

TECHNOSUP

Les FILIÈRES TECHNOLOGIQUES des ENSEIGNEMENTS SUPÉRIEURS

STATISTIQUE

Visualisation et analyse statistique de données

Concepts de base, tableur,
statistiques à 1 et 2 variables

Renaud SEIGNEURIC



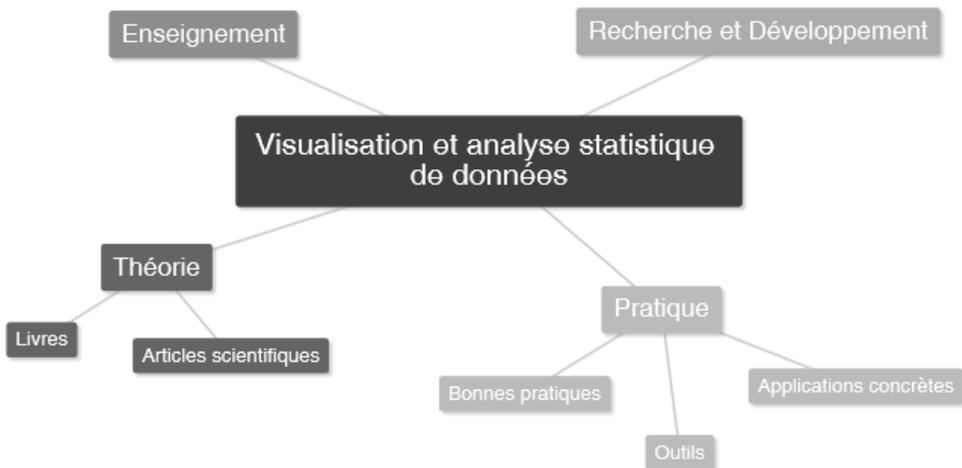
Chapitre I.

Introduction

1. Pourquoi ce livre ?

Les statistiques sont enseignées depuis longtemps dans quasiment toutes les filières post-bac (ex : comptabilité mais aussi médecine, psychologie, etc.) et il y a déjà de nombreux livres sur le sujet. Cependant, pour une majorité d'élèves et d'étudiants, elles restent une discipline difficile à apprendre, comprendre, appliquer et communiquer, même concernant les statistiques de base^{1,2}.

Pourquoi un livre de plus ?



Contenant plusieurs originalités, ce livre :

- Est une synthèse de plus de **20 livres** de statistiques (et de nombreux articles scientifiques).
- Est une synthèse de près de **20 ans** d'enseignement en mathématiques et statistiques de la 6^{ème} au doctorat, dans de nombreuses filières différentes (en collège, lycée, enseignement technique et professionnel, en IUT, étudiants de pharmacie, de biologie, au doctorat, etc.)
- Est écrit par un enseignant qui n'était *pas* un bon élève en mathématiques ou en statistiques (loin de là ...) mais qui est finalement devenu enseignant et chercheur dans ces disciplines, qui a publié un certain nombre d'articles dont des analyses statistiques dans des journaux scientifiques de niveau international, qui a notamment conçu et programmé un logiciel d'analyse de données (Nanodigit^{®3}, utilisé pour de grandes entreprises internationales) et qui a été chercheur invité à la *State University of New York* (SUNY aux Etats-Unis) dans un département de recherche spécialisé en analyse de données
- Contient des **applications pratiques avec un tableur**, mais aussi d'autres outils qui rendent les statistiques à la fois plus concrètes et plus interactives. Un constat relativement fréquent est que le temps consacré à la découverte et à l'apprentissage d'un outil informatique (tableur, éditeur de texte, mais aussi langage de programmation Scratch ou Python par exemple) est généralement trop court. Les **exemples** sont donc ici **détaillés étape par étape**. Par contre, nous n'utiliserons **pas de langage de programmation** comme Python ou R. S'ils

sont très puissants et utiles, ils ajoutent une couche de complexité à la discipline étudiée. De plus, l'apprentissage de langages informatiques est souvent relativement longue, et finalement nous butons souvent sur des problèmes plus techniques (informatiques) que statistiques. Dans cet ouvrage, nous préférons donc l'utilisation d'un tableur, qui est adaptée pour balayer un grand nombre d'applications fréquentes.

- Propose des **bonnes pratiques**. A part quelques exceptions remarquables, les enseignants ne sont pas nécessairement des utilisateurs intensifs des outils qu'ils enseignent et ils n'ont pas bénéficiés d'une formation avancée systématique. Ils ne sont donc pas forcément équipés pour mettre en place et enseigner des bonnes pratiques. Il se trouve que de par mon parcours souvent qualifié d'« atypique », j'ai été amené à en développer, et à les affiner. J'essaie de les transmettre au plus grand nombre dans cet ouvrage.
- Permet de **couvrir les programmes académiques** présents (et peut-être futurs). En effet, il couvre de nombreuses notions nécessaires pour comprendre et surtout appliquer les statistiques, que vous soyez étudiant ou bien élève, enseignant voire autodidacte.
- **Connecte recherche et enseignement**. Dans mon premier livre, « Pratique de la science »⁴, je constatais que les fruits de la recherche sont nombreux mais finalement peu accessibles. En effet, la plupart des articles scientifiques sont publiés en anglais, dans des journaux dédiés qui sont souvent payants. Globalement, les fruits de ces travaux sont utilisés par les chercheurs, les post-doctorants et les doctorants. Pour les étudiants en master ou en licence, cela est plus rare. Dans ce livre, j'essaie très humblement d'aller chercher *quelques-uns* de ces nombreux fruits de la recherche, de les traduire et de les proposer aux apprenants et enseignants qui n'ont ni le temps ni les incitations pour le faire. J'ai donc inséré dans ce livre **des résultats de recherche** (ex : théorème de Bayes, carré-proba) qui feront peut-être partie des programmes à venir. Ces résultats me paraissent à la fois **intéressants** et **éclairants**, parfois **puissants** : n'attendons donc pas plus pour les découvrir et peut-être les utiliser.

Globalement, ce petit livre vise à rendre les concepts statistiques plus simples et plus concrets¹.

Nos sociétés produisent des quantités de données comme jamais auparavant. Cependant, il me semble que les statistiques occupent une petite part des heures consacrées aux mathématiques en collège et lycée. De fait, des notions de statistiques sont abordées -niveau après niveau- mais il est **difficile** pour les apprenants **de tisser des liens entre ces notions vues de façon éparse**, et du coup, d'avoir une vue d'ensemble cohérente, et une boîte à outils fonctionnelle. De plus, il peut arriver que les enseignants, qui ont eux aussi suivi relativement peu d'heures de statistiques, soient plus ou moins à l'aise avec ces notions lorsqu'ils doivent les transmettre.

Les statistiques et l'analyse de données sont des compétences très recherchées sur le marché de l'emploi. Il paraîtrait judicieux de **former à un tableur** le plus rapidement possible le plus grand nombre de personnes : les étudiants mais aussi les élèves (dès le collège), et le personnel (en formation continue par exemple), pour en connaître les **bonnes pratiques**, les fonctions de base, puis les fonctions avancées (quelques-unes figurent dans les annexes).

Les enseignants pourraient guider les apprenants pour **générer des données avec méthode** (voir les étapes proposées dans ce livre), sur un sujet qui les intéresse voire les passionne (par binôme ou par groupe). Rien ne vaut des données que nous avons générées ou collectées sur un sujet

¹ Comme tout livre, il est imparfait. Merci de me faire part d'améliorations : renaud.seigneuric@gmail.com

qui nous tient à cœur (statistiques de joueurs sportifs pour certains, réseaux sociaux pour d'autres, musique, cinéma, fonctionnement d'un établissement ou d'une organisation, etc.). Réfléchir sur les données, essayer de leur donner du sens, de les représenter, les interpréter, d'essayer de prendre une décision à partir de ces données deviennent alors un moteur pour apprendre.

Les **statistiques et l'analyse de données** sont des **disciplines transversales** que l'on peut utiliser dans un très grand nombre de domaines. Elles permettent également de manipuler des nombres (arithmétique), quelques équations (algèbre), des logiciels (informatique), et comportent, comme la géométrie, une **partie visuelle** non-négligeable. Elles **balayent** donc **un éventail de notions mathématiques assez complet**.

2. Comment utiliser ce livre ?

Comme de très nombreux ouvrages, le mieux est probablement de lire les chapitres dans l'ordre proposé, sauf lorsque l'on souhaite étudier un point précis.

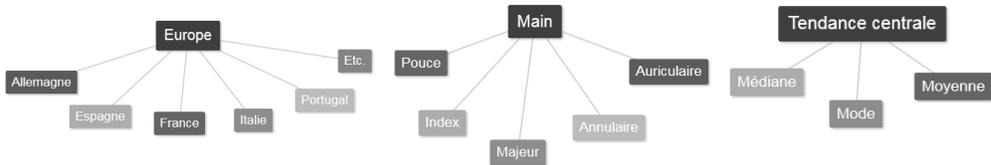
Certains chapitres au début de cet ouvrage ² ou certaines annexes à la fin peuvent paraître hors sujet car non purement dédiés aux statistiques. Chacun a son propre cursus, les notions qui ont été bien comprises, celles plus floues, mal maîtrisées, oubliées ou non vues. Il est donc recommandé de regarder ces chapitres et annexes. Gardons à l'esprit que **les choses que nous pensons savoir** (mais qu'en fait nous ne savons pas ou de travers) **nous font faire de nombreuses erreurs**. En règle général, ne pas hésiter à revoir les règles des priorités de calculs (en annexe) qui sont la cause de nombreuses erreurs.

2.1. Les conventions utilisées

Afin de clarifier au maximum, des photos, illustrations ou schémas sont insérés dans le texte. Pour décomposer certaines notions, j'ai utilisé des listes ordonnées $\begin{matrix} \text{---} \\ \text{---} \\ \text{---} \end{matrix}$ ou bien des listes non-ordonnées $\begin{matrix} \text{---} \\ \text{---} \\ \text{---} \end{matrix}$.

2.1.1. Liste non-ordonnée

Pour les listes non-ordonnées simples (exemple : à 2 niveaux de hiérarchie), nous utiliserons un diagramme qui décompose un ensemble en plusieurs de ses parties, par exemple :



L'Europe est composée de plusieurs éléments -les pays- qui sont : l'Allemagne, l'Espagne, la France, l'Italie, le Portugal, etc.

La main est composée de 5 doigts : pouce, index, majeur, annulaire et auriculaire.

La tendance centrale est principalement décrite par 3 notions : la médiane, le mode et la moyenne.

² L'image de la couverture est une illustration réalisée à partir des statistiques du texte du chapitre sur les concepts de base.

2.1.2. Liste ordonnée

Pour les listes ordonnées, nous utiliserons ce type d'illustration pour rappeler qu'il y a un ordre et qu'il est important de le respecter :



3. Apprendre à apprendre

« Vivre comme si on devait mourir demain, apprendre comme si on devait vivre une éternité »
citation anonyme

« La seule chose que je sais, c'est que je ne sais rien » Socrate, philosophe grec.

Voici quelques pistes pour apporter aux apprenants des éléments afin d'essayer de (mieux) apprendre et travailler.

En effet, les choses évoluent sans cesse (et de plus en plus vite). Tout au long de notre vie personnelle et professionnelle, nous sommes donc confrontés à apprendre pour nous adapter. Pour retenir et pouvoir appliquer, il faut apprendre, comprendre et appliquer (ce qui aide à retenir). Pourtant, souvent pris dans une course (finir le programme, passer cette épreuve, obtenir cet examen, etc.) nous consacrons peu de temps à « Apprendre à apprendre ». Quoi qu'il en soit, il est fondamental de :

1	<p>Donner un maximum de sens à ce que l'on apprend : c'est être capable d'expliquer comment cela fonctionne, et si possible pourquoi.</p>	<p>Outils : décomposer les mécanismes, Questions : pourquoi, comment, quand peut-on (doit-on) s'en servir ?</p>
---	--	---

2	<p>Revenir le plus souvent possible sur la notion : c'est essentiel pour retenir de façon durable. On peut imaginer le cerveau comme un terrain en friche (où nous essayons finalement de connecter des neurones entre eux). Apprendre représente un effort, comme de débroussailler pour créer un chemin. Mais après, plus ce chemin sera fréquenté et entretenu (plus on fera appel à cette notion) plus il sera facile de l'utiliser. Sinon, les mauvaises herbes auront vite fait de reprendre le dessus, et il ne sera plus possible de retrouver, le chemin. Tout sera à recommencer...</p>	<p>Outils : faire des fiches de synthèse (ex : par chapitre, par notion...), des formulaires, savoir expliquer aux autres, en parler avec eux.</p>
---	--	---

3	<p>Tisser des liens : relier ce que l'on vient d'apprendre avec ce que l'on sait déjà par un raisonnement logique. Ceci permet de consolider les connaissances (car on remplit les points 1 et 2). C'est aussi en voir l'utilité, les applications, les conséquences.</p>	<p>Outils : rechercher des applications du cours dans la vie de tous les jours et observer le monde qui nous entoure.</p> <p>Questions : Qui, quand, où ?</p>
---	--	---

Pour construire son propre édifice intellectuel, il faut des cours (les briques) et des exercices (le ciment qui fait le lien entre les différents cours). Il est humain d'avoir quelques lacunes, mais tout au long de son apprentissage il faut veiller à en avoir le moins possible, et de colmater aussitôt la moindre brèche (se remettre à niveau). Chaque classe du collège puis du lycée, chacune des premières années du supérieur constitue un étage qui sert de base pour l'année qui suit. **Avant** d'essayer de bâtir l'étage suivant, si les bases ne sont pas assez solides, il est judicieux de consacrer du temps à reprendre ou simplement réviser les bases nécessaires (voir donc les annexes).

Pour travailler efficacement, il faut être suffisamment concentré. **La Loi de Pareto** nous dit que 20% de temps de travail efficace vaut mieux que 80% de temps à papillonner. On préférera donc travailler en plusieurs séquences courtes et intenses (un peu comme des sprints) que passer beaucoup de temps à travailler lentement. Pour être efficace, on s'obligera donc à travailler en temps limité. Par exemple, plusieurs cycles de 25 minutes de travail chacun suivi de 5 minutes de pause permettent de travailler efficacement sans trop ressentir les effets de la fatigue.

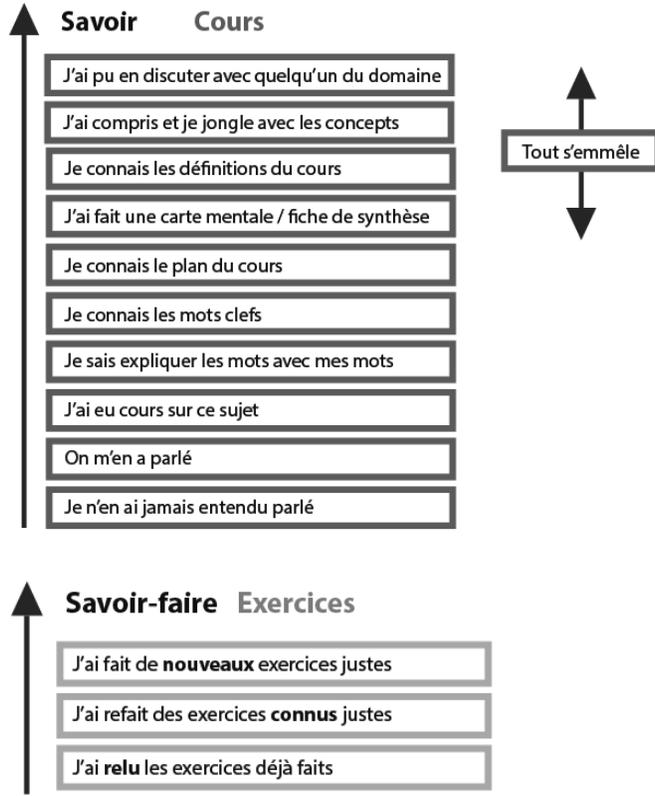
Pour progresser, il est très important d'éviter de se bloquer avec des attitudes négatives et dévalorisantes comme « Je (ne) sais pas (faire) ! », « (De toutes façons) je (n') y comprends rien ! » etc. pour se concentrer sur les aspects positifs, même s'ils ne sont pas aussi nombreux que l'on souhaiterait. Nous sommes là pour progresser, prenons le taureau par les cornes !

4. Apprentissage d'une nouvelle notion

Chaque chapitre commence par des rappels ou des notions par ordre croissant de difficulté. Il est très important dès le début d'en profiter pour prendre les meilleurs réflexes possibles (et éventuellement combler ses lacunes via les annexes ou d'autres sources) qui permettront de développer les meilleurs automatismes possibles.

Progressivement, la difficulté augmente pour englober ou compléter les notions de départ. Ceci peut provoquer un petit *flottement* (« Tout s'emmêle ») car nos connaissances sont alors un peu bouleversées entre ce que nous savions ou croyions savoir, et ce que nous sommes en train d'apprendre. Pendant cette **phase de transition**, il est important « d'enfoncer le clou » (par les points 1, 2 et 3 vus plus haut) pour ancrer solidement ces nouvelles notions.

Les schémas ci-dessous indiquent plusieurs des étapes typiques, mais ces étapes et/ou leur séquence peuvent varier pour chacun (une version plus détaillée des « savoir-faire » se trouve dans l'annexe du même nom : « Apprentissage d'une nouvelle notion »). En fonction des chapitres abordés notamment, le sentiment que « **Tout s'emmêle** » peut s'étirer davantage, ou non. Il est important de se donner les moyens de le dépasser, par exemple en utilisant d'autres sources, en en parlant à quelqu'un, en demandant de l'aide, etc.



Apprendre demande un effort, mais tout ce qui est **nouveau** n'est **pas forcément difficile** ! Souvent, les apprenants se plaignent que c'est **difficile** alors que c'est surtout **nouveau** (et avec un peu de recul, pas si difficile). Peu importe d'être plus ou moins doué, **l'essentiel est la motivation**. Il faut veiller à apprendre de la façon la plus intelligente possible, donc préférer la compréhension logique et éventuellement des moyens mnémotechniques, mais les plus directs possibles. En effet, le temps efface très vite ce que l'on n'a pas vraiment compris. L'acquisition des nouvelles notions est une étape, qui constituera un point de départ indispensable pour l'apprentissage d'une nouvelle notion. Nous chercherons à réinvestir et à transférer nos connaissances vers d'autres matières et d'autres situations (vie quotidienne, etc.) pour essayer de les ancrer efficacement et durablement.

5. Déroulement d'une séquence

L'enseignant conçoit des séquences d'interventions pour transmettre un ensemble de notions combinant **savoir** et **savoir-faire** à ses apprenants. Voici un ensemble d'étapes recommandées pour maximiser son apprentissage avant chaque nouveau chapitre ou avant chaque intervention :

1. Avant la séance

Lire son livre et/ou ce livre, effectuer des exercices préparatoires, combler ses lacunes : il faut essayer au maximum de préparer le terrain pour la nouvelle leçon en lisant le chapitre du livre correspondant.

2. Lors de la séance

Ecouter en cours est bien sûr nécessaire (et c'est autant de temps gagné par rapport au travail personnel), mais cela ne suffit pas. Il faut à tout prix être acteur en réfléchissant et en (se) posant des questions. Il faut en comprendre le plus possible. Ne pas se contenter d'un « J'ai rien compris », mais repérer la notion qui fait défaut, ou le moment où l'on ne suit plus ou plus très bien.

3. Travail personnel

Intégrer un ensemble de notions et préparer l'évaluation correspondante se prépare bien à l'avance (à chaque cours, à chaque session de travail personnel). Ceci nécessite plusieurs aspects :

- Apprendre le cours : le jour même, sans regarder le cours, sur un brouillon et en temps limité (5-10 min) : écrire ce que l'on pense avoir retenu du cours. Toujours sans regarder la leçon, relire son brouillon et essayer de se corriger. Entourer ce qui est certainement juste, entourer et barrer ce qui est certainement faux. Qu'est-ce que j'ai oublié ? Cacher le brouillon, ouvrir le cours et/ou ce livre, relire attentivement les portions concernées plusieurs fois si nécessaire. Laisser de côté le cours, reprendre le brouillon en essayant de le corriger et de le compléter. Avec le cours ou ce livre ouvert : pour que les briques soient bien en place : on identifie ses erreurs (des lacunes ?) et on cherche à les comprendre. Pour retenir, il faut lire à haute voix, en parler, écrire. On recopiera donc les formules, définitions etc. Penser à bien utiliser le **vocabulaire spécifique** -les mots clefs- du chapitre. Pour vérifier que l'on maîtrise bien une formule, on doit être capable de l'écrire avec d'autres lettres.
- Réaliser des exercices et des applications du cours, se préparer en vue de l'évaluation. Par exemple, si l'épreuve est de 2 heures, s'habituer progressivement à faire des exercices pendant 2 heures également pour éviter de se fatiguer trop vite. Ce n'est pas le temps passé à regarder sa feuille qui fait un exercice bien résolu (voir la **Loi de Pareto**) ! En s'y prenant tôt, on peut démarrer un exercice difficile, se donner un peu de temps pour s'en imprégner et y revenir plus tard. Oser ouvrir son livre, rechercher sur internet. Lorsque l'on obtient un résultat, toujours essayer de le vérifier. Après la correction, on fera une analyse de l'erreur (voir ci-dessous).
- Effectuer l'**analyse de l'erreur** (et son bilan le plus tôt possible, avant une évaluation) : il est humain de se tromper, et **l'erreur est une étape normale voire probablement nécessaire à l'apprentissage. Mais**, pour en tirer profit et progresser, il faut l'identifier, l'analyser et la dépasser (sinon, on pourrait trébucher toujours au même endroit sans même s'en rendre compte ...). Veiller à acquérir de bons automatismes (exemple pour les calculs avec le tableur) et à mettre-à-jour son analyse de l'erreur (voir le chapitre dédié).

- Concevoir des fiches de **synthèse** (définitions, formules, etc.). On n'utilisera des couleurs pour faire ressortir ce qui est essentiel, penser à réaliser des **cartes mentales**. Il est utile de se poser les questions : quel est le plan du cours ? A quoi et quand s'en sert-on ? etc.
 - Concevoir des fiches de **méthode**. Comprendre quand et comment il faut utiliser cette méthode, quand il ne faut *pas* l'utiliser. Voir comment généraliser la démarche (prendre du recul, de la hauteur). Quand nous arrivons au constat que « **c'est toujours pareil** » alors nous sommes en bonne voie pour maîtriser les notions. Finir les révisions en ayant les idées claires.
4. Passer l'évaluation : plusieurs types d'évaluation existent : QCM, devoir, etc. mais des principes sont applicables à tous, dont : lire attentivement la consigne autant de fois qu'il faut (et au moins une !), et jusqu'au bout ! Chercher sur un brouillon (propre et lisible). Qu'est-ce que je dois trouver ? Quels sont mes outils ? Penser à décomposer le problème en sous-problèmes que l'on sait résoudre. Quelles étapes choisir et dans quel ordre ? Essayer de calculer le plus longtemps possible avec le calcul littéral (« avec les lettres ») puis, seulement à la fin, remplacer les lettres par leur valeur (l'application numérique). Recourir au calcul littéral est une difficulté voire un obstacle pour beaucoup d'apprenants mais c'est un outil très puissant. En remplaçant les lettres par leur valeur à la fin (l'application numérique), il est possible d'identifier des erreurs de calculs et ainsi de gagner beaucoup de temps.
 5. Corriger l'évaluation en faisant sa propre analyse de l'erreur (voir « **Analyse de l'erreur** »)
 6. Revenir sur des exercices non compris, (se) poser des questions. Une fois corrigé il peut être nécessaire de revenir sur un exercice qui paraît encore confus, en essayant par exemple de prendre un peu de recul et de comprendre quand il faut utiliser cette méthode.

6. Résoudre un problème de statistiques

Pour résoudre un problème de statistiques, il est nécessaire de lire l'**énoncé** pour identifier ses **composantes** :

