

Présentation de l'enseignement

Pourquoi les stats ?

Suicide chez France télécom : 25 sur un an et demi, cela fait 16,6 pour 100 000 par an. La moyenne nationale des suicides chez les grandes entreprises est de 17 pour 100 000. France Télécom est (très légèrement en dessous) ! Mais les journalistes sont nuls en stat. Ajoutez à cela un reportage mélodramatique avec des interview voyeuristes ou l'on filme des gens en train de pleurer, il n'en faut pas plus pour clouer France télécom au pilori !

Les chiffres donnent une vision plus objective, moins émotionnelle de la réalité. Tous les citoyens devraient avoir une culture minimale en stat afin de ne pas se laisser bernier. D'ailleurs, au lycée, il y a de plus en plus de stat dans les programmes...

Autre exemple, la violence est-elle en augmentation ou diminution ? Sur les 4 derniers siècles, diminution constante du taux d'homicide : 40/100 000 puis 10/100 000 puis 2/100 000 puis 1/100 000 puis 0.5/100 000. Nicolas Sarkozy a beau dire, la violence diminue constamment à travers les siècles et dans tous les pays. Donc, enseigner les stat et éduquer le citoyen, c'est faire acte d'anti-Sarkozysme.

Pourquoi vous ?

Parce que vous êtes des citoyens et donc il faut vous éduquer (!) Egalement parce que vous avez un mémoire de M1 à faire... Dans votre mémoire, vous allez devoir « prouver » quelque chose : l'efficacité d'une méthode, l'impact d'une activité sportive sur tel ou tel paramètre. Il faudra le moins possible laisser la place à la subjectivité. D'où la nécessité de recourir aux statistiques.

Logiciel

Dans l'UFR, on utilise R, SPSS, SAS, Statistica, Sigmaplot, Mathematica... Il n'y a clairement pas homogénéité ! Plus généralement, il existe deux grandes familles de logiciels : ceux à menu déroulant et ceux à ligne de commande.

- Avantage des « menus déroulants » : la prise en main est rapide.
- Avantage des « lignes de commandes » : le code est ré-exécutable (par exemple si des individus sont ajoutés dans la base de données). Si le résultat est « surprenant », on peut également vérifier ce qui a été fait.

Nous utiliserons R, un logiciel à ligne de commande. Il est gratuit et téléchargeable sur internet.

Modalités de contrôle

Pour chacune des deux grandes parties du cours, je vous donnerais un jeu de données que vous aurez à analyser (chez vous) puis vous aurez un rapport à me rendre. Le premier rapport sera à rendre après le TD 7, le deuxième après le TD 12. Chaque rapport est corrigé puis noté (note N1).

Si vous avez moins de 15 : vous avez ensuite une semaine pour intégrer les corrections à votre rapport. Il est alors re-noté (note N2, N3) et ainsi de suite jusqu'à ce qu'il soit d'une qualité satisfaisante. La note finale d'un rapport est la première note obtenue (N1) modifiée par un bonus/malus qui est fonction de l'amélioration du rapport (si N2 ou N3 ≥ 15 , alors le bonus est de +2. Si N2 ou N3 < 10 , le malus est de -2).

La note finale de l'EC est la moyenne entre les notes finales du premier et du deuxième rapport.

Rapport

Un rapport doit comporter 3 parties :

1. **Analyse des données** : les données doivent être analysées et présentées clairement. Toutes les informations pertinentes doivent apparaître. Par contre, aucune ligne de code R ne doit figurer dans les résultats.
2. **Code R** : après chaque analyse, le code R que vous avez utilisé doit être donné.
3. En annexe de rapport, vous devrez présenter une **fiche résumé** du cours. Cette fiche doit contenir l'ensemble du cours tout en étant le plus synthétique possible. Son but est de vous permettre de retrouver rapidement ce que vous savez quand, dans 6 mois, vous en aurez besoin mais vous aurez tout oublié..

Un exemple de rapport est présenté sur le site. Les parties 1 et 2 s'alternent régulièrement tout au long du rapport, la partie 3 est en annexe.