

INTRODUCTION À LA STATISTIQUE BAYÉSIENNE

Problème

Les données :

Une base de données a été utilisée dans un article de recherche (<https://bmjopensem.bmj.com/content/5/1/e000543>) pour étudier l'influence de la morphologie sur la performance en natation. Une partie de ces données est disponible pour ce TP, correspondant aux meilleurs performances mondiales, pour l'année 2014, aux compétitions de nage libre, hommes et femmes, pour des distances allant du 50 à 1500m.

Les variables disponibles sont :

- Nom : nom de l'athlète,
- Pays : pays représenté par l'athlète,
- Sexe : genre de l'athlète,
- Distance : distance de course (50, 100, 200, 400, 800, 1500) en mètres,
- Taille : taille de l'athlète en cm,
- Poids : poids de l'athlète en kg,
- IMC : Indice de Masse Corporelle de l'athlète,
- Temps : performance chronométrique sur la course (en sec),
- Vitesse : vitesse moyenne sur la course (en km/h).

Travail attendu :

Pour ce travail, vous serez encore une fois encouragés à prendre en main de manière très autonome l'étude de ce jeu de données. Il consistera en la première prise en main d'outils d'analyse Bayésienne. Pour cela, vous pourrez par exemple utiliser les packages R suivant :

- BEST : <https://cran.r-project.org/web/packages/BEST/vignettes/BEST.pdf>, permettant d'effectuer des comparaisons de moyennes entre 2 distributions via l'estimation bayésienne en remplacement du traditionnel t-test.
- bayestestR <https://easystats.github.io/bayestestR/articles/example1.html>, permettant d'écrire un modèle linéaire bayésien, d'estimer ses paramètres et d'analyser les lois a posteriori obtenues.

N'hésitez pas également à procéder à des analyses exploratoires préliminaires, comme des graphiques bivariés par exemple.

Le choix des variables et modalités à étudier en priorité est *libre* et devra faire l'objet d'une problématisation bien claire. Pour ce troisième et dernier rapport, il sera attendu à la fois de la *prise d'initiative* et de la *concision*, sur lesquels nous avons insisté jusque ici. N'hésitez à proposer des méthodes originales et interprétations de votre propre initiative, les packages proposés sont simples d'utilisation, mais vous pouvez tout à fait chercher à compléter votre travail par d'autres approches. Les méthodes bayésiennes sont nouvelles pour vous mais leur interprétation est très naturelle. Il peut ainsi être intéressant de regarder beaucoup d'aspects/variables différents, même en utilisant plusieurs fois la même méthode/fonction R. Cela permettra de bien se familiariser avec l'inférence bayésienne, et d'apprécier tout son potentiel en terme d'interprétation et de prise en compte de l'incertitude.

Quelques règles de rédaction d'un rapport :

Le rapport est un document qui doit être de qualité avec une présentation réfléchie et structurée. Il a pour but de présenter le cheminement de votre étude. Ce n'est pas une liste exhaustive des travaux effectués mais une vitrine du travail effectué. Voici quelques règles simples à respecter.

Organisation du rapport

- Le rapport commence toujours par une introduction générale du sujet d'intérêt.
- Ensuite vous présenterez la base de données : contexte et but de l'étude, présentation des différentes variables, problématique.
- Viennent ensuite les résultats pertinents de statistiques descriptives. Les statistiques descriptives peuvent être résumées sous forme de tableaux et de figures et doivent être commentées (valeurs pertinentes pour l'étude, valeurs atypiques, valeurs manquantes, dissymétrie des distributions, ...).
- L'étude des relations entre les variables doit être présentée avec soin. Pour chaque relation étudiée, présentez précisément les variables étudiées, l'outil statistique utilisé, le résultat de la comparaison ou du test, la conclusion que l'on peut en tirer. Terminez par un commentaire replaçant le résultat obtenu dans le contexte de l'étude.
- N'oubliez pas d'interpréter **TOUS** vos résultats.
- Vous conclurez votre rapport en donnant deux conclusions à cette étude :
 1. La première s'adressant à un spécialiste de la statistique.
 2. La seconde étant destinée à un non spécialiste de la statistique.
- Si vous insérez des annexes veillez à ce qu'elles soient en nombre limité.