Factors influencing the levels of mercury in the hair of fishermen and non-fishermen

Urvan Christen, Amandine Goffeney, Joseph Vermeil, Lucile ViguÃ©

Introduction

Mercury is a metal present in the environment whose harmful effects on human health are well known [1]. In *Al-Majed and Preston* study [2], total mercury and methyl mercury levels in the hair of 100 fishermen of Kuwait, aged 16 to 58 years, were compared to those of a control population of 35 non-fishermen, aged 26 to 35 years. The aim of our study 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, since both variables are strongly correlated (shown in the paper). The dataset contains six numerical variables (age, height, weight, number of fish meals per week and residence time in Kuwait) and two categorical variables (being a fisherman or not, fish consumption habits). All study participants are male.

Exploratory analysis

Before fitting a model, we first look at the data. Table 1 shows the distribution of individuals according to the number of fish meals per week and the two groups. We note that for some values, there are very few people. For instance, only two people eat fish three times per week. We also see that the number of fish meals per week is completely separable by population group. This explains the strong correlation observed between these two variables in the correlation matrix (Table 2). To check whether we have multicolinear variables, we use the variance inflation factor (VIF). We set the following criteria: we keep only the variables that have a result below 5. We observe that all the variables have a variance inflation factor below 2, so we do not eliminate any variable, for the moment.

Model selection

We now use a stepwise method of model selection to select the more relevant variables to explain the *TotHg* variations within the population. The selection of the model is based on the AIC score of the model, which means that after having added or delected an explanatory variable from the model, the algorithm keeps the new model if the AIC score is better. In the end, it is an optimisation problem in which we want to find the model with the best AIC score possible. We begin our process of model selection with a formula with all the variables. After this stepwise selection, the selected model is: $TotHg = \beta_0 + \beta_1 \cdot fisherman + \beta_2 \cdot age + \beta_3 \cdot restime + \beta_4 \cdot weight + \beta_5 \cdot fishmlwk$ (Table 3). The intercept and the weight coefficient are highly significant but the others are not. However, the signs of the coefficients are not absurd: while it is not really intuitive that the weight coefficient should be positive or negative, the coefficient of *fishmlwk* has to be positive, and it is the case here. Now, to determine possible differences between the fishermen and non-fishermen, a model based on

Table 1: Distribution of the number of fish meals accross fishermen and non-fishermen populations

	0	1	2	3	4	7	14	21
non-fisherman	10	14	11	0	0	0	0	0
fisherman	0	0	0	2	12	70	5	11

Table 2: Correlation matrix

	fisherman	age	restime	height	weight	fishmlwk	fishpart	TotHg
fisherman	1.00	0.25	0.25	-0.06	-0.09	0.61	0.46	0.23
age	0.25	1.00	0.58	0.00	0.05	0.26	-0.01	0.16
restime	0.25	0.58	1.00	-0.05	0.11	0.19	0.00	0.06
height	-0.06	0.00	-0.05	1.00	0.30	-0.04	-0.03	0.19
weight	-0.09	0.05	0.11	0.30	1.00	0.04	-0.05	0.41
fishmlwk	0.61	0.26	0.19	-0.04	0.04	1.00	0.19	0.30
fishpart	0.46	-0.01	0.00	-0.03	-0.05	0.19	1.00	0.11
TotHg	0.23	0.16	0.06	0.19	0.41	0.30	0.11	1.00

the interactions between the *fisherman* variable and all the others is proposed. The best model is now $TotHg = \beta_0 + \beta_1 \cdot fisherman + \beta_2 \cdot weight + \beta_3 \cdot fishmlwk + \beta_4 \cdot fisherman \cdot weight + \beta_5 \cdot fisherman \cdot fishmlwk$.

Results and discussion

We now turn to discussing the results of the model: $TotHg = \beta_0 + \beta_1 \cdot fisherman + \beta_2 \cdot weight + \beta_3 \cdot fishmlwk + \beta_4 \cdot fisherman \cdot weight + \beta_5 \cdot fisherman \cdot fishmlwk$.

Difference between fisherman and control populations

According to our model fitting, being a fisherman or not has an impact on how other variables such as *weight* or *fishmlwk* are correlated to total Hg levels. Eventually the value of β_1 (corresponding to *fisherman* variable) is surprising: it implies that the fact of being a fisherman gives you -9.00 mg/g Hg compared to non-fishermen. Yet this comes from the increased value of *fisherman:weight*, which will make the overall Hg concentration more important among fishermen, as expected.

Number of fish meals per week

The negative value of β_5 (corresponding to *fisherman:fishmlwk* interaction) is also a surprising result: it suggests that among fishermen the number of fish meals per week has a weak impact on total Hg levels ($\beta_3 + \beta_5 = 0.1$) while it has a much stronger impact on non-fishermen ($\beta_3 = 1.53$). However, we have to remember that the number of fish meals per week is completely separable by population group. The simplest explanation could be that the relation between the number of fish meals per week and total Hg levels is positive but not

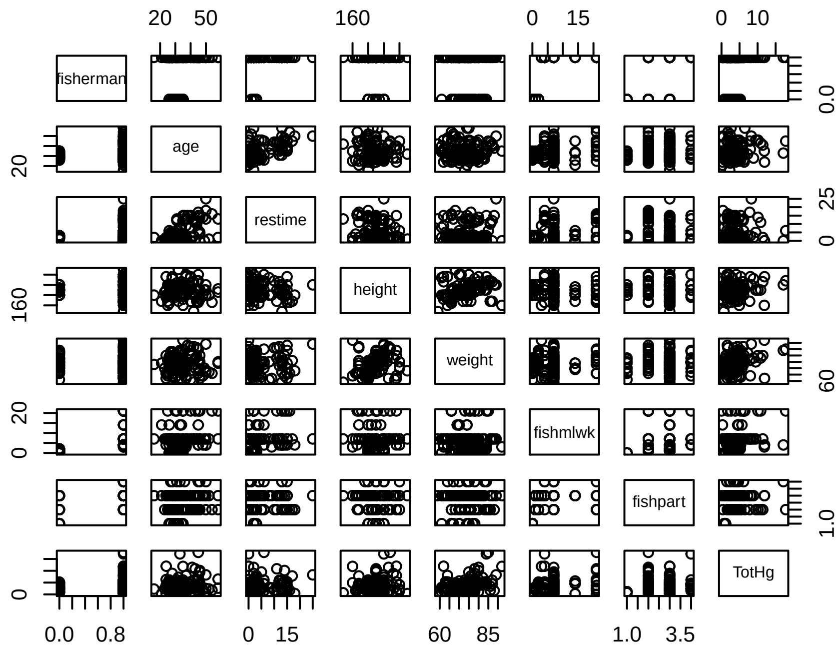


Figure 1: Pair plots

linear: it increases fast for low numbers of fish meals (i.e. for non-fishermen) and more slowly for high numbers of fish meals (i.e. for fishermen). Furthermore, the observable does not reflect entirely the quantity of fish eaten, since one can eat more or less fish per meal. The weight of fish eaten per week, might be a more accurate observable to study.

Weight

The fact that the weight has a positive influence on this concentration was unexpected, since a concentration and not an absolute quantity was measured. However, even though it was unexpected it has many possible explanations such as the fact that weight is much likely correlated with adiposity more susceptible to catch toxins than other tissues. Another explanation could be that the fatter, the more one eats and possibly ingests mercury that could fix in the hair; since hair weight isn't likely to be correlated with body weight, it could explain the high mercury concentration in hair.

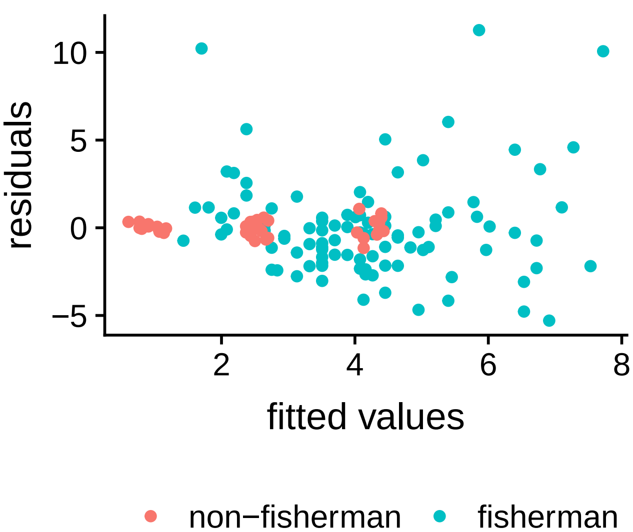
Table 3: Full model regression results

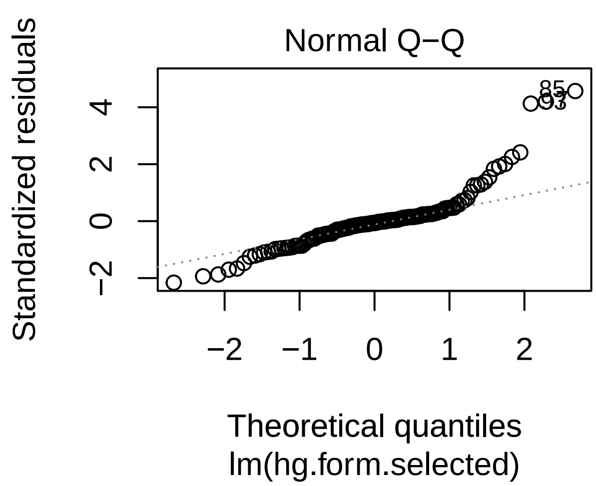
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-12.68	2.71	-4.68	7.0e-06
fisherman	1.11	0.65	1.70	9.1e-02
age	0.05	0.03	1.43	1.6e-01
restime	-0.08	0.05	-1.45	1.5e-01
weight	0.19	0.03	5.58	1.4e-07
fishmlwk	0.10	0.05	1.82	7.2e-02

Table 4: Final model regression results

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.42	6.94	-0.20	8.4e-01
fisherman	-9.00	7.43	-1.21	2.3e-01
weight	0.03	0.10	0.33	7.4e-01
fishmlwk	1.53	0.70	2.18	3.1e-02
fisherman:weight	0.16	0.11	1.48	1.4e-01
fisherman:fishmlwk	-1.43	0.70	-2.04	4.4e-02

Diagnostic plots





The plot of the residuals against the fitted values help us to assess the 3 assumptions of the residuals. For the first one, which is that the mean of the residuals is 0, the plot confirms it. For the second one which is the homoscedasticity of the model, it does not seem to be really homoscedastic. Indeed, a model is homoscedastic if the variance is the same for all the values, and here, the variance is much higher for fishermen than for non-fishermen. However, within each class, the variance is overall similar, even if it tends to be a little more spread for high fitted values. At last, for the third one, which is the uncorrelation between the X variables and the residuals, we can again confirm it with the plot.

The QQ plot shows that we have a heavy tailed distribution of residuals, with a very heavy right tail. It could be explained by a non-linear relation between the variables and the concentration of mercury.

Conclusion

We have built a simple model that can help to explain the levels of mercury observed in a fishermen population compared to a control group. It appears that the variables that have the most significant influence over the measured levels of mercury are the weight of the individual and the frequency at which they eat fish. The former can seem surprising even though some hypotheses can be formed to account for the influence of weight on mercury levels. The latter may be the main explanation for the differences observed between our two groups: fishermen eat fish much more often than non-fishermen, since fish is a well-known source of mercury it seems logical to see a positive correlation between fish meal frequency and mercury levels and thus to observe higher mercury levels in fishermen populations compared to non-fishermen.

References

[1] J.-D. Park and W. Zheng, "Human exposure and health effects of inorganic and elemental mercury," *Journal of preventive medicine and public health*, vol. 45, no. 6, p. 344, 2012.

[2] N. Al-Majed and M. Preston, "Factors influencing the total mercury and methyl mercury in the hair of the fishermen of kuwait," *Environmental Pollution*, vol. 109, no. 2, pp. 239–250, 2000.