Practical "Multiple linear regression" (P6)

1. Dataset bloodpress.sav contains blood pressure and other clinical data of 20 high blood pressure patients.

   a) Make a multi-scatterplot of all variables in the data (except for variable *Pt*) to check their relationship with each other. Give examples of strongly correlated pairs of variables. NOTE: Go to *Graphs* → *Legacy dialogs* → *Scatter* and there select *Matrix scatter*.

   b) Estimate a multiple regression model with all variables (except for *Pt* and *BP*) as covariates, and *BP* as outcome variable. Check and interpret *R–square*. Examine collinearity statistics. Note which variables cause problems and why. Remove variable BSA and re-fit the model. Are the tolerance values (or equivalently variance inflation factors) for this model acceptable? Compare *R–square* values for both models. Note that the number of parameters is high relative to the sample size, but we will ignore this for illustration purposes of the collinearity issue. NOTE: To run multiple regression go to *Analyze* → *Regression* → *Linear* and within *Statistics* select *Collinearity diagnostics*.

2. Dataset braindom.sav comes from a study into how different kinds of brain dominance (left-brained, right-brained or integrative) affect the ability to recall information of various types for a sample of 24 subjects.

   a) Are there differences in score between the three groups? Use a one-way ANOVA. NOTE: Go to *Analyze* → *Compare Means* → *One-Way ANOVA* and within *Options* select *Homogeneity of variance test*. To make a histogram of *score* for each of the three brain dominance groups go to *Graphs* → *Legacy Dialogs* → *Histogram* and insert *score* as Variable and *brain* as Columns. To make a boxplot of *score* for each of the three brain dominance groups go to *Graphs* → *Legacy Dialogs* → *Boxplot* (Simple).

   b) Create dummy variables from *brain* and estimate a linear regression model with *score* as the dependent variable and these dummies (excluding one of them, say one for integrative brain dominance, taken as reference, to avoid perfect collinearity) as independent variables. Compare results with those of the ANOVA. Interpret the coefficients of the model. NOTE: To create a dummy variable for right brain dominance go to *Transform*→ *Recode into different variables*. Then, choose *brain* to be the numeric variable, fill in the name of the output variable and click *Change*. After that, click *Old and New Values* button and there under *Old Value* fill in value 2 and under *New Value* fill in value 1. Click *Add*. Then, under *Old Value* click on *All Other Values* and under *New Value* fill in value 0. Click *Add*, then *Continue* and finally *OK*. Follow the same steps to create the second dummy for left brain dominance (remember to use 1 as the *Old Value*). Alternatively, use the *Transform* → *Compute variable* option to create dummies.

3. Dataset lifesatisfaction.sav contains information on age, gender, academic achievement, social support (specifically from spouse/partner and from children) and life satisfaction of 97 elderly individuals.

   a) Use the overall F-test in combination with individual t-tests to determine factors that influence life satisfaction of elderly.

   b) Interpret the partial regression coefficients for the significant predictors. Which of these variables is the most important in predicting life satisfaction?

   c) Check *R–square*. Are the variables useful in predicting life satisfaction? Are all potentially important predictors included in this study? If not, what other factors do you think are missing?