VARIABLES are written with capitals.

Linear regression

Use ***Supermodel.sav*** (see Field 2018, Chapter 9.17, Task 4 for more detailed explanations) to check whether the conditions for calculating a regression are fulfilled and to calculate a linear regression model. The file contains four variables: the models SALARY (the supposed outcome variable) as well as the AGE of the model, how many YEARS it worked in the industry and a BEAUTY-rating made by a panel of experts.

Begin with checking whether the assumption for calculating a regression are fulfilled:

- *Check – using scatter plots – whether the condition of linearity is fulfilled.*
- *Check whether the variables conform with a normal distribution and whether there are outliers.*
- *Check the bivariate correlation between the variables. Is there anything to raise red flags?*

Continue with building a Linear regression model and enter all variables at the same time into the model:

- *Which variable goes into which field (Dependent / Independent)?*
- *Is the model significant as a whole? Which predictor variables make a significant contribution?*
- *Use the diagnostic measures under the buttons "Statistics" and "Plots" to evaluate the quality of your model. What do you observe? Which of these measures should raise a red flag and why?*
- *Try out other methods than entering all variables at once to add variables to your model. What did you choose, why did you choose it and what is the outcome? Re-evaluate the previous two questions for the updated model.*

Linear regression (extra)

Use ***Coldwell 2006.sav*** (see Field 2018, Chapter 9.17, Task 7 for more detailed explanations) to check whether the conditions for calculating a regression are fulfilled and to calculate a linear regression model. The file contains ten variables (of which one [FAMILYID] is irrelevant): the child's adjustment SDQ (the supposed outcome variable). There are those groups of predictor variables: (1) the child's age (CHILD_AGE) and gender (CHILD_GENDER); (2) the relationship with the mother in terms of warmth (CHILD_WARMTH) and anger / hostility (CHILD_ANGER); (3) the perceived relationship between mother and child divided into positive (MUM_POS) and negative aspects (MUM_NEG); and (4) the CHAOS in the household.

- *The dataset contains a selection variable removing the families with the IDs 264 and 274. Do you agree with this decision and if so why?*
- *Assess the assumption of linearity using scatter plots. Can you use these variables?*
- *Check the bivariate correlations of the variables. Which variable should not be included and why? When looking at the correlation matrix, is there anything that raise red flags and why?*
- *Enter all predictors into the model at once. Is the model significant and which predictors make a significant contribution?*
- *Try out other methods to add variables to the model. Assess the significance, describe which predictor variables made a significant contribution and use the previously described diagnostics to evaluate the model. Summarize these observations.*