

CENG 3516

Statistical Computing

Instructions for Makeup Project

General Description

For the makeup project, you address datasets generated by the **Izmir Municipality** (IBB Acik Veri Portali <https://acikveri.bizizmir.com/dataset>) and analyze these datasets with the statistical methodology we learnt in CENG 3516.

Objective: Using at least three datasets from IBB Acik Veri Portali to analyze and report an important problem such as economics, energy, traffic etc.

1. Select a theme. Example: energy.
2. Find relevant datasets from the portal
3. Synthesize all the material from the course. Hence, it's one of the best ways to solidify your understanding of statistical methods. You must apply the following five methods. Examples are provided just to give you an idea. Please don't just repeat them and try to be creative.

a. Data visualization and descriptive stats

Example: Explore traffic violations based on region.

b. Confidence intervals

Example: You can compare confidence intervals of traffic fines for Region A and Region B.

c. Hypothesis testing

Generate at least 3 hypothesis and test them. Assume normality for each group if normality check fails.

Example: Are there significant difference between mean traffic fines of Region A and Region B?

d. ANOVA or Regression

Example: Are there significant differences between mean waiting times of different Subway locations?

Example: Can we create a model for

$\text{parking_fees} \sim \text{location} + \text{parking_type} + \text{parking_capacity}$? Can we create different models per location?

Project grading guidelines

I will be looking for the following characteristics:

- a. Completeness: Did you follow the guidelines above and provide the source code?
- b. Clarity: Is it easy for your reader to understand what you did and the arguments you made?
- c. Relevancy: Did you use statistical techniques wisely to address your question?

- d. Interest: Did you tackle a challenging, interesting question (good), or did you just collect descriptive statistics (bad)?
Important: Did you evaluate your results?

Guidelines for making an effective report

An effective report communicates your project in a clear and concise fashion. The

RMarkdown report should address the following five points:

- a. Data collection: Explain your datasets. Include and explain variables you analyzed.
- b. Analyses: Describe the analyses you did. Explain why you believe these methods are justified if needed.
- c. Results: Present relevant results of the statistical methods listed above.
- d. Discussion: What implications do your results have for the population you sampled from? What could be done to improve the study if it was done again? What types of biases might exist?

Please post your.html files. The name of the files should be **your name_surname_hwfinal** (like **zeynepfiliz_eren_hwfinal.html**) through DYS.