# Assignment 4

*Marketing Research*

---

**General Note for the following tasks**

Please do the following tasks and copy the solutions into a word file, which you then please convert into a PDF and upload to Learn@WU. Don't forget to copy the R codes into your solution. You have to be prepared to present your assignment solutions in the next session. In case you are selected to present a subtask of this assignment, please be prepared to answer questions regarding the theoretical concepts underlying the specific task, e.g., "How do you calculate the degrees of freedom of the test-statistic? How is the test-statistic distributed?"

---

Please first follow the below instructions:

- Use the file "kurs3_data_0319.csv" from the last assignment (Note: You can find an explanation of the variables in the document "readme_variables.txt").
- Set the Working Directory in R to the folder, in which you saved "kurs3_data_csv_0319.csv".
- Read in the data file.

## Task 1 Correlation (Chapter 6.1; 2 Points)

- Produce a scatterplot:
  - Number of ads (`number_ads`) on the horizontal and number of impressions (`impressions`) on the vertical axis.
  - `impressions` on the horizontal and number of clicks (`clicks`) on the vertical axis.

Comment on the correlation of the variables regarding the scatter plots. Which correlation coefficient do you suspect to be larger?

- Calculate the covariance and the correlation coefficient of the variables below:
  - `number_ads` and `impressions`
  - `impressions` and `clicks`

Comment on the correlation of the variables regarding the covariance and correlation coefficient (e.g., strength of the association, direction of the association).

Hint: Have a look at Chapter 6.1.1. The scatterplot (Figure 6.4) was created with the follwing code:

```
library(ggplot2)
attitude <- c(6, 9, 8, 3, 10, 4, 5, 2, 11, 9, 10, 2)
duration <- c(10, 12, 12, 4, 12, 6, 8, 2, 18, 9, 17,
    2)
att_data <- data.frame(attitude, duration)
att_data <- att_data[order(-attitude), ]

ggplot(att_data, aes(x = duration, y = attitude)) +
    geom_point() +
    theme_bw()
```

## Task 2: Simple Linear Regression (Chapter 6.2.1; 4 Points)

You are dubious about the relationship of the given variables in Task 1, so you decide to make a deeper investigation:

- Run regression analysis to make sure the effect of
  - Number of ads on number of impressions, i.e. `impressions~number_ads` (regression 1)
  - Number of impressions on number of clicks, i.e. `clicks~impressions` (regression 2)

- Interpret your results (e.g., number of ads has a positive (or negative) effect on number of impressions by x much; see 6.2.1.1 Estimating the coefficients).
- Prediction (6.2.1.4 Using the model):
  - How many impressions would a website owner sell on average if she published 5 ad slots on her website?
  - How many clicks would a website owner generate on average if she sold 1M impressions on her website?

## Task 3: Multiple linear regression (Chapter 6.2.2; 4 Points)

You further want to investigate the effect of both number of ads and number of impressions on number of clicks.

- Run the appropriate regression.
- Compare the obtained results with above regression 2 results and comment on whether the model fit has improved.
- Interpret the effect of number of impressions on number of clicks.
- How many impressions would a website owner sell on average if she published 5 ad slots and sold 1M impressions on her website?