

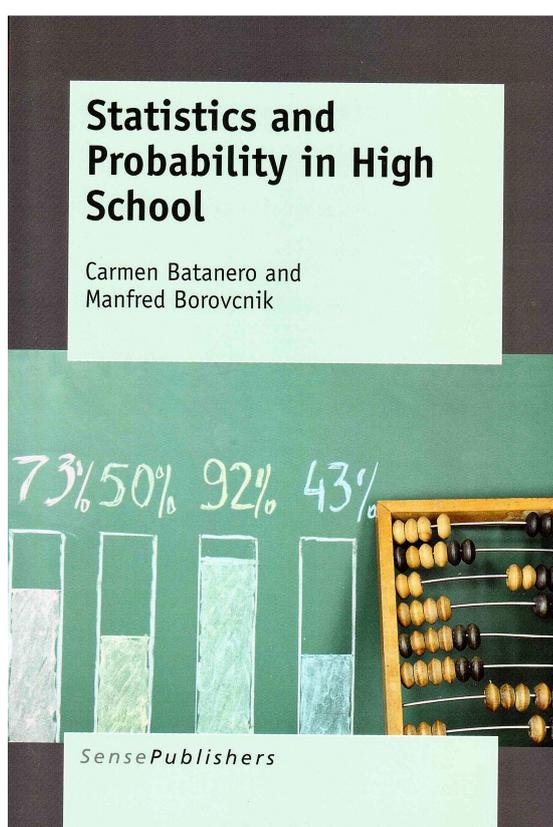
Note de lecture rédigée par Corinne HAHN<sup>1</sup>

## STATISTICS AND PROBABILITY IN HIGH SCHOOL

Carmen BATANERO and Manfred BOROVCNIK

Livre (223 pages)

Édition : Sense Publisher – 2016



La statistique et le calcul de probabilités occupent aujourd’hui une place non négligeable dans les programmes de l’enseignement secondaire, en particulier en ce qui concerne la tranche d’âge de 14 à 18 ans.

Ces enseignements ont fait l’objet de nombreuses recherches dont les résultats ne sont pas toujours facilement accessibles aux enseignants. Pourtant ces recherches permettent de mieux comprendre les difficultés rencontrées par les élèves et peuvent servir d’appui aux enseignants pour construire des activités adaptées.

---

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

*Note de lecture : « Statistics and Probability in High School » (C. Batanero and M. Borovcnik, 2016)*

Carmen Batanero et Manfred Borovcnik sont deux des figures de proue de la recherche en éducation probabiliste et statistique et sont depuis de longues années investis dans l'accompagnement des enseignants.

Cette expérience leur a permis d'écrire un livre particulièrement original et riche. En effet, cet ouvrage combine leur expérience d'enseignant et de chercheur et leur permet d'apporter un regard neuf dans le domaine.

## Les principes pédagogiques

Le premier chapitre détaille les grands principes pédagogiques qui structurent le livre.

Tout d'abord les auteurs proposent de centrer l'enseignement autour de groupes de notions fondamentales. Ils en identifient quatre : l'analyse de données exploratoire (EDA), la modélisation de l'information par le calcul de probabilités, l'exploration et la modélisation de l'association, l'échantillonnage et l'inférence. Ces quatre groupes de notions fondamentales feront l'objet des quatre chapitres suivants du livre.

Si les probabilités et la statistique sont de nature complémentaire et doivent être traitées conjointement au sein d'un programme de mathématiques, les auteurs insistent sur leur spécificité : la généralisation de résultats empiriques. Cette spécificité rend nécessaire la construction d'activités pédagogiques qui prennent en compte le cycle complet d'une étude empirique (PPDAC : Problème/Planification/Données/Analyse/Conclusion) et mettent les élèves en situation de recherche et d'investigation à partir de données réelles ou simulées. Pour cela, ils recommandent d'introduire les notions complexes (comme l'inférence) par des approches informelles avant de développer un enseignement plus formel.

Dans le cadre d'un enseignement du calcul de probabilités et de la statistique, le recours à la technologie est incontournable. Il est important néanmoins d'être conscient de ses apports et ses limites. Les auteurs rappellent que, si la technologie facilite la visualisation des données, permet d'utiliser de grandes bases de données et de simuler des processus, elle conduit aussi parfois à ne pas travailler suffisamment les concepts sous-jacents et ainsi ne pas permettre une réelle compréhension des notions enseignées.

## La structure

Les auteurs proposent pour chaque groupe de notions fondamentales un ensemble cohérent à la fois très pratique et très original.

Ainsi, pour chacun des quatre chapitres, les auteurs commencent par situer le groupe de notions d'un point de vue historique et épistémologique. Puis ils proposent une ou plusieurs activités directement utilisables dans la classe, activités qui permettent de reprendre et expliciter les concepts et les procédures concernés. Après avoir détaillé les différentes étapes de ces activités, les auteurs résument les objectifs d'apprentissage visés par l'enseignement. La dernière partie synthétise les apports des travaux de recherche concernant les notions traitées dans le chapitre, listant les principales difficultés rencontrées par les élèves ainsi que les outils susceptibles de les aider à dépasser ces difficultés. En fin de chapitre, les auteurs proposent des ressources complémentaires : revues qui permettront aux lecteurs qui le

*C. Hahn*

souhaitent d'accéder à un plus large éventail de résultats de recherches, liens internet permettant de télécharger des bases de données et des outils...

Le livre comporte par ailleurs une très large bibliographie ainsi qu'un double index, par auteur et par notion, très pratique pour faciliter la recherche au sein de cet ouvrage.

## **Le contenu**

Le premier chapitre porte sur l'analyse exploratoire de données. Comme l'expliquent les auteurs, l'objectif n'est pas seulement de présenter les outils de statistique descriptive (graphiques, synthèses numériques...). Il s'agit de mettre en œuvre un changement de paradigme : à partir d'analyses simples (sur un groupe de données mais aussi en comparant deux groupes), de détecter des organisations dans les données et d'envisager modélisations et inférences, méthodes qui seront développées dans les trois chapitres qui suivent.

Au travers des activités du second chapitre consacré à la modélisation de l'information avec les probabilités, les auteurs insistent sur l'importance de lier les trois approches des probabilités (fréquentiste, subjectiviste et Laplacienne). Ils abordent ensuite les notions de probabilité conditionnelle et composée, et enfin ouvrent vers les modèles de distribution applicables à des situations standards (modèles binomial et normal) ainsi qu'aux deux théorèmes fondamentaux (loi des grands nombres et théorème central limite) qui permettent de construire un pont entre probabilités et statistique.

Les auteurs développent l'idée que les probabilités permettent d'aborder deux types de situations : celles où l'incertitude est inhérente à la situation (dans le cas d'une loterie par exemple) et celles où l'incertitude est introduite afin de mieux appréhender et généraliser la situation. Ce second type de situation est étudié dans le quatrième et dernier chapitre qui explore les notions d'intervalle de confiance et de test d'hypothèse dans le cas d'une proportion (le cas de la moyenne est également brièvement abordé). On sait à quel point ces notions sont difficiles, quel que soit le public. Le paragraphe sur la notion de p-valeur, notion qui apparaît souvent simple à interpréter par des publics non avertis, propose de la construire par le biais de trois méthodes différentes : simulation informatique, calcul pour un nombre limité de cas et utilisation de la loi binomiale.

En ce qui concerne les tests, les auteurs différencient clairement les deux formes de tests, Fisher et Neyman-Pearson, en montrant la différence entre une situation dans laquelle on veut rejeter une hypothèse (le cas du test de thé de Fisher) et celle où l'on cherche à décider entre deux hypothèses contraires (illustré par la pratique du contrôle de qualité).

J'ai personnellement particulièrement apprécié le chapitre 3 qui porte sur l'exploration et la modélisation de l'association, thème auquel notre enseignement n'accorde pas, à mon avis, une place suffisante. Les auteurs proposent deux activités : la première autour du croisement de deux variables qualitatives et de l'analyse d'une table de contingence (distribution conditionnelle, concept d'indépendance) ; la seconde qui porte sur le croisement de deux variables quantitatives et les concepts de corrélation et de régression. Je regrette évidemment que l'on n'aborde pas le cas du croisement d'une variable qualitative avec une variable quantitative, mais c'est compréhensible vu l'objectif du livre, destiné à l'enseignement au lycée. Ce croisement permet en effet d'ouvrir des perspectives intéressantes puisqu'il donne la possibilité d'approcher les concepts de variance inter et intragroupes mais surtout de différencier corrélation et corrélation linéaire, ce que font les auteurs, ce qui est assez rare !

*Note de lecture : « Statistics and Probability in High School » (C. Batanero and M. Borovcnik, 2016)*

Pour conclure, j'ai trouvé ce livre tout à fait remarquable à la fois par son originalité et par sa dimension résolument opérationnelle puisqu'il combine résultats issus de la recherche en éducation statistique et pratiques basées sur une longue expérience de l'enseignement. Il s'adresse avant tout aux professeurs de mathématiques qui enseignent la statistique dans le secondaire. Néanmoins, étant donné que les notions qu'il explore sont enseignées également dans le supérieur, les outils qu'il propose intéresseront un large public qui y trouvera certainement matière à enrichir sa pratique et ses réflexions.

Cet ouvrage est en anglais ce qui sera sans doute un frein pour certains lecteurs, mais l'importance donnée aux applications pratiques et la structure très claire de ce livre rend sa lecture facile, même lorsqu'on ne maîtrise pas très bien la langue anglaise !