

## Les examens de statistique sous la loupe de l'analyse exploratoire des données

Pierre Cullus

Service de Biostatistique et Informatique Médicale,  
Faculté de Médecine, Université Libre de Bruxelles, CP 623  
Route de Lennik, 808 ; 1070 Bruxelles  
Adresse Courriel : [pcullus@ulb.ac.be](mailto:pcullus@ulb.ac.be)

### **Résumé.**

Le cours de base de Biostatistique dans le Pôle Santé de l'Université propose aux étudiants une introduction aux éléments fondamentaux de la statistique médicale à partir d'exemples concrets. Ces derniers sont destinés à favoriser l'appréhension, la compréhension, l'application et l'interprétation des principales techniques utilisées dans les premières investigations des études biomédicales.

L'évaluation du cours est constituée par un examen écrit s'appuyant sur des « Questions à choix multiples » pour la partie « théorie » et sur des « Questions à réponses ouvertes » justifiées pour ce qui concerne les exercices .

Cette formule expérimentée permet d'engranger aisément les réponses des étudiants, d'analyser celles-ci ( par le truchement des techniques de l'Analyse des Données ) et de dégager de précieux renseignements d'ordre pédagogique.

Le dépouillement exhaustif des résultats des copies d'examens révèle clairement les erreurs systématiques commises (au niveau de la compréhension des énoncés, du choix des tests, de l'utilisation des « formules », de la lecture des tables,...).

Ce relevé permet de cibler les lacunes et d'adapter en conséquence notre enseignement.

Des disparités importantes dans la réussite invitent à la réflexion quant aux stratégies d'apprentissage des étudiants et à l'appréhension de l'examen (report de session, gestion du temps d'examen, ciblage sur une partie...).

Nous examinons également le positionnement spécifique des notes du cours de Biostatistique parmi les autres matières dans les différentes sections du Pôle Santé.

Le propos de cet exposé n'est certes pas de fournir une interprétation globale et définitive mais plutôt de souligner l'intérêt d'une analyse docimologique.

**Mots-clés.** Analyse exploratoire des Données ; Composantes principales ; Correspondances multiples ; Classifications ; Analyse conjointe de tableaux ; Biostatistique ; Examens ; Docimologie ; Evaluation ; Enseignement ; Logiciels SPAD et SPSS

**Abstract.**

**Exploratory Data Analysis of test scores of Biostatistics : contributions to pedagogy**

This paper proposes two readings, based on exploratory multivariate analyzes, of test results of Biostatistics. This course is followed by the students of Bachelor of Faculties of Medicine and Movement Science at the Free University of Brussels.

The first reading concerns the positioning of the course among other courses during the last academic years.

The syntheses are made available through conventional methods of Data Analysis (Factor Analyses and joint tables, classifications). The connections between the results are identified, enriched by the addition of criteria "illustrative" (section, academic delay, gender, ...).

Typologies of students are made to develop opportunities for differentiated instruction.

We examine, in a second reading, the information provided by the responses to examinations of Biostatistics, involving simultaneous theoretical questions multiple choice and exercises with short answers.

The counting of the results reveals systematic errors committed by students (understanding statements, choice of tests, use of "formulas", reading tables, ...) and can target the gaps in order to adjust our course accordingly.

Disparities in the success invite reflection on strategies for student learning, apprehension of examination and adequacy of the evaluation.

This approach to test evaluation by Exploratory Data Analysis provides valuable indicators for the proper monitoring of our teaching.

**Keywords.** Exploratory Data Analysis ; Principal Component Analysis ; Multiple Correspondence Analysis ; Classifications ; Joint Analysis of tables ; Biostatistics ; Examinations ; Docimology ; Evaluation teaching ; SPAD software and SPSS

Nous proposons dans cet exposé deux lectures complémentaires d'analyses statistiques de notes obtenues par des étudiants à des examens classiques de « Biostatistique ». Les deux approches ont été obtenues chacune par la mise en œuvre de différentes techniques de l'Analyse exploratoire classique des Données. Nous montrerons l'intérêt des méthodes de traitements multivariés, non

seulement dans le cadre de l'évaluation et du positionnement d'un cours mais également dans l'optique de dégager des perspectives pédagogiques.

Au sein du Pôle Santé de l'ULB, dans les différentes sections des Facultés de Médecine et de Sciences de la Motricité, le cours ex cathedra de Biostatistique de base de Bachelier (huit séances de deux heures) propose aux étudiants une introduction aux éléments fondamentaux de la statistique médicale à partir d'exemples concrets. Ces derniers sont destinés à favoriser l'appréhension, la compréhension, l'application et l'interprétation des principales techniques utilisées dans les premières investigations des études biomédicales.

L'acquisition des notions est renforcée par des séances d'exercices pratiques (huit séances de deux heures assorties de séances complémentaires de révision, de guidances et de permanences) .

Les chapitres traités portent successivement sur les statistiques descriptives de base, les distributions de probabilité, les intervalles de confiance classiques sur une moyenne et une probabilité, les tests d'hypothèse classiques sur une ou deux moyennes et sur une ou plusieurs probabilités. Sont abordés ensuite les régression et corrélation linéaires, les tableaux de contingence et le test du chi-carré. Les tests non paramétriques de base sont présentés. Le cours se termine par une introduction aux techniques de l'Analyse de variance ( ANOVA ) à un et deux facteurs.

Les étudiants disposent de supports d'enseignements détaillés, revus et complétés annuellement : fascicule de diapositives projetées au cours, syllabus du cours théorique, syllabus d'exercices pratiques, recueil de tables statistiques et formulaire. Le site de l'enseignement sur l'Université Virtuelle leur permet d'être confrontés à de nombreux exercices supplémentaires, portant aussi bien sur la matière théorique que sur l'aspect « exercices pratiques » du cours. L'étudiant curieux pourra également y découvrir quelques notions complémentaires ainsi que l'une ou l'autre démonstration occultée au cours, souvent pour cause de volume horaire réduit.

A l'issue de cet enseignement, les étudiants « devraient » être capables, face à des données biomédicales, de procéder au choix correct des premiers traitements statistiques à effectuer et de dégager l'information fournie par ceux-ci. Une compréhension et une lecture critique des traitements statistiques « élémentaires » des publications scientifiques sont également espérées de leur part.

Le temps limité réservé à ce cours ne permet naturellement pas d'envisager l'utilisation de l'ensemble des méthodes statistiques que pourraient rencontrer les étudiants lors de leurs travaux de fin d'études. A cet effet, un cours complémentaire est dispensé en Master dans certaines sections, comportant des notions additionnelles sur l'ANOVA et une introduction à la régression multiple. Cet enseignement donne aussi l'opportunité à l'étudiant de se familiariser à l'utilisation du logiciel statistique SPSS et de le rendre ainsi plus autonome.

L'évaluation du cours de base est constituée par un examen écrit s'appuyant sur des Questions à choix multiples (QCM) pour la partie « théorie » et sur des Questions à réponses ouvertes et courtes mais cependant justifiées ( QROC ) pour ce qui concerne les exercices .

Cette formule expérimentée depuis plusieurs années et affinée graduellement, tant en première qu'en deuxième session d'examens, permet d'engranger aisément les réponses des étudiants,

d'analyser celles-ci et de dégager de précieux renseignements d'ordre pédagogique. La constitution systématique de bases de données nous suggère de tenir compte du caractère évolutif et permet de pérenniser la fonction de mémoire du cours que les examens remplissent.

Dans une première lecture, nous analysons le positionnement du cours très spécifique de Biostatistique parmi les autres matières dans les différentes sections du Pôle Santé. Nous nous limitons volontairement aux cinq dernières années académiques, afin d'écartier les risques d'anachronismes.

Les synthèses sont acquises grâce aux méthodes classiques d'Analyse des Données (Analyses factorielles, analyses conjointes des tableaux et analyses de classification, toutes réalisées essentiellement à l'aide du logiciel SPAD).

Les liaisons entre les résultats des différentes matières ( et l'influence de ces dernières dans l'évaluation finale de l'étudiant ) sont mises en exergue, enrichies par l'adjonction de différents critères « illustratifs » associés à l'étude ( section de l'étudiant, retard académique, première année d'inscription, sexe ,...).

Une comparaison des analyses multivariées est effectuée entre les sections ; à l'intérieur de celles-ci, la stabilité temporelle des résultats est étudiée par l'analyse conjointe de la succession des tableaux de résultats.

Des typologies des étudiants sont construites afin d'envisager des possibilités de pédagogie différenciée au sein de groupes homogènes (organisation de séances de révisions et de guidances ).

Nous examinons, dans une seconde lecture, l'information apportée par les réponses fournies par les étudiants aux questions posées lors des examens de Biostatistique.

Nous utilisons également ici les techniques classiques d'Analyse des Données.

Le dépouillement exhaustif des résultats des copies d'examens révèle clairement les erreurs systématiques commises par les étudiants (au niveau de la compréhension des énoncés, du choix des tests, de l'utilisation des « formules », de la lecture des tables,...). Ce relevé permet de cibler les lacunes et d'adapter en conséquence notre enseignement.

Des disparités importantes dans la réussite invitent à la réflexion quant aux stratégies d'apprentissage des étudiants et à l'appréhension de l'examen (report de session, gestion du temps d'examen, ciblage sur la partie pratique ou théorique...).

De même se pose la question de la pertinence de l'évaluation (choix des QCM, élaboration des questions pratiques orientées sur la compréhension de la matière et évitant la complication des calculs, établissement d'une grille de correction, ...).

L'approche de la docimologie par les différentes techniques d'Analyse des Données fournit ainsi régulièrement des instruments précieux aussi bien pour assurer la qualité pédagogique des séances de cours et d'exercices que pour élaborer et traiter les évaluations (dont l'objectivité se doit de répondre aux soucis de rigueur et de justice) des étudiants.

Le propos de cet exposé n'est certes pas de fournir une interprétation globale et définitive mais plutôt de souligner la volonté de poursuivre la recherche dans ce domaine, en particulier par l'examen de diverses considérations générales non abordées ici et la formulation de nouvelles hypothèses.

L'histoire de la docimologie des examens de statistique ne s'arrête d'ailleurs pas à ce stade, elle se fera avec des acteurs toujours nouveaux, des enseignements remodelés et des évaluations peut-être différenciées ! A partir de ce constat, tout reste donc sujet à interrogations ... à traiter par l'Analyse exploratoire des Données !

### **Bibliographie**

[1] Gilbert Saporta, *Probabilités, Analyse des données et Statistiques*, 3<sup>ème</sup> édition révisée, Paris, Editions Technip, 2011, 622 p. (ISBN [978-2-7108-0980-7](#)).

[2] Stéphane Tufféry, *Data Mining et statistique décisionnelle*, Paris, Editions Technip, 2010, 705 p. (ISBN [978-2-7108-0946-3](#))

[3] Brigitte Escofier, Jérôme Pagès, *Analyses factorielles simples et multiples*, 4<sup>ème</sup> édition, Paris, Dunod, 2008, 318 p. (ISBN [978-2-10-051932-3](#))

[4] Ludovic Lebart, Marie Piron et Alain Morineau, *Statistique Exploratoire Multidimensionnelle*, 4<sup>ème</sup> édition, Paris, Dunod, 2006, 464 p. (ISBN [978-2-10-049616-7](#))

[5] Christine Lavit, *Analyse conjointe de tableaux quantitatifs*, Paris, Masson, 1988, 254 p. (ISBN [978-2-225-81476-3](#))