



## Note SFdS audition du 22/11/17

L'importance d'un enseignement de la statistique et des probabilités solide dans le secondaire tient, de notre point de vue, essentiellement à deux aspects : **économique** et **citoyen**.

### D'un point de vue économique

La demande de compétences dans le domaine de la science des données se fait de plus en plus pressante, comme en témoignent le nombre toujours plus important de recrutements de « data scientists » capables de faire émerger de l'information et de la faire partager au travers de techniques de visualisation. Ces métiers nouveaux demandent **des compétences solides en statistique et probabilités** (analyse des données, statistique inférentielle, théorie de la décision), en mathématiques générales (outils de modélisation, optimisation) et en informatique. **Ces compétences ne s'acquièrent que sur le long terme et des contenus correspondants doivent s'enseigner dès le collège.**

### D'un point de vue citoyen

La statistique et les probabilités apportent des outils essentiels à la construction du jugement critique, particulièrement nécessaires à l'exercice d'une citoyenneté active dans la société du chiffre :

- notions fondamentales de proportion, de pourcentage, de probabilité ;
- sens des indicateurs statistiques de position (moyenne, médiane), de dispersion (étendue, quartiles, quantiles, écart-type) et des représentations graphiques courantes (diagrammes en barres, circulaires, cartésiens, histogrammes...) ;
- hasard et prédictibilité ;
- sensibilisation à différents biais liés à la psychologie humaine (biais d'équiprobabilité, absence de mémoire, statistique des petits nombres, importance de la taille de l'échantillon sur la variabilité...) ;
- notion de modèle, notamment probabiliste ;
- notion de différence significative, prise de décision en environnement aléatoire, notion d'intervalle de confiance, cas des sondages ;
- corrélation et causalité ;
- probabilités conditionnelles et rudiments de raisonnement bayésien ;
- ...

Pour atteindre ce double objectif, il nous semble essentiel de s'appuyer sur deux leviers :

- l'interdisciplinarité au collège ;
- le traitement de données en grande quantité, voire massives, au lycée.

**L'interdisciplinarité au collège** est le terrain idéal où mettre en situation l'intérêt de l'analyse statistique, favorisant la construction des concepts au-delà de la particularité des

contextes. La démarche de projet développe la coopération et la prise d'initiative et joue un rôle important dans la motivation des élèves.

*Exemple : travail interdisciplinaire (EPI) sur les crues dans le Sud Seine-et-Marne.*

**Le développement de l'algorithmique et de la programmation** au lycée permet, par le langage Python et la bibliothèque Pandas, l'accès à des données en très grande quantité et à leur analyse statistique. L'accès aux données est facilité par le phénomène open data.

*Exemples : fichier INSEE des prénoms à la naissance (plus de 3 millions de lignes, une dizaine de variables) ; fichier des données de location Airbnb, environ 60 000 lignes et une dizaine de variables pour Paris, possibilité de comparer avec Londres, New-York etc.*

*On a là un vaste terrain d'investigation, de data visualisation, la possibilité de prélever des échantillons aléatoires dans les jeux de données, d'élaborer et de tester des modèles, d'effectuer des analyses croisées des variables.*

*On peut envisager une exploration des données par équipes sous forme de « challenge ». Cette forme ludique de l'activité mathématique et le contexte motivant des données favorisent l'investissement des élèves.*