

Quatre conférences critiques qui peuvent détourner les non-statisticiens de la statistique

Richard Hindls, Stanislava Hronova

University of Economics, Faculty of Informatics and Statistics

nám. W. Churchilla 4

130 67 Prague 3, Czech Republic

hindls@vse.cz, hronova@vse.cz

L'enseignement de la statistique représente une composante de la plupart des programmes d'études dans les universités. On aurait du mal à trouver dans l'enseignement supérieur des spécialisations ne contenant pas au moins un cours élémentaire de statistique – le nombre de crédits étant naturellement différent. Il faut néanmoins avouer (au grand regret des statisticiens) que la statistique ne fait pas partie des matières aimées des étudiants. Même, pour certains étudiants, elle représente un cauchemar. En général, on sait bien qu'un enseignant excellent peut enthousiasmer ses étudiants pour des matières qu'ils ont refusées à l'origine. A contrario, détourner les étudiants de la statistique par un mauvais cours est facile et donc malheureusement fréquent. Prêtons attention à quelques parties du cours élémentaire de statistique où un enseignant pédagogiquement mal préparé peut tout au début complètement étouffer l'intérêt potentiel pour la statistique.

Le cours élémentaire de statistique dans les universités a un cadre commun. Commun pour toutes les universités et toutes les spécialisations dans le monde entier. Il sert à la création de l'appareil statistique, il présente les façons dont la statistique reflète la réalité. Voilà les éléments de base : I - concepts de base, II - traitement statistique élémentaire, III - statistique descriptive, IV - calcul des probabilités, V - bases de l'induction statistique, VI - étude de la liaison entre deux (ou plusieurs) variables, VII - analyse des séries chronologiques.

A ces parties-là, il convient d'en ajouter d'autres suivant les besoins des programmes d'études dans lesquels un cours de statistique est dispensé. Selon nos expériences de longue durée ces parties du cours élémentaire de statistique cachent certaines lacunes. Nous allons en présenter quatre. Elles représentent des points critiques de la démarche pédagogique et on peut dire que leur maîtrise est une condition nécessaire au succès de l'enseignement de la statistique pour les non-statisticiens. De quoi s'agit-il ? Les parties I et II ne posent pas de problèmes. Elles ont un caractère descriptif complété par des images et des graphiques. Le succès est garanti si l'on utilise des exemples actuels de la vie pratique.

Premier point critique. Le premier point critique gît dans la partie III et il est souvent insidieux. Sa mauvaise maîtrise peut facilement endormir la moitié d'étudiants. Un silence inhabituel et inattendu et des regards vitreux des étudiants représentent un premier avertissement pour un enseignant. Le problème consiste au fait que cette partie critique concerne la dispersion qui est sans aucune doute un des concepts statistiques les plus importants.

La partie III commence par l'explication des caractéristiques descriptives (les quantiles et les caractéristiques de tendance centrale). On ne peut pas dire que cette partie est très intéressante pour les étudiants, mais pour la plupart du temps ils l'acceptent et ils la comprennent bien. Et tout d'un coup, un changement brusque arrive. Il est représenté par des caractéristiques de dispersion, caractéristiques si importantes pour la statistique. Les étudiants perçoivent la nécessité de mesurer la variabilité, mais leur capacité à comprendre la notion de variance est une pierre d'achoppement. Le calcul de la variance n'est pas primordial (grâce à l'utilisation des logiciels et des calculatrices) mais malgré tout, les étudiants se concentrent trop sur les techniques et ils interprètent mal le sens de la variance.

La solution pédagogique est simple. Il faut bien expliquer (à l'aide d'exercices pertinents) que la variance est une moyenne des carrés des écarts à la moyenne et que l'écart-type est une moyenne (quadratique) des écarts à la moyenne. Mais il est évident que l'expression « écart moyen à la moyenne » n'est pas non plus facile à comprendre.

Deuxième point critique. On le trouve dans la partie IV. Cette partie commence par les notions tels que événement aléatoire, épreuve aléatoire, théorèmes de probabilité, concept de probabilité, etc. La largeur de l'explication et la surabondance des opérations formelles avec des événements aléatoires représentent ici une faute. Les étudiants reconnaissent bien que les illustrations sont souvent artificielles, leur attention et leur intérêt diminuent vite. Le risque est que dans cette situation, les étudiants ne saisissent pas bien le passage aux notions de variable aléatoire et de loi de probabilité.

Dans la partie IV, il y a encore une lacune – on étouffe les étudiants par un exposé trop complexe des lois de probabilité. A notre avis, il suffit de présenter certaines d'entre elles (binomiale, de Poisson, de Laplace-Gauss ...). En ce qui concerne les lois théoriques (χ^2 , de Student, de Fisher-Snedecor), nos expériences montrent qu'il est beaucoup plus efficace de les présenter dans la partie V. Leur explication dans la partie IV est tout à fait correct, mais les étudiants ne sont pas encore capables de comprendre leur importance pour l'induction statistique. La solution pédagogique de ce problème consiste à la réduction de la description des théorèmes de probabilité et au passage efficace à la notion de variable aléatoire.

Troisième point critique. Il figure dans la partie V et il est représenté par le nombre énorme d'estimateurs et de tests. Ce problème est le plus facile à résoudre. Il suffit d'expliquer le fond des estimations ponctuelles, des estimations par intervalle et des tests d'hypothèses. En effet, les étudiants sont obligés d'affronter une grande offensive de la statistique mathématique après laquelle il ne sont plus capables de répondre à la question élémentaire - ce qui est estimé et par quoi. Les discussions incessantes sur les conditions d'utilisation semblent être inutiles et embrouillés aux étudiants. Les expériences montrent qu'il est bon de commencer l'explication par la typologie des tests (ou des intervalles de confiance). On peut dire que dans cette partie (plus que dans les autres) notre devoir consiste à présenter aux non-statisticiens les principes de la pensée statistique, c'est-à-dire en quoi consiste la réflexion statistique du monde et comment elle se différencie des autres.

Quatrième point critique. On le trouve dans la partie VI. Les étudiants maîtrisent bien la régression et la corrélation descriptives. Mais il est discutable d'incorporer des intervalles de confiance et des tests dans le cours de régression. Les non-statisticiens ont du mal à comprendre comment il est possible que dans la situation modélisée où le coefficient de régression doit être positif (par ex. en cas de liaison positive entre des dépenses y et des recettes x où $b_{yx} > 0$) on puisse obtenir, grâce à l'application de l'estimation par intervalle, une valeur négative ($b_{yx} < 0$).

La solution pédagogique n'est pas simple. Le problème de la confusion dans les pensées des étudiants a pour la plupart du temps deux causes. Premièrement : les étudiants perdent de vue que si la taille de l'échantillon est petite, l'intervalle de confiance peut être assez large. Par conséquent, les valeurs de coefficient de régression peuvent être positifs et négatifs. Deuxièmement : les données qui devraient prouver des liaisons incontestables en théorie sont souvent « contaminées » par des effets secondaires et conduisent ainsi à un large intervalle de confiance. Il est difficile pour des non-statisticiens débutants de comprendre toutes ces situations. Ainsi préférons-nous ne pas traiter la partie des estimations et des tests en régression dans le cadre du cours élémentaire pour des non-statisticiens.

Le cours élémentaire a pour vocation de susciter l'intérêt pour la statistique auprès des étudiants non-statisticiens. Mais il est évident que le cours élémentaire fait « la publicité » pour la statistique. Malheureusement, certains enseignants l'oublient et compromettent ainsi la réputation de la statistique. Mais soyons prudents : l'effort du « marketing statistique » peut quelquefois mener jusqu'à brader à des prix trop bas, ce que ne mérite pas la statistique. Le rôle des enseignants du cours élémentaire de statistique pour des non-statisticiens est important – et plus difficile qu'on le pense souvent. Aussi ne sous-estimons pas ceux qui se consacrent pendant toute leur carrière professionnelle au cours élémentaire. Incontestablement ils rendent un service utile à la statistique.

RÉSUMÉ

This article shows critical moments of elementary statistic courses in universities. Their pedagogical mastering is the vital condition for achieving good results in teaching statistics for non-statisticians. Turning points are especially the explanation of variance, probability distributions and hypothesis testing and interval estimation in regression.