Research Experiment
Data Analytics – Project 1

In this project, we will use the scientific method and the concepts of descriptive and inferential statistics to give an answer to a given research question. The exercise was initiated in the class but now, we'll do the full exercise by using R and the methodology learnt with A/B testing (inferential statistics). You are also required to use RMarkdown and nice plotting libraries from R.

The research question that we are trying to answer is:

Are designers and producers more extrovert than programmers and artists?

To answer this question, please follow the next steps:

1. **Draw the experiment**
   Identify the independent variables and the dependent variables of this experiment.

2. **Define the hypotheses**

   Write the null hypothesis and the alternative hypothesis.

3. **Collect and explain the data**
   We will use the dataset that we collected in class, with the results of the Eysench personality test. Explain which of the available data in the dataset you will use: which datasets (years) you will use and which variables. Then, explain what instances you will allocate to each of the samples. Explain your criteria.

   The dataset is available here:

   https://docs.google.com/spreadsheets/d/11A_CCrN06Prx5NFO83YnQiG972qqqDbZPNSqs5E6AUE/edit?usp=sharing

   You need to access with the edu Tecnocampus gmail account.

   Please, do not modify the original dataset, since it needs to be accessible to everyone as it is (with all the imperfections and errors). Download the dataset to your computer and modify it in your own local folder, and inside the R program.

4. **Preprocess the data**
   Once you have decided the data that you will use, download the data in your computer and upload it into the R program. Then, preprocess the data so that it can be used for this analysis. This may require removing data that is incomplete or contains NAs. Also, decide how you will use individuals with different profiles from

the ones that are interesting for our research question. Think what you will do with multidisciplinary profiles as well.

5. **Descriptive statistics**
   Draw some plots to show the main trends in the data. Interpret the plots.

6. **Inferential statistics**
   Apply inferential statistics to answer the research question at 95% and 90% confidence level. You need to write an R program that performs all the calculations. Show the critical value, the observed value, and the p value.

7. **Confirm or reject the hypotheses**
   Based on the analysis of the previous section, conclude whether you can confirm or reject the hypotheses.

8. **Answer the Research Question**
   What is your answer for the research question? Provide a justification for your answer.

9. **Other research questions**
   Based on the data that you have in your hands, write another research question of your choice and perform an A/B test on this question. Do all the steps for the new research question and provide an answer.

10. **Summary**
    Write a summary of the main insights of this analysis.

Practical issues to perform this project:

- This project illustrates the basics of R programming for data analysis, visualization, reporting, as well as descriptive statistics and inferential statistics. Thus, it is important that you do this consciously because it serves as a learning experience for the contents of the class so far. You need to use RMarkdown and ggplot.
- You can work in teams up to 4-5 people, but be sure that everyone in the team is knowledgeable about all the contents. You can split in two sub-teams: one does the first part of the project, and the second team does the second part (the new research question). Then, together, you prepare the report, agree on the summary and revise the whole contents before submitting.
- Structure the project nicely. A part of the grade goes to the document format and the structure.
- You need to deliver the source file (Rmd) and the html/PDF output file. To do this, you'll need to use RMarkdown, as it is explained in class.
- The deadline for this project will be published in ecampus.