**Exercise 1**

The `Mroz` data frame (contained in the R add-on package **carData**) contains observations from married women from the Panel Study of Income Dynamics (PSID). The following covariates should be considered:

| | |
|---|---|
| lfp | labor-force participation; a factor with levels: no; yes |
| k5 | number of children 5 years old or younger |
| k618 | number of children 6 to 18 years old |
| age | in years |
| wc | wife's college attendance; a factor with levels: no; yes |
| hc | husband's college attendance; a factor with levels: no; yes |
| lwg | log expected wage rate |
| inc | family income exclusive of wife's income |

Fit a binary logistic regression model `mroz.glm.mod` using the `logit` link function. The variable `lfp` is your binary response. All other given variables are used as explanatory variables. Interpret $\beta_{age}$ specifically.

Estimate the confidence intervals for all regression coefficients using the function `confint()`.

**Exercise 2**

In this exercise you will study some properties of the maximum extreme value distribution (the Gumbel distribution) with distribution function

$$F_Z(z) = \text{loglog}^{-1}(z) = \exp(-\exp(-z))$$

where $\text{loglog}(p) = -\log(-\log(p))$. Hence we have $F_0(y) = \text{loglog}^{-1}(h(y))$ and $F_1(y) = \text{loglog}^{-1}(h(y) - \beta)$.

An interesting interpretation for $\beta$ is given on page 100 in the script.

(a) Explain step-by-step what is happening in the derivation of

$$\mathbb{P}(\max\{Y_1, \ldots, Y_n\} \leq y) = \prod_{i=1}^{n} \mathbb{P}(Y_i \leq y) = \prod_{i=1}^{n} F_0(y) = F_0(y)^n.$$

(b) Load the `lehmann_data.csv` into R. The dataset contains the binary indicator `sex` and the response `bmi`. Fit a Lehmann type I alternative to estimate the shift parameter $\beta$ as a log reverse-time hazard ratio (Hint: Use the `Lehmann()` function from package **tram**.)

(c) Interpret $\beta$ as on page 100 in the script.

(d) Sample $n = \lfloor \exp(\beta) \rceil$ observations from $\hat{F}_0$ and interpret the maximum of these $n$ draws. Here $\lfloor \cdot \rceil$ means rounding to the nearest integer. (Hint: You can use `simulate(as.mlt(mod), newdata = ...)`, where `mod` is the model and `newdata` is a data.frame containing $n$ rows of the covariate `sex`, to simulate from a fitted transformation model.)

(e) Repeat (d) 1000 times and visualize the empirical cdf of the resulting vector of maxima. Add the estimated distribution function $\hat{F}_1$ to the plot and interpret your results.
(Hint: You can extract the estimated cdf using `predict(mod, type = "distribution", newdata = ...)`.)

luisa.barbanti@uzh.ch
lucasheinrich.kook@uzh.ch