

# SAVOIR ACADÉMIQUE, SAVOIRS PRATIQUES : TENSIONS ET RECHERCHE D'ÉQUILIBRE

## LE COURS DE STATISTIQUE À HEC PARIS ET ESCP EUROPE

Corinne HAHN<sup>1</sup> et Gilles STOLTZ<sup>2</sup>

### TITLE

Academic knowledge, practical knowledge: Tensions and quest for an equilibrium  
The statistics courses in two French business schools, HEC Paris and ESCP Europe

### RÉSUMÉ

Nous décrivons les difficultés auxquelles nous devons faire face en tant que professeurs responsables du cours fondamental de statistique dans deux écoles de management (ESCP Europe et HEC Paris), ainsi que les stratégies mises en œuvre pour les résoudre. Nous situons ces difficultés dans le cadre général des évolutions observées ces dernières années dans l'enseignement supérieur, et en particulier, l'enseignement supérieur de gestion, puis nous explorons la complexité particulière du positionnement de la discipline. Après avoir brossé le contexte spécifique de nos écoles, nous expliquons pourquoi la tension entre savoir académique et savoirs pratiques y est particulièrement forte dans l'enseignement de la statistique. Nous décrivons alors les stratégies mises en œuvre afin de tenter de trouver un équilibre entre ces savoirs et de rendre légitime la statistique aux yeux des étudiants. Nous les illustrons par la présentation de quelques activités ou jeux de données utilisés en cours.

*Mots-clés : écoles de commerce, curriculum, expérience commentée, légitimité.*

### ABSTRACT

We, coordinators of the first-level statistics courses in two business schools (ESCP Europe and HEC Paris), describe the challenges raised by teaching our field, as well as the strategies implemented to circumvent them. We put these challenges into the broader picture of the evolutions observed in higher education in the past years—and in particular, in the world of business schools—and then detail the complex-by-nature positioning of the field of statistics. After describing the specific context of our schools, we explain why the tension between academic and practical knowledge is particularly strong therein when it comes to teaching statistics. We then present the strategies used to find an equilibrium between these two kinds of knowledge and to make statistics a legitimate field of study in the eyes of the students. For the sake of concreteness, we provide some examples of activities performed or of data sets studied in our courses.

*Keywords: business schools, curriculum, from experience, legitimacy.*

Cet article est né des échanges entre deux professeurs, en charge chacun du cours fondamental de statistique au niveau pré-master dans deux écoles de management françaises, HEC Paris et ESCP Europe. Au cours de nos discussions, nous avons été amenés à confronter les difficultés rencontrées, à les mettre en perspective, à expliciter et comparer les stratégies de résolution mises en œuvre.

<sup>1</sup>ESCP Europe, Paris, [hahn@escpeurope.eu](mailto:hahn@escpeurope.eu)

<sup>2</sup>HEC Paris, CNRS, Jouy-en-Josas, [stoltz@hec.fr](mailto:stoltz@hec.fr)

Selon nous, la tension entre savoir académique et savoirs pratiques explique en grande partie ces difficultés. Aussi les solutions expérimentées dans le cadre de nos enseignements visent-elles en priorité à rechercher un équilibre entre ces différents savoirs. Bien que nous présentions ici le cas de deux écoles françaises, il nous semble que le débat dépasse largement le cadre de nos institutions et de notre pays, et que la plupart des aspects, difficultés et solutions traités peuvent être valables dans les écoles de management d'autres pays. (Une exception à cela concerne les liens avec la formation antérieure en classes préparatoires qui, elles, sont un système de formation propre à la France.)

Dans la première partie de cet article, nous allons situer les rapports entre savoir académique et savoirs pratiques dans le cadre plus général de l'évolution du système d'enseignement supérieur, et en particulier de celui de l'enseignement de gestion (paragraphe 1.1). Nous décrivons ensuite les attentes contradictoires des étudiants face à ces savoirs (paragraphe 1.2) puis nous considérerons le cas de l'enseignement de la statistique, une discipline dont le positionnement est intrinsèquement complexe (paragraphe 1.3).

Dans la deuxième et la troisième parties, nous décrivons le contexte de nos écoles (paragraphe 2.1), ainsi que les difficultés rencontrées et les solutions mises en œuvre. Nous avons choisi de structurer nos réflexions selon les trois angles d'approche suivants : le choix du degré de formalisation et des mises en situation (paragraphe 2.2) ; parvenir à faire vivre à nos étudiants un problème statistique (paragraphe 2.3) ; rendre légitime la statistique aux yeux des étudiants (partie 3). Ce dernier point est le plus crucial et le plus spécifique peut-être aux écoles de management<sup>3</sup>, c'est pourquoi il lui est consacré une partie à part. Comme révélateur de nos objectifs et attentes (paragraphe 2.4), nous donnerons un bref aperçu de la mise en cohérence de nos systèmes de notation avec notre discours.

## 1 Tension entre savoir académique et savoirs pratiques

En période de crise, la demande de la société vis-à-vis du système éducatif est avant tout une demande d'efficacité : il faut développer la valeur d'usage du savoir. Il ne s'agit pas de se limiter à enseigner des techniques qui deviendraient rapidement obsolètes dans un environnement en constante évolution. Les individus doivent être en mesure de produire des savoirs nouveaux afin de pouvoir s'adapter en permanence (Stevenson, 2002). Pour répondre à cette demande de la société, de plus en plus de pays déclinent les curricula selon une logique de compétences et recommandent de développer des approches transversales et interdisciplinaires.

Depuis quelques années cette évolution touche également l'enseignement supérieur. Lesart et Bourdoncle (2002) ont montré que le modèle dominant de l'université de recherche (celui des grandes écoles françaises, des *graduate schools* aux Etats-Unis et de l'université allemande) était progressivement remplacé par le modèle de l'université de service. Dans ce dernier, au service de la société et non pas de la science, la formation pratique, professionnalisante, occupe une place plus importante. Le changement de modèle ne se fait pas sans heurt : la nécessité de concilier les exigences d'académisation et celles de professionnalisation conduit à

---

<sup>3</sup>Dans cet article, nous utiliserons de manière interchangeable les termes « école de management » et « école de commerce », le premier ayant tendance à se substituer au deuxième dans nos écoles alors que ce dernier est encore le terme le plus utilisé par le grand public et les journalistes.

des conflits de valeurs et de représentations (Lemaitre, 2011). Les frontières entre formation initiale et formation continue deviennent plus floues et le paysage éducatif se recompose (Dunkel et Le Mouillour, 2007). Selon certains auteurs, ce changement est à l'origine d'un phénomène de « marchandisation » du savoir plus ou moins important selon les institutions, les plus cotées conservant un haut niveau académique, et selon les disciplines, celles à forte légitimité épistémologique résistant mieux que les autres (Naidoo et Jamieson, 2005).

Cette remise en question du fort cloisonnement disciplinaire va de pair avec un recentrage sur l'apprenant. S'il n'est bien sûr pas question de nier le rôle actif que le sujet joue dans le processus d'apprentissage, il est important de distinguer savoir de bon sens et savoir scientifique (Young, 2009) et de résister à la mode du culte de l'amateur (Keen, 2007). Dans une récente interview au journal *Le Monde*, le philosophe Matthew Crawford (2013) s'inquiète des conséquences de cette mode sur la capacité à développer une pensée complexe, à l'origine de l'innovation et du progrès scientifique. Si tout savoir est situé, il est important de ne pas l'enraciner dans un contexte mais de permettre à l'apprenant de développer à la fois la dimension horizontale (la capacité à effectuer un « transfert » d'un contexte à un autre) et la dimension verticale (disciplinaire).

## 1.1 Traduction dans l'enseignement supérieur de gestion

En ce qui concerne les écoles supérieures de gestion, le mouvement que nous décrivons pourrait être considéré comme une forme de retour en arrière. En effet, les sciences de gestion qui sont des disciplines récentes (nées dans les années 1970) ont cherché à asseoir leur légitimité en développant leur dimension académique. Elles ont mis l'accent sur la recherche afin de définir des savoirs scientifiques qui seraient enseignés dans les écoles (Le Moigne, 1999). L'évolution actuelle de l'enseignement supérieur les conduit à ré-interroger leur dimension praxéologique et à reconsidérer leurs cursus de formation, conçus à l'origine pour accueillir des étudiants après deux premières années d'études supérieures (généralement, un passage par des classes préparatoires). Ces écoles choisissent soit de professionnaliser des parcours dès la première année d'école, soit de se caler sur le modèle libéral dans lequel la formation professionnalisante intervient en master après un cursus d'études plus généralistes. Dans ce cas, la première année de formation en école est composée d'enseignements dits « fondamentaux », est qualifiée de dernière année de licence et est vue comme un prolongement des enseignements de classes préparatoires. Il semble qu'une voie existe entre ces deux options mais elle n'est certainement pas facile à construire. C'est un débat en cours dans de nombreux pays, et notamment en Amérique du Nord, dont les *business schools* sont plus que jamais, à l'heure de la compétition internationale pour attirer les meilleurs étudiants, le modèle de référence de nos écoles supérieures de gestion.

Cette évolution des cursus de formation va de pair avec l'évolution des profils des enseignants. Les écoles recrutent désormais en priorité des enseignants-chercheurs, dont l'activité de recherche permettra de progresser dans les classements internationaux ; ces derniers accordent en effet un poids important aux publications scientifiques. Ces enseignants sont traditionnellement moins investis en pédagogie et dans le champ de la formation continue que ceux de la génération précédente, qui avaient un profil d'enseignants-consultants.

Les écoles dont il sera question dans cet article, HEC Paris et ESCP Europe, ont toutes

deux choisi la voie d'une année de formation dite fondamentale composée d'enseignements dont la maîtrise est une condition nécessaire à la poursuite en master (où la formation est plus appliquée).

## 1.2 Incarnation de cette tension dans les attentes des étudiants

Le calage sur le modèle libéral permet d'intégrer encore plus facilement des étudiants au niveau de la deuxième année de scolarité, à l'entrée du master. Cette voie d'entrée dans les écoles de gestion, sur concours ou sur dossier, est appelée la voie d'admission directe. Lorsque ces étudiants n'ont pas suivi, dans leur cursus antérieur, de cours de statistique, ils ont généralement l'obligation d'en suivre un.

Nous nous intéressons uniquement, dans cet article, à la formation en statistique dispensée en première année d'école. Le profil des étudiants y est alors plus homogène en apparence, puisque dans les écoles les plus renommées, les étudiants sont tous issus des classes préparatoires aux grandes écoles. Après avoir précisé ce profil, nous expliquerons comment ces étudiants vivent la tension entre savoir académique et savoirs pratiques, compte tenu du positionnement ambigu de la discipline entre forte légitimité des enseignements reçus en classes préparatoires et attentes de passage à des enseignements plus directement appliqués au management.

**Études antérieures : plusieurs types de classes préparatoires** — Le terme de classes préparatoires recouvre un type d'organisation de l'enseignement mais correspond à des formations disciplinaires qui sont fort différentes. Ainsi, environ la moitié ou un peu plus de la moitié de nos étudiants avaient choisi la filière EC/S (classes préparatoires économiques et commerciales, voie scientifique), et entre un quart et un tiers la filière EC/E (voie économique, réservée aux titulaires d'un baccalauréat ES — économique et social). Des effectifs plus réduits d'étudiants sont originaires des classes EC/T (voie technologique, réservée à la poursuite d'études après un baccalauréat STG — sciences et technologies de la gestion) ainsi que des formations littéraires A/L (lettres uniquement) et B/L (lettres et sciences sociales) ; ces dernières préparent essentiellement aux concours des écoles normales supérieures de Paris, Lyon et Cachan mais se voient offrir depuis environ quinze ans des débouchés alternatifs dans les écoles de commerce.

Les proportions d'admission par filière d'origine varient d'une année et d'une école à l'autre : contrairement aux écoles d'ingénieur, aucun quota n'est imposé, le classement des étudiants est commun à l'ensemble des filières. Cela oblige à noter de manière très homogène entre les différents sous-concours EC/S, EC/E, EC/T, B/L et A/L. La conséquence est que nous, enseignants, découvrons à la rentrée la coloration de la nouvelle promotion ; une tendance récente est la montée en puissance des filières autres qu'EC/S, et notamment, EC/T et A/L. Cela a un impact net, notamment en termes de bagage mathématique moyen des étudiants.

**Tension au sein même des attentes des étudiants** — Comme nous l'avons brièvement évoqué ci-dessus, la première année dans nos écoles est maintenant pensée comme une année de transition entre les classes préparatoires et le master en management. Comme toute période de transition, elle entraîne des changements identitaires. Ces changements sont marqués chez de jeunes adultes qui traversent une période très chaotique (Cohen-Scali, 2000). Yves Abraham (2007), dans un travail de recherche mené à HEC Paris, a étudié le processus de changement

identitaire et d'acculturation qui est mis en œuvre – changement qu'il qualifie de passage du « souci scolaire » au « sérieux managérial ». Ce processus donne parfois l'impression que le programme Grande Ecole d'HEC Paris n'est qu'une « chambre de mûrissement de bananes ». Il explique qu'en réalité prend place une action pédagogique pensée mais non explicitée, qui vise à assurer la (re)production des élites économiques. Il montre également que ce processus fonctionne plus ou moins selon le profil des étudiants (les « héritiers », les « dévots », les « égarés », les « reconvertis ») et que ces types d'étudiants ont des attentes très différentes vis-à-vis des cours. Les différences en termes de profil se combinent à celles liées à l'origine scolaire et rendent la mise en œuvre d'un cours comme celui de statistique particulièrement ardu.

### 1.3 Une tension inhérente à l'enseignement de la statistique

Dans cet environnement mouvant et face à ce public en construction de soi et aussi hétérogène dans ses attentes, comment penser un cours de statistique ? Ce n'est pas chose aisée et la place de cette discipline se réduit année après année, comme nous l'expliquerons à la fin de ce paragraphe. Le cours de statistique est réputé difficile et, compte-tenu des difficultés intrinsèques de positionnement de la discipline, est particulièrement affecté par les mouvements contradictoires observés dans l'enseignement supérieur. La tension entre professionnalisation et académisation renvoie au vieux débat entre approche à partir d'un modèle (« approche formelle ») et approche à partir des données (« approche data »), deux approches qui co-existent dans la culture française (Armatte, 2010).

**Multiplicité des épistémologies du côté des enseignants** — La multiplicité des épistémologies, le plus souvent non clarifiée, est source de difficulté pour les étudiants qui n'identifient pas d'où parle le professeur (Armatte, 2010). Or les choix pédagogiques sont nécessairement liés à la conception qu'a le professeur de l'identité de la discipline enseignée (Fine, 2010). En France, la statistique reste considérée comme une branche des mathématiques, mais les déclinaisons sectorielles se multiplient (statistique pour la biologie, la gestion, la psychologie, etc.), avec pour conséquence, comme le remarque Jeanne Fine (2012), citant Bourdieu, un fort risque d'hyper-spécialisation et d'affaiblissement de l'identité de la discipline. En pratique, dans les établissements d'enseignement supérieur de gestion, la tonalité et le contenu du cours de statistique dépendent fortement du profil de l'intervenant (mathématicien, statisticien en entreprise, enseignant-chercheur d'une autre discipline comme le marketing, par exemple) ; et ce, même au sein d'une organisation où plusieurs intervenants se répartissent la prise en charge des étudiants. Cette diversité peut également affecter un enseignant donné, dont la vision de l'identité de la discipline se met en mouvement au cours du temps au contact des intervenants de profils différents.

Tout savoir a nécessairement trois dimensions : historique, systématique et opérationnelle (Fabre, 2011). Malheureusement, dans beaucoup d'écoles de commerce, seule la troisième dimension est prise en compte. Les fondements de la statistique sont supposés acquis au cours de la scolarité antérieure ; en effet, la statistique a toujours été, à des degrés divers, au programme du collège et du lycée, et des classes préparatoires. (Dans ces dernières, les étudiants voient, rapidement, la notion de modèle statistique, l'estimation ponctuelle et par intervalle dans des modèles simples, notamment ceux de Bernoulli.) En raison de ces fondements supposés (sans doute à tort) bien acquis, l'enseignement de la statistique est souvent réduit à la présentation par

des non spécialistes de techniques mises en œuvre à l'aide de logiciels spécialisés.

Nous nous rendons compte lors des premières séances de nos cours de statistique que nos étudiants, qui se placent pourtant parmi les meilleurs de leur promotion de classes préparatoires, n'ont pas encore appréhendé correctement les concepts fondamentaux de la statistique inférentielle qui y sont enseignés (à savoir, uniquement la notion de modèle statistique et les fondements de l'estimation, ponctuelle ou par intervalle), ni même ceux de la statistique descriptive, supposés acquis depuis longtemps (représentations graphiques, mesures de centralité et de dispersion, etc.). C'est pourquoi il ne nous semble pas qu'un simple enseignement des techniques statistiques puisse atteindre l'objectif que nous donnons à nos cours : doter ces citoyens et futurs dirigeants d'une culture statistique qui leur permette de mieux comprendre le monde qui les entoure et d'approcher la dimension statistique des problèmes rencontrés dans leur vie quotidienne comme dans leur vie professionnelle.

**Multiplicité des épistémologies du côté des étudiants** — Nous avons dit que les épistémologies des professeurs sont diverses et le plus souvent non clarifiées ; il en va de même de celles des étudiants et cette multiplicité est un frein à l'apprentissage. En effet plusieurs travaux récents montrent que le processus d'apprentissage, et en particulier de l'apprentissage en statistique, est lié à l'identité de l'apprenant (Cobb et Hodge, 2002 ; Hahn, 2000). Plusieurs identités co-existent et peuvent entrer en conflit (Boaler, 2002). Dans le cas des étudiants issus de classes préparatoires, la multiplicité des identités fait que le cours de statistique n'est jamais ce qu'ils attendent : certains lui reprochent d'être « trop mathématique » (sous-entendu : trop formel), d'autres « pas assez mathématique » (sous-entendu : manquant de rigueur et/ou ne présentant pas les preuves des résultats utilisés). Le conflit est net entre la très forte identité d'étudiant de classe préparatoire, vouée à disparaître, et celle idéalisée de futur cadre dirigeant. Par ailleurs la vision très déterministe qu'ils ont du monde conduit souvent nos étudiants à rejeter la discipline pour son côté flou, source d'angoisse, et à lui reprocher son absence de rigueur, au sens où dans certaines situations, la seule chose qu'on puisse prouver, c'est qu'on ne puisse rien dire de manière tranchée. Il est alors rassurant de réduire cette discipline à un ensemble de techniques « presse-bouton », ce qui ne rend finalement pas nécessaire le suivi d'un cours. Pourtant, l'absence de certitude est une caractéristique de la réalité et il faut aider les étudiants à l'intégrer (Meletiou-Mavrotheris et Lee, 2002). Ceci passe par la maîtrise conjointe du raisonnement inductif et déductif (Fine, 2010), par la combinaison de deux perspectives : l'approche centrée sur les données et la modélisation (Peters, 2011).

Petocz et Reid (2010) proposent de considérer trois niveaux dans la formation statistique des futurs professionnels : le niveau *extrinsic technical* (« extrinsèque technique »), qui consiste à rassembler des techniques statistiques utilisables dans différentes situations (on en reste à un stade calculatoire) ; le niveau *extrinsic meaning* (« extrinsèque signifiant »), qui permet de donner du sens à des données (il s'agit de pouvoir traiter de manière active un jeu de données) ; le niveau *intrinsic meaning* (« intrinsèque signifiant »), qui conduit l'apprenant à modifier sa vision du monde et qui correspond à une culture statistique intégrée dans la vie quotidienne professionnelle et personnelle. Si l'objectif de l'éducation statistique est d'atteindre ce troisième niveau, force est de constater que les curricula en restent souvent au premier niveau « extrinsèque technique », alors qu'ils pourraient cependant ouvrir l'esprit des étudiants aux deuxième et troisième niveaux<sup>4</sup>, sans bien sûr avoir le temps de les y former complètement. Nous montrerons que c'est

<sup>4</sup>Ces niveaux peuvent d'ailleurs se retrouver dans la pratique professionnelle d'un enseignant : le premier

un des objectifs à moyen terme de nos cours de statistique, dont l'évaluation porte cependant principalement sur les compétences du premier niveau.

La typologie proposée par Petocz et Reid (2010) est intéressante parce qu'elle adopte le point de vue de l'apprenant en tant que futur professionnel (et non comme étudiant). En revanche, elle s'inscrit dans une vision linéaire du processus d'apprentissage largement contestée aujourd'hui. De nombreux auteurs lui substituent une conception en réseau (Noss et Hoyles, 1996) : ils affirment que l'apprentissage survient au travers d'un processus dialectique entre des conceptualisations dans l'action, liées au contexte dans lequel elles sont apparues, et des théories ou concepts scientifiques. En fonction des cadres théoriques auxquels se réfèrent ces auteurs et le choix qu'ils font de privilégier soit la dimension culturelle, soit la dimension cognitive, l'accent est mis sur la continuité (Noss et Hoyles, 1996) ou sur la discontinuité (Pastré *et al.*, 2006). L'apprenant construit ses connaissances soit par intégration progressive au sein d'une communauté de pratiques, soit selon un processus de ruptures et adaptations.

Concrètement, nous observons chez nos étudiants de première année un intérêt assez divers et parfois très faible pour notre discipline ; cependant, une partie significative d'entre eux revient vers nous à l'occasion d'un stage court (en fin de première année) ou long (entre la deuxième et la troisième année), lors de missions « junior entreprise » (voir le paragraphe 3.3 et l'annexe 2), lors du mémoire de fin d'études, ou même lors des premières années de pratique professionnelle, lorsqu'ils sont confrontés à un problème statistique et doivent créer puis traiter en autonomie un jeu de données. Il serait donc important de penser des environnements pédagogiques dans lesquels l'apprenant puisse confronter les savoirs empiriques construits au cours de ses expériences personnelles avec les savoirs académiques qu'il doit intégrer afin de devenir un professionnel compétent ; ceci permettrait de transmettre aux étudiants la capacité à questionner les problèmes posés et à ne pas aveuglément adopter les solutions toutes faites proposées par d'autres (Fabre, 2011). Ce vœu pourrait être réalisé dans le cadre d'un cours ou d'un séminaire au niveau master du cursus de nos écoles, mais, comme nous l'expliquerons ci-dessous, la création de cet environnement n'est pas à l'ordre du jour. Cette impossibilité de revoir nos étudiants au cours de la formation dans un cadre d'apprentissage statistique dédié explique en partie les choix pédagogiques que nous avons effectués et que nous présenterons dans les pages suivantes.

Aujourd'hui, pour lutter contre le désintérêt des élèves, la tendance est plutôt à la survalorisation de la motivation extrinsèque avec notamment le recours à des environnements ludiques (Fabre, 2011). Eveiller l'intérêt des élèves est une condition nécessaire à l'apprentissage mais ce n'est certainement pas une condition suffisante. Plusieurs dérives ont été largement documentées. Ainsi de nombreux auteurs ont étudié les effets pervers des problèmes pseudo-réels, largement utilisés dans l'enseignement professionnel (Meletiou-Mavrotheris et Lee, 2002 ; Adda, 1976 ; Hahn, 2000).

**Evolution de la place de l'enseignement de la statistique en école de management** — A partir des années soixante, les écoles de management ont accordé une place importante à la statistique afin de donner une légitimité scientifique à leurs enseignements. Au fil du temps, les

---

niveau est le mode dans lequel se place un enseignant pour la rédaction formelle du corrigé d'un sujet d'examen ; le deuxième niveau pourrait correspondre à un enseignant issu d'une formation purement mathématique devant enseigner un cours de statistique appliqué ; le troisième niveau est ce même enseignant après quelques années de telle pratique nourrie par des expériences de consultance ou des travaux de recherche appliqués.

disciplines telles que le marketing ou la finance ont acquis leur propre légitimité et le recours à la statistique a été jugé moins nécessaire (en tout cas en ce qui concerne l'enseignement : le recours à des méthodologies statistiques sophistiquées est encore incontournable pour être publié dans une revue de haut niveau).

La place de la statistique a été fortement revue à la baisse dans la dernière décennie ; c'est un phénomène général qui touche même les écoles offrant un enseignement très quantitatif. Il se traduit concrètement par une baisse du nombre d'heures de cours consacrées spécifiquement aux techniques statistiques, et par le déplacement et l'intégration de leur enseignement à l'intérieur des cours de marketing ou de finance. Il s'accompagne évidemment d'un arrêt des recrutements d'enseignants-chercheurs ayant pour spécialité la statistique.

Cette évolution drastique de la place de la statistique dans le concert des enseignements n'est pas encore arrivée à son terme, certaines écoles envisageant même la disparition pure et simple de cours de statistique dédiés. L'argument présenté est que les méthodes statistiques sont utiles dans de nombreux autres cours et qu'elles peuvent être vues au passage. Pour nous, cette ubiquité utopique de l'enseignement de la statistique risque de se transformer en absence réelle, chaque professeur pensant à tort que la réflexion sur l'usage de la statistique aura été abordée par un de ses collègues. Un des objectifs de cet article est de montrer tout l'intérêt et le bénéfice de cours de statistique dédiés et défendre ainsi leur existence.

## 2 Incarnation et gestion de cette tension dans nos écoles

En charge du cours fondamental de statistique (celui dispensé en première année) dans nos écoles, HEC Paris et ESCP Europe, nous avons été confrontés aux difficultés relatées dans la partie précédente, à savoir, la conjonction des attentes contradictoires des étudiants mises en lumière au paragraphe 1.2 et la difficulté inhérente à l'enseignement de la statistique détaillée au paragraphe 1.3. Nous allons décrire plus précisément comment celles-ci se sont incarnées dans nos cours. Nous allons ensuite expliquer les choix pédagogiques que nous avons été conduits à effectuer, choix qui ont été inspirés par les travaux de recherche que nous avons mentionnés. Ces choix ont parfois été déclinés de manière différente car ils ont été adaptés au public et à l'esprit de deux écoles de traditions différentes.

Avant cela, toutefois, nous précisons le contexte particulier de nos deux écoles, ESCP Europe et HEC Paris.

### 2.1 Eléments de contexte supplémentaires

Nous rappelons ici les connaissances antérieures de nos étudiants en probabilités et statistique, précisons l'esprit commun de nos cours, puis décrivons leur contenu et organisation matérielle respectifs.

**Niveau et habileté en probabilités et statistique par filière d'origine** — Suite aux changements de programme récents dans les classes de lycée, les programmes des classes préparatoires économiques et commerciales sont eux-mêmes en cours de refonte (une version presque définitive circule) ; les lignes de séparation qui suivent n'ont cependant pas été affectées.



C. Hahn et G. Stoltz

Nous regardons les programmes à travers le prisme de deux résultats fondamentaux en théorie des probabilités, la loi (faible) des grands nombres et le théorème limite central. Le premier nécessite en effet également l'étude de suites de variables aléatoires indépendantes et identiquement distribuées, de même que celle de la convergence en probabilité. Le second requiert quant à lui la définition de la loi normale et de la convergence en loi et conduit à la construction, déjà en classes préparatoires, d'intervalles de confiance sur le paramètre d'une loi de Bernoulli. Les deux résultats sont présents dans les programmes des classes EC/S et EC/E, et, dans une moindre mesure, des classes B/L, même si bien sûr, les étudiants issus des séries EC/E et B/L ont en moyenne une habileté et un goût pour les mathématiques moins prononcés que ceux de la série EC/S. Le programme des voies technologiques EC/T ne contient que la loi des grands nombres, essentiellement à titre culturel d'ailleurs. Enfin, les étudiants des classes A/L n'ont pas de cours de mathématiques et ne disposent donc que du bagage mathématique de leur sortie du lycée ; ils ont pu y préparer un baccalauréat S (scientifique) ou L (littéraire). En termes d'intérêt pour les mathématiques, force est de constater qu'en général, une (petite) partie seulement des étudiants, issus d'EC/S, l'éprouvent. La plupart de ceux originaires d'autres filières ne les apprécient généralement que peu et seraient peinés de les voir revenir en force (et à vrai dire, voient la statistique, quoi que l'on fasse, comme ce retour des mathématiques).

**Objectif commun de nos cours de statistique** — Nos cours ont une vocation introductive : il s'agit de former l'esprit des étudiants, qui iront plus loin plus tard et seront appelés à apprendre d'autres méthodes statistiques dans le futur (lors de leur scolarité ou lors de leur premier poste en entreprise). Notre objectif est ainsi de leur apprendre à apprendre la statistique.

**HEC Paris : un cours sur un format particulièrement court** — Le cours de statistique s'y déroule en environ 20 heures et par petits groupes d'une quarantaine d'étudiants (en parallèle, à l'instar de tous les autres cours) ; il comporte également une séance de travaux pratiques sur machine avec le logiciel SPSS, en demi-groupes d'une vingtaine d'étudiants. Il mêle à des degrés divers, un peu de statistique descriptive et de la statistique inférentielle : estimation ponctuelle et par intervalle, tests d'hypothèses (comparaison à une valeur de référence, tests du  $\chi^2$ ), régression linéaire, simple et multiple (dans un cadre gaussien admis : l'étude des résidus ne peut être abordée faute de temps). Il est essentiellement dispensé par des professeurs permanents ou affiliés d'HEC Paris (ayant une activité de recherche plutôt mathématique), et à l'occasion par des intervenants extérieurs, souvent mathématiciens universitaires, plus rarement statisticiens issus du monde de l'entreprise. Tous les enseignants utilisent le même support de cours, un polycopié d'environ 200 pages de cours rédigé et 400 pages d'exercices d'annales, et réalisent leurs séances de cours au tableau, en réservant la projection de documents à celle des sorties de logiciels.

Ce cours de statistique, sauf exception pour une petite partie de la promotion préparant en parallèle une licence à l'université, a lieu au premier semestre, en même temps que les cours de finance (*financial economics* puis *financial markets*) et d'économie (quantitative). Les étudiants de première année sont pour ainsi dire plongés dans un bain quantitatif. Par ailleurs, lors d'une réforme du cursus en 2007, la direction d'HEC Paris a instamment souhaité que le contenu du cours de statistique soit revu, selon les mots du directeur de l'Ecole HEC d'alors, en évitant toute présentation « presse-bouton », c'est-à-dire dans le sens d'un plus grand formalisme (avec en particulier, la démonstration rigoureuse de certains résultats). Ce faisant, l'institution prenait parti pour l'une des épistémologies décrites au paragraphe 1.3. Les difficultés engendrées en

2007 par une mise en œuvre maladroite de sa part de cette commande institutionnelle ont guidé l'un des auteurs sur la voie de compromis et solutions raisonnables décrits ci-dessous.

**ESCP Europe : un cours fondamental court, suivi d'un cours avancé au choix** — A ESCP Europe, le cours de statistique de pré-master est un cours de 45 heures réparties en 15 séances de 3 heures. Le cours est dispensé à des groupes de 40 à 50 étudiants, à l'exception de deux séances de travaux pratiques avec logiciels (SPAD et SPSS) dispensés en groupes de taille plus faible. Les professeurs qui dispensent ces cours ont des formations très diverses (doctorat d'économie, de mathématiques, de statistique, de gestion, diplôme d'ingénieur) et exercent des métiers différents (enseignants-chercheurs à l'université, professeurs d'école de commerce, professionnels en poste dans une entreprise). Le programme du cours est strictement défini : même ouvrage de référence, mêmes jeux de transparents, mêmes exercices (sur une plate-forme *e-learning* associée au livre). Le cours se compose de deux parties : une partie commune de 30 heures dont la moitié est consacrée à la statistique descriptive (uni- et bi-variée, séries chronologiques), et l'autre moitié à la statistique inférentielle (estimation, tests d'hypothèses et modèle de régression linéaire simple) ; puis un module de 15 heures au choix entre celui consacré aux méthodes descriptives d'analyse des données (ACP, ACM, classification) et celui qui porte sur la modélisation (modèle de régression linéaire multiple, modèle de régression logistique). L'objectif est double : proposer des outils immédiatement opérationnels et poser les fondations d'une culture statistique qui pourra être développée par la suite dans les différentes spécialisations du master.

## 2.2 Equilibre (instable) entre présentation formelle et mises en situation

Nous sommes donc face à un dilemme permanent dans le cadre du cours de statistique. La ligne de partage floue entre savoir académique et savoirs pratiques (voir partie 1), l'intérêt ou le désintérêt pour les mathématiques, les hétérogénéités de niveaux (voir paragraphe 2.1) rendent l'enseignement de notre discipline particulièrement ardu. C'est ce que nous allons développer maintenant.

### 2.2.1 Le nœud du problème : où placer le curseur ?

La juste place du niveau de formalisation et de détail mathématiques peut être déterminée à l'aune des trois remarques suivantes. Premièrement, nos étudiants ne seront bien évidemment jamais amenés à développer des théories statistiques. Deuxièmement, ils auront peut-être, en revanche, à mettre en œuvre telle ou telle procédure spécialisée que nous n'aurions pas vue en cours, auquel cas il faudra qu'ils soient en mesure de comprendre et mettre en œuvre les méthodes présentées dans les manuels. (Il s'agit donc pour nous, comme nous l'avons déjà dit, de leur apprendre à apprendre la statistique.) Troisièmement, les utilisateurs de la statistique qu'ils seront doivent surtout savoir comment interpréter les résultats obtenus. Ces trois constats plaident en faveur d'une solution paradoxale : une grande rigueur mathématique dans la présentation de la méthodologie, avec omission des démonstrations complètes (mais, éventuellement, avec la fourniture de quelques éléments de démonstration). Cette solution tranche sans doute avec des équilibres différents choisis dans les enseignements de sciences humaines à l'université (où le degré de formalisation est moindre) et ceux des écoles d'ingénieurs (où effectuer les démonstrations permet de continuer à former l'esprit).

**Que faire lorsque les étudiants réclament des démonstrations ?** — Lorsque l'enseignant ne prend pas les devants quant à l'absence de démonstrations, une (petite) partie des étudiants issus d'EC/S insistera pour les voir effectuées. Il est difficile de déterminer si cette demande se fonde sur l'habitude acquise en classes préparatoires de voir prouvés la plupart des résultats mathématiques et représente une défiance face à l'enseignement jugé plus utilitaire en école, ou si elle relève d'un intérêt véritable pour les mathématiques. La stratégie adoptée dans nos deux écoles consiste à reconnaître la légitimité de cette revendication mais à ne pas y accéder autrement qu'avec des moyens détournés empiétant le moins possible sur la présentation de la méthodologie statistique et ses applications. L'objectif est ainsi de ne pas décourager ni d'effrayer les autres étudiants (la majorité).

### 2.2.2 Solutions mises en œuvre

Nous décrivons maintenant plus précisément comment nous pensons avoir trouvé un juste milieu, une harmonie raisonnable entre ces postures extrêmes qui seraient de tout prouver mathématiquement ou, au contraire, de ne rien prouver.

**ESCP Europe : accompagner le changement de perspective** — Nous savons que les étudiants qui arrivent de classes préparatoires ont des attentes contradictoires qu'il est impossible de satisfaire. A ESCP Europe, l'objectif que nous nous sommes fixé est de proposer un cours à la fois très rigoureux, mais également déstabilisant, afin de les préparer au changement de perspective et de culture. Comme nous l'avons écrit, un certain nombre de ressources communes ont été mises en place. Nous offrons par ailleurs des compléments adaptés aux différents publics (cours complémentaires pour les étudiants littéraires, démonstrations disponibles en ligne pour les étudiants scientifiques qui les souhaiteraient). Le renversement de perspective s'opère par le lien que nous faisons avec les pratiques réelles en entreprise, en présentant des problématiques authentiques, en comparant systématiquement les pratiques scolaires et les pratiques professionnelles, en intégrant l'usage de logiciels professionnels. Il ne s'agit évidemment pas de se limiter à donner des recettes. Au contraire, nous montrons aux étudiants que, pour pouvoir prendre du recul par rapport à sa pratique de manager, il est primordial de savoir comment a été construite la « boîte noire technologique ».

Ce renversement de perspective est particulièrement important sur des thèmes déjà traités dans le curriculum passé. Ainsi les fondements de la statistique descriptive ont été vus à plusieurs reprises au cours de la scolarité antérieure et ce thème est donc supposé maîtrisé. En réalité, nous avons pu constater que le plus souvent les connaissances antérieures ne permettent pas de mettre en pratique de manière efficace les notions fondamentales de statistique descriptive. Nous devons donc réinvestir leur champ en suscitant l'intérêt des étudiants. Comme cela a été observé à d'autres niveaux de formation, les étudiants connaissent des formules par cœur mais ne leur attribuent pas vraiment de sens pratique. De même, il est toujours surprenant de constater que certains étudiants issus de la voie scientifique sont incapables d'interpréter un histogramme tracé par un logiciel de statistique.

C'est pourquoi nous proposons tout d'abord de revisiter des mesures qui leur semblent familières, telles la moyenne et la variance. Plusieurs moyens simples sont utilisés, comme la comparaison avec d'autres outils d'analyse d'une distribution qui leur sont moins familiers (quartiles et boîtes à moustaches). D'autres techniques donnent toujours aux étudiants le sen-

timent d'une vraie valeur ajoutée par rapport à leur scolarité antérieure : le lien avec la représentation physique (la moyenne comme centre de gravité et la variance comme inertie) et son application à la lecture d'un tableau individus-variables grâce au centrage et à la réduction. Ce dernier point, l'analyse d'un tableau centré-réduit, est d'autant plus déterminant que les étudiants d'ESCP Europe sont directement concernés. En effet, à partir des notes obtenues dans l'ensemble des matières, ils se voient chacun attribuer un score centré réduit qui permet de les situer par rapport aux autres étudiants de leur promotion. Ce score est décisif lors de leurs choix (notamment, de campus ou d'échanges à l'étranger) et permet de les classer à leur sortie. La représentation physique et le centrage-réduction présentent également l'avantage d'introduire l'analyse en composantes principales, au programme du cours du second semestre.

En ce qui concerne l'estimation, une notion au programme officiel des classes préparatoires, nous devons permettre aux étudiants d'effectuer un lien entre l'approche probabiliste et la pratique statistique en entreprise. Les étudiants ont été introduits à la théorie mathématique, qui consiste à déterminer des estimateurs d'un paramètre dans le cadre d'un modèle simple donné (mais pas nécessairement raisonnable, par exemple, un modèle de lois normales à variance connue) et ensuite, après calcul de la loi suivie par ces estimateurs, de construire des intervalles de confiance pour des niveaux de confiance donnés (et pouvant varier). Ils ont par conséquent une vision d'établissement de formules.

En entreprise, la pratique courante consiste à appliquer deux formules simples, toujours les mêmes, car les seules quantités qui sont considérées sont des moyennes d'une variable quantitative ou des proportions d'un événement ou d'un caractère dans une population, estimées *via* des intervalles symétriques à partir d'échantillons de grandes tailles (pour lesquels on effectue l'approximation normale proposée par le théorème limite central), et à un niveau de confiance systématique de 95 %. Encore ne s'intéresse-t-on généralement pas à l'intervalle, mais plutôt à sa demi-longueur ou marge d'erreur. C'est pourquoi les professionnels ont recours à deux formules qui leur permettent de calculer directement ces marges d'erreur (par exemple  $1.96\sqrt{p(1-p)/n}$  pour une proportion dans une population, lorsque  $p$  est la proportion observée dans l'échantillon de taille  $n$ ), et éventuellement ensuite, mais rarement, de donner un intervalle de confiance. Cette approche de « boîte » devient de plus en plus « noire » à mesure que l'usage d'outils technologiques grand public se généralise (par exemple la « calculette » de Médiamétrie occulte la formule ci-dessus permettant d'obtenir la marge d'erreur). A ESCP Europe, nous avons fait le choix de partir des pratiques en entreprise, de les « détricoter », afin de les relier aux pratiques scientifiques vues en classes préparatoires. Nous expliquons ensuite aux étudiants que ponctuellement, dans certains domaines, ils seront éventuellement amenés à dépasser et élargir ces pratiques et que leur connaissance de la méthodologie scientifique leur permettra alors d'accéder sereinement à cette nouvelle dimension.

**HEC Paris (solution actuelle) : parenthèses culturelles, musique mathématique** — Le compromis en vigueur est que nous faisons entendre la musique mathématique des démonstrations aux étudiants mais leur rappelons que nous les évaluerons sur leur capacité à appliquer les formules obtenues et à interpréter les résultats obtenus ; il s'agit donc d'une parenthèse culturelle. Par musique mathématique, nous entendons ici les grandes lignes des preuves. Par exemple, pour obtenir la formule d'intervalle de confiance asymptotique sur une moyenne de population lorsque la variance est estimée, nous présentons tout d'abord les limitations de l'application du seul théorème limite central ; nous énonçons alors (sans le démontrer) le lemme de Slutsky ; nous en tirons la convergence en loi d'une certaine suite de variables aléatoires et en déduisons

les intervalles souhaités. A la fin de la parenthèse culturelle, nous rappelons aux étudiants que nous avons effectué ce cheminement essentiellement pour qu'ils gardent bien à l'esprit, d'une part, que les intervalles obtenus sont asymptotiques, et d'autre part, pour qu'ils voient qu'au moins trois types d'intervalles sont naturels (ceux qui respectivement, encadrent, majorent et mineurent le paramètre d'intérêt). La présentation de la théorie sous-jacente tend vers un but concret.

Il peut là encore arriver que certains étudiants se plaignent du caractère incomplet de la démonstration, qui ci-dessus met en jeu le lemme de Slutsky. Ce sont souvent les mêmes qui sont fort déçus que nous n'ayons pas le temps de prouver le théorème limite central qu'ils ont également admis en classes préparatoires. A ceux-ci, la réponse invariable est que nous ne sommes pas en faculté de mathématiques ; c'est une réponse d'autant plus pertinente qu'elle est assortie du rappel qu'ils avaient le choix, en début d'année, de s'inscrire en parallèle de leur première année de scolarité à HEC Paris à une licence universitaire en mathématiques appliquées, à l'Université Paris-Sud ou à l'Université Panthéon–Sorbonne, où ils auraient pu étudier les outils nécessaires à la réalisation de leur désir de démonstrations complètes.

Le compromis exposé ci-dessus permet une conciliation raisonnable des attentes diverses des étudiants, pour peu que les parenthèses culturelles soient courtes (de l'ordre de dix minutes) et que leur début et leur fin soient clairement annoncées.

Il trouve également sa traduction dans la rédaction du polycopié de cours, qui contient, outre une version rédigée *in extenso* du cours magistral : des fiches de synthèse reprenant en une trentaine de pages l'ensemble des cours et des formules à appliquer ; des compléments de cours facultatifs très mathématisés ; et de nombreux exercices, avec notamment plusieurs années d'Annales d'examens et de quiz. Cela permet aux étudiants de définir eux-mêmes, librement, leur degré d'implication dans ce cours.

**HEC Paris et ESCP Europe (solution expérimentée indépendamment un temps) : mise en place de groupes de niveaux** — En revanche, une mauvaise solution aux attentes et capacités différentes des étudiants a été, pendant quelques années, de constituer des groupes dits de niveaux, sur des fondements objectifs comme la filière d'origine. L'échec de cette stratégie de différenciation des cours de statistique est à mettre au crédit d'une forte revendication d'égalité, plus puissante que la potentielle soif de savoirs théoriques ou pratiques. En effet, les évaluations des cours par les étudiants soulignaient alors l'injustice criante qu'il y avait à devoir travailler davantage, pour une note égale, lorsque l'on avait été affecté aux groupes avancés en statistique.

### 2.3 Faire vivre le problème

Dans ce paragraphe, nous allons commenter une phrase synthétique prononcée par Jacques Obadia, professeur honoraire à HEC Paris, juste avant son départ à la retraite : « Le plus difficile pour les étudiants est de comprendre l'énoncé ». Or, les énoncés fournis à nos étudiants se présentent généralement sous la forme rédigée suivante :

Sur 100 fumeurs motivés qu'il a choisis au hasard pour tester son nouveau traitement anti-tabac, un laboratoire pharmaceutique constate un arrêt au moins temporaire de la consommation pour 64 participants. Peut-il, dès lors, garantir formellement une efficacité au moins égale à 60 % ?

On peut donc, dans un premier temps, être perplexe face à la difficulté soulevée, tant l'histoire décrite et la question posée semblent naturelles : elles auraient presque pu être prononcées dans une conversation mondaine. En réalité, il s'agit en grande partie de comprendre que faire de l'énoncé, de cerner ce qui est attendu. Ainsi, les difficultés sont-elles de deux ordres : les premières, connues et étudiées par d'autres chercheurs, relèvent de la formalisation statistique de la question posée (identification de la population et de l'échantillon, reformulation de la question comme la détermination du caractère significatif ou non de la déviation par rapport à la valeur de référence). La dernière difficulté, essentielle à résoudre pour nos étudiants en vue de leur vie professionnelle future, concerne l'interprétation et l'exploitation du traitement statistique.

Nous allons d'abord décrire en détails chacune de ces trois difficultés, puis nous présentons les solutions retenues pour les résoudre.

### 2.3.1 L'inversion du point de vue : du problème à la modélisation

Nous détaillons et incarnons ici l'opposition et les complémentarités entre raisonnements inductif et déductif, et approches centrée sur les données et par modèle décrites au paragraphe 1.3.

**Une inversion de point de vue particulièrement délicate après les enseignements de classes préparatoires** — Nos étudiants ont ceci de particulier qu'ils se situent au milieu d'un gué. En effet, ils ont de bonnes connaissances des résultats fondamentaux du calcul des probabilités et doivent effectuer le changement de point de vue typique de ceux qui ont suivi un cours de probabilités avant d'étudier la statistique inférentielle. Concrètement, ils ont l'habitude d'énoncés d'exercices commençant ainsi :

Soient  $X_1, \dots, X_n$  des variables aléatoires indépendantes et identiquement distribuées selon une loi de Bernoulli de paramètre  $p_0 \in [0, 1]$ . [...]

La loi est fixée, le paramètre  $p_0$  est connu, et l'inconnue est le comportement d'une certaine fonction des  $X_1, \dots, X_n$ .

Or, dans notre exercice initial, on part des réalisations de  $X_1, \dots, X_{100}$ , c'est-à-dire de données, et l'inconnue est  $p_0$ . Cette mesure d'efficacité  $p_0$  qu'il s'agit de comparer à 60% dans l'exercice ci-dessus est la proportion d'arrêt que l'on observerait sur les clients, des fumeurs motivés pour s'arrêter de fumer, si on commercialisait le traitement. Il s'agit donc d'une proportion qui sera, éventuellement, constatée dans une population, la difficulté supplémentaire de cet exercice étant qu'il s'agit d'une population future et fictive.

Ce renversement de point de vue est déjà délicat pour des étudiants de mathématiques, mais ces derniers n'ont, dans leurs examens, que très peu affaire à des données, ce qui n'est pas du tout le cas de nos étudiants, destinés à (savoir) traiter des données. L'autre rive du gué est formée par les étudiants en sciences de gestion de l'université, pour qui les cours de probabilités et statistique sont menés en parallèle au long de leur formation, et non de manière séquentielle comme c'est le cas pour nos étudiants.

Notre difficulté est donc d'enseigner à nos étudiants la démarche inverse de celle qu'ils ont suivie jusque-là : non plus la démarche déductive, qui est de partir d'une loi, considérer des variables aléatoires indépendantes et identiquement distribuées selon cette loi, et décrire leur

comportement ; mais au contraire, la démarche inductive, qui consiste à partir d'un comportement, de données, pour construire un modèle, et en estimer les paramètres. Nous ne disons pas que les écoles de commerce sont le seul lieu de l'enseignement supérieur où cette double démarche est enseignée (c'est le cas, également, dans une moindre mesure, dans les cursus de mathématiques), mais nous voudrions souligner qu'elle l'est en revanche, et cela peut-être est unique, de manière fortement séquentielle, avec une rupture temporelle claire. La démarche déductive est en effet étudiée en classes préparatoires, où les savoirs théoriques ont une légitimité forte car ils sont le fondement des concours d'entrée aux écoles ; la démarche inductive n'est considérée qu'en école, où ces savoirs théoriques, comme nous l'avons expliqué en partie 1, n'ont déjà plus qu'une valeur relative. Cette démarche inductive peut alors se trouver récusee pour son manque de rigueur et de précision.

Une autre difficulté liée à l'inversion de la démarche est que, par habitude, les étudiants sont tentés de penser que tout est encore précisé dans l'énoncé, en particulier, le paramètre d'intérêt  $p_0$ , fréquence d'arrêt future dans la population, et en déduire faussement que ce paramètre est égal à la fréquence observée dans l'échantillon, ici 64 %. La distinction entre population et échantillon, dont on sait qu'elle est en général peu aisée, est compliquée encore une fois par les habitudes déductives.

**Une occurrence de l'inversion de point de vue : l'adverbe « significativement »** — L'étape initiale de la compréhension de l'énoncé consiste à reformuler la question posée en termes des données et de la valeur de référence : « La fréquence d'arrêt de 64 % observée dans l'échantillon est-elle significativement<sup>5</sup> supérieure à la valeur de référence 60 % ? » L'adverbe « significativement » fait évidemment référence au fait que la proportion observée dans l'échantillon, connue, diffère en général de celle, inconnue, qui correspond à la population, mais que la question posée par l'exercice porte sur cette dernière quantité, révélée partiellement seulement par la proportion dans l'échantillon.

Il est parfois possible de vérifier la compréhension de la démarche inductive : on peut constater la présence ou l'absence de l'adjectif « significatif » ou de l'adverbe « significativement » dans la réponse proposée par un étudiant. Ceux ne saisissant pas cette démarche se borneront à comparer, de manière laconique, la proportion  $\bar{x}_n$  observée dans l'échantillon (ici,  $\bar{x}_{100} = 64\%$ ) à la valeur de référence  $p_{\text{ref}}$  (ici,  $p_{\text{ref}} = 60\%$ ), en omettant les problèmes d'inférence, et se demanderont où donc veut en venir l'énoncé. Ceux à l'aise avec cette démarche verront que trois cas étaient possibles, que  $\bar{x}_n$  soit significativement supérieure ou inférieure à  $p_{\text{ref}}$ , ou qu'au contraire, elle n'en diffère pas significativement.

**Solution mise en œuvre à HEC Paris : un canevas de prise en main** — Une séance entière de cours est désormais consacrée à la démarche de modélisation et de formalisation statistique. Un canevas de prise en main d'un énoncé est proposé, décomposé en six temps ; un exemple étant souvent plus parlant qu'une longue description théorique, nous écrivons ce que ce canevas donne pour l'exercice ci-dessus de traitement du tabagisme.

1. Population visée : tous les fumeurs motivés pour arrêter de fumer, ressentant le besoin

<sup>5</sup>Nous expliquons bien sûr en cours que cet adverbe « significativement » se réfère implicitement à un niveau de confiance standard de 95 %, mais que ce niveau peut et doit être déterminé par le commanditaire de l'étude. Ce qui nous intéresse ici, c'est de formuler une phrase qui soit accessible au plus grand nombre, à charge pour nos étudiants d'en connaître et d'en méditer le sens caché.

- d'une aide paramédicale et qui achèteront à cet effet ce traitement s'il est commercialisé (on a recruté au hasard un échantillon de 100 volontaires pour représenter cette population<sup>6</sup>);
2. Table de correspondance : on note dans la suite 1 lorsqu'un arrêt au moins temporaire a été constaté, 0 sinon ;
  3. Données disponibles :  $x_1, x_2, \dots, x_{100} \in \{0, 1\}$ , où  $x_j$  désigne le résultat du traitement du  $j$ -ème membre de l'échantillon ;
  4. Statistique d'échantillon : la fréquence d'arrêt au moins temporaire observée dans l'échantillon est de  $\bar{x}_{100} = 64\%$  ;
  5. Modélisation : on peut considérer que les données  $x_j$  sont la réalisation de variables aléatoires  $X_1, X_2, \dots, X_{100}$  indépendantes et identiquement distribuées selon une loi de Bernoulli de paramètre  $p_0 \in [0, 1]$  ;
  6. Interprétation : le paramètre d'intérêt  $p_0$  est la proportion d'arrêts au moins temporaires que l'on observerait, si on commercialisait le traitement, sur les clients l'achetant.

Le formalisme retenu jette un pont entre celui du calcul des probabilités vu en classes préparatoires et le nouveau point de vue de traitement des données. Nous cherchons à montrer la profonde continuité des outils mathématiques employés malgré l'inversion de point de vue. Le point le plus crucial de ce canevas, et celui qui révèle la bonne compréhension de l'énoncé, est l'interprétation du paramètre ; dans cet exemple, elle est nécessairement au conditionnel, parce que l'on n'est pas sûr de lancer le produit. Une partie significative des étudiants proposant une interprétation laconique ou énoncée au présent («  $p_0$  est la proportion d'arrêt au moins temporaire du tabagisme »), ou pire, au passé, se révèlent confondre échantillon et population, lorsque nous les interrogeons plus avant.

**ESCP Europe : de la résolution de problèmes à la problématisation** — A ESCP Europe, nous identifions trois niveaux de compétences statistiques, que nous ne hiérarchisons pas, même si nous développons plus longuement le premier d'entre eux.

*Niveau 1 : résolution de problèmes.* Nous développons la forme et la méthodologie qui nous semblent susceptibles de soutenir la construction des concepts. Ainsi, l'insistance sur l'analyse de l'énoncé d'un problème, avec une formalisation de cette analyse par une représentation imagée, vise à opérationnaliser les concepts requis tout en identifiant les éléments-clés. Par exemple, dans un test d'hypothèses, il s'agit de préciser la population étudiée, la variable sur laquelle porte l'étude, le type de cette variable, le mode de sélection de l'échantillon, les informations dont on dispose sur cet échantillon. De même, le choix des notations vise à insister sur ce qui nous semble primordial : paramètre sur la population (lettre grecque), valeur obtenue sur l'échantillon (lettre romaine), estimateur (lettre capitale). En effet, derrière les notations, se trouvent des concepts souvent difficiles à construire pour les étudiants. Il n'est pas évident

<sup>6</sup> On verra à la page suivante une distinction entre une population mère, dans laquelle on tire aléatoirement l'échantillon, et la population visée ou population cible, cible étant entendu au sens du marketing. On aimerait que les deux soient confondues ; dans la plupart de nos exercices, nous supposons implicitement que c'est le cas, en imaginant qu'aucun biais d'échantillonnage n'a eu lieu, qu'on a bien eu accès à l'ensemble de la population visée, etc. C'est évidemment très difficile en général pour les études de terrain que de garantir cela.



C. Hahn et G. Stoltz

qu'une mesure simple à calculer et à obtenir, la moyenne, puisse prendre différents statuts en inférence : une valeur fixe mais inconnue, une valeur connue mais qui peut varier selon l'échantillon, une variable aléatoire.

*Niveau 2 : identification de problèmes.* Il faut naturellement maîtriser la résolution de problèmes, mais le plus souvent nos étudiants auront à analyser des résultats obtenus avec des logiciels professionnels ou à comprendre des résultats d'études. Dans ce cas, ils devront prendre une décision ou juger de la pertinence de l'analyse à partir de quelques traces statistiques. Qu'y a-t-il derrière une  $P$ -valeur ? Quelle a été la démarche mise en œuvre ? Est-elle pertinente ? Justifiée ? L'analyse des informations sur l'échantillon est-elle suffisante ? A-t-on mis en œuvre les méthodes adaptées ? L'inférence a-t-elle été correctement menée ? N'a-t-on pas confondu population mère (à partir de laquelle on a construit l'échantillon) et population cible (celle sur laquelle on veut conclure en réalité) ? Etc. L'objectif principal est ici de les faire ouvrir la boîte noire et comprendre que sans réelle maîtrise des méthodes d'analyse, le risque est grand d'aboutir à des erreurs.

*Niveau 3 : formulation de problèmes.* Dans la vraie vie, et notamment dans la vie du cadre dirigeant, les problèmes sont rarement donnés. Face à une situation posant question, il devra mobiliser des ressources pour résoudre le problème qu'il a identifié et construit. Deux individus ne vont pas construire le même problème puisqu'ils ne feront pas appel aux mêmes ressources. L'objectif du cours est aussi de les aider à percevoir la dimension statistique dans une situation problématique rencontrée. Nous explorons ce niveau lors de deux séances de travaux pratiques pendant lesquelles les étudiants prennent une décision à partir de jeux de données, mais aussi, et surtout, lors d'un projet personnel qu'ils réalisent en binôme.

Comme nous l'avons dit, il n'est pas simple de mettre les étudiants en situation « réelle ». Il s'agit de démarches très coûteuses en temps. Les problèmes authentiques comportent du « bruit » que le professeur de statistique n'a ni le temps ni souvent la compétence pour traiter. Une fois en situation professionnelle, les étudiants découvrent avec surprise que la difficulté principale n'est pas tant d'appliquer les méthodes vues en cours que de trouver l'information, une information susceptible d'être traitée statistiquement afin de répondre à la question qu'ils se posent. Une des activités que nous utilisons s'inspire d'un dispositif de recherche (Hahn, 2013). Nous proposons à des groupes d'étudiants de prendre une décision managériale (le choix d'une zone commerciale parmi trois possibles) à partir d'un extrait d'un fichier clients. Le contexte de l'activité et la problématique sont l'objet d'une courte présentation vidéographique. Le fichier est constitué de plusieurs variables dont les distributions ont été construites de manière à susciter des questions statistiques et à conduire les étudiants à mettre en œuvre des méthodes vues en cours. Il n'y a pas bien sûr de zone uniformément meilleure que les autres, et le choix final dépend du point de vue adopté.

### **2.3.2 Vivre le problème jusqu'au bout : l'exploitation stratégique du résultat statistique**

Nous allons décrire et commenter dans ce paragraphe une autre difficulté propre à l'enseignement de la statistique en école de commerce, qui se pose avec moins d'acuité dans des enseignements de statistique en filières de sciences de gestion ne formant pas des cadres dirigeants (et ne se pose pas du tout dans les cursus de mathématique). La résolution d'un exercice ne peut pas

seulement consister, pour nos étudiants, en l'établissement d'une conclusion statistique (souvent prudente et nuancée). Elle doit aussi et surtout mener à la formulation d'une conclusion stratégique (claire et tranchée), à savoir, l'action que va mener celui qui est en position de décider une fois qu'il a été éclairé par le traitement statistique mené. Ceci est naturel au niveau 3 de compétence statistique décrit au paragraphe précédent (formulation de problèmes), mais nous voulons également que les étudiants suivent cette démarche aux niveaux 1 et 2 (résolution et identification de problèmes, respectivement).

Nous insistons, pour les motiver à y réfléchir, sur le fait que c'est leur valeur ajoutée par rapport à ce que pourrait faire un logiciel de statistique... et que c'est cette valeur ajoutée interprétative et stratégique qui distingue un technicien de la statistique muni d'une licence professionnelle en statistique d'un cadre dirigeant formé au niveau master. En particulier, notre objectif est de les faire passer d'une rationalité technique à une implication vécue comme une préparation à leurs futures responsabilités en entreprise. Plus précisément, nous voulons convertir leur appréhension de la matière, obtenir qu'ils dépassent la simple maîtrise des outils et leur mise en œuvre dans un contexte et pour un objectif purement académiques (la note finale du semestre) pour se tourner vers une mise en situation. Nous insistons alors également sur le fait que l'interprétation du résultat statistique nécessite du recul sur l'obtention de ce dernier, recul permis entre autres par le cadre rigoureux dans lequel la méthodologie statistique a été introduite.

Concrètement, si l'on reprend l'exemple ci-dessus de traitement anti-tabagisme, et que l'on teste les hypothèses  $H_0 : p_0 \leq 60\%$  contre  $H_1 : p_0 > 60\%$  en recourant à une statistique de test fondée sur le théorème limite central, on obtient une  $P$ -valeur de 20.6%. On en déduit (c'est la conclusion statistique) que les données disponibles sont compatibles avec l'hypothèse  $H_0$ , qu'on ne peut donc pas rejeter. Au lieu d'une conclusion stratégique, les étudiants formulent alors trop souvent un constat statistique, c'est-à-dire une pure redite de la conclusion statistique dans le contexte de l'exercice : ici, que le laboratoire ne peut pas garantir formellement la mesure d'efficacité de 60%. Les conclusions stratégiques elles-mêmes, lorsqu'elles sont énoncées, peuvent être plus ou moins intéressantes. Celle la plus fréquemment rencontrée (lors de l'examen court de 30 minutes à HEC Paris duquel est tiré notre exemple) était de la forme suivante : « Il faut que le service R&D améliore le produit pendant que le service marketing prend des vacances. » Cela révèle d'ailleurs, en creux, leur idée de la division du travail et des responsabilités au sein d'une entreprise, pas encore vue comme un lieu de mise en commun des talents orientée vers la réalisation d'objectifs communs. A cet égard, la conclusion stratégique que nous leur avons proposée (et à laquelle personne n'avait pensé) était que le service marketing devrait plutôt revoir son slogan publicitaire et communiquer sur une efficacité garantie de 50% ou 55%, selon qu'il pense que le grand public sera plus réceptif à des chiffres ronds ou à une fréquence strictement supérieure à 1/2.

De la même manière, lorsqu'il s'agit d'intervalles de confiance, nous avons constaté que les étudiants sont souvent embarrassés par le choix d'un des trois types d'intervalles les plus couramment utilisés, ceux qui sont symétriques et précisent donc une marge d'erreur associée à une estimation ponctuelle, ceux qui majorent le paramètre d'intérêt, ceux qui le minorent. Se placer dans la perspective du commanditaire de l'étude statistique, comprendre sa démarche et ses objectifs permet cependant de déterminer la forme la plus appropriée ; ce recours à un élément de choix subjectif car stratégique trouble les étudiants, habitués à la vision plus manichéenne des mathématiques. On peut évoquer des hésitations et perplexités similaires face au choix des

hypothèses  $H_0$  et  $H_1$  à tester dans un problème concret. Là encore, une conversion du point de vue, de l'étudiant vers le décideur, est nécessaire, et c'est à nous de l'accompagner.

## 2.4 Un révélateur : l'évaluation des compétences en cohérence avec notre discours

Nous expliquons en conclusion de cette partie comment nous mettons en cohérence l'évaluation des compétences des étudiants avec le discours développé ci-dessus sur le cours de statistique, montrant en particulier sa différence de nature profonde avec le cours de mathématiques de classes préparatoires et devant faire vivre des problèmes d'un bout à l'autre (de la modélisation des données à la formulation de conclusions stratégiques).

**HEC Paris : importance de la rédaction, choix des mises en situation** — Selon les habitudes de la maison, plusieurs contrôles courts d'une trentaine de minutes sont régulièrement organisés au sein de notre série de séances de cours, ce qui oblige les étudiants à travailler régulièrement (et permet une meilleure assimilation du cours). Ces contrôles sont toujours annoncés à l'avance, avec mention précise du programme de révisions. Contrairement à de nombreux autres cours, ils ne consistent cependant pas en des questionnaires à choix multiples. Il s'agit plutôt d'exercices dont il faut rédiger la solution, comme par exemple celui reproduit à l'annexe 2. Ceci est évidemment motivé par l'importance accordée à la rédaction de conclusions stratégiques et à la mise en situation de cadre dirigeant, qui s'allient mal avec la simple exécution de tâches calculatoires d'application de formules. Certains étudiants ont tendance à rédiger de manière lapidaire et à se concentrer sur les valeurs numériques (réalisations des intervalles de confiance,  $P$ -valeurs, etc.), en pensant à tort que si celles-ci sont exactes, ils auront la note maximale. Nous corrigeons ce trait de contrôle en contrôle, ce qui leur permet de mieux réaliser les attendus du cours de statistique : savoir rédiger un rapport d'étude professionnel, à la fois précis et accessible à des non-initiés.

Enfin, nous rebondissons sur les moyens que se donnent les étudiants pour préparer les contrôles, à savoir, s'intéresser aux annales (fournies dans le photocopié de cours), pour leur y soumettre les situations auxquelles nous voulons qu'ils réfléchissent. Plus précisément, nous portons un soin extrême à la rédaction de ces sujets de contrôles (qui deviennent les annales des années suivantes) et tâchons soit qu'ils aient pour contexte une situation vécue en entreprise par un consultant ou un ancien élève, soit qu'ils soient en rapport avec la vie citoyenne. Nous profitons de l'attention maximale à laquelle nous avons droit pendant ces éléments notés pour non pas seulement tester des compétences techniques, mais placer le cours de statistique dans l'image plus large dans laquelle il prend tout son sens. Ainsi, le sujet de contrôle correspondant à la figure 2 en annexe 2 a été jugé, dans les évaluations du cours par les étudiants, comme présentant une situation concrète et utile, qu'il y a tout lieu d'étudier. Un sujet d'examen récent revenait quant à lui, dans la partie de régression linéaire et dans l'esprit des interventions médiatiques de Nate Silver (voir par exemple l'article de Louise Couvelaire, 2013), sur la prévision du vainqueur aux élections présidentielles américaines de 2012. Il s'agissait, dans le cadre du modèle linéaire gaussien, de prévoir pendant l'été le score du parti au pouvoir lors des élections de novembre, de 1944 à 2012, à partir de données réelles consistant en des variables macro-économiques (taux de chômage, taux de croissance, niveau d'endettement fédéral, etc.) et politiques (cote de popularité en juin, résultats des élections à mi-mandat, existence de scan-

dales politiques pendant le mandat). Le jeu de données utilisé a été construit par Sinha *et al.* (2012), qui en proposent une analyse statistique méthodologiquement critiquable (ce qu'essayait d'ailleurs d'illustrer le sujet d'examen, extraits de cet article à l'appui).

**ESCP Europe : une évaluation à plusieurs niveaux** — Compte tenu de ce qui a été décrit ci-dessus, nous avons conçu un système d'évaluation à plusieurs niveaux, correspondant aux différents niveaux de lecture et d'appréhension du cours, une note minimale de 8/20 étant requise à chaque niveau et une note globale de 10/20 étant nécessaire pour valider le cours :

- [Niveau 1] un premier contrôle, long (2 heures), « traditionnel », composé de problèmes à résoudre, qui représente 55% de la note finale ;
- [Niveau 2] un second contrôle, court (1 heure), composé de questions de compréhension de résultats statistiques, pour 15% de la note finale ;
- [Niveau 3] un projet réalisé à deux (mais avec une soutenance orale individuelle), pour lequel les étudiants doivent, accompagnés par un professeur, définir une problématique, trouver des données et les analyser avec un logiciel de statistique (30% de la note finale).

Ce système d'évaluation en cohérence avec la philosophie du cours vise à les obliger à modifier leurs habitudes de travail. Trop souvent, leur conception des exigences scolaires les conduit à se limiter à un entraînement intensif avant l'examen final à partir d'annales (en général, quelques jours auparavant).

### 3 La question de la légitimité de la matière aux yeux des étudiants

L'intégration d'une école de commerce devrait marquer, aux yeux de la plupart des étudiants, le début d'un nouveau cycle, fondé sur des enseignements pratiques en rupture avec les enseignements théoriques des classes préparatoires dont le but le plus apparent, du moins, le plus immédiat, a été de permettre la sélection à l'entrée de l'école. C'est ainsi que des matières non abordées dans le cursus antérieur, telles que la comptabilité, la finance d'entreprise ou le marketing, sont nimbées d'une légitimité sans faille : elles incarnent par leur nouveauté l'aspiration nouvelle, non plus celle de savoir académique, mais celle de savoirs pratiques, en lien avec le monde de l'entreprise. D'autres matières, comme l'économie et le droit, ont un statut particulier : elles ont pu être étudiées à des degrés divers selon les filières dans les classes préparatoires économiques et commerciales mais bénéficient malgré tout d'une bonne légitimité. Sans doute sont-elles considérées comme une culture générale utile à tout honnête cadre d'entreprise ou entrepreneur évoluant dans un monde incertain. Tous ces cours, comme nous l'avait fait remarquer un collègue d'HEC Paris partant à la retraite, « racontent des histoires » : tout en développant un contenu technique, ils proposent au passage des éclairages, anecdotes et autres récits de nature à être placés dans une conversation mondaine.

Le cas du cours de statistique est différent. D'une part, en apparence, ou tout du moins dans l'imaginaire des étudiants, il n'est pas en position de rupture nette par rapport aux cours de mathématiques des classes préparatoires. Ces cours de mathématiques leur semblaient d'ailleurs

*C. Hahn et G. Stoltz*

essentiellement destinés à être une matière de sélection, non pas une matière ayant un intérêt intrinsèque ou immédiat pour un futur cadre dirigeant, ce qui est naturel au vu des sujets qui y sont traités (algèbre linéaire, analyse réelle, fondements du calcul des probabilités, une touche de statistique inférentielle mais sans contact avec les données). D'autre part, la statistique est considérée comme une matière outil, tant pour son utilisation en entreprise qu'en vue des autres cours : l'impression est que le traitement statistique peut être délégué, qu'il est accessoire, et qu'il suffit d'en parler au détour d'un autre cours, lors de la formation en école. (Cette conception de la statistique se retrouve d'ailleurs dans d'autres filières et face à d'autres publics.) C'est enfin le dernier cours où l'on s'attend à éprouver le plaisir de la découverte et à pouvoir faire partager ses nouveaux savoirs, académiques ou pratiques, lors de conversations mondaines : la discipline semble cantonnée dans un espace restreint.

Nous allons étudier plus précisément dans ce qui suit les facteurs sapant, dans l'esprit des étudiants, la légitimité d'un enseignement en statistique, et nous expliquerons surtout comment nous nous attachons à les contrer. Ces facteurs et solutions peuvent être en lien avec le futur proche (les autres cours dispensés) ou un peu plus lointain (la vie professionnelle), avec le passé (l'enseignement suivi en classes préparatoires), ou avec le présent (la vie en école). Nous les détaillons dans cet ordre.

### **3.1 Futur (proche) : expliquer l'intérêt pour un manager de posséder une culture statistique minimale**

Nous commençons par le futur le plus lointain, mais qui est celui où se projettent le plus les étudiants : la future vie professionnelle. Nous passerons ensuite au futur plus proche constitué par les autres cours dispensés en école.

**Comment permettre aux étudiants de dépasser leurs représentations fausses des métiers du management** — Le polycopié de cours distribué à HEC Paris commence par quelques articles de la presse écrite illustrant la pertinence du traitement statistique en entreprise. Deux articles emblématiques de cette revue de presse sont « Le métier de statisticien sera le plus “sexy” de la décennie » (Robin, 2010) et « Les stats ont la cote » (Maes, 2011) ; ils ont trait respectivement, au traitement de données massives, comme celles issues du monde numérique, et à la sélection des joueurs de football professionnels.

Cependant, sans nécessairement affirmer que la statistique est inutile en entreprise, nos étudiants ont souvent une image idéalisée de leurs futurs postes. En réalité, le discours ambiant et leurs propres ambitions les poussent davantage à envisager le genre de poste auquel ils n'accéderont qu'après avoir fait leurs gammes, c'est-à-dire, après de nombreuses années d'expérience professionnelle. Ils pensent donc, à tort, qu'ils se trouveront immédiatement dans une position de cadre dirigeant pouvant s'appuyer sur une petite équipe à laquelle il confiera, en particulier, les tâches annexes, calculatoires, et sans intérêt : notamment, l'ensemble des traitements statistiques. Cette vision est doublement fautive, par le côté non annexe et potentiellement essentiel du traitement statistique, mais également par l'absence en général, lors du premier poste, de subordonné à qui déléguer la tâche ! A vrai dire, cette vision est tellement fautive et loin de la réalité qu'il est même difficile de savoir comment commencer à la réfuter. Une autre difficulté est que, comme décrit en partie 1, une part significative des intervenants de nos cours a une formation et une expérience exclusivement académiques et peut ressentir une absence de légitimité

à effectuer cette réfutation.

C'est pourquoi, pour à la fois partir de l'idéalisation qu'ont en tête les étudiants et donner un argument simple à exploiter à tous les intervenants, une première remarque de bon sens est que lorsque l'on délègue une tâche, il faut malgré tout avoir une idée du temps à lui impartir et savoir évaluer son accomplissement ; à cet effet, il est en général bon de savoir effectuer soi-même la tâche. Apprendre les rudiments du traitement statistique est alors vu plus positivement comme le développement d'une faculté nécessaire au manager. Cet angle d'approche peut sembler extrêmement complaisant et flatteur pour l'*ego* des étudiants ; c'est évidemment sans compter que dans un deuxième temps, nous essayons de détailler l'intérêt et l'importance intrinsèques de la statistique en entreprise. Ce deuxième temps est évidemment plus délicat pour tous les intervenants n'ayant pas une expérience concrète du monde de l'entreprise (par exemple, pour les universitaires n'ayant même jamais effectué de mission de consultance).

Là encore, des éléments de langage suffisamment généraux et découlant du bon sens peuvent cependant être fournis ; l'objectif est que l'intervenant puisse se les approprier pour les reprendre avec autorité, contrairement à des exemples plus précis et plus factuels que seuls ceux qui les ont vécus pourraient utiliser. Un premier élément est qu'il faut toujours présenter les chiffres avec du recul, c'est-à-dire avec une marge d'erreur, et en particulier, ne pas fournir d'estimations trop précises. A cet égard, on gardera à l'esprit le bon mot de l'économiste français Georges Elgozy : « Dans toute statistique, l'inexactitude du nombre est compensée par la précision des décimales. » Un bon mot que l'on peut conseiller aux étudiants, selon leur tempérament, de prononcer dans un souffle lorsque l'un de leurs futurs collègues commettra l'erreur de fournir de telles estimations trop précises... Un second élément est de rappeler le côté parfois subjectif voire politique du traitement des données : par exemple, dans le choix du couple d'hypothèses  $H_0, H_1$  dans un test d'hypothèses, qui est rarement le même pour les associations de consommateurs et les industriels. Du recul et un peu de pratique pendant les études sont là encore nécessaires, même et surtout au cadre dirigeant qui doit prendre la décision en lisant simplement la méthodologie et le résultat du traitement statistique. Il doit en particulier comprendre la différence entre conserver une hypothèse  $H_0$  (ce que peut faire un test) et valider cette hypothèse  $H_0$  (ce qu'en général un test ne peut pas faire).

La question est de savoir s'il faut un troisième temps d'argumentation, davantage sur le fond (et requérant donc un peu d'expérience pratique de la part de l'intervenant). Notre expérience est mitigée. Il semble difficile de persuader les étudiants par avance, par un simple discours, aussi rationnel soit-il, de l'intérêt et de l'utilité de la statistique en elle-même dans le monde de l'entreprise. En réalité, c'est ce dernier lui-même qui convainc une partie significative de nos étudiants, lors de leurs stages de première année ou de césure. Cette conversion s'effectue parfois dans la douleur lorsque certains étudiants constatent, avec désarroi, l'utilité qu'aurait pu avoir de suivre plus attentivement le cours de statistique ; ceux-ci reviennent alors vers nous, par courriel, pour nous poser des questions sur des sujets que nous avons pourtant traités en détails en cours (et nous les renvoyons bien évidemment à leurs notes de cours).

En conclusion, il semble nécessaire de détailler et justifier l'intérêt d'étudier notre discipline, qui souffre d'un déficit de notoriété préalable. Cela passe par un discours bien rôdé sur le cours, ce qui contraste fortement avec les cours de mathématiques et de calcul des probabilités suivis en classes préparatoires, où une notion était intéressante parce qu'il en était ainsi (parce qu'elle était au programme).

**La statistique, discipline outil, discipline pont ?** — Nous avons expliqué ci-dessus comment rectifier la vision de la statistique comme une matière simplement outil pour l'entreprise ; en fait, le problème se pose également à l'intérieur même du cursus de formation. Des disciplines comme la finance ou le marketing sont friandes de traitements statistiques. D'ailleurs, comme nous l'avons expliqué en partie 1, le contenu du cours de statistique a été en grande partie déterminé en accord avec les professeurs de ces disciplines et en fonction de leurs besoins. Cependant, une tentation de plus en plus grande d'externaliser l'enseignement de la statistique pour favoriser des disciplines jugées plus au cœur des orientations de nos écoles (stratégie, gestion des ressources humaines, marketing, finance) instille l'idée que la statistique est là encore un outil, au service des autres disciplines, et n'a pas besoin d'être étudiée pour elle-même. En un sens, les étudiants baignent, inconsciemment, dans une ambiance, un bain culturel, un discours intégré qui placent certaines matières au-dessus d'autres. Nonobstant ceci, nous nous efforçons de trouver une position harmonieuse pour notre discipline dans le concert des matières étudiées ; plutôt que de l'asservir aux autres matières, nous tâchons qu'elle soit un pont entre elles. Ainsi, certains jeux de données présentent des problématiques en lien avec les autres cours.

A HEC Paris, où le cours est plus resserré, ces liens se font surtout avec l'économie. On peut ainsi mentionner un modèle de régression linéaire du prix du forfait de ski en fonction de diverses caractéristiques des stations, dont leur taille (en nombre de lits). Cette variable est statistiquement significative et son caractère économiquement significatif procède quant à lui de la loi de l'offre et de la demande ; à l'équilibre économique, la taille de la station révèle son attractivité, de sorte que le forfait a tendance à être d'autant plus cher que la station est grande (attractive). Un autre jeu de données, issu d'un article de recherche (Michalski et Stoltz, 2013), permet de tester l'authenticité des déclarations comptables des Etats dans leur balance des paiements, *via* un test d'ajustement du  $\chi^2$  à la loi de Benford.

A ESCP Europe le lien se fait avec l'ensemble des différentes disciplines de la gestion. Nous introduisons chaque séquence de cours avec un exemple authentique d'utilisation de la statistique dans les sciences de gestion, par exemple en marketing pour le cours de statistique descriptive univariée (une étude pour une association professionnelle qui cherche à comprendre les raisons de la baisse de fréquentation aux réunions qu'elle organise) ; en droit pour la statistique descriptive bivariée (l'effet de la loi sur le travail dominical) ; en gestion des ressources humaines pour l'estimation (le *turn-over* des consultants d'un cabinet de conseil) ; en finance pour la régression linéaire (le modèle du MEDAF<sup>7</sup>). Puisque le marketing et la finance sont les deux spécialisations principalement choisies par les étudiants en master, les cours du second semestre sont présentés comme des cours de pré-spécialisation. Le cours sur les méthodes descriptives d'analyse des données, intitulé « traitement d'enquêtes », est explicitement lié au domaine du marketing, tandis que le cours de modélisation est lié au domaine de la finance. Les exemples et jeux de données utilisés dans ces cours sont tous issus de problématiques du domaine de référence.

---

<sup>7</sup>Le MEDAF est un modèle, employé largement par les analystes financiers, qui permet de déterminer la sensibilité d'un actif par rapport aux fluctuations du marché.

### 3.2 Passé : rupture trop nette ou absence de rupture

Nous partons tout d'abord du sentiment des étudiants, qui est variable mais le plus souvent tranché : beaucoup jugent le cours de statistique pas assez en rupture avec les mathématiques de classes préparatoires, tandis que d'autres (lors de l'évaluation du cours par les étudiants) déplorent par exemple une absence de rigueur dans (l'enseignement de) la discipline, soulignant en réalité un manque de lien avec les mathématiques qu'ils connaissaient jusque-là. Dans tous les cas, il nous faut susciter le désir de notre discipline, et nous détaillerons dans un second temps comment nous nous y employons.

**Critique finale (parfois) : rupture trop nette** — Les évaluations écrites par les étudiants à la fin du cours pointent parfois un prétendu manque de rigueur du cours de statistique par rapport aux cours de mathématiques suivis en classes préparatoires (ce qui est assez savoureux, étant donné que de nombreux résultats intéressants mais difficiles, comme le théorème limite central, y sont admis). Ces critiques disparaissent dès que l'intervenant explique que l'objet même de la statistique n'est pas de formuler des assertions nécessairement exactes, mais simplement sûres (ou alors, de s'abstenir de conclure lorsque cela n'est pas possible). A cet égard, les difficultés dont nous avons dressé la liste au paragraphe 1.3 surgissent, mais elles ne sont pas vraiment propres à notre public : elles concernent tous les étudiants qui ont eu une formation minimale en mathématiques. C'est pourquoi nous développerons davantage le point suivant, propre aux étudiants des écoles de management.

**Crainte initiale (fréquente) : absence de rupture** — C'est peu de dire que les attentes de changement des étudiants sont déçues lorsqu'ils découvrent sur leur emploi du temps des cours de statistique. Les plus honnêtes d'entre eux soulignent d'ailleurs dans leur évaluation du cours combien ils redoutaient le « retour des mathématiques » après les années de classes préparatoires. (Pourquoi parler de retour après seulement quelques semaines de pause estivale depuis les classes préparatoires ? C'est sans doute parce que dès les résultats du concours proclamés, ils avaient fait une croix sur tout enseignement ressemblant aux mathématiques.) Pour eux, il y a identité *a priori* entre mathématiques et calcul des probabilités d'une part, statistique d'autre part (avec, hélas, une charge négative très forte mise sur les mathématiques, ressentie comme matière pivot de la sélection académique tout au long des études antérieures). Bien évidemment, le changement de point de vue décrit plus haut et les objectifs poursuivis, notamment, de simple application de formules, mais d'interprétation de leurs résultats et de formulation de conclusions stratégiques, pourront leur montrer, rétrospectivement, la différence de nature entre probabilités et statistique.

Mais l'expérience montre qu'il ne suffit pas d'attendre que cette prise de conscience rétrospective ait lieu, quand elle a lieu. Il faut devancer l'expression des attentes déçues face au « retour » des mathématiques, susciter le désir, et grossir le trait de la rupture de point de vue par rapport aux enseignements de mathématiques précédents. La solution retenue tant à ESCP Europe qu'à HEC Paris consiste en une scénarisation du cours, à rebours des cours de mathématiques académiques offerts aux étudiants jusque-là. Trois volants d'adaptation peuvent être employés ; tous trois mettent en jeu des données, qui forment le point d'ancrage de la statistique et sont le marqueur et le révélateur de sa différence par rapport aux mathématiques.

Le premier consiste à systématiquement relier les méthodes statistiques présentées à des contextes managériaux authentiques, comme nous l'avons déjà illustré au paragraphe 2.2.2



(dans la description de l'accompagnement du changement de perspective) et à la fin du paragraphe 3.1. C'est l'approche plutôt employée par ESCP Europe, et on la retrouve dans une moindre mesure dans le cours donné à HEC Paris. Le second vise à ouvrir la méthodologie statistique au vaste monde, loin de l'univers confiné des mathématiques, par exemple en la reliant à la vie d'honnête citoyen, *via* des données correspondant à des situations politiques ou sociales. Le troisième, traité dans le paragraphe 3.3, consiste à s'inscrire dans la culture de l'école. A HEC Paris, ces deux dernières approches sont particulièrement développées, mais elles sont également intégrées ponctuellement par ESCP Europe.

### **Eveiller les citoyens d'aujourd'hui à prendre les informations chiffrées avec du recul —**

Nous voulons montrer que la statistique apporte un éclairage inattendu sur des problématiques citoyennes<sup>8</sup> ou concernant la société dans son ensemble ; tout en traitant des données, nous pouvons ainsi, nous aussi, « raconter des histoires » et illustrer que les chiffres ne parlent pas d'eux-mêmes. Nous sommes alors à des années-lumières des cours de théorie probabiliste des classes préparatoires car nous nous efforçons de donner une culture générale de réception des chiffres. C'est là notre manière d'ancrer le cours de statistique dans le présent des étudiants, dans un réel vécu ; nous ne le faisons pas, comme cela peut être le cas dans des universités ou écoles de management du monde anglo-saxon, en utilisant des jeux de données collectées sur les étudiants (temps de sommeil, consommation d'alcool, etc.). Il nous semble en effet préférable de les ouvrir au monde plutôt que de les prendre comme référence.

L'axe le plus évident en ce sens concerne la lecture des sondages, mais d'autres jeux de données, en lien avec les tests du  $\chi^2$  peuvent être employés avec profit. Nous décrivons quelques anecdotes et histoires traitées en cours dans l'annexe 1.

### **3.3 Présent : inscription dans la culture de l'école**

Une troisième manière, plus marginale, de rendre le cours de statistique légitime est de l'inscrire dans la vie et la culture de l'école, c'est-à-dire, de l'ancrer dans le présent des étudiants. Nous pouvons utiliser à cet effet l'implication des étudiants dans une des associations les plus prestigieuses des écoles de management, la « junior entreprise » (JE), dans laquelle le côté professionnel est très marqué. Cette dernière effectue des missions, souvent d'ordre statistique comme la réalisation et le traitement d'enquêtes, pour le compte d'entreprises commanditaires n'ayant pas nécessairement de liens avec nos écoles. Les étudiants sont alors appelés à mettre en œuvre de manière autonome les compétences développées dans nos cours de statistique, y compris leur capacité à consulter et comprendre des manuels de statistique pour accéder à des techniques non vues dans ces cours mais utiles pour leur étude. Davantage de détails sur cet axe sont proposés dans l'annexe 2.

---

<sup>8</sup>Cet angle d'approche se fonde sur une maxime formulée par l'écrivain britannique H.G. Wells : « Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write ».

## 4 Conclusion

Dans cet article, nous avons décrit les difficultés auxquelles nous sommes confrontés dans notre pratique d'enseignants et de responsables d'un cours fondamental de statistique en école de management. Nous avons montré que ces difficultés s'inscrivaient dans un cadre très mouvant, ayant pour origine les évolutions observées dans l'enseignement supérieur ces dernières années. Ces évolutions exacerbent les tensions, récurrentes dans les écoles de management, entre savoirs académiques et savoirs pratiques. Ces tensions sont particulièrement fortes dans notre discipline dont le positionnement épistémologique est complexe.

Nous avons commenté certaines des stratégies mises en œuvre afin de tenter de résoudre ces difficultés, en nous focalisant sur trois angles d'approche, communs au-delà des différences liées aux cultures de nos écoles ainsi qu'à nos épistémologies personnelles.

### 1. *Le choix du degré de formalisation et des mises en situation*

Il s'agit d'un cours introductif qui vise d'une part à permettre aux étudiants de prendre conscience des limitations de la discipline (on ne peut pas toujours conclure mais on sait quand on ne peut pas conclure), à dépasser les aspects « presse-bouton » tout en ne rejetant pas la discipline pour autant. Notre monde est incertain, c'est inconfortable mais c'est ainsi : on peut quantifier les incertitudes, et parfois les réduire (mais en aucun cas les faire disparaître) en y consacrant un temps et/ou un montant qu'on peut calculer à l'avance (et qui n'est pas toujours raisonnable). D'autre part, il s'agit surtout d'apprendre aux étudiants à apprendre les techniques statistiques : de leur donner un bagage minimal, notamment en termes de formalisme, afin qu'ils puissent comprendre par eux-mêmes des méthodes plus sophistiquées plus tard, selon leurs besoins.

### 2. *L'objectif de faire vivre les problèmes de statistique*

Notre cours de statistique succède aux cours de théorie des probabilités suivis en classes préparatoires (et s'appuie sur eux), ce qui est une approche très française. Nous devons alors souligner les différences avec les mathématiques des classes préparatoires, et notamment, la nécessité de dépasser l'aspect calculatoire pour aller jusqu'à la prise de décision. Nous voudrions favoriser le passage d'un statut d'étudiant passif face à un énoncé qu'il subit à celui d'un décideur actif prenant en main un problème (ce qui est une des dimensions du phénomène de conversion décrit par Abraham, 2007).

### 3. *La volonté de rendre légitime la discipline*

Pour cela, nous relierons systématiquement les méthodes statistiques aux cours de mathématiques des classes préparatoires pour souligner les différences de perspective, ainsi qu'avec leur utilisation en entreprise ou dans la vie citoyenne pour mettre en avant leur intérêt pratique. Nous devons toujours et encore substituer à l'argument d'autorité du professeur mathématicien des classes préparatoires (« nous étudions cet objet parce que je vous dis qu'il est intéressant » ou parce qu'il est au programme) une démarche inductive, partant de situations concrètes mais revenant également à la théorie mathématique.

Nos réflexions nous conduisent aujourd'hui à envisager plusieurs pistes d'amélioration en ce qui concerne nos enseignements. Ainsi, il nous apparaît désormais qu'il faudrait repenser la place du cours de statistique dans l'ensemble de la formation en fonctionnant au moins en

C. Hahn et G. Stoltz

partie en post-requis et non plus seulement en pré-requis : à ESCP Europe, le cours de master à destination des admis directs, assuré par les mêmes professeurs, est un cours plébiscité par les étudiants ; à HEC Paris, les questions des étudiants sont nombreuses lors des stages en entreprise (ou des missions JE décrites en annexe 2). Nous pourrions par exemple construire un cours électif de master complémentaire du cours fondamental existant (par exemple sur le modèle du projet de cours décrit en annexe 2 pour valoriser l'expérience acquise lors des missions JE).

Ces mêmes réflexions nous ont également conduits à considérer de nouvelles questions de recherche. Nous explorons notamment actuellement celle de l'influence des différentes épistémologies des professeurs en école de management sur leur enseignement et les conceptions des étudiants.

**Sources pour ESCP Europe** — Le cours dispensé à ESCP Europe a été publié : *Méthodes statistiques appliquées au management*, Corinne Hahn et Sandrine Macé, Pearson, 2012. Le site web de l'éditeur permet notamment aux étudiants d'accéder aux corrections des exercices.

**Sources pour HEC Paris** — La page <http://www.hec.fr/stoltz> reprend tous les matériels pédagogiques disponibles pour le cours « Eléments de statistique pour citoyens d'aujourd'hui et managers de demain », et en particulier, les photocopiés distribués aux étudiants (un tome de cours, un tome d'exercices contenant de nombreuses annales) et les enregistrements vidéographiques réalisés lors du premier semestre de l'année universitaire 2012–2013.

## Références

- [1] Abraham, Y. (2007), Du sérieux scolaire au souci managérial, ou comment devenir un « HEC », *Revue française de sociologie*, **48**(1), 37–66.
- [2] Adda, J. (1976), Difficultés liées à la présentation des questions mathématiques, *Educational Studies in Mathematics*, **7**, 3–22.
- [3] Armatte, L. (2010), Le rôle de l'histoire dans l'enseignement de la statistique, *Statistique et enseignement*, **1**(2), 23–47.
- [4] Beber, B. et A. Scacco (2009), The devil is in the digits, *The Washington Post*, 20 juin 2009.
- [5] Boaler, J. (2002), The development of disciplinary relationship: Knowledge, practice and identity in mathematics classrooms, *For the Learning of Mathematics*, **22**(1), 42–47.
- [6] Cobb, P. et L. Hodge (2002), Learning, identity and statistical data analysis, in B. Phillips, éditeur, *Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS)*.
- [7] Cohen-Scali, V. (2000), *Alternance et identité professionnelle*, PUF, Paris.
- [8] Couvelaire, L. (2013), Et Nate créa le data, *M le magazine du Monde*, 25 mai 2013.
- [9] Crawford, M. (2013), Dans un espace public saturé de technologies, l'attention s'épuise, *Le Monde*, interview par Jean-Baptiste Jacquin, 25 juillet 2013.

- [10] Dunkel, T. et I. Le Mouillour (2007), “Through the looking glass.” Diversification and differentiation in vocational education and training in higher education, in P. Descy et M. Tessaring, éditeurs, *Modernising Education and Training: fourth report on vocational training research in Europe*, pages 242–268, Cedefop.
- [11] Fabre, M. (2011), *Eduquer pour un monde problématique. La carte et la boussole*, PUF, Paris.
- [12] Fine, J. (2010), Probabilités et statistique inférentielle : approche sondage versus approche modèle, *Statistique et enseignement*, **1**(2), 5–21.
- [13] Fine, J. (2012), Statistique, informatique, mathématiques et interdisciplinarité, *Statistique et enseignement*, **3**(2), 33–59.
- [14] Hahn, C. (2000), Teaching mathematics to shop-assistant apprentices: Exploring content and didactical situations, in A. Bessot et J. Ridgeway, éditeurs, *Education for Mathematics in the Workplace*, pages 159–166, Kluwer Academic Publishers, Dordrecht.
- [15] Hahn, C. (2013), Linking professional experiences with academic knowledge: The construction of statistical concepts by sale managers apprentices, in A. Damlamian, J. Rodrigues, et R. Strässer, éditeurs, *Educational Interfaces between Mathematics and Industry: report on an ICMI–ICIAM study*, pages 137–146, Springer.
- [16] Keen, A. (2007), *The Cult of the Amateur: How Today’s Internet Is Killing Our Culture*, Nicholas Brealey Publishing.
- [17] Le Moigne, J. (1999), *La modélisation des systèmes complexes*, Dunod, Paris.
- [18] Lemaitre, D. (2011), Professionalisation et modèles professionnels dans les grandes écoles françaises, *Recherche et formation*, **66**, 93–106.
- [19] Lessart, C. et R. Bourdoncle (2002), Qu’est qu’une formation professionnelle universitaire ? Conceptions de l’université et formation professionnelle, *Revue française de pédagogie*, **139**, 131–153.
- [20] Maes, A. (2011), Les stats ont la cote, *20 Minutes*, 16 novembre 2011.
- [21] Meletiou-Mavrotheris, M. et C. Lee (2002), Teaching students the stochastic nature of statistical concepts in an introductory statistics course, *Statistics Education Research Journal*, **1**(2), 22–37.
- [22] Michalski, T. et G. Stoltz (2013), Do countries falsify economic data strategically? Some evidence that they might, *The Review of Economics and Statistics*, **95**(2), 591–616.
- [23] Naidoo, R. et I. Jamieson (2005), Empowering participants or corroding learning? Towards a research agenda on the impact of student consumerism in higher education, *Journal of Education Policy*, **20**(3), 267–281.
- [24] Noss, R. et C. Hoyles (1996), *Windows on mathematical learning*, Kluwer Academic Publishers, Dordrecht.

C. Hahn et G. Stoltz

- [25] Pastré, P., G. Vergnaud, et P. Mayen (2006), La didactique professionnelle, *Revue française de pédagogie*, **154**, 145–198.
- [26] Peters, S. (2011), Robust understanding of statistical variation, *Statistics Education Research Journal*, **10**(1), 52–88.
- [27] Petocz, P. et A. Reid (2010), On becoming a statistician; a qualitative view, *International Statistical Review*, **78**(2), 271–286.
- [28] Piketty, T. (2011), Réguler (enfin) les sondeurs, *Libération*, 8 mars 2011.
- [29] Robin, J.-P. (2010), Le métier de statisticien sera le plus « sexy » de la décennie, *Le Figaro*, 18 octobre 2010.
- [30] Saporta, G. (2011), *Probabilités, analyse des données et statistique*, Technip, Paris.
- [31] Sinha, P., A. Sharma, et H. V. Singh (2012), Prediction for the 2012 United States presidential election using multiple regression model, *The Journal of Prediction Markets*, **6**(2), 77–97.
- [32] Stevenson, J. (2002), Concepts of workplace knowledge, *International Journal of Educational Research*, **37**(1), 1–15.
- [33] Tenenhaus, M. (2007), *Statistique : méthodes pour décrire, expliquer et prévoir*, Dunod, Paris.
- [34] Young, M. (2009), Education, globalization and the “voice of knowledge”, *Journal of Education and Work*, **22**(3), 193–204.

## Annexe 1 — Statistique et vie citoyenne

L’axe le plus évident du lien entre statistique et vie citoyenne concerne la lecture des sondages, de mieux en mieux cernée par les journalistes même si l’on revient de loin ; c’est au tour du grand public désormais de mieux les appréhender et nos étudiants peuvent être les ambassadeurs d’une exploitation réfléchie des études d’opinion, tant en entreprise que dans la société (c’est-à-dire : une exploitation avec des marges d’erreur à plus ou moins 3 % des fréquences observées sur l’échantillon, lorsque ce dernier contient environ 1 000 participants). Notre matériau d’étude en cours est extrait de la presse. Par exemple, un article de Piketty (2011) rappelle l’existence de marges d’erreur dans les estimations ponctuelles fournies par une enquête et plaide également pour un meilleur encadrement des instituts et une plus grande transparence de leurs méthodes. Une « minimare » du *Canard enchaîné*, reproduite à la figure 1, montre qu’en août 2010 le volatile n’avait pas encore intégré mentalement l’existence de ces marges d’erreur (« pan sur le bec ! ») ; commenter cette brève a formé un exercice de quiz d’HEC Paris, plutôt bien réussi dans l’ensemble. Enfin, la troisième et plus frappante illustration montre que parfois, les sondeurs se trompent (sans que cela ne soit à attribuer aux fameux retournements soudains



FIGURE 1 – *Une minimare du Canard enchaîné, publiée en août 2010.*

d'opinion, qui constituent l'explication privilégiée des journaux télévisés<sup>9</sup>). Au vu des niveaux de confiance typiques de 95 %, il suffit de réaliser simultanément une vingtaine de sondages pour être (presque) assuré qu'au moins l'un d'entre eux ne capture pas bien la réalité. Les soirées d'élections municipales forment un moment privilégié pour ce faire, puisque justement, une vingtaine de communes (ou d'arrondissements parisiens) sont suivis avec suffisamment de moyens pour que soient formées des estimations peu après la clôture du vote. A HEC Paris, nous illustrons ce fait par l'exemple réel suivant : dans le 5<sup>e</sup> arrondissement de Paris, un institut de sondage a ainsi prévu, à 20h45, que Lyne Cohen-Solal avait gagné l'élection, avant de changer sa prévision à 21h15 en faveur de Jean Tibéri (dont on a pu savoir à 22h avec certitude, une fois le dépouillement effectué, qu'il l'avait effectivement emportée). En réalité, nous montrons aux étudiants qu'une fraction des sondages est vouée, par construction, à se tromper, et comme nos sociétés modernes (et leurs entreprises) tendent à ne plus accepter l'erreur, il s'agit donc de présenter leurs résultats avec les pincettes nécessaires : « On peut dire, avec confiance 95 % et avec une marge d'erreur de plus ou moins 3 %, qu'aujourd'hui, telle proportion de la population pense que... »

A HEC Paris, le cours sur les tests du  $\chi^2$  présente d'autres exemples naturels pour rejeter l'adage que « les chiffres parlent d'eux-mêmes ». Des politologues (Beber et Scacco, 2009) ont étudié les résultats de l'élection présidentielle de 2009 en Iran ; ils voulaient tester l'hypothèse de report sincère des résultats obtenus par les quatre candidats dans les 29 régions du pays. On dispose donc ici de  $29 \times 4 = 116$  données, qui sont des nombres de voix. Ces nombres sont typiquement de l'ordre de plusieurs dizaines ou centaines de milliers de voix ou plus. Leurs chiffres des unités ne semblent pas devoir être gouvernés par une loi autre que la loi uniforme : ils ne procurent pas d'information (contrairement aux premiers chiffres significatifs par exemple). Ces

<sup>9</sup>Sur France 2, la présentation des résultats des sondages électoraux est assortie, depuis quelques années, de la formule : « Nous vous rappelons que ce sondage ne constitue qu'un instantané de l'opinion publique et ne préjuge pas des résultats des élections. »

C. Hahn et G. Stoltz

politologues ont donc posé pour hypothèse  $H_0$  celle d'uniformité de la répartition des chiffres des unités des scores des quatre candidats. L'idée était que  $H_0$  correspond au cas de données non modifiées de manière flagrante, alors que le cas  $H_1$  où l'un des chiffres serait sur- ou sous-représenté serait le signe d'une altération humaine de la réalité des votes, par un être humain manipulant les données et ne les reportant pas de manière sincère. (En particulier, on sait que les humains sont de mauvais inventeurs de chiffres aléatoires.) Or, le raisonnement des politologues se fonde sur l'étude de fréquences individuelles de chiffres bien choisis *a posteriori* au vu des données : le 7 par exemple, pour qui l'écart au comportement uniforme est le plus grand. Cela revient, mathématiquement, à considérer la déviation maximale par rapport à la fréquence théorique, qui ne suit évidemment pas la même loi que la déviation d'un chiffre donné, choisi *a priori*. A cause de cette erreur méthodologique, ils parviennent à rejeter catégoriquement  $H_0$  et remettent donc sévèrement en cause la sincérité du scrutin. Or, le traitement rigoureux des données, comme l'ont fait remarquer dans les jours qui suivent des statisticiens, consiste par exemple à mettre en œuvre un test du  $\chi^2$  d'ajustement à la loi uniforme ; en l'espèce, la  $P$ -valeur obtenue sur les  $29 \times 4 = 116$  données est d'environ 7%, ce qui est une valeur qui ne permet pas de conclusion tranchée.

## Annexe 2 — Liens entre le cours de statistique et la « junior entreprise »

Nous décrivons dans cette annexe le fonctionnement interne de la « junior entreprise » (JE) et les deux moments où nous pouvons être appelés, en tant que responsables du cours de statistique, à interagir avec ses membres. Nous proposons également, à HEC Paris, un projet de cours de statistique avancé en lien avec la JE.

**Fonctionnement interne.** — La JE est dirigée par un bureau, élu, dont les membres sont appelés « administrateurs » ; les missions sont, elles, offertes à l'ensemble des étudiants ayant cotisé à la JE. Parmi les nombreux avantages et opportunités avancés par les étudiants pour être un tel administrateur, on peut noter, outre la constitution d'un réseau professionnel, l'introduction régulière et soutenue de travaux concrets en lien avec l'entreprise dans le cursus (une autre bonne raison d'être administrateur est que ce poste est rémunéré sur une base mensuelle).

Les contacts avec la JE en lien avec le cours de statistique peuvent avoir lieu à deux moments : lors de la période électorale ou lors de l'exercice du mandat d'administrateur. Nous commençons par ce dernier.

**Suivi et interactions postérieures au cours de statistique** — Les JE d'HEC Paris et d'ESCP Europe mettent en avant la qualité de leur expertise en statistique et il leur est donc souvent confié des missions de cette nature, avec parfois des demandes très précises, mais pas toujours adaptées, de la part du commanditaire (« Je voudrais un graphique d'analyse factorielle ») ou des incompatibilités entre la ligne budgétaire accordée et la masse de travail demandée (un vœu de précision de sondage incompatible avec le nombre de sondés possible). Il se peut alors que les étudiants se tournent vers nous pour obtenir un conseil, qui reste le plus souvent vague de notre part pour deux raisons concordantes : les étudiants ne nous présentent que très partiellement le problème rencontré par mail (et de la manière, bonne ou plus critiquable, dont ils ont déjà com-

mencé à le prendre en main) ; nous nous refusons à réaliser l'étude à leur place et nous contenons de leur indiquer des références, souvent issues de manuels classiques, comme Saporta (2011) et Tenenhaus (2007).

Cette situation est frustrante pour tous et l'idée de créer à HEC Paris un cours électif de statistique avancée émerge. Ce cours mêlerait à degrés divers trois activités : des présentations par les étudiants de techniques statistiques non vues ou trop rapidement évoquées lors du cours fondamental (ANOVA, régression logistique, analyse en composantes principales, etc.) ; des exposés sur des missions JE déjà effectuées, afin d'avoir un retour sur la pertinence statistique du traitement effectué ; et enfin, de courtes discussions sur les missions en cours, pour suggérer des pistes d'étude. Si ce cours est monté, un retour sera effectué dans cette même revue sur les écueils et intérêts rencontrés.

**Cours de statistique et épreuves professionnelles de la campagne JE.** — Une des particularités des écoles de commerce par rapport à d'autres cursus de l'enseignement supérieur touche à l'implication dans la vie associative. La tension entre apprentissage de savoirs académiques et préparation aux futurs métiers de cadres d'entreprise est particulièrement bien incarnée par le comportement des étudiants lors des campagnes menées pour les élections des deux associations les plus prestigieuses, la « junior entreprise » et le bureau des élèves (BDE). Ces campagnes elles-mêmes, et pas seulement les responsabilités exercées par ceux qui seront élus, constituent déjà pour nos étudiants un apprentissage des futures qualités dont on voudra qu'ils soient parés (capacité à travailler en équipe, habilité à lever des fonds, à nouer des partenariats avec des entreprises extérieures, mises en œuvre au service de l'organisation de diverses animations sur le campus et de la constitution d'un réseau parmi les étudiants influents de deuxième ou troisième année). L'étude sociologique de Abraham (2007), qui prend pour cas d'étude HEC Paris, et d'autres travaux détaillent le déroulement de cette campagne et son impact sur la vie de l'école.

Nous voulons nous focaliser, pour notre part, sur les épreuves dites professionnelles de la campagne JE : généralement, des enquêtes à effectuer et à présenter. Chaque année, une partie des candidats à l'administration de la JE se manifeste aux enseignants de statistique afin de poser des questions pour mettre en œuvre les techniques vues en cours. Il s'agit souvent de réfléchir plus avant à la constitution des échantillons de sondés ou à la manière de présenter ses résultats (à quelles représentations de statistique descriptive ou à quelles mesures de centralité et de dispersion recourir). L'avantage à HEC Paris est que cette campagne JE se déroule en octobre–novembre, pendant le cours de statistique. Nous pouvons donc rebondir sur elle pour motiver l'intérêt de notre cours.

Par exemple, lors de la campagne JE à HEC Paris en novembre 2012, et avant que nous ne traitions des tests du  $\chi^2$  d'ajustement à une loi en cours, a émergé une question de la part d'une liste quant au caractère « représentatif » de son échantillon (ce mot « représentatif » est discuté ci-dessous, nous reproduisons la question originellement posée par les étudiants). Bien évidemment, nous n'avons pas traité les données à leur place et les avons simplement renvoyés vers le photocopié (qu'ils ont pu lire avec profit à l'avance). Les données fournies ont malgré tout été, par leur pertinence, la matière à un exercice de contrôle continu court, que nous reproduisons ci-dessous.

Un groupe d'étudiants de la liste H interroge des étudiants sur le campus afin de réaliser une enquête sur la culture à HEC ; les 200 étudiants sondés se répartissent de la manière suivante, selon le programme et le sexe :



C. Hahn et G. Stoltz

	Hommes	Femmes	Total
Grande Ecole	78	73	151
MS / MSc	17	12	29
MBA	10	7	17
Doctorat	1	2	3
<b>Total</b>	<b>106</b>	<b>94</b>	<b>200</b>

Sophie Garnichat, de la Direction de la communication, indique que les 2 983 étudiants en cours de scolarité se répartissent quant à eux de la manière suivante dans ces catégories :

	Hommes	Femmes	Total
Grande Ecole	39.4%	31.3%	70.7%
MS / MSc	9.4%	7.0%	16.4%
MBA	6.5%	3.6%	10.1%
Doctorat	1.5%	1.3%	2.8%
<b>Total</b>	<b>56.8%</b>	<b>43.2%</b>	<b>100%</b>

Déterminer si l'échantillon de sondés est représentatif de l'ensemble des étudiants en cours de scolarité à HEC Paris. On rédigera la réponse sous la forme d'un texte expliquant la méthodologie retenue, ses limites, ses résultats. On imaginera que l'on s'adresse à un public large.

Suivaient alors des tableaux d'analyse tels que fournis par SPSS, comme celui reproduit à la figure 2 ci-dessous.

Catégorie			
	Effectif observé	Effectif théorique	Résidu
Grande Ecole / Homme	78	78,7	-,7
Grande Ecole / Femme	73	62,6	10,4
MS-MSc / Homme	17	18,8	-1,8
MS-MSc / Femme	12	14,1	-2,1
MBA / Homme	10	13,1	-3,1
MBA / Femme	7	7,2	-,2
Doctorat / Homme	1	2,9	-1,9
Doctorat / Femme	2	2,7	-,7
<b>Total</b>	<b>200</b>		

  

Test	
	Catégorie
Khi-deux	4,368 <sup>a</sup>
ddl	7
Signification asymptotique	,737

a. 2 cellules (25,0%) ont des effectifs théoriques inférieurs à 5. L'effectif théorique minimum d'une cellule est 2,7.

FIGURE 2 – Un exemple de tableau fourni en rapport avec la question de la représentativité de l'échantillon.

Cet exercice a suscité quelque émotion, d'une part parce que c'était là un cadeau inattendu d'une liste JE à l'ensemble du campus, mais également, d'autre part, parce qu'il prouvait que

*Le cours de statistique à HEC Paris et ESCP Europe*

dès ces premières semaines de cours, le contenu des enseignements de statistique pouvait trouver une application concrète et pertinente. Nous avons été heureux de pouvoir rebondir efficacement sur les préoccupations des étudiants pour les rejoindre là où ils en sont, et pas là où nous voudrions qu'ils se projettent.

Evidemment, il ne s'agissait pas dans cet exercice, posé en temps limité, de définir ce que l'on entendait par « représentatif ». En l'occurrence, la question initiale des étudiants est qu'ils avaient essayé, du mieux possible, de construire un échantillon aléatoire en sondant des étudiants au hasard sur le campus, et ils voulaient mettre le résultat obtenu à l'épreuve du feu, selon des critères, éventuellement croisés, de sexe et de programme de scolarité. Ils voulaient donc tester si les proportions observées dans leur échantillon dans les différentes catégories définies étaient compatibles, ou non, avec celles de la population. Leur souhait était de s'assurer qu'aucun biais manifeste d'échantillonnage n'avait eu lieu, favorisant des étudiants d'un sexe ou d'un programme au détriment d'autres.