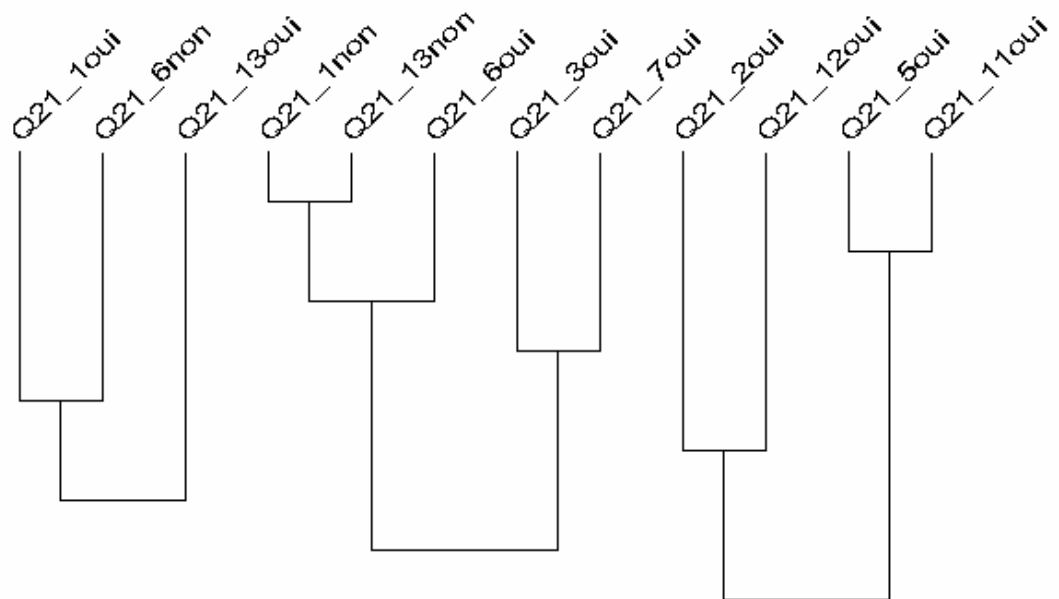
Nous avons, tout d'abord, croisé toutes ces variables dans un arbre de similarité et nous avons obtenu le graphique suivant :



Arbre de similarités de la question 21

Ce graphique est très peu lisible, nous allons donc traiter ce graphique en plusieurs parties. Nous voyons bien, quand nous regardons le graphique, qu'il se sépare en plusieurs « petits » arbres distincts (le dernier n'étant formé que de la variable Q21_12non car elle est nulle, ce qui signifie que tous les étudiants se sentent capables de traiter un questionnaire).

5.3.21.3. Etude de la première partie de l'arbre des similarités



Première partie de l'arbre de similarités

Nous voyons, grâce à la première partie du graphique, que la plus forte similarité entre variables est celle entre « Q21_1non » et « Q21_13non », cela signifie que les étudiants en 2^{de} année STID de Lyon qui ne se sentent « Pas capable d'appliquer des méthodes inférentielles » sont aussi « Incapables d'utiliser un logiciel statistique » au seuil de 0,94.

La seconde plus forte similarité que l'on observe, qui est de 0,80, nous indique que les étudiants qui se sentent « Capables d'élaborer un questionnaire » sont aussi « Capables de réaliser seul une enquête ».

La troisième similarité à mettre en valeur est celle entre « Q21_6oui » et « (Q21_6non ; Q21_13non) », elle est de 0,78. Elle signifie que les étudiants qui ne se sentent capable ni d'appliquer des méthodes inférentielles, ni d'utiliser un logiciel statistique, se sentent capable de faire de la cartographie.

5.3.22. Etude de la seconde partie de l'arbre des similarités

Nous allons étudier maintenant la seconde partie de l'arbre de similarité : nous voyons donc bien deux sous arbres distinct. Dans le premier, nous observons une similarité parfaite entre « Q21_9oui » et « Q21_10oui » car l'indice de similarité vaut 1, et une similarité très forte (0,93) entre ces 2 dernières variables et « Q21_8oui ».

Ceci signifie que les étudiants se sentant « Capable de mettre en place un plan d'expérience » se sentent aussi « Capable de mettre en place un SPC » et « Capable de mettre en place des cartes de contrôles ». Cette conclusion semble logique car ces étudiants sont très sûrement ceux qui ont des missions touchant à la qualité industrielle dans leur entreprise.

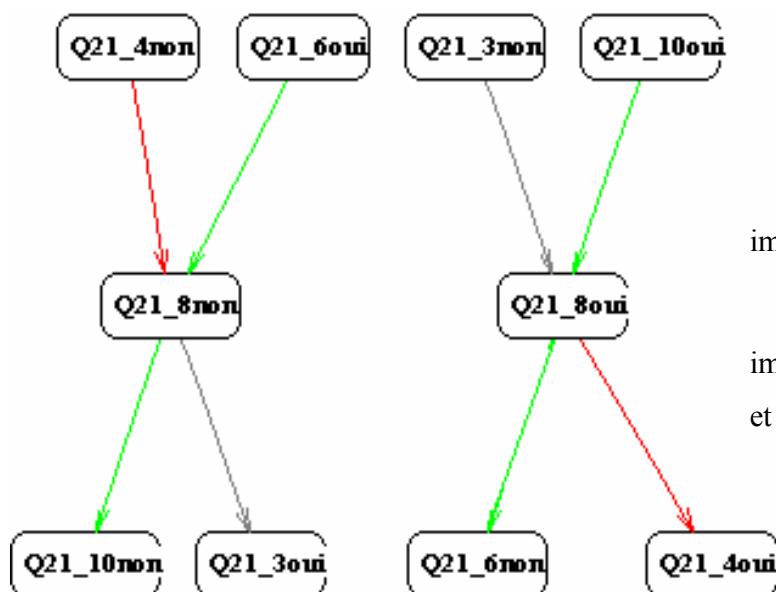
On remarque aussi, dans ce sous arbre, une forte similarité entre « Q21_5non » et « Q21_11non », cette relation est l'inverse de celle vu dans la première partie de l'arbre.

Dans le second sous arbre, nous observons que la plus forte similarité (0,86) est entre les étudiants se ne sentant « Pas capable de réaliser un contrôle de qualité » et ceux ne se sentant « Pas capable de mettre en place des cartes de contrôles ».

5.3.22.1. Etude de l'implication

5.3.22.1.1. Graphique implicatif

Nous allons, dans un second temps, interpréter le graphique implicatif obtenu grâce au croisement de toutes les modalités de la question 21.



Niveau d'implication :

→ En rouge : les implications supérieures à 95 %

→ En vert : les implications comprises entre 85 et 90 %

→ En gris : les

Graphique implicatif de la question 21

5.3.22.1.2. Analyse

Nous observons que le graphique est divisé en deux parties distinctes qui s'opposent, elles s'interprètent donc de la même façon. Nous allons donc interpréter seulement la première partie du ce graphique.

L'implication la plus importante représentée par la flèche rouge signifie que « Q21_4non » implique très fortement « Q21_8non » (niveau d'implication : >95%). Cela signifie que les étudiants ne se sentant pas capable d'effectuer des contrôles de qualité ne se sentent pas capable non plus de mettre en place des cartes de contrôles.

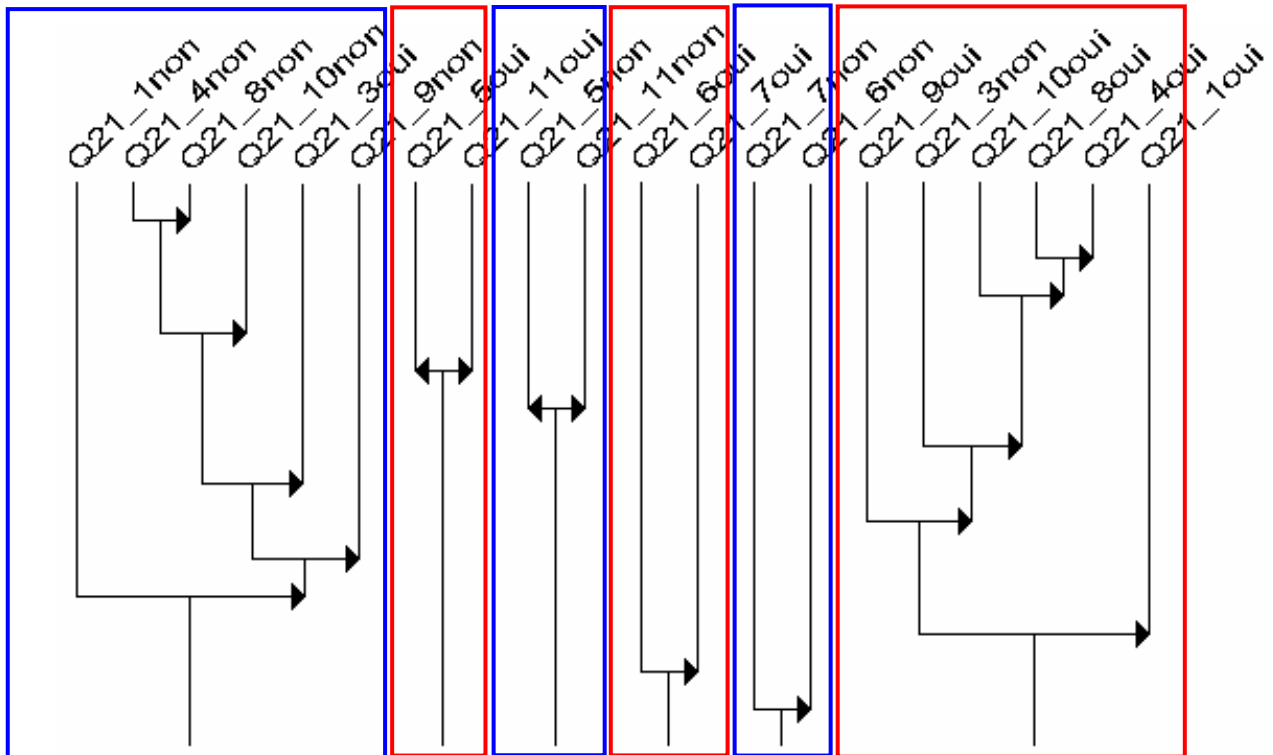
On peut également conclure que les étudiants ne se sentant « Pas capable de mettre en place des cartes de contrôles » ne sentent aussi « Pas capables de mettre en place un SPC » (niveau : >90%).

Mais on peut également compléter : les étudiants qui se sentent « Capable de faire de la cartographie » ne se sentent « Pas capable de mettre en place des cartes de contrôles » (niveau : <85%), ce qui semble logique car la cartographie et la qualité sont des domaines opposés.

5.3.22.2. Etude de la cohésion

5.3.22.2.1. Arbre de Cohésion

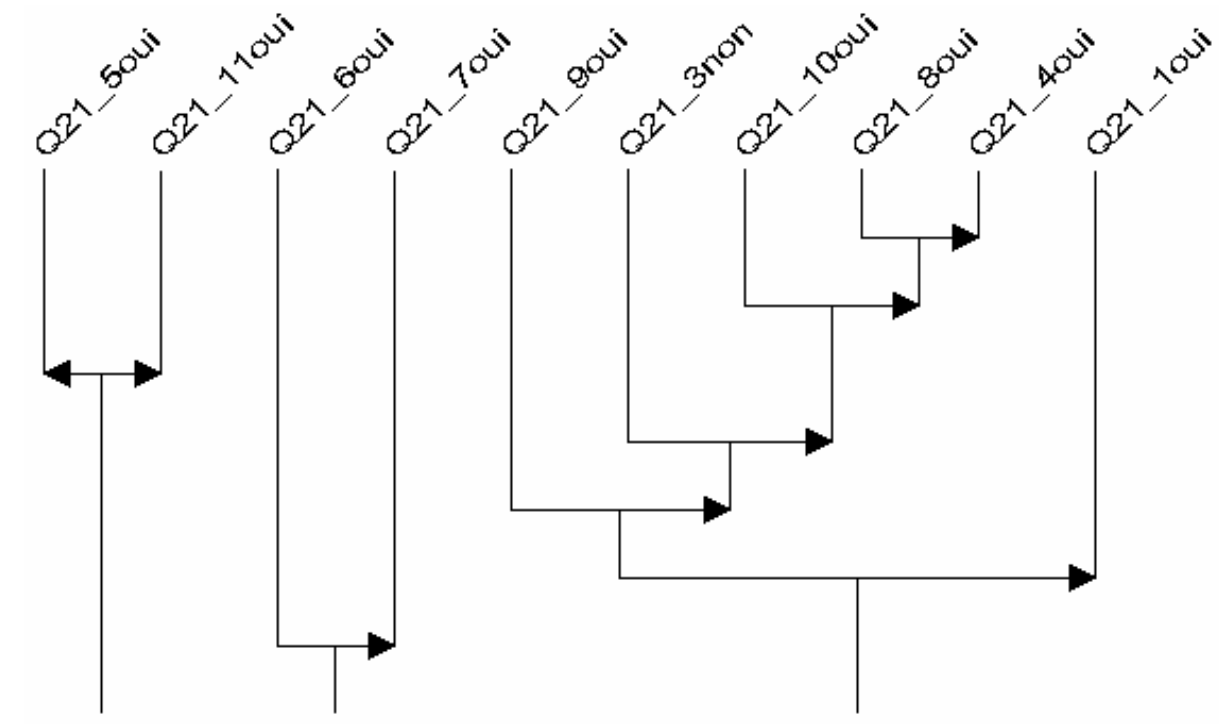
Nous allons maintenant étudier l'arbre de cohésion, nous n'avons laissé sur l'arbre que les modalités qui avaient un rapport de cohésion avec une autre modalité.



Arbre de cohésion de la question 21

5.3.22.2.2.Analyse

Nous observons que cet arbre est « symétrique », comme le graphique d'implication.



Les sous arbres encadrés en rouges s'opposent aux sous arbres encadrés en bleus. Nous allons traiter ci-dessous, seulement les parties encadrées en rouge.

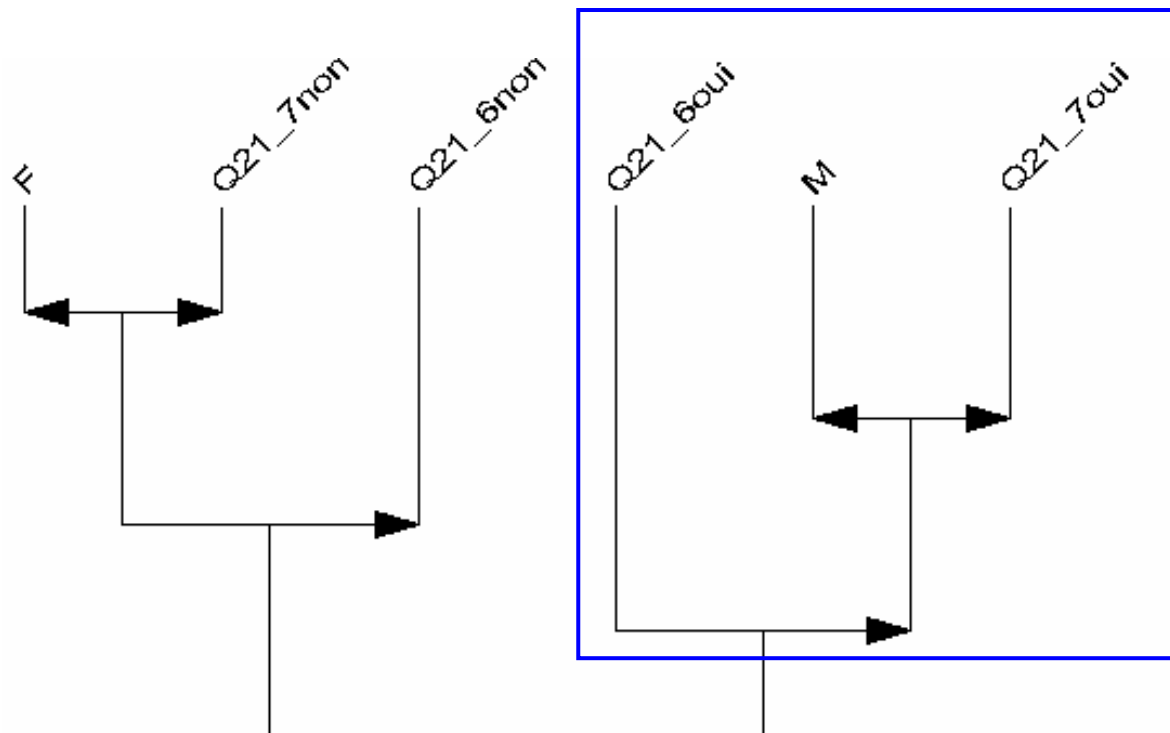
Nous observons que la cohésion la plus forte sur cet arbre est celle entre les variables « Q21_8oui » et « Q21_4oui » (niveau de cohésion de 0,98), cela signifie que les étudiants en 2^{de} année de DUT STID à Lyon qui se sentent capables de mettre en place des cartes de contrôle, se sentent également capables d'effectuer un contrôle de qualité. Nous observons aussi une cohésion entre « Q21_10oui » et « (Q21_oui ;Q21_4oui) » qui nous informe que les personnes se sentant capables de mettre en place un SPC se sentent capables à la fois de mettre en place des cartes de contrôle mais aussi d'effectuer un contrôle de qualité (niveau de 0,88).

Nous pouvons également voir, à gauche de l'arbre, une cohésion réciproque entre les variables « Q21_5oui » et « Q21_11oui » avec un niveau de 0,85. Cela signifie que les étudiants qui se sentent capables d'élaborer se sentent presque tous capables de traiter seul une enquête et inversement.

5.3.23. Croisement de la question 21 avec les variables sexe M et F

Nous allons dans cette partie tenter de découvrir des relations entre le sexe d'un étudiant et le genre de mission qu'il se sent capable de réaliser en entreprise. Nous allons pour cela, dans un premier temps, interpréter l'arbre de cohésion obtenu grâce à ce croisement, puis, dans un second temps, le graphique implicatif obtenu.

5.3.23.1. Etude de la cohésion

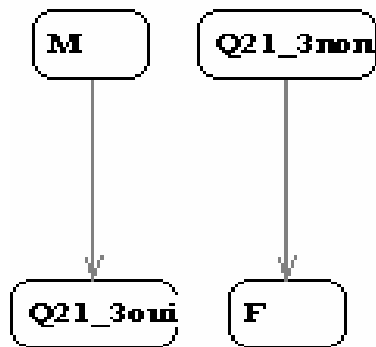


Parties de l'arbre de cohésion lors du croisement de la question 21 et du sexe

Nous observons que cet arbre est composé de deux sous arbres opposés, nous allons donc traiter seulement la partie encadrée.

Nous observons une forte cohésion réciproque entre « M » et « Q21_7oui » (niveau de cohésion : 0,71) ce qui veut donc dire que les étudiants garçons en 2^{de} année de DUT STID ont répondu majoritairement qu'ils se sentaient capables de faire du datamining. Mais nous voyons aussi que les étudiants se sentant capables de faire de la cartographie sont principalement des étudiants garçons qui se sentent capables de faire du datamining en entreprise.

5.3.23.2. Etude de l'implication



Niveau d'implication :
 → En gris, les implications comprises entre 75 et 80 %

Parties du graphique d'implication résultant du croisement entre la question 21 et le sexe

En regardant ce graphique, nous pouvons voir que celui-ci est composé de deux sous graphiques opposés.

Le premier sous graphique représente une implication forte (niveau : 80 %) de « M » sur « Q21_3oui », le second représente une implication de même importance de « Q21_3non » sur « F ».

Ces deux implications signifient que les étudiants garçons se sentent, pour la plupart, capables de créer des tableaux de bord et que les interrogés qui ne se sentent pas capables de créer des tableaux de bord sont principalement des filles.

Conclusion générale concernant le traitement de la question 21

Nous avons pu voir, grâce au traitement de cette question, les liens qui pouvaient exister entre l'idée que se font les étudiants de ces différentes missions. Par exemple, les étudiants se sentant capables d'effectuer un contrôle de qualité se sentaient également capables de réaliser les différentes autres missions du domaine de la qualité comme mettre en place un SPC ou des cartes de contrôle. Nous avons pu également constater que des personnes qui se sentaient capables de faire des missions assez spécifiques comme de la cartographie ne se sentaient pas capables de faire des missions d'un domaine spécifique différent comme la qualité. Enfin, grâce au croisement de la question avec le sexe, nous avons pu nous rendre compte que les garçons se sentaient très majoritairement capables de faire du datamining ou de créer des tableaux de bord. Et inversement, nous avons vu que les étudiantes ne se sentaient pas, pour la plupart, capables de faire du datamining, ni de la cartographie.

5.3.23.3. Conclusion

Cette étude nous a permis de mieux cerner les caractéristiques de notre population. Nous avons remarqué des implications intéressantes, telles que celle entre le sexe et le bac obtenu. Nous avons axé notre étude sur la vision de la statistique des étudiants STID 2nde année, et même si notre première idée était de voir quelle était l'implication de l'alternance dans l'apprentissage de la statistique – idée intéressante mais non réalisable en raison d'un biais trop fort, mais sans doute les résultats les plus intéressants sont à chercher dans le croisement de la variable sexe et des autres variables. Ainsi les garçons semblent plus facilement se projeter dans la cartographie ou le SPC que les filles. Autre constat, et cela est peut être un effet de l'alternance, les champs de compétences semblent déjà bien choisis (de la cartographie mais pas de la qualité par exemple).

5.3.24. Annexe 1 : Le questionnaire 2006-2007

Année universitaire 2006/2007

IUT Lumière de l'Université Lyon 2

Etude des difficultés et des facilités d'apprentissage de la statistique à l'université dans le cadre de la formation STID (Statistiques et Traitement Informatique des Données)

L'objet de cette enquête par questionnaire est de construire un corpus de données à partir duquel nous étudierons la question de la vision de la statistique selon le type de la formation STID à savoir en cursus normal ou en cursus en alternance.

Cette enquête est réalisée sous la responsabilité de Jean-Claude Oriol, chef du département STID (Statistique et Traitement Informatique des données) de Lyon

Nous vous remercions de votre coopération en tant qu'informateur, sans laquelle les informations nécessaires ne peuvent être obtenues.

Le questionnaire est à nous renvoyer à l'adresse mail suivante : missmy73@hotmail.com

Le groupe d'étudiants STID de l'IUT Lumière chargés de cette enquête

Questionnaire auprès des étudiants

1. Nom :

2. Prénom :

3. Sexe : F M (garder la mention utile)

4. Année de naissance :

5. IUT : Grenoble Lyon

6. Baccalauréat : Série : (précisez)

7. Année d'obtention :

8. Dernier diplôme préparé :

9. A l'entrée dans cette formation en IUT, qu'évoquait pour vous la statistique ? (Donnez trois mots)

-
-
-

10. A ce jour, qu'évoque pour vous la statistique ? (Donnez trois mots)

-
-
-

11. Pensez-vous que les termes suivants sont : très représentatif, assez représentatif, peu représentatif ou pas du tout représentatif de la statistique ? (cochez une case par ligne)

Termes	Très	Assez	Peu	Pas du tout
Contrôle qualité				
Enquête				
Estimation				
Probabilité				
Sondage				
Tests d'hypothèse				
Traitement des données				

12. Dans la vie sociale actuelle, quelle utilité attribuez-vous à la statistique ? (cocher une réponse)

- Très utile
- Assez utile
- Peu utile
- Pas du tout utile

13. Parmi les notions visées par le cours de statistique, quelles sont celles que vous pensez avoir (citez au plus six) :

...assez bien comprises

...plutôt mal comprises

14. Quelles sont les trois difficultés majeures que vous pensez devoir affronter dans le monde du travail en statistique ?

- Réaliser seul une étude complète
- Appliquer les acquis théoriques
- Avoir la capacité d'augmenter ses connaissances
- Choisir la bonne méthode statistique
- Faire de l'inférence
- S'adapter aux logiciels inconnus
- Manquer de crédibilité face aux manques d'expérience
- Autre : (précisez)

15. Quels sont les trois apports majeurs que vous attendez de la formation STID ?

-
-
-

16. Combien de temps aurez-vous passé en entreprise à la fin de votre formation ?

- Alternance Stage : précisez le nombre de semaines :

17. Quelles sont les trois suggestions majeures que vous feriez a priori pour faciliter l'apprentissage de la statistique dans cette formation STID ?

-
-
-

18. De quels supports de cours disposez-vous ? (cocher une ou plusieurs réponses)

Notes personnelles
Ouvrages
Consultation de sites Internet

Papier Web

Polycopiés
Diaporamas (PowerPoint...)
Autres (précisez) :

19. Pensez-vous qu'il soit nécessaire d'aller chercher d'autres ressources que les documents de cours distribués ?

- Très voir toujours nécessaire
- Assez nécessaire
- Peu nécessaire
- Pas du tout nécessaire

20. Par quoi pourriez-vous les compléter ?

Ouvrages
Consultation de sites Internet

Papier Web

Polycopiés
Diaporamas (PowerPoint...)
Autres (précisez) :

21. Si vous aviez, demain, un poste en statistique, dans une entreprise, pensez-vous être capable de :

Oui Non

Appliquer des méthodes inférentielles
Créer des indicateurs
Créer des tableaux de bord
Effectuer un contrôle qualité
Elaborer un questionnaire
Faire de la cartographie
Faire du datamining
Mettre en place des cartes de contrôle
Mettre en place un plan d'expériences
Mettre en place un SPC
Réaliser seul une enquête
Traiter un questionnaire
Utiliser un logiciel de statistique

22. Sur chaque ligne, indiquez, en cochant la case appropriée, si vous êtes :

- Tout à fait d'accord = A
- Plutôt d'accord = B
- Plutôt pas d'accord = C
- Pas du tout d'accord = D

	A	B	C	D
Une formation en statistique à l'IUT devrait rendre capable de...				
Expliciter les questions d'une problématique dont les réponses relèvent d'une approche statistique.				
Décrire, traiter, analyser des données de manière pertinente dans le cadre d'une étude en particulier dans le domaine professionnel choisi				
Faire le lien entre la réflexion analytique sur des questions relevant du champ professionnel, leur formalisation et leur traitement quantitatif				
Lire avec un regard critique et distancié, les conclusions de diverses études statistiques apparaissant dans des rapports de recherche de votre domaine professionnel				
Poursuivre de façon autonome et personnalisé un apprentissage en statistique afin d'enrichir ses acquis				
Poser un regard distancié à l'égard du domaine de la statistique, largement exploité dans les médias, dans le sens de ne pas considérer les résultats dans l'ordre du tout ou rien mais on les replaçant judicieusement dans leur domaine de validité				
Exploiter des notions et des démarches mathématiques à des fins d'outils, et de ce fait, modifier dans un sens positif le rapport souvent négatif que certains entretiennent avec la statistique				
Mettre en œuvre un raisonnement intégrant l'idée de risque d'erreur dans l'énoncé de ses conclusions				
Mettre en œuvre l'interprétation de phénomènes sur la base de données statistiques sur des faits et sur des relations entre ces faits				
Communiquer des résultats d'analyses de données en distinguant clairement le modèle utilisé de la réalité qu'il est supposé représenter, et en séparant bien les traitements menés à l'intérieur du modèle des interprétations reformulées dans le contexte du problème.				

Merci de votre coopération !

5.3.25. Annexe 2 : Le recodage de chaque question

Question 3 :

M	Masculin
F	Féminin

Question 6 :

BAC_ES	ES
BAC_S	S

Question 9 :

Math	Mathématiques
Graph	Graphiques et tableaux
Difficulté	difficulté
Enqu	Enquêtes et études de marché
%	pourcentage
Previs	Prévision
Proba	Probabilité et moyenne

Question 10 :

Test	Tests
Enqu	Enquêtes et études de marché
Inf	Statistiques inférentielles
Math	Mathématiques
ADD	Analyse de données
Desc	Statistiques descriptives
Proba	Probabilité et lois
Pop	Population et échantillon

Question 11 :

Termes	Très	Assez	Peu	Pas du tout
Contrôle qualité	CQT	CQA	CQP	CQPT
Enquête	ENT	ENQA	ENQP	ENQPT
Estimation	ESTT	ESTA	ESTP	ESTPT
Probabilité	PRT	PRA	PRP	PRPT
Sondage	SOT	SOA	SOP	SOPT
Tests d'hypothèse	THT	THA	THP	THPT
Traitement des données	TDDT	TDDA	TDDP	TDDPT

Question 13 :

Notions assez bien comprises	Ok_tests	tests
	Ok_anova	ANOVA
	Ok_inf	Statistique inférentielle
	Ok_ADD	Analyse Des Données
	Ok_chrono	Séries chronologiques

	Ok_desc	Statistique descriptive
Notions plutôt mal comprises	No_chrono	Séries chronologiques
	No_proba	probabilité
	No_inf	Statistique inférentielle
	No_ADD	Analyse Des Données
	No_anova	ANOVA
	No_tests	tests

Question 14 :

Q14_1	Réaliser seul une étude complète
Q14_2	Appliquer les acquis théoriques
Q14_3	Avoir la capacité d'augmenter ses connaissances
Q14_4	Choisir la bonne méthode statistique
Q14_5	Faire de l'inférence
Q14_6	S'adapter aux logiciels inconnus
Q14_7	Manquer de crédibilité face aux manques d'expériences
Q14_8	Autre

Question 15 :

Stat	Connaissance en stat
2comp	Double compétence
Inf	Connaissance en info
Recon	Diplôme reconnu
Alter	Insertion professionnelle + alternance
TD	Méthode
Recul	Prise de recul
Epa	Epanouissement personnel

Question 16 :

Q_16alt	Alternance
Q_16stage	Stage
Q_16_7s	7 semaines de stage
Q_16_13s	13 semaines de stage

Question 17 :

Apli	Plus d'applications pratiques et concrètes
------	--

Struct	Cours plus structurés
Theo	Avoir davantage de théories
Recap	Avoir des cours récapitulatifs
Log	Utiliser plus régulièrement des logiciels de statistique
Péda	Faire plus de sorties pédagogiques

Question 21 :

-
- 1 Appliquer des méthodes inférentielles
 - 2 Créer des indicateurs
 - 3 Créer des tableaux de bord
 - 4 Effectuer un contrôle qualité
 - 5 Elaborer un questionnaire
 - 6 Faire de la cartographie
 - 7 Faire du datamining
 - 8 Mettre en place des cartes de contrôle
 - 9 Mettre en place un plan d'expériences
 - 10 Mettre en place un SPC
 - 11 Réaliser seul une enquête
 - 12 Traiter un questionnaire
 - 13 Utiliser un logiciel de statistique

Chaque variable est nommée de la façon suivante : Q21_ + numéro de la mission + oui pour capable ou non pour incapable.

Conclusion

Tout au long de ce travail nous avons essayé de montrer comment l'enquête par questionnaires d'une part, et la simulation d'expériences aléatoires d'autre part s'inscrivent dans une perspective constructiviste de la connaissance statistique s'appuyant sur les concepts de schèmes et de situations. L'enquête par questionnaires mobilise la presque totalité des compétences en action dans le métier de statisticien : collecte des données, nettoyage des données, vérification des cohérences, choix techniques, choix technologiques, traitements, validation auprès des commanditaires. Mais l'appareillage mis en place pour effectuer une enquête est assez conséquent et ne peut donner une proximité suffisante de la variabilité qu'après de multiples enquêtes et des années d'expérience. La simulation permet aux étudiants de dégager des invariants de la variabilité, de comprendre ce que recouvre la convergence, de s'approprier la statistique inférentielle, de concevoir les fluctuations d'échantillonnage.

L'opération est délicate et nécessite d'avoir à batailler sur tous les fronts : celui des signifiés, celui des signifiants et celui des situations de référence. Dans la première enquête présentée dans la partie V de ce travail on constate que les représentations de la statistique, pour les étudiants du département QLIO qui font peu de statistique, tournent essentiellement autour du calcul numérique et qu'en revanche elles sont multi référentielles pour les étudiants du département STID qui ont un emploi du temps chargé en statistique, enquêtes et calcul de probabilité.

A l'issue de ce travail, nous savons qu'administrer des enquêtes par questionnaires est une tâche fondamentale pour l'appropriation des techniques et du champ de la statistique ; elle ne se résume pas au simple traitement des données et en ce sens permet d'enrichir l'ensemble des situations de référence ; mais nous savons aussi que construire des simulations est une activité fondamentale où le sujet apprenant est obligé de construire et manipuler des signifiants d'objets inconnus, et de ce fait d'accorder une place, à ce que j'ai appelé le non savoir, dans le champ même du savoir.

Table des figures et schémas

Figure 1 : Schème - situation selon Vergnaud	40
Figure 2 : Schématisation de l'activité des élèves selon Guy Brousseau (BROUSSEAU, 2003, p.8)	43
Figure 3 Qu'est ce qu'apprendre ? (Régnier, 2000a, p. 44).....	54
Figure 4 : Quatre types d'alternance selon Régnier, 2000a, p. 48	55
Figure 5 : Extrait d'une lettre du 12 janvier 1877 de Galton.....	72
Figure 6 : Simulation de la planche de Galton	77
Figure 7 : Pluie aléatoire et courbe de la loi normale centrée réduite.....	78
Figure 8 : Valeurs expérimentales et intervalle de confiance.....	79
Figure 9 : Simulation d'un tirage de 10 échantillons de 30 individus sur une population de 10000	80
Figure 10 : Tableau de bord de la simulation.....	81
Figure 11 : Dix points au hasard ayant des coordonnées entre 0 et 1.....	83
Figure 12 : Répartition de 6344 valeurs du coefficient de détermination de 10 points au hasard (coordonnées entre 0 et 1)	84
Figure 13 : Trois points au hasard ayant des coordonnées entre 0 et 1.....	84
Figure 14 : Répartition de 2400 valeurs du coefficient de détermination de 3 points au hasard (coordonnées entre 0 et 1)	85
Figure 15 : Fluctuations d'échantillonnage de R coefficient de corrélation.....	89
Figure 16 : Intervalles de confiance sur R après la transformation de Fischer.....	90
Figure 17 : Comparaison des valeurs de q et des valeurs de Phi correspondantes	93
Figure 18 : Indices d'implication et intensités d'implication	94
Figure 19 : Comparaison des valeurs simulées et théoriques.....	95
Figure 20 : Comparaison des valeurs de l'intensité d'implication (loi de Gauss et loi de Poisson).....	95
Figure 21 : 100 valeurs « au hasard » de l'indice d'implication.....	96
Figure 22 : 500 valeurs de l'intensité d'implication	96
Figure 23 : Exemple de graphe implicatif.....	104
Figure 24 : Exemple d'un arbre cohésitif.....	105
Figure 25 : Exemple d'un arbre des similarités.....	106
Figure 26 : Exemple d'un dendrogramme obtenu à l'aide du logiciel StatBox	107

Bibliographie

- Académie des Sciences, (2000) *La statistique Rapport sur la science et la technologie n°8*, Paris Éditions TEC&DOC.
- ADLER I., (1969), *Statistique et probabilités pour aujourd'hui*, OCDL, Paris.
- AÏVAZIAN S., (1970), *Etude statistique des dépendances*, Editions Mir.
- AÏVAZIAN S., ENUKOV I., MECHALKINE L., (1986), *Eléments de modélisation et traitement primaire des données*, Editions Mir.
- ALBERT, J. H. (1993) "Teaching Bayesian Statistics Using Sampling Methods and MINITAB", *The American Statistician*, 47, 182-191.
- ALLAL, L. (1979) « Stratégies d'évaluation formative : conceptions psychopédagogiques et modalités d'application », in L. Allal, J. Cardinet et P. Perrenoud dir., *L'Évaluation formative dans un enseignement différencié*, Peter Lang, Berne.
- ANTOINE, F., GROOTAERS D., TILMAN F., (1988), *Manuel de la formation en alternance*, Lyon, Chronique sociale : Bruxelles, Vie ouvrière.
- APMEP, *Analyse des données*, Tome 1 (Publication n° 28) & Tome 2 (Publication n°40).
- ARBIB, KFOURY & MOLL, (1981) *A basis for theoretical computer science*, Springer-Verlag.
- ARDILLY P., 1994, *Les techniques de sondage*, Editions Technip, Paris.
- ARISTOTE, *Constitution d'Athènes*, traduite par B. Haussoullier, (1891), E. Bouillon, Paris , Bibliothèque de l'École des hautes-études. Sciences philologiques et historiques. Fascicule 89.
- ARISTOTE, *Derniers analytiques*, Traduction de Jules Barthélemy-Saint-Hilaire.
- ARMATTE M., *Histoire du modèle linéaire. Formes et usages en statistique et économétrie jusqu'en 1945.-* Thèse de doctorat, EHESS, Paris.
- ARTIGUE, M. (1989) Ingénierie didactique *Recherches en Didactique des Mathématiques*. 9(3), 281-308.
- ARTIGUE, M. (1990) Epistémologie et didactique, *Recherches en Didactique des Mathématiques*, 10-2.3, 241-286.
- ARTIGUE, M., GRAS R., LABORDE C., TAVIGNOT P. (1994), *Vingt ans de didactiques des mathématiques en France*, Grenoble, La Pensée Sauvage Éditeurs.
- Association Pénombre (1999). *Chiffres en folie. Petit abécédaire de l'usage des nombres dans le débat public et les médias*. Paris : La Découverte.
- AUBRY B., (2002), *La Statistique en pratique*, collection L'esprit des Sciences, Ellipses, Paris.

- BACHELARD, G. (1938), *La formation de l'esprit scientifique*, Paris Vrin.
- BACHER, F. (1973) « La Docimologie », in M. Reuchlin, *Traité de psychologie appliquée*, vol. VI, P.U.F., Paris.
- BALACHEFF, N. (1981), Une approche expérimentale pour l'étude des processus de résolution de problème. *Communication présentée à la Vème Conference for the Psychology of Mathematics Education*.
- BALACHEFF, N. (2003), « Cours de l'école d'été de didactique des mathématiques », ARDM
- BARTH, I. (2001), « L'utilisation de la situation de travail dans la formation d'un cursus diplômant, cas du DUT GEA de l'IUT Lumière », *Deuxième rencontre des acteurs de la formation technologique d'Europe et d'Amérique Latine, Pointe à Pitre*.
- BATANERO, C. (2001), *Didáctica de la Estadística GIEE Universidad de Granada* ISBN 84-699-4295-6.
- BATARENO, C. ET GODINO J., (2001), *Análisis de Datos y su Didáctica*, Universidad de Granada.
- BENFORD, F. (1938) *Proceedings of the American Philosophical Society*, 78 , 551.
- BENICHOU P., BENICHOU R., BOY N., POUGET J.-P., (1992), *Statistique - Probabilités, formations supérieures technologiques et tertiaires*, Dunod, Paris.
- Bible, Ancien Testament, (1956), NRF, Gallimard, La pléiade, Paris.
- BLANCHARD-LAVILLE, C. (1981), Les dimensions affectives de l'apprentissage des statistiques, *Éducation Permanente* (61) pp.41-62.
- BLANCHARD-LAVILLE, C. (1980) *Les étudiants de psychologie face à l'enseignement de statistiques (analyse des réponses à un test de mathématiques et à des questionnaires d'opinion.)* Thèse de Doctorat de 3ème cycle, Université Paris VII.
- BLOOM, B. et al, (1971) *Handbook on Formative and Summative Evaluation of Student Learning*, McGraw-Hill, New York.
- BLOOM, B. et al., (1956) *A Taxonomy of Educational Objectives, vol. I : The Cognitive Domain*, Longmans Green, New York.
- BONNET, C., HOC, J.-M., TIBERGHEN, G. (1986), *Psychologie, intelligence artificielle et automatique*, Bruxelles, Mardaga.
- BONORA, D. (1985), *Les Représentations des professeurs à l'égard des objectifs pédagogiques dans l'enseignement des sciences naturelles* (thèse), Paris, E.P.H.E.
- BONSOIR, A. (1985), *La Méthode des tests en pédagogie : problèmes d'évaluation et psychométrie*, P.U.F.
- BOREL, É., (1926), *Probabilités Erreurs*, Librairie Armand Colin, Paris.
- BOREL, É. (1924). *Probabilité et certitude*. Paris : PUF.
- BOROVKOV, A., (1987), *Statistique mathématique*, Editions Mir.
- BOULEAU, N. (2000), *Processus stochastiques et applications*, collection Méthodes, Hermann, Paris.
- BOULEAU, N. (2002), *Probabilités de l'ingénieur, Variables aléatoires et simulation*, collection Méthodes, Hermann, Paris.
- BOULIGAND, G. (1944), *Les Aspects Intuitifs de la Mathématique*, N.R.F. Gallimard, Paris.
- BOURBAKI, N. (1984), *Eléments d'histoire des mathématiques*, Masson, Paris, 376 p.
- BOURDIEU, P. (1980), *Le Sens pratique*, Éditions de Minuit, Paris.
- BOURSIN J.-L., (1986), *Les structures du hasard, les probabilités et leurs usages*, collection Points Sciences, Editions du Seuil, Paris.
- BRÉMAUD P., (1988), *Introduction aux probabilités, modélisation des phénomènes aléatoires*, Springer.
- BROSSIER G., DUSSAIX A.-M., (1999), *Enquêtes et sondages, Méthodes, modèles, applications, nouvelles approches, 2^e et 3^e cycle*, Dunod, Paris.

- BROUSSEAU G. (1993), *Fiches de Statistiques non paramétriques pour la didactique*, LADIST Université de Bordeaux 1 et IUFM d'Aquitaine.
- BROUSSEAU G., (1993), *Stratégies de l'analyse statistique, cours et aide-mémoire à l'intention des professeurs en formation*, LADIST Université de Bordeaux 1.
- BROUSSEAU, G. (1974), "L'enseignement des probabilités à l'école élémentaire" in *Compte-rendu de la XXVIe rencontre de la C.I.E.A.E.M. IREM de Bordeaux*.
- BROUSSEAU, G. (1986), "Fondements et méthodes de la didactique des mathématiques", in *Recherches en didactique des mathématiques 7.2*, Grenoble, La pensée sauvage.
- BROUSSEAU, G. (1998), *Théorie des situations didactiques*, La Pensée Sauvage Editions, Grenoble.
- BROUSSEAU, G., BROUSSEAU N., WARFIELD Virginia, (2002) "An experiment on the teaching of statistics and probability" *Journal of Mathematical Behavior*, 20 (2002) 363-441.
- BROUSSEAU, G. (2003), "Cours de l'école d'été de didactique des mathématiques », *Actes de la XIIe école d'été de didactique des mathématiques* (Corps, 20-29 août 2003), ARDM.
- BRY X., (1996), *Analyses factorielles multiples*, collection Poche Techniques quantitatives, Economica, Paris.
- CALLENS, S. (1997), *Les maîtres de l'erreur*, PUF, Paris
- CARDANO, J. (1663), *De Ludo Alea Liber*, http://www.filosofia.unimi.it/cardano/testi/operaomnia/vol_1_s_10.pdf.
- CARDINET, J. (1988), *Pour apprécier le travail des élèves*, DeBoeck-Wesmael, Bruxelles.
- CHASSÉ J.L., PAVÉ A., (1975), *Probabilités statistiques et biologie*, CEDIC, Paris.
- CHAUVAT G., LEDOUX F., REAU J.-P., (1993), *Mathématiques BTS et DUT industriels, Algèbre, géométrie, probabilités, statistiques*, Tome 2, Armand Colin, Paris.
- CHEVALLARD, Y. (1978), *Notes pour la didactique de la statistique I.R.E.M. d'Aix-Marseille*. 395 p. + annexes.
- CHEVALLARD, Y. (1980), *Mathématiques, Langage, enseignement : la réforme des années soixante*, http://yves.chevallard.free.fr/spip/spip/IMG/pdf/Mathematiques_langage_enseignement.pdf.
- CHEVALLARD, Y. (1985) (1991), *La transposition didactique*, Grenoble, La Pensée Sauvage.
- CHEVALLARD, Y. (1994), *Situations recherches*, Ecole d'été du CNRS.
- CHEVALLARD, Y. (2003a), « *Cours de l'école d'été de didactique des mathématiques* », *Actes de la XIIe école d'été de didactique des mathématiques* (Corps, 20-29 août 2003), ARDM.
- CHEVALLARD, Y., & WOZNIAK, F. (2003b),. Enseigner la statistique au secondaire. Entre genre prochain et différence spécifique, *Actes de la XIIe école d'été de didactique des mathématiques* (Corps, 20-29 août 2003).
- CIBOIS P., (2000), *L'analyse factorielle*, collection Que sais-je n° 2095, PUF, Paris.
- CLOT Y.,(1999), *Avec Vygotski (sous la direction)*, La Dispute, Snédit, Paris. (Collectif, (1991), *Le hasard aujourd'hui*, collection Points sciences, Editions du Seuil, Paris.
- COMBELLES, C (2005), Application du théorème central limite. Exemples. Le cas du surbooking dans les réservations des places de transport aérien, *Rapport d'activité IREM Aix-Marseille*.
- COMBES, M.-C. (1996), *Coordination de L'alternance : enjeux et débats*, Paris, La documentation française.

- Commission inter-IREM Statistique et Probabilités, (1997), *Enseigner les probabilités au lycée*, publié par le réseau des IREM avec le soutien de la Direction des Lycées et Collèges.
- COUTROT B., DROESBEKE J.-J., (1995), *Les méthodes de prévision*, collection Que sais-je n° 2157, PUF, Paris.
- d'ERRICO, F. (2006) « Evolution ou révolution ? L'origine des comportements modernes en Afrique et en Europe selon l'archéologie », *conférence au Cercle Genevois d'Archéologie, 4 avril 2006*.
- DACUNHA-CASTELLE D., (1996), *Chemins de l'aléatoire, le hasard et le risque dans la société moderne*, Flammarion, Paris.
- DAGNELIE P.,(1998), *Statistique théorique et appliquée, Tome 1 : Statistique descriptive et bases de l'inférence statistique*, De Boeck Université, Paris – Bruxelles.
- DAGNELIE P., (1998), *Statistique théorique et appliquée, Tome2 : Inférence statistique à une et à deux dimensions*, De Boeck Université, Paris – Bruxelles.
- DAUDIN J.-J., ROBIN S., VUILLET C., (1999), *Statistique inférentielle, Idées, démarches, exemples*, Société Française de Statistique et Presses Universitaires de Rennes.
- DE LANDSHEERE, G. (1971), *Évaluation continue et examens. Précis de docimologie*, éd. Labor, Bruxelles.
- DE LANDSHEERE, G. (1979), *Dictionnaire de l'évaluation et de la recherche en éducation*.
- De MOIVRE, A (1733), *The Doctrine of Chance*, 3rd ed., <http://www.ibiblio.org/chance/>
- DEHEUVELS P., (1990), *La probabilité, le hasard et la certitude*, collection Que sais-je n° 3, PUF, Paris.
- DESROSIÈRES A. et THÉVENOT L., (1992), *Les catégories socioprofessionnelles*, La Découverte, Paris.
- DESROSIERES, A., (1985), *Histoires de formes : statistiques des sciences sociales avant 1940. Revue française de sociologie, XXVI, 277-310*.
- DESROSIERES, A., (1993), *La politique des grands nombres*. Paris : La découverte.
- DESROSIERES, A., MAIRESSE J., & VOLLE, M., (1977), *Les temps forts de la statistique française depuis un siècle. Pour une histoire de la statistique, tome 1 (pp. 509-517)*. Paris : Economica.
- DHOMBRES, N. & DHOMBRES, J., (1989), *Naissance d'un nouveau pouvoir: sciences et savants en France 1793-1824*, Ed. Payot, coll. "Bibliothèque historique", Paris.
- DHOMBRES, J. (1992), *L'Ecole Normale de l'An III, Leçons de mathématiques, Laplace-Lagrange-Monge*, Dunod, Paris.
- DIEUDONNE, J. (1978) (sous la direction de), *Abrégé d'histoire des mathématiques*, Hermann, tome 1 392 p. et tome 2 472 p.
- DROESBEKE J.-J., TASSI P., (1975) (1997), *Histoire de la statistique*, collection Que sais-je n° 2527, PUF, Paris.
- DROESBEKE J.-J., (1992), *Éléments de statistique*, collection Statistique et Mathématiques appliquées, Editions de l'Université de Bruxelles et Editions Ellipses, Paris.
- DROESBEKE J.-J., FINE J., SAPORTA G., (1997), *Plans d'expériences Applications à l'entreprise*, Editions TECHNIP, Paris.
- DROESBEKE J.-J., FINE J., SAPORTA G., (2002), *Méthodes bayésiennes en statistique*, Editions TECHNIP, Paris.
- DUBOS J., (1980), *Statistique descriptive en science économique*, Dunod, Paris.

- DUCROCQ A. & WARUSFEL A., (2001), *Les mathématiques, Plaisir et nécessité*, Vuibert, Paris.
- DUPONT-FERRIER, G. (1930-1932) *Études sur les institutions financières de la France*, Paris, Librairie Firmin-Didot, Paris.
- DUTARTE, P KERN C., (1998), *Simulation d'expériences aléatoires*, Université Paris Nord,.
- ENGEL A., (1975), *L'enseignement des probabilités et de la statistique*, Volume 1, éditions CEDIC, Paris.
- ENGEL A., (1979), *L'enseignement des probabilités et de la statistique*, Volume 2, éditions CEDIC, Paris.
- ENGEL A., (1990), *Les certitudes du hasard*, Aléas éditeur, Lyon.
- ESCOFIER B., PAGES J., (1997), *Initiation aux traitements statistiques, Méthodes, méthodologie*, collection Didact Statistique, Presses Universitaires de Rennes.
- FERMAT (de), P. (1679), *Varia Opera mathematica*, éd. Samuel de Fermat, Toulouse, rééd. Berlin, 1861.
- FERMAT (de), P. (1891-1912), *Œuvres de Fermat*, publiées par les soins de MM. Paul Tannery et Charles Henry sous les auspices du Ministère de l'instruction publique, The University of Michigan Historical Mathematics Collection, <http://quod.lib.umich.edu/cgi/t/text/text-id?c=umhistmath&cc=umhistmath&type=simple&rgn=full+text&q1=fermat&cite1=fermat&cite1restrict=author&cite2=&cite2restrict=author&Submit=Search>.
- FERRALL, C. (1995) « Interactive Statistics Tutorials in Stata » *Journal of Statistics Education* [Online], 3(3). (www.amstat.org/publications/jse/v3n3/ferrall.html).
- FISCHBEIN, (1975), *The intuitive sources of probabilistic thinking in children*, D. Reidel Publishing Company, Dordrecht (Hollande) & Boston. (Consultable à l'IREM de Paris 7).
- FOUCART T., (1997), *L'analyse des données, mode d'emploi*, collection Didact Statistique, Presses Universitaires de Rennes.
- FOURDRINIER D., (2002), *Statistique inférentielle*, Cours et exercices corrigés 2^e cycle, collection Sciences Sup, Dunod, Paris.
- FOURGEAUD C., FUCHS A., (1972), *Statistique*, Dunod, Paris.
- FRANKLIN, L.A. (1992) "Using Simulation to Study Linear Regression," *The College Mathematics Journal*, 23, 290-295.
- FREUD, S. (1925), *Freud présenté par lui-même*, Gallimard, Paris, 1984, 143 pages.
- FREUD, S. (1969), *Psychopathologie de la vie quotidienne* (Zur Psychopathologie des Alltagslebens), Payot, rééd. 1989.
- GALILEO GALILEI, (1718), « *Sopra le scoperte de i dadi* », *Opere*, VIII, 1898 (1933), pp. 591-594).
- GALTON F., (1869), *Hereditary genius. An inquiry into its laws and consequences.*- London : Macmillan.
- GERBAUD, J. (2006), Département Biosciences Végétales, Module : Statistique 1, ENSAT, Toulouse.
- GOMEL, C., (1896), *Histoire financière de l'Assemblée Constituante*, Paris, Guillaumin & Cie, t. 1 (1789).
- GOODY, J. (1993), *La science sauvage, des savoirs populaires aux ethnosciences*, Seuil, Paris.
- GOULD, S.H. (1953) « *Cardano, The Gambling Scholar* ».
- GRAS R., (1996), *L'implication statistique, nouvelle méthode exploratoire de données*, La Pensée Sauvage Editions, Grenoble.

- GRIFFITHS, M. (1986) *Techniques algorithmiques pour l'intelligence artificielle*, Hermès Publishing.
- GUERBER L., HENNEQUIN P-L., (1968), *Initiation aux probabilités*, Bibliothèque d'Enseignement Mathématique, APMEP.
- HACKING I., (1990), *The Taming of chance.*- Cambridge : Cambridge University Press.
- HAGEGE M., (1967), *Eléments de calcul des probabilités et applications*, OCDL, Paris.
- HALD, A. (2003) *History of Probability and Statistics and their applications before 1750*, Wiley, Hoboken, New Jersey.
- HENRY, M. (1994) *Les premiers apprentissages en géométrie et en probabilités: des processus de modélisation comparables*, IREM de Franche-Comté.
- HENRY, M. (2004) *La notion de probabilité: évolution historique et applications contemporaines*, IREM de Franche-Comté, www.univ-lille1.fr/irem/manifs/jhasard/conf_henry.pdf
- HESTERBERG, T. C. (1998) "Simulation and Bootstrapping for Teaching Statistics", *American Statistical Association Proceedings of the Section on Statistical Education*, Alexandria, VA: American Statistical Association, 44-52.
- HILL, T., (1996), *Statistical Science*, 10 , 354,
- HILL, T., (1999), Le premier chiffre significatif fait sa loi, *La recherche hors série*, 316, p.72
- HOWELL D. C., (1998), *Méthodes statistiques en sciences humaines*, DeBoeck Université, Paris.
- IFRAH, G., (1981), *Histoire universelle des chiffres*, Seghers, Paris.
- INSEE, (1977), collectif, *Pour une histoire de la statistique t.1*, Economica, Paris.
- JENSEN, P. A. (1983) "Microcomputer Forum: Microcomputers in the Instruction of Operations Research," *Interfaces*, 13, 18-24.
- JOULE, R.-V. (1987) *Petit traité de manipulation à l'usage des honnêtes gens*, Grenoble, PUG.
- Journal européen de psychologie de l'éducation, (1987), « *Psychologie de l'évaluation et éducation* », vol. XXI, no 2, (no spécial).
- KENNEDY, K., OLINSKY, A., and SCHUMACHER, P. (1990) "Using Simulation as an Integrated Teaching Tool in the Mathematics Classroom", *Education*, 111, 275-296.
- KIMBERLEY, K. (1984) « Évolution de la méthode d'évaluation en Angleterre et au pays de Galles », in *Perspectives*, no 3.
- LACAN, J. (1991) *Le Séminaire- Livre XVII - L'envers de la psychanalyse*, (production des quatre discours) Seuil, Paris.
- LAHANIER-REUTER, D. (1999), *Conceptions du hasard et enseignement des probabilités et statistiques*, Puf, Paris.
- LAPLACE, Pierre-Simon (marquis de) (An III), *Leçons*, in *Leçons de Mathématiques*. École Normale de l'An III (Laplace, Lagrange et Monge). Rééd. sous la dir. de J. Dhombres. Paris, Dunod, 1992.
- LAPLACE, P.S. (1810), *Essai philosophique sur les probabilités*, Texte de la 5ème édition de 1825, préface R. Thom, postface B. Bru. Ed. C. Bourgeois, Paris, 1986)
- LAPLACE, P.S. (1814), *Théorie analytique des probabilités*, Courcier Imprimeur-Libraire, Paris.
- LE NIR, M., MICHEL, J-A., NERIEUX, C. (1998), «Deux outils d'insertion par et pour l'alternance», *communication présentée dans le cadre d'un colloque sur L'Enseignement professionnel court post-baccalauréat : enjeux sociaux, enjeux territoriaux*, Le Mans.
- LE NIR, M., ORIOL, J-C (2000), L'enquête : élément central de l'enseignement de statistique en première année GLT in *CNR'UT 2000 Recherche et innovation*, Presses Universitaires d'Orléans, pages 157 à 169.

- Dictionnaire historique de la langue française LE ROBERT , (1993), sous le direction de Alain Rey, Nouvelle édition 1993.
- Lettre au Duc de Saxe-Cobourg et Gotha, citée par Dreesbeke J.-J. et Tassi Ph. In Histoire de la statistique, PUF, 1997, p. 5.
- LEVY, (1994), *Algorithmique combinatoire*, Dunod.
- LEY, E. (1996), *The American statistician* (Am. stat.), vol. 50, no 4, pp. 311-313.
- LIOZOU, A. (1979), *Initiation pratique à la statistique*, Paris, Eyrolles.
- MAGER, R. F. (1986) Comment mesurer les résultats de l'enseignement, Bordas.
- MALGLAIVE, G., WEBER A., (1982), « Théorie et pratique : approche critique de l'alternance en pédagogie », *Revue française de pédagogie*, n°61, pp. 17-27.
- MANOUKIAN E. B., (1986), *Guide de statistique appliquée*, collection Méthodes (série grise), Hermann, Paris.
- MARASINGHE, M-G., MEEKER, W. Q., COOK, D., AND SHIN, T. (1996) "Using Graphics and Simulation to Teach Statistical Concepts," *The American Statistician*, 50, 342-351.
- MARVIN & MINSKY, (1967), *Computation : Finite and infinite machine*, Prentice-Hall.
- MATHIEU, J., THOMAS R. (1985), La compréhension, in *Manuel de Psychologie*, Vigot, p.154.
- MILLS, J. D. (2002) Using Computer Simulation Methods to Teach Statistics: A Review of the Literature, *Journal of Statistics Education* Volume 10, Number 1.
- Ministère de l'Éducation, (1981), *Politique générale d'évaluation pédagogique : secteur du préscolaire, du primaire et du secondaire*, gouvernement du Québec, Québec..
- MOINE, N. (2002), « Le miroir des statistiques, Inégalités et sphère privée au cours du second stalinisme » in *Cahiers du Monde russe*, 44/2-3, Avril-septembre 2003, p. 481-518.
- MOLES A., (1995), *Les sciences de l'imprécis*, collection Points Sciences, Editions du Seuil, Paris.
- MORLAT, G. (2000) Encyclopaedia Universalis version 6, Article « Statistique ».
- MURPHY K. R., MYORS B., (1998), *Statistical power analysis, A simple and general model for Traditional and Modern Hypothesis Tests*, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, London.
- NAKACHE, J.-P. ET CONFAIS J. (2005) *Approche pragmatique de la classification*, Editions Technip, Paris.
- NIGRINI, M. (1996) *Journal of the American taxation Association*, 18 , 72.
- NOIZET G. CAVERNI J-P., (1978) *Psychologie de l'évaluation scolaire*, P.U.F.
- OLINSKY, A., and SCHUMACHER, P. (1990) "Illustration of Regression Results Using Simulation and Graphical Presentation," *Working Paper Series*, Smithfield, RI: Bryant College.
- ORIOU, J-C. ROBERT C. (1981), *Matchinettes 3*, Irem de Grenoble.
- ORIOU, J-C. (1985) *Intégration de démarches et de programmes d'intelligence artificielle dans des séquences didactiques*, Mémoire de DESS d'informatique, Université Joseph Fourier, Grenoble 1.
- ORIOU, J-C. (1986) *Approche de la récursivité pour une transposition éventuelle*, Mémoire de DEA de didactique, Université. Grenoble 1, 75 pages.
- ORIOU, J-C. (1986) *Projets et programmes : introduction à la science de l'informatique*, Erpi, Editions du nouveau pédagogique inc., Montréal(Canada), 272 pages.
- ORIOU, J-C. (1999) L'évaluation en alternance, *conférence in colloque sur l'alternance, IUT Lumière*, Lyon.
- ORIOU, J-C. (2002) "Réaliser une enquête par questionnaires : un outil didactique pour la statistique inférentielle à l'Université", in *3ème Rencontre Europe-Amérique-Latine sur la formation technologique et professionnelle*, La Havane, Cuba.

- ORIOU, J-C. (2003) « Petites chroniques d'une complication annoncée... ou un grand cabotage du contrôle à l'évaluation », *Intervention Journées Nationales STID*, Niort.
- ORIOU, J-C. REGNIER, J-C. (2003b) « Fonctionnement didactique de la simulation en statistique : Exemple de l'enseignement du concept d'intervalle de confiance », 35^{èmes} *Journées de Statistique, SFDS Lyon*, tome 2 Pages 743 à 750.
- ORIOU, J-C. REGNIER J-C., (2003c) « Fonctionnement didactique de la simulation en statistique dans l'enseignement du concept de corrélation », *Espace Mathématique Francophone 2003*, Tozeur, Tunisie.
- ORIOU, J-C. REGNIER, J-C. (2004), Loi de Benford et Pratique d'enquêtes par questionnaires : un exemple de situations didactiques en statistique, 36^{èmes} *Journées de Statistique, SFDS Montpellier*
- ORIOU, J-C. REGNIER, J-C. (2006) Formation en statistique en DUT STID et Transposition didactique, 38^{èmes} *Journées de Statistique, SFDS Clamart*.
- ORIOU, J-C. REGNIER, J-C. (2007) Conceptualisation de l'analyse statistique implicite, 39^{èmes} *Journées de Statistique, SFDS Angers*.
- ORIOU, J-C. REGNIER, J-C. (2007b), Enseignement-apprentissage de l'ASI en 1er cycle universitaire, *ASI 4, Castellon (Espagne)*.
- PAGES G., BOUZITAT C., (2000), *En passant par hasard ... Les probabilités de tous les jours*, Vuibert, Paris.
- PAGES, J. (2005) *Statistiques générales pour utilisateurs*, Rennes, PUR.
- PARSYSZ, B., *L'enseignement des probabilités et de la statistique en France depuis 1965*, http://www.caib.es/ibae/esdevenement/jornades_10_01/doc/prostat-F.doc
- PASCAL, B. (1665) *Les lettres de Blaise Pascal : accompagnées de lettres de ses correspondants*, Ed. 1922, NUMM-69975, consultable <http://gallica.bnf.fr>.
- PASCAL, B. (1954) *Œuvres complètes*, Edition Gallimard.
- PEARSON K., (1912), *La grammaire de la science*.- Paris, Alcan,(trad. L. March).
- PEDHAZUR, E. J. (1997) *Multiple Regression in Behavioral Research: Explanation and Prediction (3rd ed.)*, Orlando: Harcourt Brace College.
- PELTIER M., ROUCHE N., MANDERICK M., (1982), *Contremanuel de statistique et probabilité*, Vie Ouvrière, Bruxelles.
- PERRENOUD, P. (1984), *La Fabrication de l'excellence scolaire*, Droz, Genève-Paris.
- PIEDNOIR, J-L, DUTARTE, P. (2001), *Enseigner la statistique au lycée : des enjeux aux méthodes*, Commission Inter-Irem Lycées technologiques, Université Paris 13.
- PIERON, H. (1963), *Examens et docimologie*, P.U.F.
- PIGEIRE, J. (1932), *La Vie et l'œuvre de Chaptal, 1756-1832*, Paris, Éditions Spès,
- PULLEY, L. B., and DOLBEAR, F. T. (1984) "Computer Simulation Exercises for Economics Statistics", *Journal of Economics Education*, 3, 77-87.
- QUETELET, A. (1835), *Sur l'homme et le développement de ses facultés ou essai de physique sociale*.- Paris : Bachelier.
- RÅDE L., (1970), *The teaching of Probability & Statistics*, Proceedings of the first CSMP International Conference co-sponsored by Southern Illinois University and Central Midwestern Regional Educational Laboratory, John Wiley & Sons, Almqvist & Wiksell (Uppsala – Suède). (Disponible à l'IREM de Paris 7).
- RAIMI, R. (1969) *Scientific American* , 109, 119.
- RAYNAL, F. RIEUNIER, A. (1987), *Méthodes, techniques, stratégies pédagogiques*, IPNETP, Abidjan.
- RAYNAL, F. REUNIER, A. (1997), *Pédagogie : dictionnaire des concepts clés*, ESF Editeur, Paris, 408 p.
- RE, J. (1932), *La Vie et l'œuvre de Chaptal, 1756-1832*, Éditions Spès, Paris.

- REGNIER, J-C. (1983), *Étude didactique d'un test autocorrectif en trigonométrie*, Strasbourg : IREM-Université L Pasteur, Tome 1 : 307 pp, Tome 2 : 171 p Thèse de doctorat de 3ème cycle en didactique des mathématiques.
- REGNIER, J-C. (1988), *Étude didactique d'une méthode d'apprentissage fondé sur le tâtonnement expérimental de l'apprenant*, Annales de Didactique et de Sciences Cognitives, séminaire de Didactique des Mathématiques de Strasbourg, pp 255-279.
- REGNIER, J-C. (1994), *Statistique documents auto-correctifs et auto-évaluatifs d'aide à l'apprentissage*, ISPEF, Université Lyon2.
- REGNIER, J-C. (1998), De la vérité autoproclamée à la vraisemblance reconnue, in J-C Girard, D. Gros, P. Planchette, J-C Régnier, R. Thomas, *Enseigner la Statistique du CM à la Seconde, Pourquoi ? Comment ?*, I.R.E.M. de Lyon, pp 107-118.
- REGNIER, J-C. (2000a), L'alternance: des évidences (en) aux questions. *Forum, revue de la recherche en travail social*, n°93, pp 38-47 ISSN 0988.6486.
- REGNIER, J-C. (2000b) Auto-évaluation et autocorrection dans l'enseignement des mathématiques et de la statistique. Note de synthèse pour l'obtention de l'Habilitation à Diriger des Recherches. Université Marc Bloch (13 décembre 2000) 240 p.
- REGNIER, J-C., (2002), A propos de la formation en statistique Approches praxéologiques et épistémologiques de questions du champ de la didactique de la statistique, in *Questions éducatives. L'école et ses marges. Revue du Centre de Recherche en éducation de l'Université Jean Monnet de Saint-Étienne*, n°22-23 décembre 2002 Didactique des mathématiques. pp. 157-201.
- REGNIER, JC, (2006) *Formation de l'esprit statistique et raisonnement statistique. Que peut-on attendre de la didactique de la statistique ?* in C.Castela et C.Houdement (Dir.) Actes du séminaire national de Didactique des Mathématiques. Année 2005. Editeurs: ARDM & IREM de Paris 7 (pp.13-37)
- REUHLIN, M. (1973), « Problèmes d'évaluation », in M. Debesse & G. Mialaret, *Traité des sciences pédagogiques*, vol. IV, P.U.F., 1973.
- REVUZ, A. (1980) *Est-il impossible d'enseigner les mathématiques ?*, Puf, Paris.
- ROBERT C., (1995), *L'empereur et la girafe, Leçons élémentaires de statistiques*, Diderot Editeur, Paris.
- ROMEY, J-L. (1995) "Statistical Education Via Simulation," *American Statistical Association Proceedings of the Section on Statistical Education*, Alexandria, VA: American Statistical Association, 205-210.
- RONAN, C. (1988) *Histoire mondiale des sciences*, Paris Seuil, 704 p.
- ROUSSET, P. (1997) «L'IUT Lumière : une expérience d'IUT en alternance », *communication présentée dans l'atelier formations professionnelles du secteur tertiaire et alternance aux journées d'études sur la professionnalisation dans l'enseignement supérieur*.
- SACHS L., (1984), *Applied Statistics, A Handbook of Techniques*, 2^{ème} édition, Springer Series in Statistics, Springer.
- SAPORTA, G., (1990), *Probabilités, analyse de données et statistique*, Editions Technip, Paris.
- SCHWARTZ, C. (2006), *Pratiques de statistique*, Vuibert, Paris.
- SCHWARTZ, D., LAZAR P., PAPOZ L., (1985), *Statistique médicale et biologique*, Médecine - Sciences Flammarion, Paris.
- SCHWARTZ, D. (1994), *Le jeu de la science et du hasard*. Paris : Flammarion.
- SENSEVY, G. (1998), *Institutions didactiques*, Paris, PUF.
- SERRES, M. (1989), *Éléments d'histoire des sciences*, Bordas.
- SPIEGEL M. R., STEPHENS L.J., (2000), *Statistiques*, troisième édition, Série Schaum, McGraw-Hill.

- STIGLER S. M., (1986), *The History of Statistics, The Measurement of uncertainty before 1900*, The Belknap Press of Harvard University Press, Cambridge (Massachusetts) et Londres.
- STIGLER S. M., (1999), *Statistics on the table, The History of Statistical Concepts and Methods*, Harvard University Press, Cambridge (Massachusetts) et Londres.
- TRIGNAN J., (1991), *Probabilités, statistiques et leurs applications*, BTS et IUT Techniques, Bréal, Paris.
- TROTIGNON, N. (2006) « Pascal, Fermat et la géométrie du hasard », Université Paris 1, <http://ces.univ-paris1.fr/membre/Trotignon/articles/pascal.pdf>.
- van der WAERDEN B.L., (1967), *Statistique mathématique*, Dunod, Paris.
- VANDERMEERSCH, B. (1986), (Coord.) L'homme de Néandertal., *Volume 7: l'extinction. Actes du Colloque International de Liège (4-7 Décembre 1986)*.
- VENTSEL H., (1973), *Théorie des probabilités*, Editions Mir.
- VENTSEL H., OVTCHAROV L., (1988), *Problèmes appliqués de la théorie des probabilités*, éditions Mir.
- VERGNAUD, G. (1990), La théorie des champs conceptuels. *Recherches en Didactique des Mathématiques*. 10(2-3), 133-170.
- VERGNAUD, G. (1994) Le rôle de l'enseignant à la lumière des concepts de schème et de champ conceptuel, in Artigue, M., Gras R., Laborde C., Tavinot P. (dir.), *Vingt ans de didactiques des mathématiques en France*, Grenoble, La Pensée Sauvage Editeurs, pp. 177-191.
- VERGNAUD G., (2000), *Lev Vygotski. Pédagogue et penseur de notre temps*, Hachette, Paris.
- VERLANT B., SAINT-PIERRE G., (2001), *Statistique et probabilités*, BTS industriels Groupements B, C, D, Tome 2, Editions Foucher, Paris.
- VERLANT B., SAINT-PIERRE G., (2002), *Statistique et probabilités*, BTS industriels Groupements B, C, D, Tome 2, Corrigé, Editions Foucher, Paris.
- VYGOTSKY L., (1985), *Pensée et langage*, (éd. originale 1934), La Dispute, Adage, Paris.
- WONNACOTT, R.J. et WONNACOTT, TH. (1972), *Introductory Statistics*, New York John Wiley & Sons.
- WOZNIAK, F. (2005), *Conditions et contraintes de l'enseignement de la statistique en classe de seconde générale un repérage didactique*, Thèse, Lyon 1.
- ZENDRERA, N. (2003). Difficultés d'apprentissage liées aux tests statistiques. Le cas des tests paramétriques auprès des étudiants en sciences humaines. *Actes des XXXVèmes Journées de Statistique. Société Française de Statistique et Université Lumière-Lyon-2. Lyon, France, 2-6 juin 2003*.
- ZENDRERA, N. (2004). Difficultés et obstacles rencontrés dans l'apprentissage des tests paramétriques : objets d'une recherche en didactique de la statistique in *Actes des XXXVIèmes Journées de Statistique. Société Française de Statistique*.

Table des matières

REMERCIEMENTS	1
INTRODUCTION.....	3
1. PARTIE I UNE APPROCHE HISTORIQUE DE LA STATISTIQUE : DES PREMIERS PAS A LA	
« GEOMETRIE DU HASARD » ET AUX « MAITRES DE L'ERREUR ».....	7
1.1. DES DEBUTS PROMETTEURS.....	8
1.1.1. <i>Les systèmes artificiels à mémoire (SAM)</i>	9
1.1.2. <i>Des actes fondateurs : les listes, les nombres</i>	9
1.1.3. <i>Une origine de la statistique descriptive</i>	11
1.1.4. <i>D'autres approches de la statistique</i>	13
1.2. LE CADASTRE	13
1.2.1. <i>Liminaire lapidaire</i>	13
1.2.2. <i>Les périodes antique et médiévale</i>	14
1.2.3. <i>De la Renaissance à la Révolution française</i>	15
1.2.4. <i>De la Révolution française à la fin du XIXème</i>	15
1.2.5. <i>En guise de conclusion de cet aperçu de l'histoire du cadastre</i>	16
1.3. DE LA THEOCRATIE.....	17
1.4. CALCULER LE FUTUR SE POSE COMME UN OBSTACLE EPISTEMOLOGIQUE	19
1.4.1. <i>Une pensée débarrassée des superstitions ?</i>	19
1.4.2. <i>Un paradoxe par rapport à la pensée d'Aristote : prendre et rejeter à la fois</i>	20
1.4.3. <i>Franchir l'obstacle épistémologique</i>	21
1.5. ALEAE GEOMETRIA	22
1.5.1. <i>Introduction</i>	22
1.5.2. <i>La géométrie du hasard</i>	22
1.5.3. <i>Cardano et Ludo Aleae (1663) ; Galilée et le « Problème du Duc de Toscane » (1620)</i>	24
1.6. LES MAITRES DE L'ERREUR.....	26
1.7. CONCLUSION	27
2. PARTIE II CADRE CONCEPTUEL ET CONTEXTE DE LA RECHERCHE.....	29
2.1. CADRE CONCEPTUEL	30
2.1.1. <i>Où l'on expose que voyager peut permettre d'apprendre, mais que l'on ne sort pas indemne d'une initiation</i>	31
2.1.2. <i>Désir de savoir : une bien mauvaise traduction</i>	37

2.1.3. <i>Apprentissage de la statistique et conceptualisation</i>	38
2.2. ENSEIGNEMENT ET DIDACTIQUE DE LA STATISTIQUE	41
2.2.1. <i>La théorie des situations didactiques de Guy Brousseau appliquée à la statistique</i>	41
2.2.2. <i>La transposition didactique</i>	43
2.3. ENSEIGNEMENT DE LA STATISTIQUE ET INGENIERIE DIDACTIQUE	44
2.3.1. <i>L'enquête par questionnaire</i>	44
2.3.2. <i>La simulation</i>	45
2.3.3. <i>Evaluation</i>	45
2.4. CONTEXTE DE LA RECHERCHE : L'IUT LUMIERE ET LE DEPARTEMENT STID	51
2.4.1. <i>Le département STID de l'IUT Lumière</i>	51
2.4.2. <i>L'alternance à l'IUT Lumière</i>	53
2.4.3. <i>Approche taxinomique de l'alternance</i>	54
2.4.4. <i>La place des tuteurs</i>	56
3. PARTIE III PRATIQUE D'ENQUETE PAR QUESTIONNAIRE : UN OUTIL D'APPROPRIATION DE LA STATISTIQUE.....	57
3.1. ENQUETE PAR QUESTIONNAIRES REALISEE PAR LES ETUDIANTS D'UN DEPARTEMENT GEA : UNE INGENIERIE PEDAGOGIQUE DE L'APPRENTISSAGE DE POUR LA STATISTIQUE INFERENTIELLE	58
3.1.1. <i>Présentation</i>	58
3.1.2. <i>L'organisation de l'enseignement de la statistique dans le département GEA</i>	58
3.1.3. <i>Mise en place du travail d'enquête par questionnaires</i>	58
3.2. LA LOI DE DE BENFORD ET LA PRATIQUE D'ENQUETES : UN EXEMPLE DE SITUATIONS DIDACTIQUES EN STATISTIQUE.....	61
3.2.1. <i>Présentation</i>	61
3.2.2. <i>La loi de Benford</i>	61
3.2.3. <i>Faire réaliser un travail d'enquête par les étudiants</i>	62
3.3. L'ENQUETE : ELEMENT CENTRAL DE L'ENSEIGNEMENT DE STATISTIQUE EN PREMIERE ANNEE DU DEPARTEMENT GLT	63
3.3.1. <i>Présentation</i>	63
3.3.2. <i>La SLTC</i>	63
3.3.3. <i>Les enquêtes à la SLTC</i>	64
3.3.4. <i>Vers la réalisation de dix enquêtes simultanées</i>	65
3.3.5. <i>Le projet SLTC comme outil pédagogique multiple</i>	65
3.3.6. <i>Un constat, des a priori, une stratégie</i>	66
3.3.7. <i>L'organisation du cours</i>	67
3.3.8. <i>La place de l'outil informatique</i>	67
3.4. PROJETS D'ETUDES STATISTIQUES STID 2EME ANNEE.....	68
3.5. CONCLUSION	68
4. PARTIE IV SIMULATION EN STATISTIQUE ET CONCEPTUALISATION	70
4.1. INTRODUCTION.....	71
4.1.1. <i>Une spécificité endogène au champ de la statistique</i>	71
4.1.2. <i>Un illustre précurseur</i>	71
4.1.3. <i>Un usage bien répandu</i>	72
4.2. LES INVARIANTS DE LA VARIABILITE.....	73
4.3. QUELQUES DISTINCTIONS CONCERNANT LES SIMULATIONS UTILISEES EN STATISTIQUE ET NOS CHOIX CONCERNANT LA CONSTRUCTION DES SIMULATIONS	74
4.4. LES OUTILS DE BASE POUR CONSTRUIRE UNE SIMULATION	75
4.5. PREMIER NIVEAU DE SIMULATIONS : CONSTAT DE LA STABILITE ET DE LA CONVERGENCE	76
4.6. SIMULATION METTANT EN PLACE UN CALCUL DE CONVERGENCE.....	77
4.7. SIMULATION CONCERNANT LE CONCEPT D'INTERVALLE DE CONFIANCE D'UNE MOYENNE	79

4.7.1. <i>Quelques travaux concernant les intervalles de confiance et la simulation</i>	79
4.7.2. <i>La situation problème et la construction de la simulation</i>	80
4.8. SIMULATION ET CONCEPT DE CORRELATION.....	81
4.8.1. <i>Quelques travaux concernant la corrélation et la simulation</i>	81
4.8.2. <i>La situation problème et la construction de la simulation</i>	81
4.8.3. <i>Conclusion</i>	85
4.9. ÉTUDE DE LA TRANSFORMATION DE FISHER DANS L'ANALYSE DU COEFFICIENT DE CORRELATION : D'UNE PRATIQUE FONDÉE SUR L'ARGUMENT D'AUTORITÉ ET LA LECTURE D'ABAQUES A UNE ARGUMENTATION FONDÉE SUR LA SIMULATION.	85
4.9.1. <i>Introduction</i>	86
4.9.2. <i>Une remarque préalable</i>	86
4.9.3. <i>Petit retour sur l'état du problème pédagogique</i>	86
4.9.4. <i>Quelques solutions apportées dans des ouvrages de base</i>	87
4.9.5. <i>Une séquence didactique d'observation des fluctuations d'échantillonnage de R</i>	88
4.9.6. <i>L'introduction de la transformation de Fisher et de la variable Z</i>	89
4.9.7. <i>Conclusion</i>	90
4.10. SIMULATIONS CONCERNANT L'ANALYSE STATISTIQUE IMPLICATIVE.....	90
4.10.1. <i>La place de la simulation</i>	90
4.10.2. <i>La situation proprement dite</i>	91
4.10.3. <i>La simulation outil d'appropriation de concepts statistiques</i>	91
4.10.4. <i>La simulation proposée</i>	92
4.10.5. <i>Recherche de bénéfices collatéraux : les valeurs trouvées du Khi deux, le test de Mc Némar⁹⁴</i>	
4.10.6. <i>Recherche de bénéfices directs concernant l'ASI : q est symétrique, il existe une relation entre le Khi^2 et $q(a, \text{non}(b))$</i>	95
4.10.7. <i>Comparaison du calcul de Phi avec la loi de Poisson et avec la loi normale</i>	95
4.10.8. <i>Observation de 100 valeurs de q, de 500 valeurs de Phi</i>	96
4.10.9. <i>Conclusion sur les simulations concernant l'ASI</i>	97
4.11. CONCLUSION DE LA PARTIE IV	97
5. PARTIE V : ETUDE DES IMPACTS DE LA PRATIQUE D'ENQUETE SUR LES REPRESENTATIONS DE LA STATISTIQUE CHEZ LES ETUDIANTS DE DUT	99
PRESENTATION	100
5.1. QUELQUES BREFS RAPPELS THEORIQUES CONCERNANT L'ANALYSE STATISTIQUE IMPLICATIVE. 100	
5.2. COMPARAISON DES REPRESENTATIONS DE LA STATISTIQUE CHEZ DEUX POPULATIONS D'ETUDIANTS.....	108
5.2.1. <i>Introduction</i>	109
5.2.2. <i>Présentation de l'étude</i>	109
5.2.3. <i>Synthèse des résultats</i>	185
5.2.4. <i>Annexes : Le questionnaire</i>	187
5.3. ÉTUDE DES REPRESENTATIONS DE LA STATISTIQUE CHEZ LES ETUDIANTS STID EN ALTERNANCE DE LYON.....	205
5.3.1. <i>Présentation</i>	205
5.3.2. <i>Travail sur un premier fichier d'essai</i>	206
5.3.3. <i>Présentation de l'étude</i>	207
5.3.4. <i>Résultats</i>	208
5.3.5. <i>Le cursus scolaire des interrogés (questions 1 à 8)</i>	209
5.3.6. <i>La vision de la statistique avant et en fin de DUT</i>	211
5.3.7. <i>Pendant la seconde année de DUT :</i>	212
5.3.8. <i>Conclusion</i>	214
5.3.9. <i>Question 11 : termes « représentatifs » de la statistique</i>	216
5.3.10. <i>Etude de la question en général</i>	217

5.3.11. Etude de la première partie de l'arbre des similarités (arbre A)	218
5.3.12. Etude de la deuxième partie de l'arbre de similarité (arbre B)	220
5.3.13. Etude de la dernière partie de l'arbre des similarités (arbre C)	222
5.3.14. Graphe implicatif pour toutes les modalités de la question 11	223
5.3.15. Conclusion générale concernant le traitement de la question 11	224
5.3.16. Question 13 : notions plus ou moins bien comprises dans le cours de statistique	225
5.3.17. Croisement avec les variables du sexe M (masculin) et F (féminin)	228
5.3.18. Question 14 : difficultés majeures	230
5.3.19. Question 15 : les apports majeurs que les étudiants attendent d'une formation STID	233
5.3.20. Question 17	237
5.3.21. Question 21	240
5.3.22. Etude de la seconde partie de l'arbre des similarités	243
5.3.23. Croisement de la question 21 avec les variables sexe M et F	247
5.3.24. Annexe 1 : Le questionnaire 2006-2007	250
5.3.25. Annexe 2 : Le recodage de chaque question	255
CONCLUSION	258
TABLE DES FIGURES ET SCHEMAS	260
BIBLIOGRAPHIE	261
TABLE DES MATIERES	271

Formation à la statistique par la pratique d'enquêtes par questionnaires et la simulation : étude didactique d'une expérience d'enseignement dans un département d'IUT

Résumé : Notre travail est une étude sur l'enseignement de la statistique en IUT, le statut et la mise en place d'une part d'enquêtes par questionnaires réalisées par les étudiants et d'autre part de simulations d'expériences aléatoires construites par les étudiants.

Il expose que l'enseignement de la statistique porte en lui une césure à la fois primitive et épistémologique, entre d'une part la collecte et le traitement des données, et d'autre part une approche fondée sur le calcul des probabilités plus ou moins éloignée des pratiques sociales, et que la réalisation d'enquêtes est un moyen d'atténuer la trace de cet écart primitif, permettant au sujet de construire le sens de ses apprentissages. D'une façon complémentaire nous étudierons comment la construction de simulations par les apprenants leur permet de dégager des invariants de la variabilité présente dans toute situation statistique.

Notre travail s'appuie sur les travaux de Gérard Vergnaud concernant la conceptualisation, sur la théorie des situations didactiques de Guy Brousseau, et sur la transposition didactique exposée par Yves Chevallard, ainsi que sur les travaux de Jean-Claude Régnier concernant la didactique de la statistique et le tâtonnement expérimental de l'apprenant, tout en essayant de déterminer la fonction de l'alternance ainsi que la dialectique entre projet et processus.

En termes d'ingénierie pédagogique notre exposé présentera un certain nombre de situations didactiques et désignera les divers obstacles rencontrés ; nous présenterons également deux enquêtes utilisant l'analyse statistique implicite (Régis Gras).

Mots clés : apprentissage de la statistique, simulation, enquêtes, enquêtes par questionnaires, champs conceptuels, situations didactiques, analyse statistique implicite, ingénierie pédagogique, statistique, IUT, alternance.

Title: Training to the statistics by the practice of inquiries by questionnaires and the simulation: didactic study of an experiment of education in a department of IUT (polytechnics, a two-year course)

Abstract: Our work consists in carrying out a survey whose purpose is to focus on the teaching of statistics at the IUT (polytechnics, a two-year course). We aim at explaining the characteristics of the teaching-learning situations specific to the field of statistics, and thus attempt to analyse the status, the effects and the conditions of the implementing of both "real size" surveys performed by the students, and simulations of random experiments designed by the students. Our analysis of statistics in terms of object-subject highlights the fact that its teaching holds in itself a primitive and epistemological breach between data collecting and processing on the one hand, and, on the other hand, an approach based on the calculation of probabilities more or less remote from social practices. It seems therefore that questionnaire-surveys constitute a means of reducing this primitive gap, by allowing the subject to build a learning meaning. Furthermore, we have studied how the constructions of simulations performed by the learners enables to obtain the invariants of the variability present in phenomena whose study is at the core of statistic situations. Our work is based on Gérard Vergnaud's works, as for conceptualization, on Guy Brousseau's theory of educational situations, as well as the educational transposition displayed in Yves Chevallard and Jean-Claude Regnier's works concerning educational statistics and the experimental trials of the learner.

Simultaneously, we have been trying to determine the function of a sandwich course training, as well as the dialectical difference between project and process. In terms of pedagogical engineering, our study displays educational situations, while attempting to point out the various obstacles we have managed to identify; we also show two surveys using statistical implicative analysis (Régis Gras).

Key-words: statistics training, questionnaire surveys, simulations, conceptual fields, educational situations, statistical implicative analysis, pedagogical engineering, statistics, IUT (polytechnics, a two-year course), sandwich courses.