

Statistique et déontologie scientifique

Résumé

L'image sociale souvent peu flatteuse de la Statistique est illustrée à travers quelques problèmes de déontologie statistique et plus généralement scientifique. Des exemples historiques ou plus récents mettent en évidence falsifications et dissimulations triviales de données. D'autres sont le résultat d'une utilisation abusive ou du détournement de méthodes statistiques, notamment des tests d'aide à la décision. Ce dernier point est illustré par les controverses autour du maïs OGM. Ces pratiques douteuses et bien trop répandues ont poussé les associations des statisticiens les plus exposés à rédiger des codes de déontologie ou de bonne pratique.

Retour à l'[introduction](#).

1 Introduction

1.1 Image "historique" de la Statistique

Depuis ses débuts ou tout du moins depuis le début du XIX-ième siècle, la Statistique bénéficie d'une renommée plus ou moins douteuse, certes par méconnaissance d'une discipline aux concepts difficiles mais aussi à cause de nombreux abus.

Ainsi, en 1906 Mark Twain attribuait, sans doute à tort, à un premier ministre britannique la trop célèbre maxime : *There are three kinds of lies: lies, damned lies, and statistics*.

Évidemment les statisticiens académiques se sont fortement alarmés des mauvais usages de leur discipline et donc des mauvais procès qui leur étaient intentés. Dès décembre 1938, Jerome B. Cohen publie un article dans le *Journal of the American Statistical Association*¹ qui décrit : *The misuse of statistics*.

Darrel Huff (1954) [4] est l'auteur d'un *bestseller* très souvent réédité (figure

How to Lie with

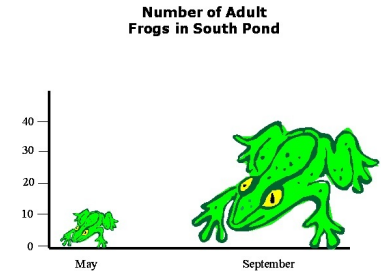


FIGURE 1 – Le livre illustré de Darrel Huff (1954) et un exemple de graphique “créatif” : la réalité (la hauteur de la grenouille) est faussée par l'impression visuelle (sa surface). Proscrire d'autant plus les graphiques faussement en perspective !

1) dans lequel il identifie et illustre plusieurs types de mensonges : falsification de données, biais d'échantillonnage, extrapolation, graphiques “créatifs”, hypothèses frauduleuses, incompetence...

Les exemples historiques montrent que le débat sur les mauvais usages innocents ou intentionnels de la Statistique est ancien. Il ne faudrait néanmoins pas penser qu'avec le temps et l'accroissement des connaissances, la place et l'image de cette discipline aient favorablement progressé. Strasak et al. (2007) de l'Université médicale d'Innsbruck ont publié *Statistical errors in medical research – a review of common pitfalls*² ; titre auquel nous aimerions ajouter : *pitfalls or fakes ?*

1.2 Motivations

La rédaction de ce texte est motivée par une concentration d'“affaires” qui ont secoué ou qui régulièrement secouent l'environnement politico-médiatique : affaire du Mediator de Servier, conflits récurrents autour des OGM, crise financière omniprésente depuis 2008. Contribuant à former des professionnels de la Statistique actifs dans des domaines très divers (figure 2), il me semble important de les sensibiliser aux difficultés, voire aux pressions, auxquelles ils vont se trouver professionnellement confrontés :

1. volume 33 N°204 pp 657-674

2. *Swiss Med Wkly*, 2007, 137, 44-49



FIGURE 2 – Les statisticiens travaillent dans des domaines “sensibles” d’un point de vue éthique et sociétale : autorisation de médicaments et veille sanitaire, critères d’affectation de crédits, produits et risques financiers...

- pression académique de publication (*publish or perish*) pour valider des résultats et conclusions d’expériences,
- pressions commerciales pour justifier des produits plus ou moins “éthiques” : crédit renouvelable ou *revolving* (en cause dans 80% des dossiers de surendettement), instruments financiers douteux...,
- justifier de l’innocuité ou de l’efficacité de produits industriels (médicaments, agroalimentaire, bio et nanotechnologies...),
- pressions politiques (sondages, statistiques publiques).

Un seuil d’indignation étant largement dépassé, notamment avec le scandale du Mediator, il nous a semblé important de mieux formaliser par écrit le contenu de ces questions déontologiques en plus des remarques et commentaires oraux intervenant en présence des étudiants. L’objectif est donc de faire réfléchir sur le rôle de la Statistique et les responsabilités de ses principaux acteurs quand il s’agit d’aider à l’évaluation des risques ou à la prise de décision.

La déontologie statistique n’est qu’un élément du débat beaucoup plus large sur l’éthique de la recherche scientifique. Le texte de ce document est volontairement ciblé sur le bon usage de la Statistique et également daté par certains événements de l’année 2011 mais son contenu restera malheureusement d’actualité compte tenu du sujet. Ainsi, dans le numéro 36 de mars 2012, le magazine *La Recherche* évoque le premier séminaire *Parlons Éthique* organisé par l’Institut Curie/Inserm (Unité 932). Cet article cite Melissa Anderson, intervenue lors de ce séminaire à propos de sa participation à une enquête anonyme auprès de 3247 scientifiques. Cette enquête a révélé que 33%

des interrogés avaient manqué à la déontologie. Parmi le “top 10” des mauvaises pratiques se classent en premier : *Falsifier ou “arranger” les données* et en septième : *Omettre de présenter des données parce qu’elles seraient en contradiction avec vos précédentes recherches*. Dans le même ordre d’idée, Fannelli (2009) [3] propose dans cet article une méta-analyse de 18 enquêtes sur les falsifications de la recherche et principalement des données. Pour résumer brièvement, 2% de scientifiques admettent avoir eux-mêmes falsifié ou modifié des données et ce pourcentage monte à 14% à propos du comportement des autres collègues. Cette étude retrouve le chiffre de 33% de manquement général déontologique, chiffre porté à 72% à propos des autres collègues.

1.3 Contenu

Une première section rappelle quelques fraudes célèbres et évidentes de falsifications ou dissimulations de données parmi celles qui ont évidemment été décelées dans différents domaines de la recherche ; celles-ci sont relativement faciles à mettre en évidence ou prouver dès que la supercherie est soupçonnée. En revanche, une section suivante aborde un exemple de malversation plus habile concernant une mauvaise utilisation de méthodes statistiques (i.e. les tests) ou leur utilisation hors limite de validité ou encore sans contrôle explicite d’une marge d’erreur. Alors que les falsifications de données ne laissent plus guère de doute et se terminent même devant les tribunaux, les mauvais usages de la Statistique peuvent générer des querelles d’experts lourdes et sans conclusion tranchée ne serait-ce que par le coût des expérimentations qu’il faudrait entreprendre pour arriver à une réponse significative. Une dernière section tâche de faire le tour des démarches d’associations professionnelles soucieuses de ces difficultés et qui ont proposé des codes de déontologie statistique.

2 Falsifications et dissimulations

Internet et plus précisément [Wikipedia](#) regorgent d’informations sur les falsifications et fraudes scientifiques historiques ainsi que sur les mauvais usages (*misuse*) de la Statistique.

2.1 Statistiques publiques

Nos collègues britanniques rappellent que leurs *Politicians use statistics like drunkards use lampposts : not for illumination, but for support*. Nos homo-

logues français n'ont rien à leur envier. Les exemples sont malheureusement très, trop nombreux. Citons une affaire qui montre nettement les relations difficiles entre monde politique et statistique officielle publique. Le 25 mai 2011, Claude Guéant, Ministre de l'Intérieur, s'exprime sur *Europe 1* :

les deux tiers des échecs scolaires, c'est l'échec d'enfants d'immigrés.

Il a fallu un mois et une menace sérieuse de grève de la part des syndicats de l'INSEE pour obliger la direction de cet organisme public à rétablir les chiffres à travers un communiqué officiel le 27/06/2011 :

La proportion d'enfants d'immigrés parmi les élèves sortis sans qualification de l'enseignement secondaire peut être estimée à environ 16% pour les enfants de familles immigrées. Si on y ajoute les enfants de familles 'mixtes', cette proportion passe à environ 22%.

“ Le ministre, accusé par l'opposition de stigmatiser les immigrés pour séduire les électeurs d'extrême droite, avait affirmé s'appuyer sur des études de l'Insee et a envoyé plusieurs droits de réponse à des médias qui écrivaient que ses chiffres étaient faux.” (*Le Monde* du 28/06/2011).

2.2 Sciences humaines

Dans ce domaine aussi, les exemples seraient nombreux. Citons celui historique de l'[affaire Burt](#). Sir Cyril Burt, professeur de Psychologie à Londres en 1931 fut Président de la *British Psychology Society* puis éditeur en chef du *Journal of Statistical Psychology*. Il reste célèbre pour des contributions à la Statistique (les tableaux de contingences multiples portent son nom) et aussi pour son influence sur l'organisation du système éducatif britannique : Burt ayant “prouvé” l'influence prépondérante de l'hérédité sur l'intelligence (mesurée par le QI de Binet), cela justifierait d'orienter par un examen (*eleven plus*) les élèves dès 11 ans vers une filière primaire, moyenne ou supérieure (grammar-modern-technical) en fonction de leurs notes.

Burt affirme, au cours de 3 études (1955, 1958, 1966), que la corrélation des QIs entre couples de jumeaux homozygotes élevés dans des milieux différents dès leur enfance est significativement plus élevée que la corrélation entre les QIs de couples de jumeaux hétérozygotes élevés dans un même milieu social. Il montrerait ainsi l'importance de la notion de quotient intellectuel et leur héritabilité.

Ce n'est qu'au début des années 70 que de sérieux doutes furent émis ou de violentes accusations prononcées sur la pertinence de ces travaux : référence à des publications inexistantes ou introuvables et surtout des corrélations identiques à la 3ème décimales entre des études séparées de plus de 20 ans ! Les carnets d'expérience de Burt n'ont pas été retrouvés et des doutes plus que sérieux ont été soulevés sur la réalité des expériences notamment celles de 1966.

2.3 Sciences du vivant

L'exemple suivant illustre la pression considérable qui pèse sur la recherche scientifique, son financement, la renommée de ses acteurs et leur carrière. Il déborde largement le cadre d'un mauvais usage de méthodes statistiques.

[Hwang Woo-suk](#) est un sud-coréen qui est (fut) très réputé dans la recherche sur le clonage (le chien Snuppy) et les cellules souches. Lui et son équipe publient 2 articles dans *Science* en 2004 sur le clonage d'un embryon humain pour la production de cellules souches, et en 2005 sur la production de 11 lignées de cellules souches à partir de cellules de peau de patients adultes. Les conséquences thérapeutiques de tels résultats seraient considérables. Ces résultats lui valurent le titre de “*Supreme Scientist*” assorti de prix financiers très substantiels et de dons très généreux de fondations pour ses recherches.

A la suite de nombreuses péripéties scientifico-médiatiques, une commission d'enquête de l'Université de Séoul dénonce la falsification (fabrication de données) en décembre 2005 et ce chercheur est inculpé le 12 mai 2006 pour “fraude, détournements de fonds et violation des lois sur la bioéthique”. Les montants des détournements, des corruptions, des achats illégaux d'ovocytes se chiffrent en centaines de millions de dollars. Il a été déclaré coupable en octobre 2009 de détournements de fonds et violation des lois sur la bioéthique, mais pas de fraude, puis condamné à 2 ans de prison avec sursis.

2.4 Sciences de la terre

Claude Allègre fut ministre de l'Éducation nationale, de la recherche et de la technologie de 1997 à 2000 (gouvernement Jospin). Il est membre de l'Académie des Sciences et a reçu la médaille d'or du CNRS (1994) pour des travaux en Géochimie. Il est l'auteur de très nombreux livres de vulgarisation scientifique dont un en 2010 [1] d'entretiens avec le journaliste Dominique de Montvalon.

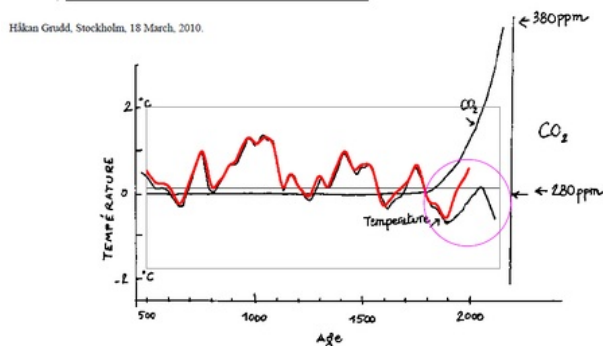


FIGURE 3 – En noir, courbe de température “dessinée” par C. Allègre à partir des Travaux de Grudd et rectifiée (en rouge) par Håkan Grudd lui-même. Il a également extrapolé très approximativement une concentration de CO2

Dans ce dernier livre, il formule de graves accusations contre le GIEC (Groupe d’experts intergouvernemental sur l’évolution du climat) qu’il qualifie de “système mafieux ayant conspiré pour faire passer aux yeux de l’ensemble du monde un mythe pour un fait scientifique”. Ce livre suscite de nombreuses controverses et accusations de “mensonges” ou “erreurs factuelles”.

Reprenons une seule de ces accusations qui concerne le graphique de la figure 3, accusation détaillée sur le blog [{Sciences²}](#) de Sylvain Huet, journaliste à *Libération*.

Håkan Grudd est un paléoclimatologue de l’Université de Stockholm spécialiste de l’étude de l’évolution du climat à partir de l’observation des anneaux des arbres (dendroclimatologie). Il publie en 2008 un article³ proposant une courbe des températures estivales au nord de la Suède pour la période 500–2004. Claude Allègre s’approprie cette courbe, la “complète”, et enrôle implicitement l’auteur dans des conclusions formellement rejetées par l’universitaire suédois qui dénonce les supercheries. Celles-ci consistent à faire prendre cette courbe locale en temps (été) et lieu (nord de la Suède) pour une courbe

3. Tornetraask tree-ring width and density AD 500–2004 : a test of climatic sensitivity and a new 1500-year reconstruction of north Fennoscandian summers, *Climate Dynamics* (2008) 31, pp843-857.

générale, à en modifier le tracé à partir de 1900 puis à l’extrapoler très approximativement jusqu’en 2100 sans aucune justification et avec le seul objectif de soutenir ses arguties plus polémiques que scientifiques.

Bien sûr, à la seule vue de la courbe rouge, un statisticien familier des séries chronologiques identifie une tendance, plutôt à la baisse, et une périodicité de l’ordre de 250 ans. Tout laisse à penser que Claude Allègre a prolongé à la “louche” cette courbe sans évidemment tenir compte de tous les travaux de modélisation entrepris par la communauté scientifique active en climatologie et les paramètres de forçage (concentration de CO2) ou covariables susceptibles d’intervenir. Cette communauté s’est d’ailleurs mobilisée (600 signataires) pour réclamer l’organisation d’un vrai débat sur le climat (*Le Monde* du 9 avril 2010).

La démarche des climat-sceptiques est aussi très développée aux États-Unis. Un livre traduit en français fait une analyse détaillée (Oreskes et Conway, 2012) [6] des motivations de “scientifiques” qui ont successivement mis en cause la nocivité du tabac, l’origine du trou dans la couche d’ozone, les pluies acides et maintenant la réalité et les causes du changement climatique.

2.5 Industrie pharmaceutique

Avant d’être autorisé à la vente, un médicament suit une procédure très réglementée d’essais cliniques.

- **Préclinique** : recherche et identification de molécules ayant potentiellement un effet thérapeutique. Essais sur des animaux (souris) pour évaluer ces effets et déterminer la dose active (modèles de pharmacocinétique). Une fois qu’une molécule est identifiée pour une cible thérapeutique donnée, des essais sont engagés sur l’homme.
- **Phase I** : Groupe de 20 à 80 sujets sains volontaires pour évaluer la tolérance et d’éventuels effets secondaires.
- **Phase II** : Groupe de 200 à 300 sujets malades pour commencer à évaluer l’efficacité, rechercher la dose optimale et toujours des effets secondaires possibles.
- **Phase III** : Ce sont plusieurs milliers de sujets malades qui participent à des études comparatives en double aveugle. L’effet du médicament candidat est comparé à celui d’un placebo et à celui du traitement de référence. La preuve statistique doit être apportée d’un effet significatif du médica-

ment par rapport au placebo et celle d'une amélioration par rapport au traitement de référence.

- A l'issue de ces essais, l'Afssaps prononce ou pas l'Autorisation de Mise sur le Marché (AMM)
- **Phase IV** (post-marketing) : Les visiteurs médicaux assurent la promotion des nouveaux médicaments et "recrutent" des médecins prescripteurs pour la détection d'effets secondaires ou de complications tardives afin que ces éventuels problèmes soient archivés par le laboratoire.

Dans le cas du Mediator, plusieurs dysfonctionnements ont été attestés, dysfonctionnements donnant suite à des poursuites judiciaires. Le Mediator, comme toute amphétamine, a des propriétés bien connues de "coupe faim" (anorexigène). Or ce médicament a obtenu une autorisation comme antidiabétique ; le député Gérard Bapt, qui a présidé la commission parlementaire sur le Mediator note : "A aucun moment (...) on n'a pu trouver un compte rendu de commission pour une autorisation de mise sur le marché qui valide cette indication pour le diabète" (*Libération* du 05/08/2011). Servier est poursuivi pour escroquerie envers l'Assurance Maladie et les Mutuelles qui ont porté plainte. En effet, considéré comme antidiabétique, celui-ci fut remboursé au maximum alors que seules ses capacités de "coupe faim" sont attestées.

Beaucoup plus grave, ce "médicament" a probablement causé 3100 hospitalisations et une estimation d'au moins 1300 morts entre 1976 et 2009 en France (article paru dans *Pharmacoepidemiology & Drug Safety* cité dans un article de *Libération* du 9-02-2012) ; il a fallu attendre novembre 2009 pour que l'Afssaps décide de retirer le Mediator du marché alors que de nombreux acteurs avaient attiré l'attention des pouvoirs publics sur les risques cardiaques et que le même type de médicament, l'Isoméride, a été interdit dès 1997 (*Le Monde* du 16/11/2010). Deux informations judiciaires ont été ouvertes, l'une pour "tromperie aggravée", l'autre pour "blessures et homicides involontaires".

Indépendamment des problèmes de conflits d'intérêts au sein de l'Afssaps et de commissions parlementaires, nous pouvons nous interroger sur la façon dont ont été traitées et analysées les données normalement recueillies lors de la phase iv des essais cliniques.



FIGURE 4 – L’environnement politico-scientifico-médiatique autour de l’autorisation de cultiver le maïs transgénique MON810

3 Abus méthodologique

Les exemples précédents mettent en évidence des cas de manquements déontologiques évidents par falsifications et/ou dissimulations de données qui conduisent à des procédures judiciaires. Dans d’autres cas, la “preuve” de la fraude ou du manquement éthique n’est pas aussi simple à apporter et relève d’une évaluation experte afin de mettre en évidence une utilisation abusive d’une méthode statistique. L’exemple du maïs OGM MON810 va permettre d’illustrer cette situation. Le site de [Marc Lavielle](#), membre du [Haut Comité pour les Biotechnologies](#), fournit des détails précis sur cette affaire politico-scientifico-médiatique (fig. 4).

3.1 Un exemple de maïs OGM

La chenille d’un papillon, la pyrale du maïs, est responsable de dégâts considérables sur les cultures de cette céréale (fig. 5). Une bactérie présente naturellement dans le sol sait produire une protéine, insecticide “naturel” et autorisé en agriculture biologique qui permet de lutter contre la pyrale.



FIGURE 5 – La pyrale du maïs, la chenille, ses dégâts et *Bacillus thuringiensis*.

Principe d'un OGM : le maïs MON810

Dans le cas du maïs MON810, l'objectif recherché par la production de ce maïs obtenu par transgénèse (fig. 6) est assez simple à appréhender. Une fois identifié le ou les gènes susceptibles de produire la protéine "insecticide", ceux-ci sont extraits de l'ADN de la bactérie et intégrés par des techniques de génie génétique à l'ADN de cellules de maïs. Les cellules obtenues sont cultivées jusqu'à l'obtention de semences d'une variété de maïs OGM "résistante" à la pyrale.

Les bénéfices attendus sont une meilleure production et une réduction de l'utilisation des pesticides de synthèse.

Les risques sont :

- des effets potentiels sur d'autres organismes, en particulier les insectes pollinisateurs (abeilles),
- la dissémination, contamination d'autres plantes par ces mêmes gènes,
- l'accroissement potentiel de la résistance de la pyrale par mutations,
- des éventuels problèmes de santé humaine à long terme.

C'est en particulier ce dernier point qui a été étudié par Monsanto et divers organismes de recherche publics ou non.

3.2 Tests de toxicité sub-chronique

Ciblons l'illustration des études sur l'éventuelle toxicité à moyen terme de ce maïs OGM. Le protocole expérimental est relativement simple à mettre en œuvre mais nécessite des moyens conséquents et bien dimensionnés. Celui mené par Monsanto peut se résumer de la façon suivante :

1. Plusieurs groupes de rats sont nourris pendant 13 semaines :
 - des groupes *tests* (effectif n) sont nourris avec le maïs MON810,

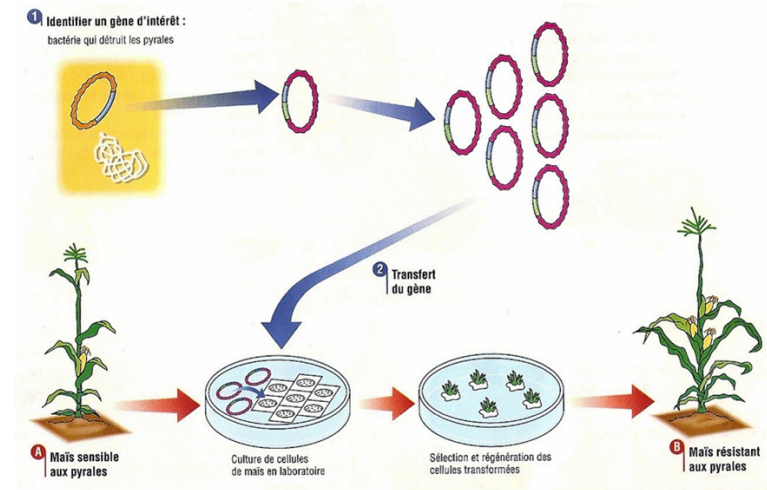


FIGURE 6 – Processus de transgénèse.

- des groupes *témoins* (même effectif n) avec un équivalent sans OGM.
2. Différentes mesures (poids des rats, de chacun des organes, paramètres biochimiques...) sont réalisées au cours de l'expérience et à la fin de celle-ci après avoir sacrifié les rats.

Question biologique

La nourriture des rats (OGM ou pas) a-t-elle un effet sur le développement de ceux-ci ? Ou, pour prendre un exemple, le maïs MON810 affecte-t-il le poids du foie de ces rats ?

On note μ_1 le poids moyen du foie des rats d'un groupe *témoin*, et μ_2 celui des rats d'un groupe *test*. Sans discuter le bien-fondé de ces hypothèses car celles-ci ne sont pas "critiques" dans cette étude, les variables de poids des rats sont supposées gaussiennes ou normales et les variances sont supposées égales à une même valeur σ^2 .

3.3 Test statistique

La question biologique revient alors à tester l'égalité de ces moyennes. Si celles-ci sont jugées *significativement* différentes, nous concluons à une influence de l'alimentation sur le développement des rats et à l'absence d'influence si l'égalité des moyennes est acceptée. Dans le cas considéré et avec les hypothèses mentionnées (normalité des variables de poids, égalité des variances qu'il est d'usage de vérifier préalablement par un test de Fisher), le **test de Student** est un usage courant.

Test de Student d'égalité des moyennes

Ce test s'intéresse à l'hypothèse dite $H_0 : \mu_1 = \mu_2$ ou $\delta = \mu_1 - \mu_2 = 0$ et à l'hypothèse alternative $H_1 : \mu_1 \neq \mu_2$.

Attention, la façon de poser l'hypothèse H_0 n'est pas "neutre", elle présume implicitement l'"innocence" du maïs et c'est aux données, aux expériences, d'apporter la preuve du contraire, le rejet de l'hypothèse H_0 au profit de H_1 . En effet, le risque (noté α) de se tromper ou de rejeter à tort l'hypothèse H_0 au profit de l'hypothèse alternative H_1 et celui (noté β) d'accepter à tort cette même hypothèse ne sont pas identiques. Ceci se résume de façon générale dans le tableau ci-dessous :

	Accepter H_0	Rejeter H_0
H_0 vraie	Probabilité $1 - \alpha$	Probabilité α
H_1 vraie	Probabilité β	Probabilité $1 - \beta$

- Si H_0 est vraie, la probabilité $(1 - \alpha)$ est celle d'une décision correcte et α est appelé risque ou erreur de *première espèce*.
- Si H_1 est vraie, la probabilité β est le risque ou erreur de *deuxième espèce* d'accepter à tort H_0 tandis que la probabilité $(1 - \beta)$ est celle d'une décision correcte, elle est appelée *puissance* du test.

Risque α de première espèce

Le risque ou seuil α est celui qui est plus particulièrement contrôlé par l'expérience. Une fois observé un échantillon, les moyennes μ_1 et μ_2 sont respectivement estimées par les moyennes empiriques (calculées sur l'échantillon) $\widehat{\mu}_1$ et $\widehat{\mu}_2$. La variance σ^2 est également estimée par la variance empirique $\widehat{\sigma}^2$. La question se pose sous la forme suivante :

L'écart aléatoire (car dépendant de l'échantillon) observé entre $\widehat{\mu}_1$ et $\widehat{\mu}_2$ est-il significativement grand ou peut-il être considéré comme dû au seul hasard ?

Sous l'hypothèse que H_0 est vérifiée la quantité

$$t = \sqrt{\frac{n}{2}} \frac{\widehat{\mu}_1 - \widehat{\mu}_2}{\widehat{\sigma}}$$

appelée *statistique de test*, est la réalisation d'une variable aléatoire qui suit une loi dite de Student à $(2n - 2)$ degrés de liberté dont la fonction de densité est représentée figure 7. La zone rouge représente $\alpha = 5\%$ de la surface ou probabilité tandis que le reste représente $(1 - \alpha) = 95\%$. Si la statistique de test observée sur l'échantillon est, en valeur absolue, suffisamment grande pour "tomber" dans la zone rouge de cette densité, la différence entre les moyennes est jugée significativement différente de 0 ; l'hypothèse H_0 est *rejetée*. Ou encore, la probabilité de se tromper en rejetant H_0 est inférieure au seuil préalablement déterminé α , et le test est jugé significatif.

Puissance de test

Une autre question concerne le contrôle de l'erreur β d'accepter à tort l'hypothèse H_0 (innocuité du maïs MON810) alors que c'est H_1 qui est vraie (influence de la consommation de MON810). Cette erreur (figure 8) ou risque de 2ème espèce dépend de plusieurs paramètres :

- du choix du seuil α (en général $\alpha = 5\%$),
- de l'écart réel $\mu_1 - \mu_2$ (hypothèse H_1 vérifiée),
- de la dispersion des valeurs ou variance σ^2 ,
- et aussi de la taille n des groupes ou échantillons.

Plus précisément, il est possible de montrer que plus n croît, plus β décroît. Ou encore, plus n est grand et plus il est possible de détecter significativement une différence réelle mais faible entre les moyennes μ_1 et μ_2 .

Voilà toute la difficulté du débat autour d'un test statistique lié à une expérimentation. Si la taille n de l'échantillon est suffisamment grande, il est possible d'obtenir un test statistiquement significatif alors que la différence observée n'a pas de significativité biologique. Réciproquement, si la taille n est trop petite, la puissance de test devient insuffisante et grand est le risque β d'accepter à tort H_0 .

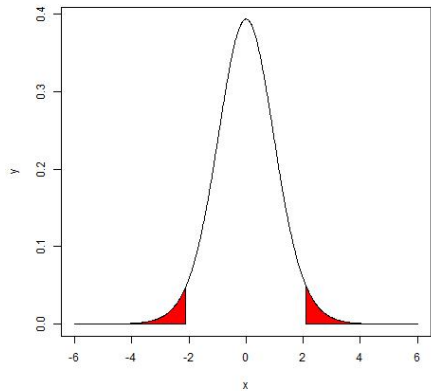


FIGURE 7 – Si H_0 vraie : Densité d'une loi de Student (18 ddl) et valeurs seuils ou quantiles de cette loi déterminant un risque $\alpha = 5\%$ de première espèce en rouge de rejeter à tort H_0 .

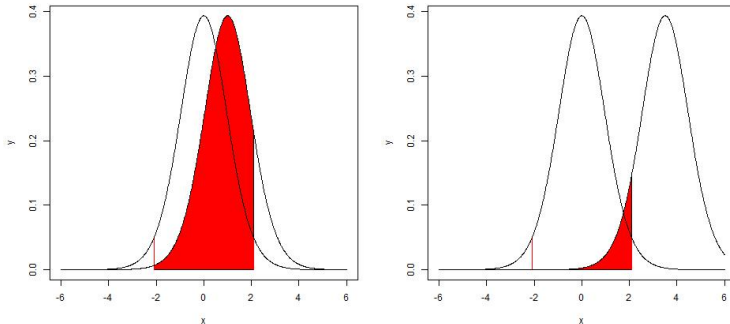


FIGURE 8 – Si H_1 vraie : Erreur de 2^{ème} espèce (en rouge) de rejeter à tort H_1 pour un test de Student (18 ddl) de seuil $\alpha = 5\%$; à gauche lorsque les moyennes sont peu séparées (1 écart-type) et à droite lorsqu'elle sont bien séparées (3,5 écarts-types).



FIGURE 9 – Devise Shadok : la vérité est qu'il n'y a pas de vérité (y compris celle-ci).

Dans le cas du MON810, la principale critique énoncée est justement que le test a été posé (H_0) en considérant implicitement l'innocuité de ce maïs alors que la taille des groupes ($n = 10!$) était bien trop faible pour assurer une puissance de test raisonnable. Seules de grosses différences auraient pu être statistiquement détectées.

Tests multiples

Dans l'autre camp, le Criigen (Comité de recherche et d'information indépendantes sur le génie génétique) a proposé une autre étude selon un protocole similaire mais sur une durée (5 ans et 14 semaines) beaucoup plus longue et avec l'observation de très nombreuses variables (500). L'étude qui a été publiée dans l'*Int. Jour. of Biological Sciences* souffre également d'une insuffisance. En effet, ayant calculé 500 statistiques de test sur le même échantillon, on s'attend à ce qu'en moyenne, $5 \times 500/100$ conduisent à des tests "statistiquement significatifs" alors que ce sont très probablement des faux positifs dus au seul fait du hasard. En effet, lorsqu'un correctif classique de tests multiples (Bonferroni) est appliqué, aucune des différences observées ne restent significatives.

Conclusion sur le MON810

Les exemples présentés illustrent des utilisations injustifiées de pratiques statistiques :

- le test est posé de façon favorable (hypothèse H_0) au maïs MON810 en contradiction avec le principe de précaution qui voudrait que les données de l'expérience apportent la preuve⁴ de l'innocuité,
- risque de 2ème espèce inconnu d'accepter à tort H_0 alors que la très faible taille de l'échantillon rend très élevé ce risque ;
- mise en œuvre incontrôlée de tests multiples.

En plus de ces problèmes, [Marc Lavielle](#) souligne d'autres manquements dans l'utilisation de données de référence ou historiques, la gestion des données atypiques, l'analyse de données répétées (courbes de poids longitudinales) ou des approximations gaussiennes abusives. Ils justifient l'article du *Monde* du 10/02/2010 qui titre : *OGM : opposants et promoteurs renvoyés dos à dos*. Les études statistiques publiées sont insuffisantes pour permettre de conclure, avec une confiance raisonnable, dans un sens ou dans l'autre. Le chercheur doit avoir l'humilité de reconnaître qu'il sait juste qu'il ne sait pas (figure 9).

3.4 D'autres abus

L'ingénierie financière

Au moment de la crise financière de 2008 provoquée par l'éclatement de la bulle immobilière aux États-Unis (*subprimes*), Michel Rocard dit dans une interview du *Monde* (01/11/2008) :

Des professeurs de maths enseignent à leurs étudiants comment faire des coups boursiers. Ce qu'ils font relève, sans qu'ils le sachent, du crime contre l'humanité.

La violence de cette accusation est-elle justifiée ? Ce type de débat soulève la question classique du discernement entre l'invention ou la création de l'outil d'une part et l'usage qui en est fait d'autre part : bombe atomique, internet... Prenons l'exemple du modèle dit de Black & Scholes (1973) reposant sur l'intégration d'équations différentielles stochastiques (formule d'Itô, 1940). Ce modèle permet le calcul de la valeur théorique du prix d'un produit financier (une option) à partir du prix de l'action sous-jacente.

Ces travaux ont valu en 1997 à Robert Merton et Myron Scholes le prix ("Nobel d'Économie") de la banque de Suède en honneur à Alfred Nobel. Ce modèle et ses extensions ont permis le développement considérable du marché des options sur les places financières mais ils nécessitent explicitement des hypothèses très simplificatrices dont certaines relatives à la *volatilité* (variance), la normalité (mouvement brownien), la rationalité des investisseurs... ne sont pas vérifiées. Dans ses livres sur l'approche *fractale* des marchés, Benoît Mandelbrot (2009, [5]) dénonce ces simplifications abusives qui rendent les modèles, certes calculables, mais déconnectés des réalités financières. La succession des crash boursiers semblerait bien justifier ces critiques sur des modèles qui sous-estiment gravement les risques.

Dans la séquence d'activités suivantes, à partir de quel moment une responsabilité "morale", si ce n'est pénale, est-elle engagée ?

1. Travailler sur les propriétés des équations différentielles stochastiques,
2. construire de nouveaux modèles plus sophistiqués d'ingénierie financière,
3. utiliser des modèles connus faux (violation des hypothèses simplificatrices de calcul) en relativisant leur pertinence,
4. écrire des programmes de *trading* automatique (*algotrading*) appliquant ces modèles sans se soucier de leur pertinence,
5. construire de nouveaux instruments financiers en les habillant d'équations sophistiquées pour vendre leur "scientificité",
6. ...

Les sondages

Depuis de nombreuses années, les statisticiens regroupés au sein de la [Société Française de Statistique](#) et dans d'autres associations professionnelles (administrateurs INSEE) tentent de faire évoluer la loi sur la publication des sondages. Ce *lobbying* a permis, par le passé, d'obliger les instituts à préciser le nombre de personnes enquêtées et le mode d'échantillonnage. Il serait en effet très fallacieux de chercher à faire croire qu'une enquête auprès de lecteurs d'un journal, de visiteurs d'un site internet ou au mieux d'abonnés du téléphone acceptant de répondre... à la même précision qu'un sondage sur un échantillon aléatoire tiré dans une liste électorale.

Les médias n'ont toujours pas obligation de publier conjointement aux résultats d'un sondage, les modes de redressement et les marges d'incertitude.

4. Mettre en œuvre un test dit de comparaison dans lequel l'hypothèse H_0 pose $\mu_1 \neq \mu_2$

En mars 2011, deux Sénateurs (Hugues Portelli UMP et Jean-Pierre Sueur PS) ont déposé une proposition de loi visant à mieux encadrer les sondages politiques. Le débat qui s'en est suivi a été pour le moins tendancieux (*Le Monde* du 17/03/2011). La loi a été votée à l'unanimité par le Sénat mais jamais inscrite à l'ordre du jour de l'Assemblée Nationale. Il est certes difficile de préciser une marge d'incertitude lors d'un échantillonnage par la méthode des quotas mais il serait instructif d'indiquer *a minima* celle qui serait obtenue par un sondage aléatoire simple sur le même effectif. Pourquoi ne pas ajouter à la publication d'un sondage une mention rédigée sous la forme "*étant obtenus à partir d'un échantillon de taille 1000, les résultats sont affectés d'une marge d'incertitude ; dans les conditions optimales d'un sondage aléatoire de même taille, la marge d'incertitude est de 3 points, c'est-à-dire, la proportion exprimée en pourcentage plus ou moins 3%*". La plupart des sites détaillant explicitement les sondages annoncent une marge d'incertitude entre 2 et 3%, sans doute optimiste, mais restent très discrets sur les pratiques de redressement jugées relevant du secret de "fabrication". Certes le discours journalistique qui commente longuement des évolutions d'1 ou 2 points d'intentions de vote en serait fortement relativisé.

4 Codes de déontologie statistique

Pour réagir aux dérives constatées dans l'usage de leur discipline, les statisticiens professionnels se sont mobilisés pour rédiger des codes de bonne pratique. Une des principales motivations de ce travail est aussi la structuration d'un collectif afin que chaque statisticien puisse mieux se protéger des pressions auxquelles il pourrait être soumis de la part de sa hiérarchie, de son commanditaire, du pouvoir politique. Dans ce contexte, il faudrait en fait distinguer la Statistique publique⁵ dont la crainte principale est d'être inféodée au pouvoir politique, la Statistique académique en relation avec la recherche scientifique et enfin celle utilisée dans la sphère économique, industrielle ou commerciale.

Voici une sélection de quelques éléments d'informations.

5. Voir à ce propos le [compte-rendu](#) du séminaire organisé par la Société Française de Statistique sur l'Indépendance de la Statistique publique le 15/12/2008

4.1 En france

En France, l'association des administrateurs de l'INSEE (AIS) et l'association des statisticiens économistes anciens élèves de l'ENSAE (ASEC) ont rédigé un [code déontologique](#) accessible sur le site du syndicat CGT de l'INSEE et introduit par Padiou (1991, [7]). Ce texte a été rédigé pour répondre aux pressions exercées par le pouvoir politique sur la définition des chiffres très sensibles du chômage dans les années soixante-dix. C'est pour défendre les positions de ce texte que ce syndicat s'est mobilisé contre les allégations de Claude Guéant (cf. section 1.1) sur la scolarité des enfants d'immigrés.

L'[Autorité de la Statistique publique](#) veille au respect des principes d'indépendance professionnelle, d'impartialité, d'objectivité, de pertinence et de qualité dans son élaboration et sa diffusion.

De son côté, l'association des épidémiologistes de langue française propose des [Recommandations de Déontologie et bonnes pratiques en Épidémiologie](#) (2007) dans un document de 37 pages.

4.2 Niveau européen

[Eurostat](#) est une direction générale de la Commission Européenne chargée de la mise en œuvre de la politique communautaire dans le domaine statistique et de fournir une information statistique de qualité.

L'Europe est également dotée d'un

- Conseil consultatif européen pour la gouvernance statistique créé par le Parlement européen et le Conseil en 2008 chargé d'évaluer, en toute indépendance, le Système statistique européen ;
- Comité consultatif européen de la Statistique ;
- [Code de bonnes pratiques de la statistique européenne](#) adopté par le Comité du programme statistique le 24 février 2005 et promulgué dans la recommandation de la Commission du 25 mai 2005 sur l'indépendance, l'intégrité et la responsabilité des services statistiques nationaux et communautaires.

L'existence de tous ces comités, conseils et directions montrent les enjeux des statistiques au sein de la gestion de la Communauté européenne.

4.3 Niveau international

Pour les acteurs économiques, la Chambre de Commerce Internationale (ICC), organisation mondiale des entreprises, s'est associée avec l'organisme ESOMAR qui regroupe les fournisseurs d'études de marchés et enquêtes d'opinion. Ils se rallient au [Code International ICC/ESOMAR de pratiques loyales en matière d'études de marché et d'opinion](#) (1995).

Pour les acteurs académiques et publiques, l'Institut International de Statistique a remis à jour sa [Déclaration d'Éthique professionnelle](#) (2010).

5 Conclusion

Bien entendu, les questions de déontologie statistique s'inscrivent dans le cadre plus large d'une éthique de la recherche scientifique. Cette réflexion difficile devient plus active ces dernières années à la suite de quelques scandales retentissants comme celui de l'affaire Hwang Woo-suk de Corée du Sud qui à la particularité de violer beaucoup de principes éthiques ; falsification de résultats et manquements bioéthiques graves. La recherche sur les cellules souches impliquant des embryons humains, le clonage du vivant, soulèvent en effet des questions politiques et éthiques nettement plus complexes que celles du seul bon usage de la Statistique. Notons simplement que cette dernière problématique a émergé chronologiquement plus tôt par le rôle très spécifique que cette discipline joue dans les sciences expérimentales comme support de preuve de la significativité d'un test d'hypothèse.

Ces questions éthiques sont évidemment d'autant plus pressantes que la mondialisation fait éclater les frontières de la recherche dans une compétition internationale exacerbée et des concentrations de moyens financiers jamais atteints. Des conférences sont organisées spécifiquement sur la mise en place d'un code d'éthique mondial de la recherche scientifique (Anderson, 2011 [2]) :

- A Amsterdam en 2011 : ALLEA. European Code for Conduct on Research Integrity. European Federation of National Academies of Sciences and Humanities
- A Singapour en 2010 : Anon. Second World Conference on Research Integrity

Derrière cette réflexion s'en cache une autre sur les modes d'évaluation de la

recherche et donc sur son *financement* et la promotion des chercheurs. L'évaluation de la recherche peut-elle reposer sur de seuls *indicateurs bibliométriques* ? Conduisant au seul objectif de publier (*publish or perish*) parfois en contradiction ou en conflit avec celui de l'avancement des connaissances pour le bien commun.

Références

- [1] Claude Allègre, *L'imposture climatique ou la fausse écologie*, Plon, 2010.
- [2] Melissa Anderson, *Vers un code d'éthique mondial de la recherche scientifique*, 2011, [Accès site](#).
- [3] Daniele Fanelli, *How Many Scientists Fabricate and Falsify Research ? A Systematic Review and Meta-Analysis of Survey Data*, PLoS ONE **4** (2009), n° 5, e5738, <http://dx.doi.org/10.1371/journal.pone.0005738>.
- [4] Darrel Huff, *How to lie with Statistics*, Norton & Company, 1954.
- [5] Benoît Mandelbrot, *Une approche fractale des marchés : risquer, perdre et gagner*, deuxième éd., Odile Jacob, 2009.
- [6] N. Oreskes et M. Conway, *Les Marchands de doute, ou comment une poignée de scientifiques ont masqué la vérité sur des enjeux de société tels que le tabagisme et le réchauffement climatique*, Le Pommier, 2012.
- [7] René Padieu, *La déontologie des statisticiens*, Sociétés Contemporaines **7** (1991), 35–61.