# Quantitative methods of business research 2023

# Hometask 1 Report (individual assignment)

Student Names

WANG GENG

For the hometask we will use the pre-made SPSS dataset created out of BEEPS data collection.

**Download the file** BEEPS\_2019\_homework1 - CohХ.sav **and copy it to your virtual machine.**

**Please follow the questions below to explore and analyze the data.**

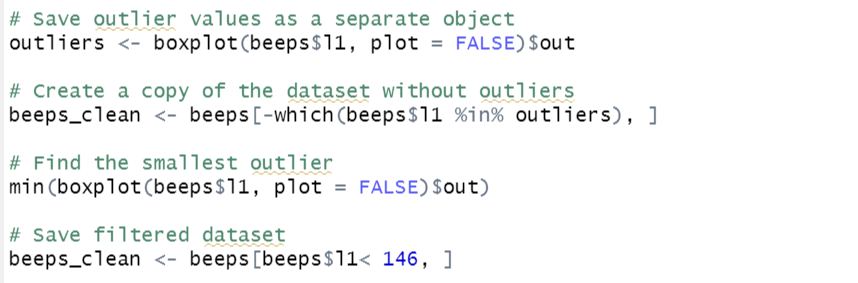
**Q1. Getting acquainted with the data (10%)**

**Use variable view and descriptive statistics functions to explore the data in general and answer the questions:**

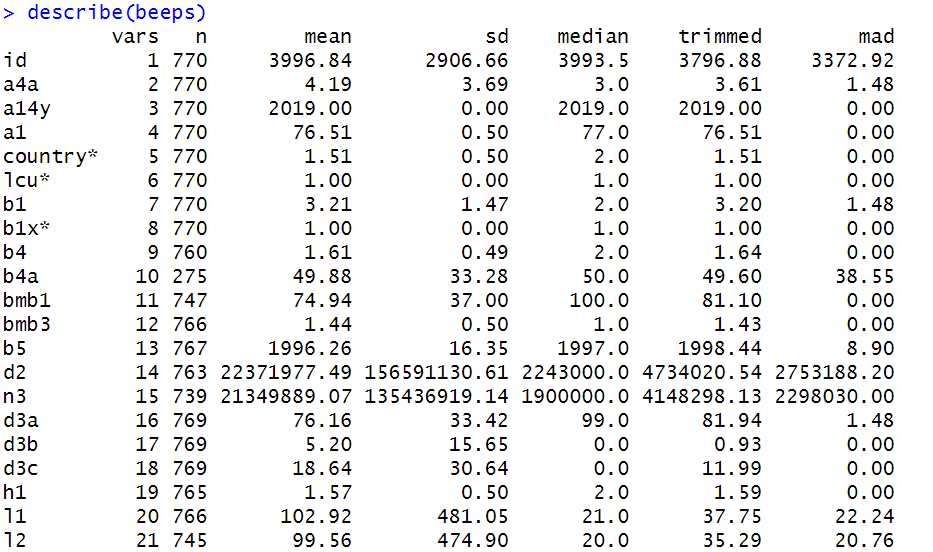
How many variables are included in the original dataset? How many of them are **nominal**?

The number variables is 21

What and how should be corrected in the dataset (check the variable view tab). Report on all corrections you make to the dataset.



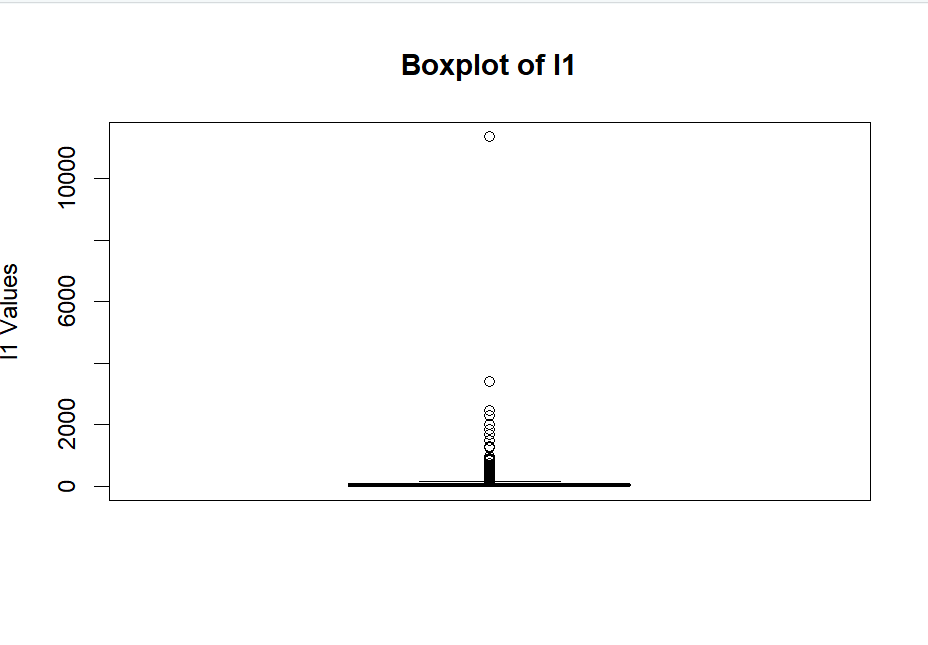
Produce a summarizing report on key descriptive measures for the variables in the dataset taking into account their measurement levels.



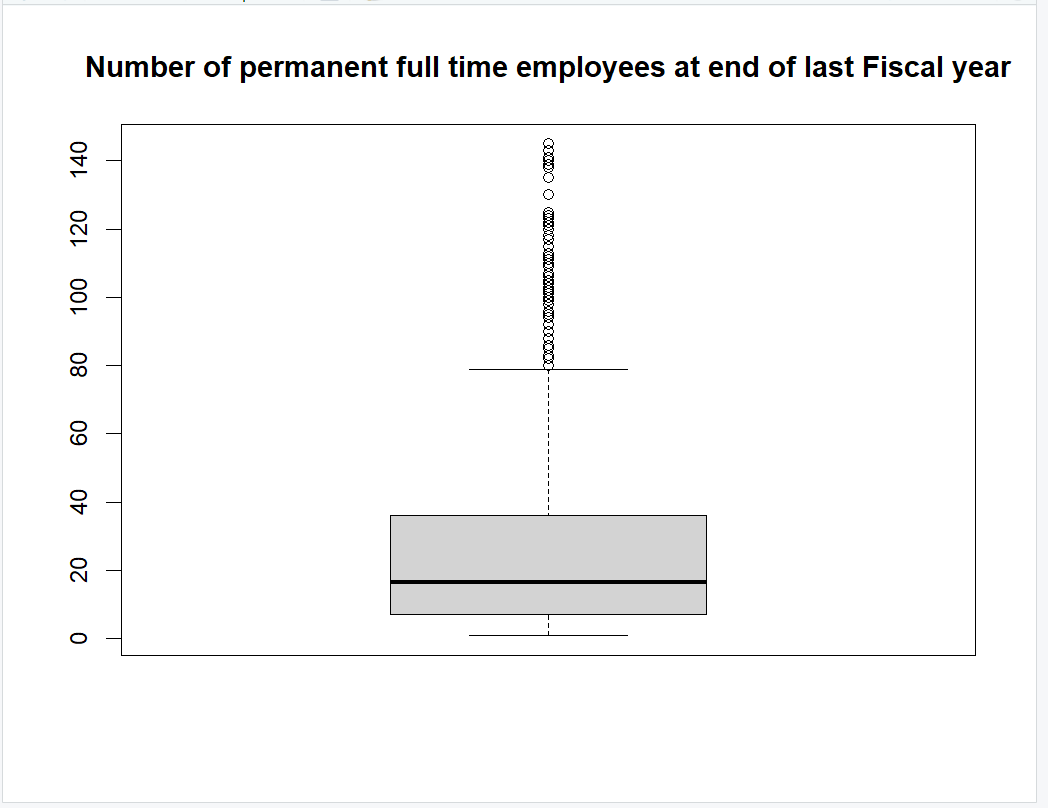
The number of permanent full-time employees at the end of last fiscal year mean is 102.92. The standard deviation is 481.05. The median is 21. The number of permanent full-time employees 3 years ago mean is 99.56.The standard deviation is 474.90. The median is 20.0

**Q2. Analyzing for outliers and normality (15%)**

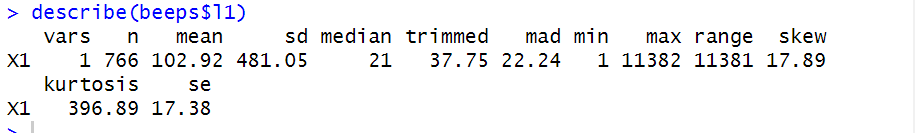
For variable ***l1*** build the boxplot. What can you say from the chart regarding presence of outliers and extreme values? **Paste the chart in the report and comment on it:**

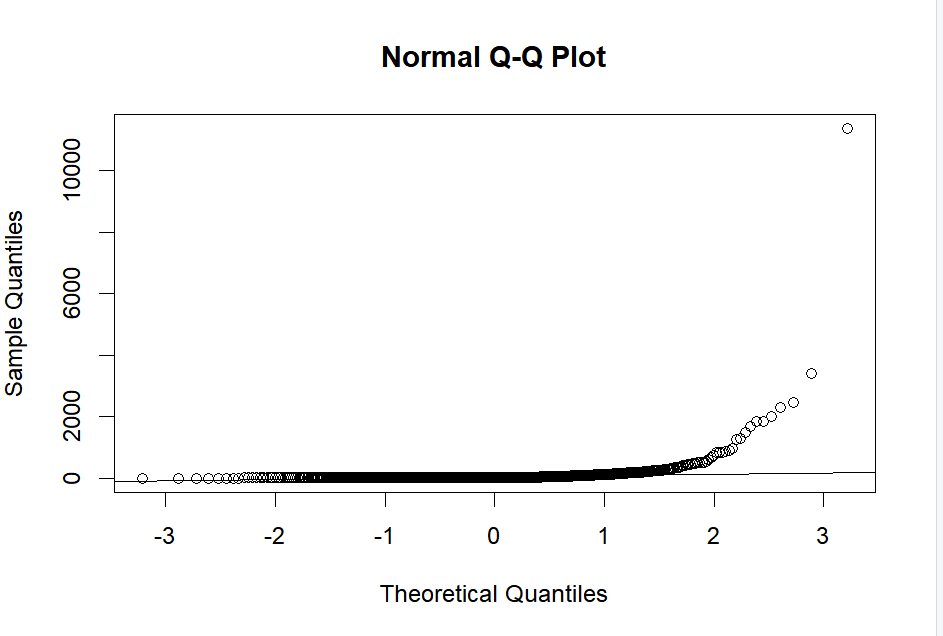
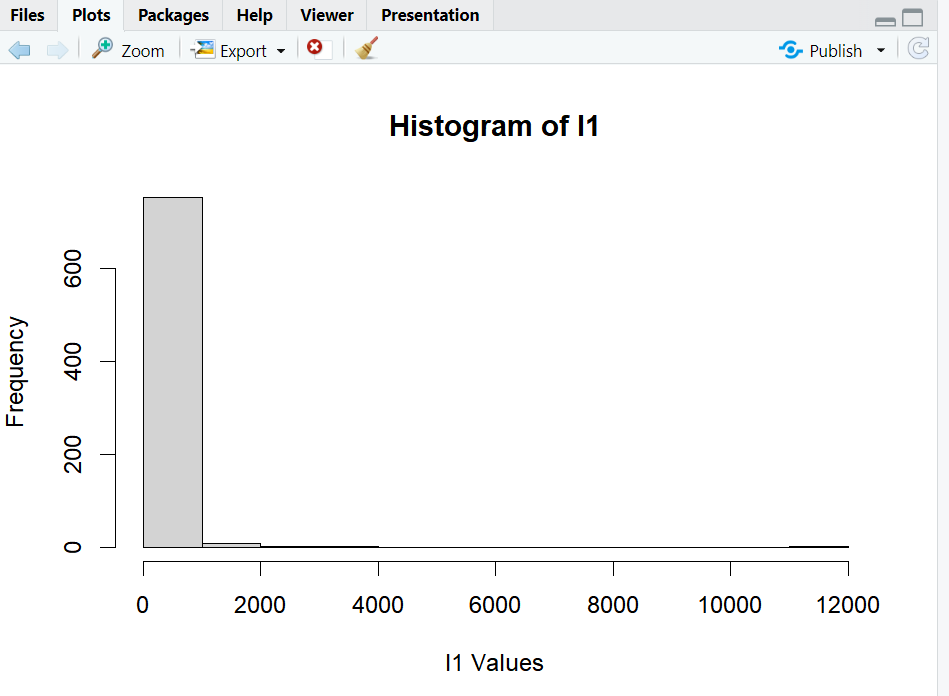


If there are outliers and/or extreme values, what will be your decision on them and why? Report your decision with arguments and necessary output tables | graphs from RStusio.



What can you say about the normality assumption for the variable ***l1*** based on descriptives, Q-Q plot, histogram and statistical tests? What will be your conclusion on normality of this variable? Why? **Paste the graphs and output tables in the report and comment on them and the final solution below:**

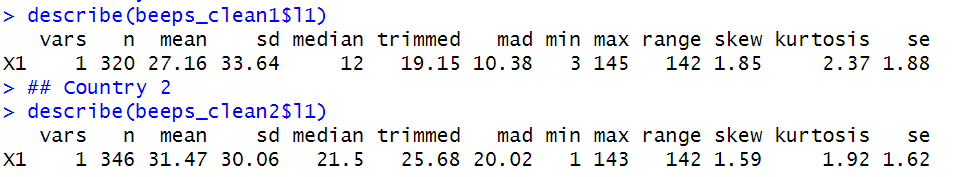




In terms of the p-value, it is evident that the data provides substantial evidence to reject the null hypothesis. Typically, a p-value less than 0.05 is considered statistically significant, indicating the ability to reject the null hypothesis. A smaller p-value suggests a stronger rejection, often observed in large datasets or with pronounced effects.

**Q3. Find the mean number of permanent, full-time employees for the sample in total. Use an appropriate test and check if the mean number of permanent, full-time employees working in companies from COUNTRY 1 is equal to this reference value. How about COUNTRY 2? (10%)**

Provide descriptive statistics for the two countries regarding the number of permanent, full-time employees. What can you say based on this data?



**Country 1** :Sample Size (n): 320

Mean: 27.16

Standard Deviation (sd): 33.64

Median: 12

Trimmed Mean: 19.15 (This is the mean when a percentage of the highest and lowest values are removed)

Median Absolute Deviation (mad): 10.38 (A robust measure of variability)

Minimum (min): 3

Maximum (max): 145

Range: 142

Skewness: 1.85 (positive value indicates a tail on the right side of the distribution)

Kurtosis: 2.37 (values > 0 indicate a "peaked" distribution compared to a normal distribution)

Standard Error (se): 1.88

**Country 2:**Sample Size (n): 346

Mean: 31.47

Standard Deviation (sd): 30.06

Median: 21.5

Trimmed Mean: 25.68

Median Absolute Deviation (mad): 20.02

Minimum (min): 1

Maximum (max): 143

Range: 142

Skewness: 1.59 (slightly less than Country 1, but still indicating a right tail)

Kurtosis: 1.92 (less "peaked" compared to Country 1)

Standard Error (se): 1.62

Country 2 generally has more employees on average than those in Country 1. However, there's greater variability in the number of employees in Country 1. While both countries have some companies with an exceptionally high employees, the extremes are more evident in Country 1. Notably, despite these differences in distribution patterns, both countries have organizations with a similar range in the number of employees.

What are the null and alternative hypotheses for this test?

**For country 1:** Null Hypothesis (H0): The mean number of permanent, full-time employees in country 1 is equal to the reference value.

Alternative Hypothesis (H1): The mean number of permanent, full-time employees in country 1 is not equal to the reference value.

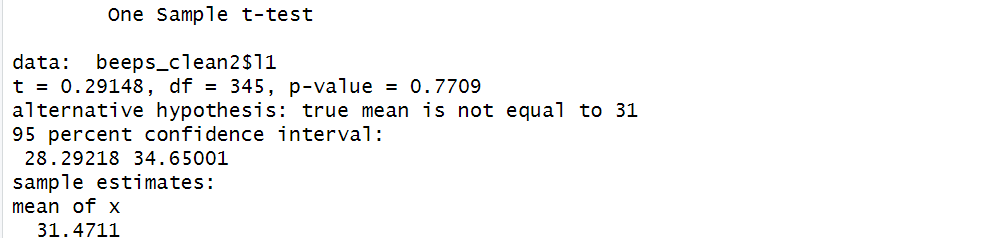
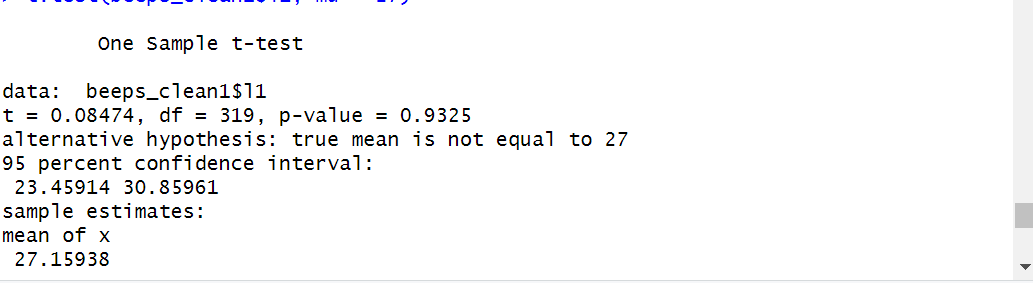
**For country 2:** Null Hypothesis (H0): The mean number of permanent, full-time employees in country 2 is equal to the reference value.

Alternative Hypothesis (H1): The mean number of permanent, full-time employees in country 2 is not equal to the reference value.

Which test should be used and why?

One-sample t-test. Because the one-sample t-test is used to determine if the mean of a single sample is statistically significantly different from a known or given value.

## Paste the output table(s) from RStudio and interpret the results. Do not forget the meaningful explanation (not only the statistical one!), i.e., the answer to the initial question.



**Q4. Now use an appropriate test and check if the average number of permanent, full-time employees working in companies of COUNTRY 1 and the similar number in COUNTRY 2 are equal. (10%)**

What are the null and alternative hypotheses for this test?

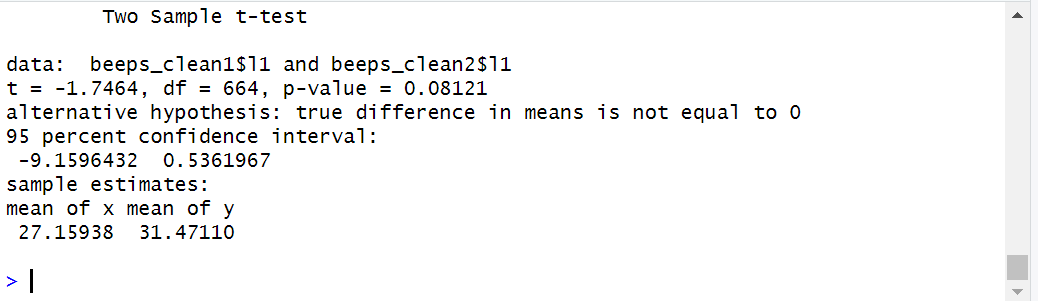
Null Hypothesis (H0): The average number of permanent, full-time employees working in companies from country 1 is equal to the average number of permanent, full-time employees working in companies from country 2.

Alternative Hypothesis (H1): The average number of permanent, full-time employees working in companies from country 1 is not equal to the average number of permanent, full-time employees working in companies from country 2.

Which test should be used and why?

The two-sample t-test

**Paste the output table(s) from RStudio and interpret the results**. Do not forget the meaningful explanation (not only the statistical one!)



The test t is -1.7464. This negative value suggests that the mean of the first group is less than that of the second group. With the p-value being below the threshold, it indicates a significant difference between the group means at the 10% level. Analyzing the provided data, companies in the first group have an average of 27.16 employees, which is fewer than the 31.47 average employees in the second group. Statistically speaking, at a 5% significance level, there isn't sufficient evidence to assert these differences are significant. However, at a 10% significance level, this difference is considered significant.

**Q5. Use an appropriate test and check if the average number of permanent, full-time employees does not differ for companies with different legal status. (15%)**

What are the null and alternative hypotheses for this test?

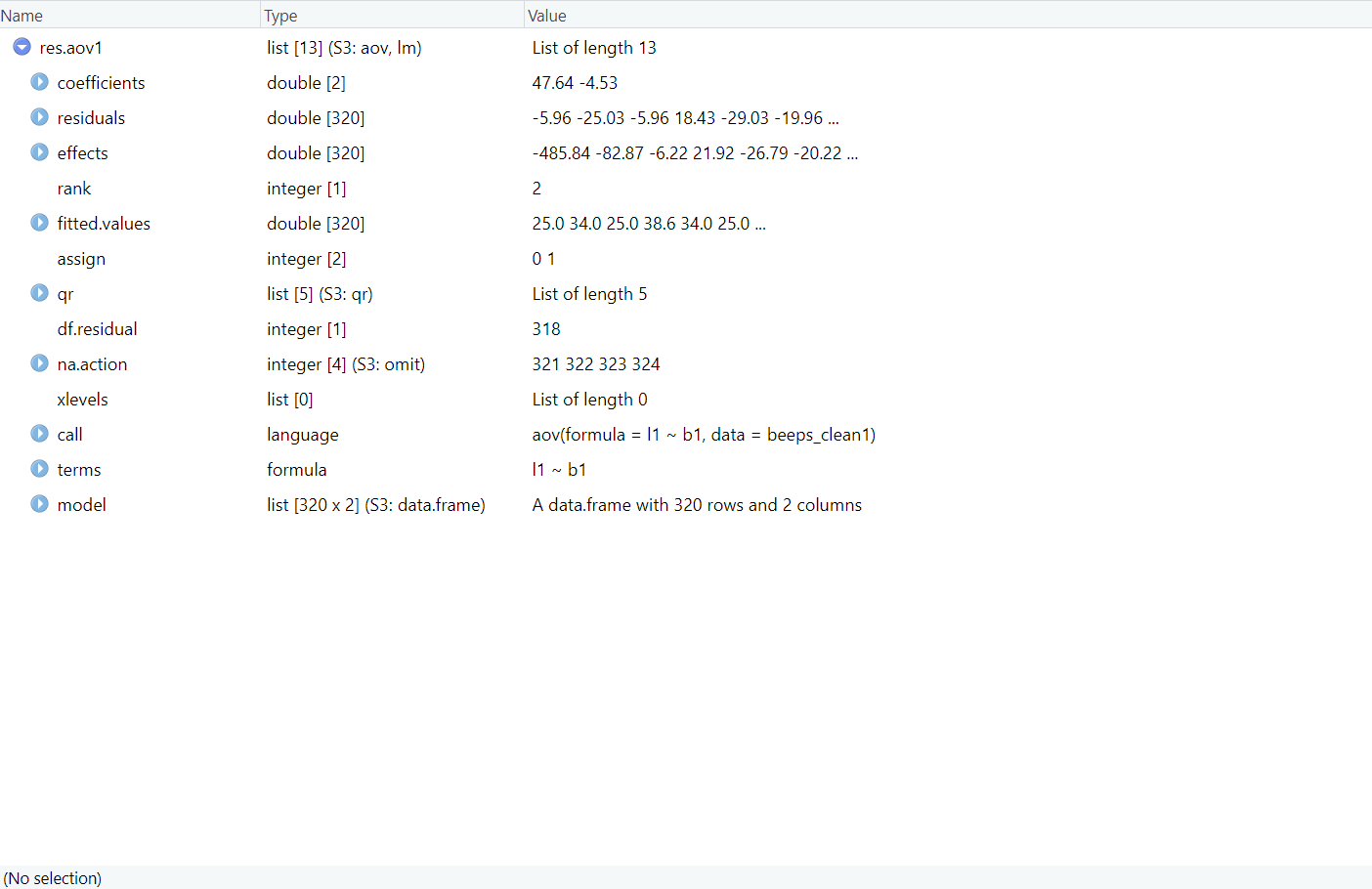
Null Hypothesis (H0): The average number of permanent, full-time employees is the same across all legal status categories. In other words, there is no significant difference in the means of permanent employees among the legal status groups.

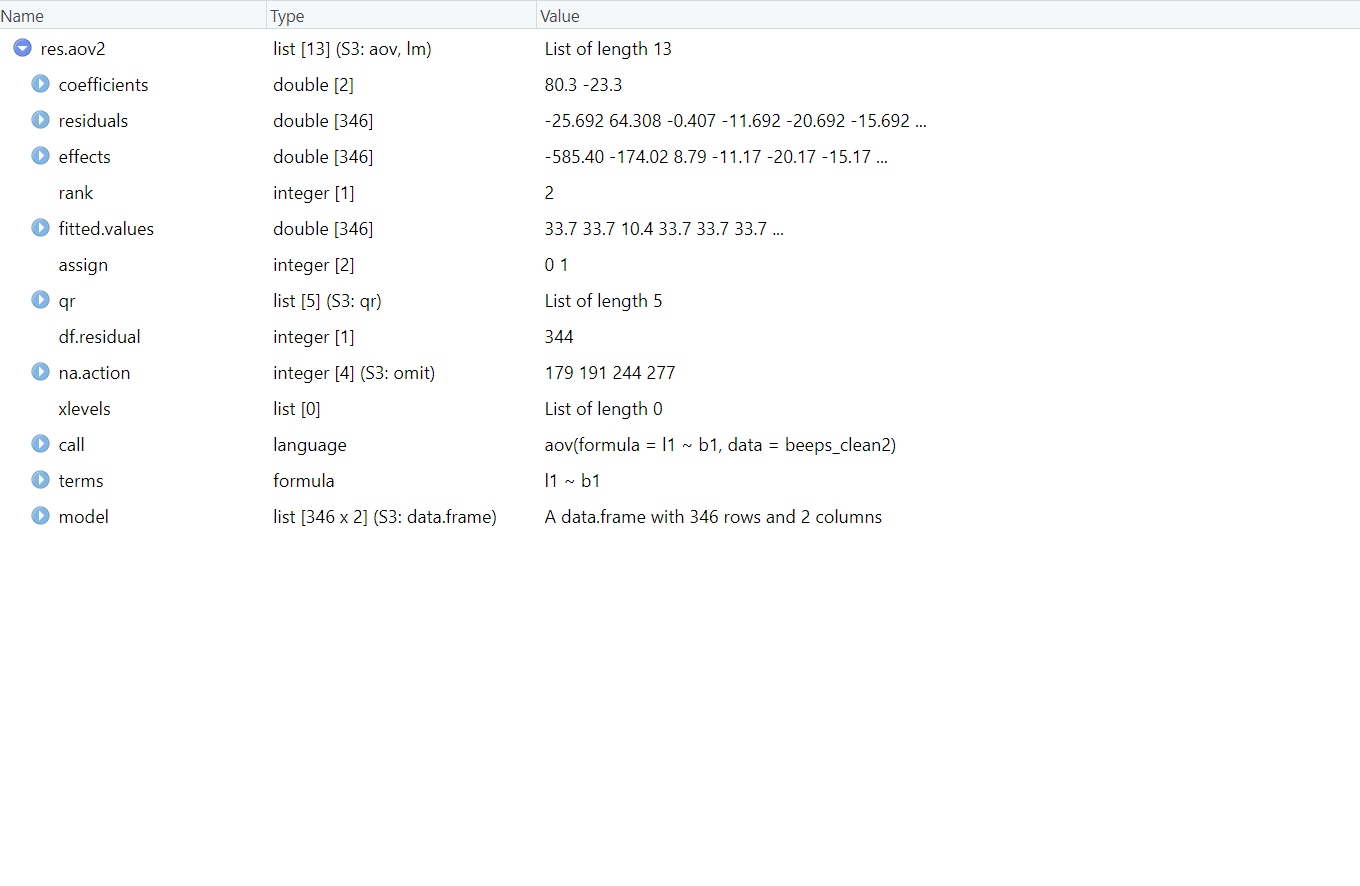
Alternative Hypothesis (H1): The average number of permanent, full-time employees is not the same across all legal status categories. There is a significant difference in the means of permanent employees among the legal status groups.

Which test should be used and why?

ANOVA

**Paste the output table(s) from RStudio and interpret the results**. Do not forget the meaningful explanation (not only the statistical one!)





**Q6. Now create a new variable representing exporter status of the firm (similar to the Seminar 1 case). Add values labels to the values and provide frequency analysis for this new variable. (10%)**

**Q7. Choose grouping variables that you have not used for analysis yet and analyze the differences in total sales between companies from different groups. Perform at least 2 tests with different groupings. (30% total)**

Explain your choice of grouping variables – why are they interesting?

Provide full report on test 1

Provide full report on test 2

What is your overall managerial conclusion based on the performed analysis?