

Construction d'un curriculum concernant l'analyse statistique implicative pour des étudiants BAC +2

Jean-Claude Oriol

► **To cite this version:**

Jean-Claude Oriol. Construction d'un curriculum concernant l'analyse statistique implicative pour des étudiants BAC +2. 42èmes Journées de Statistique, 2010, Marseille, France, France. 2010. <inria-00494854>

HAL Id: inria-00494854

<https://hal.inria.fr/inria-00494854>

Submitted on 24 Jun 2010

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.

CONSTRUCTION D'UN CURRICULUM CONCERNANT L'ANALYSE STATISTIQUE IMPLICATIVE POUR DES ETUDIANTS BAC +2

Jean-Claude Oriol
Cerral & IUT Lumière Université Lyon2

Résumé : ce texte a pour objet de décrire les premiers pas d'un apprentissage de l'analyse statistique implicative, théorie développée par Régis Gras. Cet enseignement est fait en direction d'étudiants de deuxième année de l'université. Nous décrivons un parcours d'apprentissage et quelques obstacles rencontrés par les étudiants.

Abstract : this text aims to describe the first steps of learning the implicative statistical analysis, a theory developed by Régis Gras. This teaching is done in the direction of second year students of the university. We describe a learning path and some barriers faced by students.

Mots clés : apprentissage de la statistique, analyse statistique implicative

1. 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.

1.1 Le département STID de l'IUT Lumière

Le département STID (STatistique et Informatique Décisionnelle) 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.

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.

1.2. Les projets d'études

Dans ce contexte les étudiants de deuxième année, par groupes de 5, mènent un projet d'étude statistique ordonnancée par un commanditaire d'une durée d'environ 60 heures entre janvier et juin et dont la moitié seulement est encadrée par un animateur (en général un enseignant ou un enseignant-chercheur du département).

Chaque année, depuis plusieurs années, un groupe d'étudiants mène un projet utilisant l'analyse statistique implicite comme outil. Cela nous a amené à observer, dans ce cadre, les apprentissages mis en jeu et à construire des outils afin de faciliter l'acquisition des concepts, et ce que nous allons développer ci-dessous.

2 Un apport concernant la théorie et un résumé

Nous avons fait plusieurs expérimentations concernant les éléments de base de la théorie :

- Cours de 2 fois 2 heures ou bien
- Travail autonome sur les 60 premières pages du livret GRAS (1996)

Dans les deux cas ce travail débouche sur l'écriture d'un compte rendu d'une dizaine de pages exposant les principaux points du B-A BA de l'analyse statistique implicite. Lors de l'exposé nous reviendrons sur ces productions écrites et les problèmes rencontrés.

3 Travail sur un fichier type

Le travail précédent est suivi de deux activités menées en parallèle

- A. D'une part l'étude approfondie du mode d'emploi du logiciel CHIC
- B. D'autre part le traitement d'un fichier type afin qu'ils appliquent l'analyse statistique implicite et qu'ils fassent fonctionner CHIC. Le fichier comporte 26 individus et 15 variables.

Beaucoup d'observation de cette situation montrent un certain nombre de difficultés qui se répètent comme par exemple de ne pas décomposer la variable Sexe en deux variables H et F. Voici d'ailleurs l'arbre des similarités où l'on constate ce phénomène :

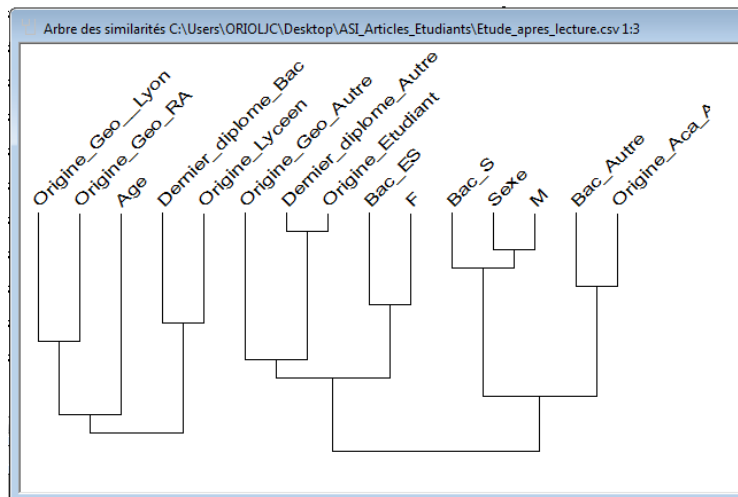


Figure 1 : arbre des similarités

Mais d'autres interprétations vont venir compléter les premiers constats par d'une part le graphu implicatif :

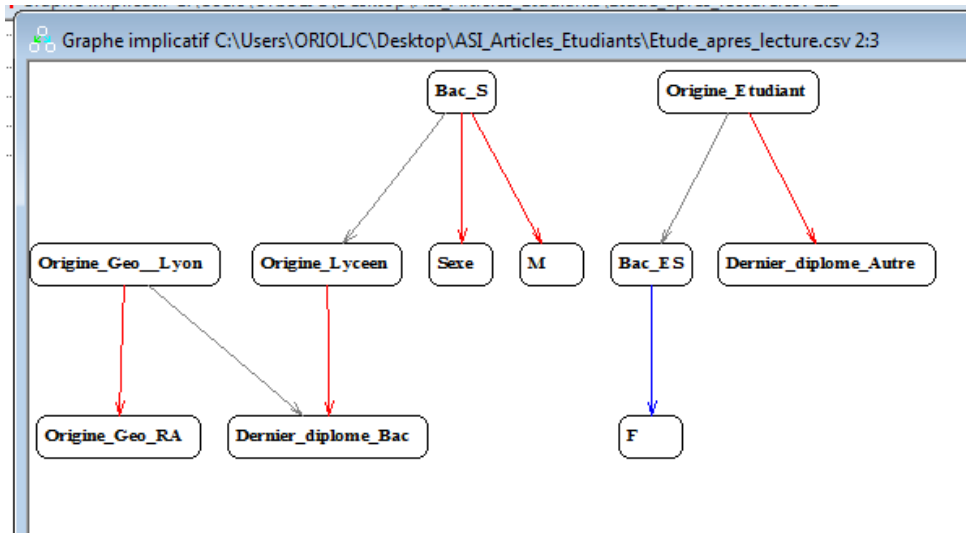


Figure 2 : graphe implicatif

Et d'autre part l'arbre cohésitif :

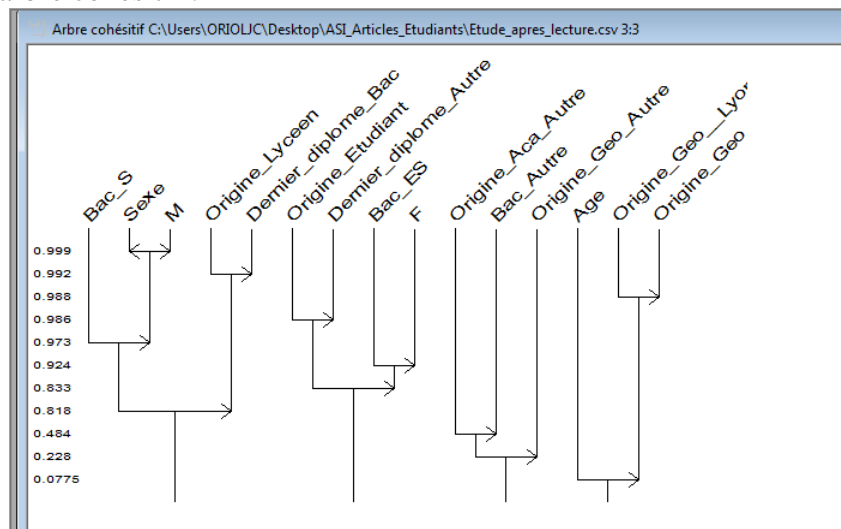


Figure 3 : Arbre cohésitif

Mais que l'on ne méprenne pas la production des graphes grandement facilitée par le logiciel CHIC est parfois porteuse de représentations peu conformes et l'on reviendra souvent avec les étudiants aux tables qui vont avec ces représentations.

3. Simulations concernant l'analyse statistique implicative

3.1. La place de la simulation

Notre a priori est de croire que 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, par l'intermédiaire des invariants opératoires, un éclairage sur le signifié.

3.2. 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 licence professionnelle. D'autres études sont menées par divers groupes pour d'autres commanditaires. Citons par exemple une étude sur des données médicales longitudinales pour la médecine du travail, une concernant les usages des NTIC pour l'Espace Numérique Entreprises, une autre sur la fréquentation des maisons du Rhône pour le Conseil Général du Rhône et une enfin pour le groupe Epures. Les méthodes utilisées dans ces études peuvent être libres ou bien imposées par le commanditaire.

3.3. Ressources mises à disposition du groupe d'étudiants : ordinateurs avec logiciels (Excel, CHIC, etc.), en tout 60 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 (notée), 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 cidessous.

3.4 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.

3.5 .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 R, 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...

Lors de l'exposé nous montrerons les réalisations des étudiants concernant à ces simulations.

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

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.

Comparaison de valeurs expérimentales de q et de Phi

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.

3.7. Recherche de bénéfices directs concernant l'ASI : q est symétrique, il existe une relation entre le Khi² et q(a,non(b))

Nous avons recherché à vérifier les relations entre q(a,non(b)) et q(non(b),a). Cela permet aux étudiants de mieux comprendre les formules utilisées.

3.8 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 puisque les deux sont possibles avec le logiciel utilisé.

La comparaison des valeurs de l'intensité d'implication calculée d'une part avec la loi de Gauss et d'autre part avec la loi de Poisson montre un écart d'environ 0.05 sur des valeurs proches de 0.65. D'une expérience à l'autre l'écart varie peu et il est de l'ordre de 0.05.

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

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

Nous indiquerons dans la présentation les dialogues entre les étudiants lors de la lecture de 100 valeurs de q et de 500 valeurs de Phi

4. Conclusion

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 différentes représentations, entre signifié et signifiant, à construire des invariants opératoires, à construire le sens des situations, bref à conceptualiser l'analyse statistique implicative.

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.

Bibliographie

- [1] GRAS, R. (1996) L'implication statistique, nouvelle méthode exploratoire de données, La Pensée Sauvage Editions, Grenoble.
- [2] GRAS R., REGNIER J.-C., GUILLET F. (Eds) (2009) Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités. RNTI-E-16 Consultable : http://www.cepadues.com/livre_details.asp?l=897
- [3] ORIOL, J-C. REGNIER, J-C. (2003a) « 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.
- [4] ORIOL, J-C. REGNIER J-C., (2003b) « Fonctionnement didactique de la simulation en statistique dans l'enseignement du concept de corrélation », *Espace Mathématique Francophone 2003*, Tozeur, Tunisie.
- [5] ORIOL, J-C. REGNIER, J-C. (2006) Formation en statistique en DUT STID et Transposition didactique, *38^{èmes} Journées de Statistique, SFDS Clamart*, <http://hal.inria.fr/aut/jean-claude+oriol/>
- [6] ORIOL, J-C. REGNIER, J-C. (2007) Conceptualisation de l'analyse statistique implicative, *39^{èmes} Journées de Statistique, SFDS Angers*, <http://hal.inria.fr/aut/jean-claude+oriol/>
- [7] ORIOL, J-C. REGNIER, J-C. (2007b), Enseignement-apprentissage de l'ASI en 1er cycle universitaire, *ASI 4, Castellon (Espagne)*. <http://hal.inria.fr/aut/jean-claude+oriol/>
- [8] ORIOL, J-C. (2007) 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, Thèse sous la direction de JC Régnier, <http://tel.archives-ouvertes.fr/tel-00191166/fr/>
- [9] 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., Tavignot P. (dir.), *Vingt ans de didactiques des mathématiques en France*, La Pensée Sauvage Editeurs, pp. 177-191, Grenoble.