# Quantitative methods of business research Seminar 2

Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The goal of the second seminar is to study the procedure of correlation and regression analysis in R while applying to dataset.

Let’s look at the example data relating to exam anxiety: a psychologist was interested in the effects of exam stress and revision on exam performance. She had devised and validated a questionnaire to assess state anxiety relating to exams. This scale produced a measure of anxiety scored out of 100. Anxiety was measured before an exam, and the percentage mark of each student on the exam was used to assess the exam performance. She also measured the number of hours spent revising.

For the seminar we will follow the protocol below, and it will be the same for the group home task (although the datafile will be different).

**Protocol for Seminar 2   
(Correlation and regression. Part 1)**

**Q1. Getting acquainted with the data.**

1. How many variables are in the data?

5

1. What can we say about the Exam variable (form of distribution, descriptive statistics)?
2. How many female/male participants we have?

**Q2. Check linear model assumptions and try to spot potential sources of bias for the variables under analysis (Exam, Revise, Anxiety).**

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

**Exam is normal, Revise and anxiety are not normal**

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

**Probably not but linear models can still be used for prediction**

1. What types of correlation coefficients can be used to analyze relationships between these variables? Why?
2. What else you can say about the variables under analysis? (optional)

Non-normality? Outliers?

**Q3. Apply Pearson’s correlation coefficient to the variables under analysis (Exam, Revise, Anxiety**). **Conclude on the output.**

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 is the coefficient of determination (r2) for significant correlations?
5. Using *t*-statistic conclude on whether the relationship between exam anxiety and exam performance is stronger than the relationship between revision time and exam performance.

**Q4. Report on Pearson’s correlation coefficients.**

*Example report:*

A Pearson product moment correlation coefficient was conducted to examine the relationships between the exam performance, the time spent revising, and the exam anxiety. The exam performance is negatively related to the exam anxiety, r = -.38 [−.522, −.194], p < .001, and positively relates to the time spent revising, r = .36 [.188, .481], p < 0.001. The time spent revising is also strongly correlated with exam anxiety, *r* = −.67 [−.830, −.417], p < 0.001. A complete list of correlations is presented in Table 1. The effect size of the exam anxiety (*R2*=0.14) indicates that the level of exam anxiety accounted for a small portion (14%) of the variability in the exam performance.

Table 1 Pearson’s correlations coefficients for the exam performance data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Exam Performance | Time Spent Revising | Exam Anxiety |
| Exam Performance | **1** | **.36\*\***  **[.188, .481]** | **-.38\*\***  **[−.522, −.194]** |
| Time Spent Revising | **103** | **1** | **-0.67\*\***  **[−.830, −.417]** |
| Exam Anxiety | **103** | **103** | **1** |

\*\* Correlation is significant at the 0.01 level (2-tailed).

**Q5. Apply Spearman’s correlation coefficient to the variables under analysis (Exam, Revise, Anxiety**). **What can you say about the output?**

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.**

*Example report:*

A Spearman’s correlation coefficient was conducted to examine the relationships between the exam performance, the time spent revising, and the exam anxiety. The exam performance is negatively related to the exam anxiety, *rs* = -.34 [-.496, -.161], p < 0.001 and positively related to the time spent revising, *rs* = .32 [.129, .491], p < 0.001. And the time spent revising is also related to the exam anxiety, *rs* = −.56 [-.736, -.359], p < 0.001. A complete list of correlations is presented in Table 2. The effect size of the exam anxiety (*Rs2* = 0.10) indicates that the level of exam anxiety ranks accounted for a small portion (10%) of the variability in the exam performance ranks.

Table 2 Spearman’s correlations coefficients for the exam performance data

|  |  |  |  |
| --- | --- | --- | --- |
|  | Exam Performance | Time Spent Revising | Exam Anxiety |
| Exam Performance | **1** | **0.32\*\***  **[.129, .491]** | **-0.34\*\***  **[-.496, -.161]** |
| Time Spent Revising | **103** | **1** | **-0.56\*\***  **[-.736, -.359]** |
| Exam Anxiety | **103** | **103** | **1** |

\*\* Correlation is significant at the 0.01 level (2-tailed).

**Q7. Build the first regression model (model 1) with Exam as the output variable and the variable with the strongest correlation with the Exam variable as the regressor. What can you say about the fit of the first model and about its coefficients?**

1. Produce regression equation for the first model.

**Exam = 98.57 – 0.57 \* Anxiety**

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.

**Example:**

Anxiety (b = -0.57): This value indicates that as exam anxiety increases by one unit, exam performance decreases by 0.57 units.

1. What proportion of variance is explained by the first model? Is the first model overall significant? Why?

**Q8. Report on the first linear regression model.**

*Example report:*

A bivariate regression was conducted to examine how well exam anxiety level could predict the exam performance. A scatterplot showed that the relationship between the exam anxiety and the exam performance was negative and linear and did not reveal any outliers. The regression equation for predicting the exam performance was Exam = 98.57 – 0.57 \* Anxiety. The R2 for this equation was 0.14. That is 14% of the variance in the exam performance was predictable from the level of exam anxiety. More information is presented in Table 3.