

2.1 Two-way ANOVA

1. In this task we analyse the dataset of a controlled experiment in which the cholesterol level LDL was measured in [mg/L]. The study investigated two different factors: The effect of medical drugs (3 groups: Placebo, Drug A and Drug B) and the factor lifestyle (2 groups: 1 hour of exercise and low carb diet). The data in Excel is labelled as follows:

Import the Excel dataset from the first sheet. Arrange the dataset in a from (data.frame) that is manageable by R. This means create a three-by-N structure with N the total number of samples. The second columns contains the label of the factor the label of the factor. Print the first few entries of your dataset in the report.

	Placebo	Drug A	Drug B
Exercise	PE	AE	BE
Diet	PD	AD	BD

2. Conduct a two-way ANOVA using the function `lsteinlineaov`. Analyse for the group, factors and the interaction between factors and groups. Print and Analyse the results.
3. Make a boxplot using `ggplot`. Plot the numeric depended values on the y-axis and different groups along the x-axis. Plot data from different factors with different colors. (You should have 6 boxes.) Analyse the results of the plot with the aid of interactions plot discussed in the lecture.
4. Conduct a post-hoc test. Which groups do significantly differ from each other? Discuss the results.

2.2 Linear regression

5. In this task we analyse the dataset of an analysis in which the cholesterol level LDL was measured in [mg/L] in dependence of the weight of an animal species in [mg]. Import the Excel dataset from the second sheet "Linear regression". Compute and report the means and variance of the explanatory and dependent variable. Show the scatter plot.
6. Compute the parameters for the standard linear model ($y = mx + b$) using R's `glm`-command. Report and analyse the output. Can we predict values beyond the x-values from the dataset? Or in other words is the linear model valid for any value of x beyond the dataset? Justify your answer.
7. Compute the Confidence intervals with $\alpha = 5\%$. Report the confidence intervals for both slope and y-intercept.
8. Show the scatter plot with regression line and confidence interval band, using the. Add the two confidence interval regression lines with the parameters you have obtained in the previous question.
9. Print the scatter plot of the residuals and analyse the plot of the residuals. Remember the analysis of residuals conducted in the lecture. What do you observe?