

## Home exam TAU R workshop 2015

The purpose of the exam is to summarize everything you learned during the workshop and to test the level of your understanding of R and statistical analysis. Please write the R codes in a script (.r file), and the answers to the theoretical questions and the plots in a word file. In addition you are asked to add the .csv of your data when sending us your exam. The last date for submission is a week from the end of the course on the 17.02.15 at 23:59. Your exam should be submitted to [marianovosolov@gmail.com](mailto:marianovosolov@gmail.com)

The exam is divided to two parts:

1. Questions based on data that is found in the website.
2. Questions based on your data

Please read the exam carefully before starting your work.

You have one week to finish it and send it to us

### Part I- Our data (50 points)

1. **a.** For the `final_exam_data.csv` data, construct a linear model with the ‘brood’ variable as response, and the ‘clutch’, ‘insularity’ and the interaction between them, as predictors. Explain the output of the model (estimates, R squared and p-values). Shortly explain the interaction term and its meaning. (10 points)  
**b.** Compare the previous model to a model without the interaction term, using the AIC. Explain the result shortly. Which model would you choose? (5 points)  
**c.** Test the normality assumptions of the model; add an appropriate plot and a p-value. (5 points)  
**d.** Create a scatterplot of ‘brood’ Vs. ‘clutch’. Add the variables ‘insularity’, ‘productivity’, and ‘family’ to the plot. Finally, add the lines from the linear model in (1. a) with colors matching to the ‘insularity’ variable (\*hint hint\*). (15 points)
2. Create a dataset with the same columns as the original dataset. Each row should contain the mean values of all numerical variables, for each categorical variable combination in the data set (hint: `plyr`). (15 points)

**Part II- Your data (50 points)**

1. Preparing the data (10 points)
  - a. Describe your data shortly (how did you obtain it, what does it mean etc.)
  - b. What is your research hypothesis?
  - c. What variables did you use and what is their type (ordinal, categorical etc.)
  - d. What statistical test did you choose to test the hypothesis and what are its assumptions?
2. Analyzing the data (20 points)
  - a. Write the code for performing your chosen statistical test.
  - b. Write the output of the test, and explain the meaning of the parameters.
  - c. Explain the results of the test.
3. Visualizing the data (20 points)
  - a. Choose one type of visualization for your results and explain why you think this visualization is the most suitable.
  - b. Write the code for the visualization you have created.
  - c. Attach the visualization and add an appropriate legend.

**Good Luck!**