Multivariate Statistics: Exercise 4

November 6, 2014

Robust regression:

Load the package `robustbase` and use the data `milk` from this package.

1. Use variable `X4` as response and `X5` as explanatory variable in linear regression.
   (a) Plot the data and add the least-squares regression line (`abline()` of the result object).
   (b) Perform LTS-regression (`ltsReg()`), and add the resulting regression line to the plot. Show in color which observations receive a weight of zero.
   (c) Perform MM-regression (`lmrob()`), and add the resulting regression line to the plot. Visualize the resulting weights of the observations by symbol size. Compare the weights from MM- and LTS regression.
   (d) Plot the result objects of LS-, LTS- and MM-regression. How can you interpret these plots?
2. Use variable `X4` as response and all remaining variables as explanatory variables in linear regression.
   (a) Estimate the regression parameters for LS-, LTS- and MM-regression. Show for each method separately the fitted values of the model versus the response variable. What do you conclude?
   (b) Compute the multiple R^2 measure for each outcome. The R^2 is defined by

   $$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

   For the robust case it can be defined as

   $$R_w^2 = 1 - \frac{\sum_{i=1}^{n} w_i(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} w_i(y_i - \bar{y}_w)^2}$$

   (Renaud and Victoria-Feser, 2010), with weights w_i, and $\bar{y}_w = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i y_i$.
   (c) Apply `summary()` on the result objects of LS-, LTS- and MM-regression. What do you conclude?
   (d) Plot the result objects of LS-, LTS- and MM-regression. How can you interpret these plots?

Save your (successful) R code together with short documentations and interpretations of results in a text file, named as *Familyname4.R*. Send the file as an email attachment to *mehmet.mert@tuwien.ac.at*, at latest Tuesday (04.11).