

# Avec des données réelles: pouvons-nous nous initier au Statistique ?

Lara-Porras, Ana María

Román-Montoya, Yolanda

*Department of Statistics and O.R., University of Granada*

*Granada (18071), Spain*

*E-mail: [alara@ugr.es](mailto:alara@ugr.es), [yroman@ugr.es](mailto:yroman@ugr.es)*

## Introduction

L'objectif de cette étude est l'enseignement et l'apprentissage des Statistiques. Plus particulièrement, nous nous intéressons à la problématique spécifique que cette matière présente aux étudiants qui effectuent un cours introductif aux statistiques dans le contexte d'analyse de données réelles et d'interprétation de résultats sur ordinateur.

L'analyse de données réelles est une activité qui requiert un certain niveau de connaissances même à un niveau exploratoire, exigeant des connaissances diverses sur les problèmes et concepts mis en jeu dans les procédés numériques, graphiques, descriptifs et d'inférence en relation avec l'association. Cela nécessite le choix d'instruments statistiques et de représentations graphiques les mieux adaptés aux données, la flexibilité pour changer les procédés, une interprétation adéquate des résultats, une capacité à mettre ceux-ci en relation avec le problème et l'évaluation de la validité et de la fiabilité des conclusions.

Lorsque nous travaillons avec des données dans la classe, généralement nous n'indiquons pas la méthode de recueil des données utilisées. De plus, nous ne pouvons pas travailler manuellement avec de grandes quantités de données. Par conséquent, nous construisons des résumés de données qui ne sont pas analysables sur la base de leurs éléments constitutifs et ils restent déconnectés des données originales. L'une des clés de ce projet est l'utilisation de données réelles et l'aide de la technologie informatique. Les moyens informatiques permettent de visualiser des grandes séries de données dans un temps relativement petit. De cette manière, nous nous centrons sur le raisonnement probabiliste.

## Contexte/Justification

Le contexte de cette étude est le nouveau panorama académique qui se profile avec l'implantation d'un Espace Européen de l'Education Supérieure (EEES), dans lequel l'acteur didactique du cours passe du professeur à l'élève, de telle sorte que c'est ce dernier qui décide de son propre apprentissage à l'intérieur de l'éventail de possibilités de chaque matière; et ce, dans le sens où l'objectif n'est pas que le professeur détecte dans l'élève la connaissance, mais que le professeur accompagne, oriente et soutient l'étudiant, pour qu'il acquiert la capacité d'apprendre la base cognitive, les habilités transversales et spécifiques et les attitudes exigées par la communauté européenne, le monde du travail, et son orientation existentielle. La complexité du processus enseignement-apprentissage a, en plus du professeur et de l'étudiant, un troisième objet d'attention: la communication entre ces deux acteurs.

Notre proposition méthodologique s'est concentrée sur les statistiques, du troisième cours comportant six crédits (soixante heures concernées), du Master de Sciences Environnementales réalisé dans l'Université de Grenade (Andalousie – Espagne). La plupart des étudiants qui suivent cette matière n'a pas

encore étudié les statistiques, ou dans le meilleur des cas, ils ont seulement quelques notions rudimentaires des concepts de base des statistiques descriptives.

Le paradoxe est que le cours introductif de la statistique de niveau universitaire a un caractère introductif (c'est la première fois que l'étudiant étudie sérieusement la statistique) et dans le même temps un caractère final, puisque on attend que l'étudiant obtienne la validation de la matière avec des connaissances suffisantes pour lui donner la capacité de les utiliser au cours de sa future vie professionnelle.

### **Méthodologie/Description de l'expérience**

Le projet que nous proposons est orienté sur la proposition d'une meilleure compréhension et une meilleure assimilation des concepts et des méthodes statistiques. Pour cela, nous proposons l'utilisation de données réelles, en mettant l'accent sur le programme de l'étude, dans la réalisation de l'analyse et de l'interprétation des résultats. Depuis le début du cours, les étudiants ont accès à l'information sur les données, les variables de travail et les objectifs qu'ils s'engagent à atteindre.

Nous encourageons le travail de groupe entre les étudiants en faisant des recherches sur la question posée par les étudiants, ce sont eux qui formulent des hypothèses, dessinent l'étude, recueillent et analysent les données, interprètent les résultats et en obtiennent des conclusions. Les activités structurées renforcent les concepts et les habilités statistiques requises pour réussir à résoudre avec succès le problème. Les étudiants utilisent des outils graphiques ou numériques appropriés et expliquent leur choix. En général, ce choix se base sur les aspects intrinsèques, comme le type de données ou la méthode induite par la question de la recherche. Ils sont capables de modifier et transformer ses représentations, et à travers elles, ils interprètent et justifient la question posée par le problème et interprètent ses résultats.

Nous envisageons d'introduire les concepts et les techniques statistiques de la matière «Statistiques» de 3<sup>ème</sup> cycle grâce à un projet statistique basé sur les données du Département d'Ecologie de l'Université de Grenade. Ces données proportionnent différentes mesures sur les feuilles de chênes verts sur cinq secteurs différents de la province de Grenade. Pour cela, on dispose de différentes mesures de feuilles (longueur, largeur à gauche, largeur à droite) réalisées sur différentes parties de l'arbre (Canopi: Cime des arbres et Sprouts: Pousses, feuilles nouvelles qui poussent sur la partie inférieure de la plante), provenant de différentes zones de la province de Grenade et durant trois ans consécutifs.

L'objectif du projet est l'étude de l'asymétrie des feuilles. L'asymétrie se définit en fonction de la largeur à gauche, à droite et totale, et sera le principal centre d'intérêt de notre projet. Il s'agit d'observer si les situations de manque de nutriments, de sécheresse, etc (appelées situations de stress) affectent ou non cette asymétrie. Les hypothèses qu'émet le département d'écologie sont: «Les situations de stress modifient la symétrie des feuilles». L'objectif de notre cours est que les élèves parviennent à accepter ou rejeter ces hypothèses en se servant pour cela de tous les outils statistiques nécessaires.

Depuis le début des cours, l'élève va se confronter à un ensemble de données réelles. Le professeur décrit les données et les variables avec lesquelles il travaille. On incite les élèves à se demander ce qu'ils vont faire avec ces données. La prochaine démarche est de laisser les statistiques «naître» des questions qui se posent en classe de la part des élèves et des professeurs. Notre objectif est de motiver les élèves de telle façon que ce soit eux qui appliquent les concepts théoriques avec lesquels se résout une situation réelle. Nous, professeurs, nous nous convertissons en simples conseillers d'orientation de l'évolution de nos élèves en

fournissant l'information qu'ils sollicitent.

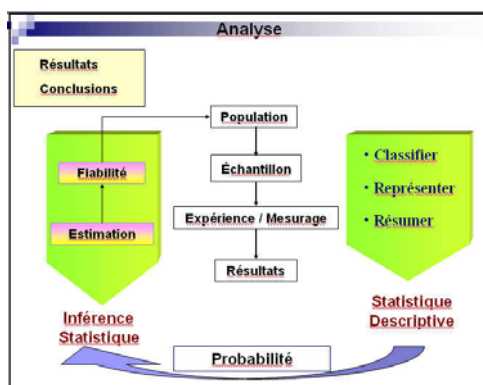
En premier lieu, et à partir d'un fichier de données qui compte 12 variables mesurées sur 2100 individus, s'introduisent les concepts basiques: le nombre total de chênes verts de Grenade permet de définir le concept de *Population*; chacun des arbres susceptibles d'être analysés définit ce que nous appelons *Individu* et les individus sélectionnés pour être analysés constituent l'*Échantillon*; la symétrie des feuilles nous amène de manière naturelle au concept de *Variable statistique*. Les données qualitatives sont prises en compte au moment d'observer l'analyse des différentes parties de la plante; les données quantitatives seront prises en compte comme résultat de l'observation de données numériques telles que l'année, la longueur, les distances... Avec ces mêmes variables et en fonction des caractéristiques des valeurs qu'elles prennent, on distingue les *Variables discrètes* comme la variable année, et les *Variables continues* comme les variables longueur et hauteur.

A la vue d'un grand nombre de données avec lesquelles les élèves devront travailler et de l'impossibilité de manier individuellement une telle quantité d'informations, on incite les élèves à se poser la question de la nécessité d'organisation des données: classer, représenter et résumer. Il apparaît donc le concept de *Statistiques Descriptives*. À mesure que les élèves s'investissent dans les Statistiques, il semble nécessaire d'utiliser un logiciel statistique qui leur permettrait de manipuler et d'analyser de grands nombres de données.

Le professeur se doit d'aider les élèves à organiser un décompte et un tableau de fréquences en leur montrant l'utilité de ce type de tableaux pour résumer les informations. Des questions d'organisation se posent donc, obligeant les élèves à voir la nécessité d'avoir une visualisation des données, des représentations graphiques et des mesures qui résument les données d'un tableau statistique en valeurs numériques qui les représentent: c'est l'*Analyse descriptive quantitative*.

Des questions se poseront: quelles relations y a-t-il entre les variables? Peut-on prévoir le comportement d'une variable en fonction du comportement d'une autre? A-t-on bien établi la relation entre les variables? Les prévisions sont-elles acceptables? Comment trouver une fonction qui représente les données? Et entre toutes les fonctions possibles, comment savoir si la fonction a été bien choisie. En existe-t-il une autre qui explique mieux les données? Nous sommes en train d'introduire les concepts de régression et de corrélation.

Avec le schéma suivant et en incitant constamment les élèves à participer, nous montrons comment va se dérouler la matière et la nécessité d'un lien entre les *Statistiques descriptives* et les *Statistiques d'inférence* que nous appelons *Probabilité*.



Nous construisons entre professeurs et élèves les concepts de probabilité; nous devons trouver un

moyen d'apprentissage adéquat pour aider les étudiants à transformer leurs croyances en probabilité en intuitions effectives et à obtenir une compréhension conceptuelle solide de la stochastique. Nous analyserons l'origine des difficultés qu'ont les étudiants en stochastique, nous construirons des moyens d'apprentissage qui encouragent le développement de conceptions correctes, et qui promeuvent des attitudes positives vers la discipline. Wilensky (1995) définit l'*Anxiété épistémologique* comme le sentiment de confusion et d'indécision que ressent la majorité des étudiants face aux différentes possibilités ou voies de résolution d'un problème stochastique. Cette *anxiété* est renforcée par la culture mathématique «protectrice» qui n'encourage pas le dialogue entre le personnel d'enseignement et les élèves ni entre les élèves eux-mêmes pour avancer dans la construction du concept.

Le développement du cours suivra toujours le même schéma. Une fois identifiés le problème et la solution, les résultats des diverses techniques statistiques se présenteront à l'élève comme des issus du programme statistique choisi qui, dans notre cas, est SPSS; postérieurement, les élèves répèteront le processus sur un échantillon et travailleront avec différents exercices. Depuis le début, il résulte de grande utilité comparer les résultats obtenus au niveau de la population et au niveau de l'échantillon, de manière à ce que l'élève, depuis le début du cours, soit familiarisé avec le concept d'aléatoire et sa signification. En même temps et quasiment de façon immédiate, une question se pose: les conclusions que nous obtenons pour n'importe lequel des échantillons sont-elles valables, peuvent-elles s'extrapoler à tous les individus de la population? La nécessité de l'inférence statistique se pose ainsi de manière naturelle.

Nous arrivons ainsi à la fin du cours et nous sommes capables de répondre aux questions que nous avons posées au moment d'énoncer l'analyse: la sécheresse, le manque de nutriments,... augmentent-ils l'asymétrie? La variation dans l'asymétrie des feuilles peut-elle être considérée comme un indicateur de stress pour les plantes, concrètement chez les chênes verts? Nos questions initiales ont des réponses statistiques à travers les *Contrastes d'hypothèses*.

## RÉFÉRENCES

Lara-Porras, A.M. "Estadística para Ciencias Biológicas y Ciencias Ambientales. Problemas y Exámenes resueltos". Ed. Proyecto Sur. Granada (España), 2002.

Lara-Porras, A.M.; Román-Montoya, Y.; Jiménez-Lara, P.; Marín-Cantero, R.; Tarifa-Blanco, J.A. "Como introducimos en la Estadística mediante unos datos? Métodos Multimedia". Ed. ALARA.España, 2006 (CD ROM).

Lara-Porras, A.M.; Román-Montoya, Y.; García-Leal, J. "I have a set of data, what can I do with them?" 7è *International Conference on Teaching Statistics (ICOTS7)*. Salvador de Bahia, Brasil, 2006.

Wilensky, U. "Paradox, programming, and learning probability: A case study in a connected mathematics framework". *Journal of Mathematical Behavior*, 14, 253-280, 1995.

## RÉSUMÉ

*Adoptons-nous en cours une attitude de réflexion et de recherche autour d'un travail à effectuer? Faisons-nous en sorte que les élèves se sentent plus motivés et plus intéressés? Effectuons-nous le lien entre ce qui est expliqué en cours et la réalité? Encourageons-nous une attitude de critiques des recherches en classe? L'objectif suivant vient donc en réponse à ces questions et à d'autres inquiétudes. En effet, nous nous proposons d'introduire la statistique au moyen d'un projet statistique basé sur des données réelles en rapport avec la symétrie des feuilles de chêne dans la province de Grenade (Espagne).*