# Quantitative methods of business research Home assignment 2 report for cohort 1

Names of the students

**Download the file** HA2\_data\_coh1.sav **and open it in RStudio.**

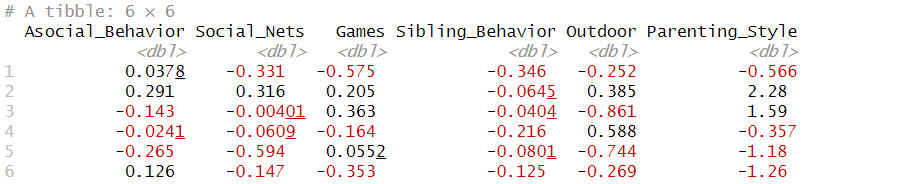
A researcher explored the relationship between children level of **Asocial Behavior** and several potential predicting factors in 511 children who had a sibling. Variables measured were **Parenting Style** (high score = bad parenting practices), **Games** (high score = more time spent playing computer games), **Social Networks** (high score = more time spent in social networks), **Time Spent Outdoor** (high score = the child spends more outdoor), and **Sibling Behavior** (high score = more asocial behavior seen in their sibling). Past research indicated that parenting style, time spent in playing computer games and sibling level of asocial behavior were good predictors of the level of asocial behavior in the child. All other variables were treated in an exploratory fashion.

**Q1. Getting acquainted with the data. (3%)**

1. How many variables are there?

511 obs. Of 6 variables

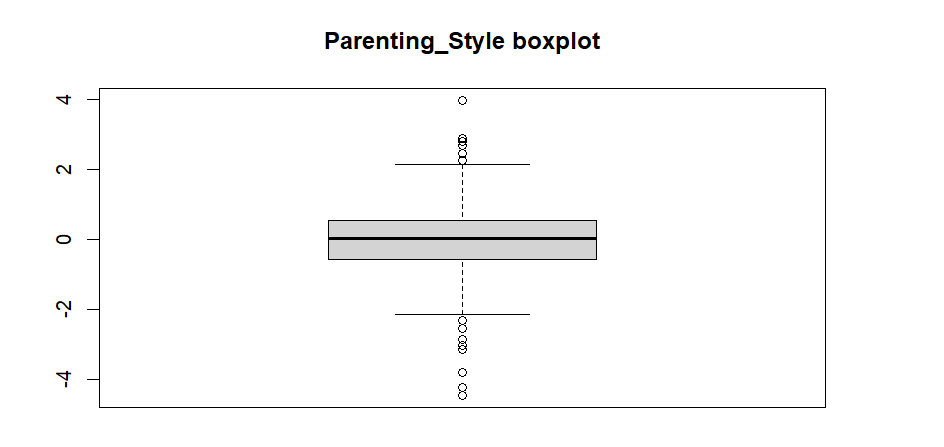
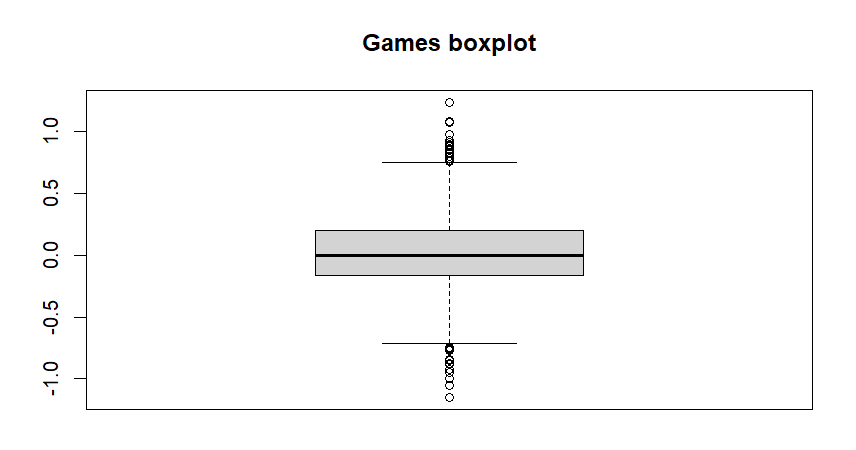
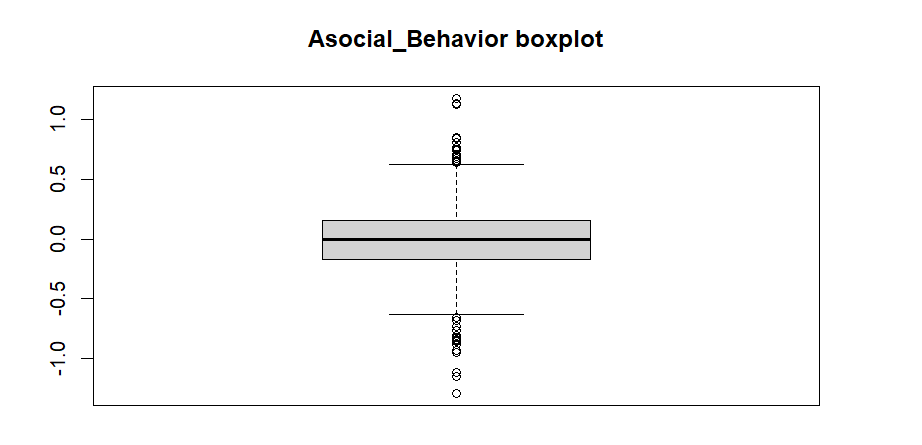
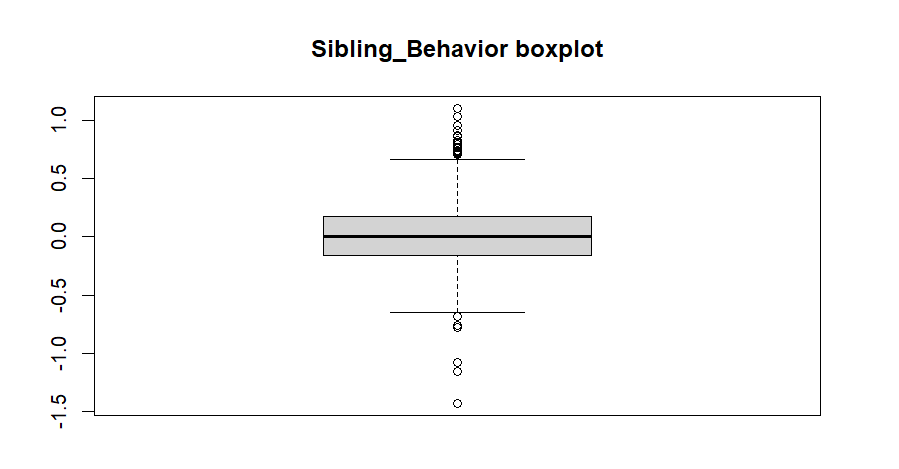
1. What can we say about **Asocial Behavior** variable (distribution, descriptive statistics)?



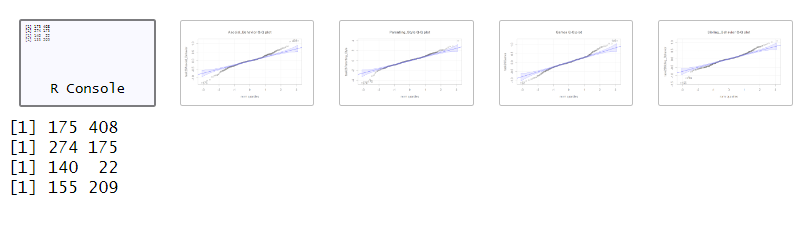
1. How many children with positive sibling level of asocial behavior are there?

**Q2. Check linear model assumptions and try to spot potential sources of bias for the variables under analysis (Asocial Behavior, Parenting Style, Games, Sibling Behavior)**. **(11%)**

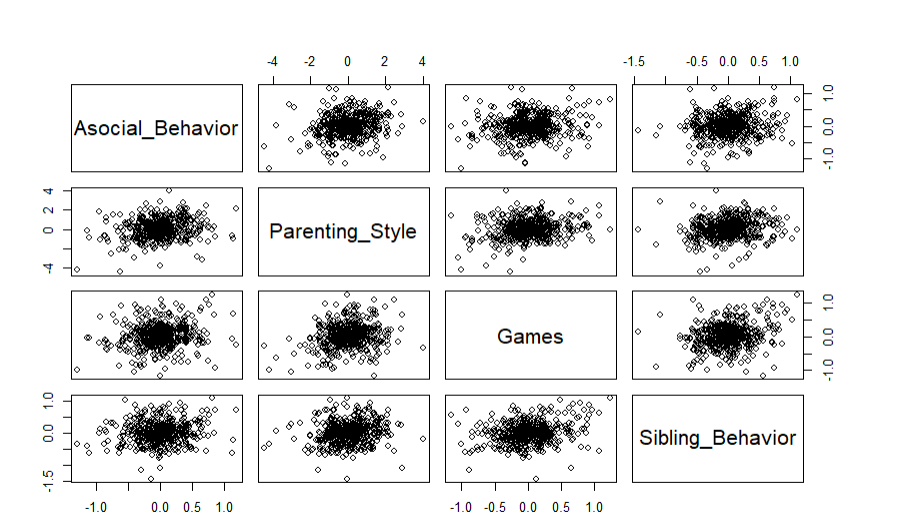
1. Using boxplots, try to conclude on the separate variables’ outliers. Do you find any? Do you need to eliminate them? In which cases? Why?

****

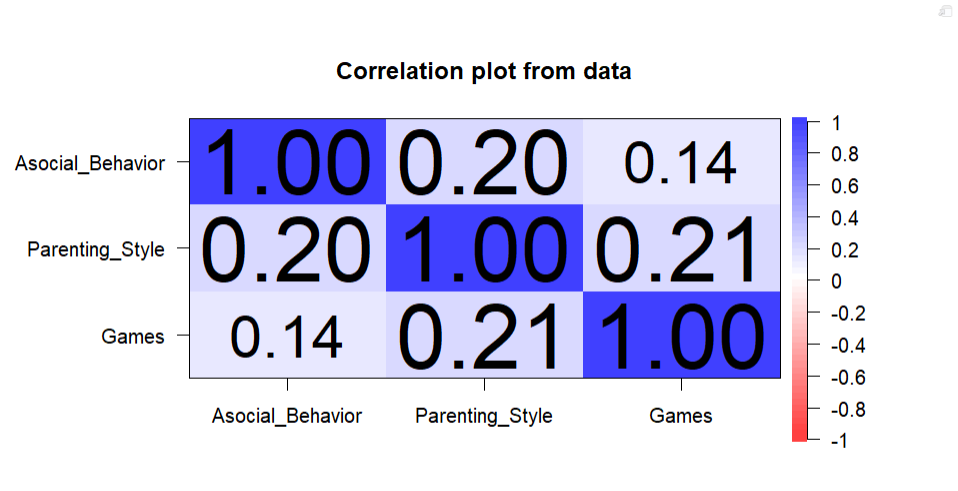
1. Using Q-Q plots, try to conclude on variables normality.

****

1. Using instruments available for you, try to conclude whether the **Asocial Behavior** variable relates to other variables in linear way.

****

1. What types of correlation coefficients can be used to analyze relationships between these variables? Why?



1. What else you can say about the variables under analysis? (optional)

**Q3. Apply Pearson’s correlation coefficient to the variables under analysis (Asocial Behavior, Parenting Style, Games, Sibling Behavior**). **Conclude on the output. (13%)**

1. Which variables are strongly/weakly positively/negatively related?
2. Which correlation coefficients are significant at the 0.01 level (2-tailed)?
3. Which correlation coefficients are significant at the 0.05 level (2-tailed)?
4. What are the effect sizes (r2) for significant correlations?
5. Using *t*-statistic, conclude on whether the relationship between level of asocial behavior and parenting style is stronger than the relationship between level of asocial behavior and the time spent playing computer games.

**Q4. Report on Pearson’s correlation coefficients. (4%)**

**Q5. Apply Spearman’s correlation coefficient to the variables under analysis (Asocial Behavior, Parenting Style, Games,** **Sibling Behavior)**. **What can you say about the output? (10%)**

1. Which variables are strongly/weakly positively/negatively related?
2. Which correlation coefficients are significant at the 0.01 level (2-tailed)?
3. Which correlation coefficients are significant at the 0.05 level (2-tailed)?
4. What are the effect sizes (r2) for significant correlations?

**Q6. Report on Spearman’s correlation coefficients. (4%)**

**Q7. Build two regression models with Asocial Behavior as the output variable using two blocks and Enter method. For the first model use the variable with the strongest correlation with the Asocial Behavior variable. For the second model use all other variables from the list in Q2. (3%)**

1. Produce regression equation for the first model.
2. Produce regression equation for the second model.

**Q8. What can you say about the fit of the first model and about its coefficients? (12%)**

1. Conclude for each predictor variable, if it has made a significant contribution to predicting the outcome. At what level?
2. Interpret the standardized and unstandardized coefficients of the model.
3. What proportion of variance is explained by the first model? Is the first model overall significant? Why?
4. Conclude whether the first model generalizes well based on multivariate R2.

**Q9. Report on the first linear regression model. (3%)**

**Q10. What can you say about the quality of the second model and its coefficients? (18%)**

1. Conclude for each predictor variable, if it has made a significant contribution to predicting the outcome. At what level?
2. Interpret the standardized and unstandardized coefficients of the second model.
3. Which variable has higher degree of importance in the second model according to the standardized beta values and/or the magnitude of the *t*-statistics?
4. What proportion of variance is explained by the second model? Is the second model overall significant? Why?
5. Conclude whether the second model generalizes well based on adjusted R2.
6. How much additional variance is explained by the second model comparing with the first one?
7. Based change statistics, conclude whether the change in R*2* is significant for the second model.

**Q11. Check the second model for the potential sources of bias. (17%)**

1. Using all available plots and residual statistics, try to identify potential influential cases. Explain it.
2. Using VIF and/or Tolerance statistics, conclude on multicollinearity of the predictors.
3. Using test plots or any other technique you know, check assumptions of errors independence, linearity and homoscedasticity of the residuals. Comment on that.

**Q12. Report on the second linear regression model. (4%)**