Statistics with Python

Project

Alejandro Martinez (CentraleSupelec)

Project

Project: general description

- Pose a question that you would like to investigate. If possible, choose something related to your major or to the ongoing research in ESIEE!
- Find or collect data that will help you answer this question (you may need to edit your question based on available data)
- Use statistics and Python to answer your question
- 2 students/project!

Project

Project: more precisely

The result will be a **three to five pages report** including

- Description of the data collection method
- Descriptive statistics (summary stats, visualization) + comments
- Hypothesis testing (p-value)
- Optional: Hypothesis test for the mean value of groups
- Python commands

Project due to 26/01/2018
pdf. or other typewritten form files must be send at dina.irofti@cea.fr

Project

Recommended structure for your project

- PART I: Introduction [2 points]
  - Clearly state what question you want to answer
- PART II: Data Collection [4 points]
  - How was the data collected?
  - Did you use existing data or you did collect it?
  - What commands did you use to import it in Python?
  - What is/are the variable(s)?
  - What is the sample and what is the population?
- PART III: Exploratory Data Analysis [6 points]
  - Perform relevant descriptive statistics, including summary statistics and visualization of the data
  - Make sure to interpret these results in your own words, don't just give the Python output!

Project

Recommended project structure

- PART IV: Hypothesis Test [6 points] (+Bonus: 3 points)
  - Conduct a hypothesis test
  - If your question does not yield obvious hypotheses, just make some up
  - Be sure to clearly state your hypotheses, and interpret all results in context
  - Check the normality
  - Use p-value(s) and give confidence intervals!
  - Minimum: comparison of one group to a fixed value
  - For the bonus: comparison of two groups with respect to each other
- PART V: Conclusion [2 points]
  - What have you learned? Interpret and discuss your results
  - Answer your original research question posed in Part I

Project

Data for the Project

- Collect your own data
  - Representative sample?
  - Randomized experiment?
- Use existing data
  - Internet!
  - Ongoing research in your one ESIEE department (e.g. René Natowicz paper)