Project 1: regression 1

Urvan Christen, Amandine Goffeney, Joseph Vermeil, Lucile Vigué

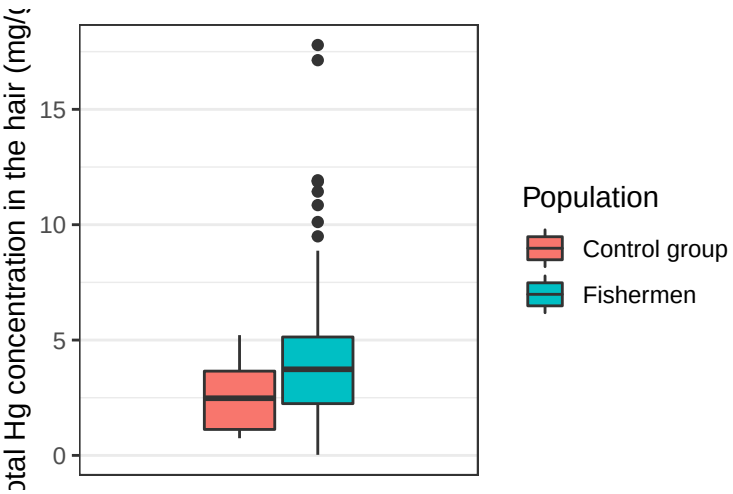
March 20, 2019

Introduction

Mercury is a metal present in the environment whose harmful effects on human health are well assessed (Park and Zheng 2012). In a study led in 2000 (Al-Majed and Preston 2000), scientists collected data on total mercury and methyl mercury levels in the hair of 100 fishermen of Kuwait, aged 16 to 58 years, comparing them to those of a control population of 35 non-fishermen, aged 26 to 35 years. The aim of this report is to analyse the factors influencing the levels of mercury in both populations. For the sake of simplicity, we will only focus on total Hg, leaving out methyl mercury, as both variables are strongly correlated.

The dataset gathers information about six numerical variables (age, height, weight, number of fish meals per week and residence time in Kuwait) and two categorical ones (being a fisherman or not, fish consumption habits). There is no additional information about gender as all the participants in the study are males.

A first insight at the data shows that the fishermen population exhibits higher levels of mercury in their hair. The significance of the difference between the means of both distributions is assessed by a *Welch Two Sample t-test* with the alternative hypothesis that the fishermen population has a average greater level of mercury than the control population (p-value = 7.473e-05).



Exploratory analysis (Amandine)

```
table(dataset$fisherman, dataset$fishmlwk)
```

```
##
##       0  1  2  3  4  7 14 21
##   0 10 14 11  0  0  0  0  0
##   1  0  0  0  2 12 70  5 11
```

Unbalanced (les non-fishermen sont <= 2, les fishermen sont >= 3), catégories avec pas du tout le même nombre d'individus (la catégorie 3 n'a que 2 personnes)

L'étude est déséquilibrée entre les deux groupes sur à peu près toutes les explanatory variables (restime, age,...).

VIF du model linéaire et squared, dire pourquoi on ne prend pas le squared.

Model selection (Joseph)

Stepwise selection (inclure * fishermen). (Dire qu'on a vérifié avec forward et backward et qu'on a trouvé la même chose => à coder quand même !)

Results and discussion (Urvan)

Fit model (* fishermen)

Commenter les résultats, valeurs des coefficients, font-elles du sens, p-values

Les diagnostic plots pour weight + dire que c'est pareil pour tout le monde

Conclusion (Lucile)

3 lignes

Les principales variables explicatives sont:

- La consommation de poisson => logique
- Le poids => plus étonnant mais on sait que les cellules graisseuses stockent les toxiques

References

Al-Majed, NB, and MR Preston. 2000. "Factors Influencing the Total Mercury and Methyl Mercury in the Hair of the Fishermen of Kuwait." *Environmental Pollution* 109 (2). Elsevier: 239–50.

Park, Jung-Duck, and Wei Zheng. 2012. "Human Exposure and Health Effects of Inorganic and Elemental Mercury." *Journal of Preventive Medicine and Public Health* 45 (6). Korean Society for Preventive Medicine: 344.