

CENG 3516

Statistical Computing

Instructions for Final Project

General Description

For the final project, you address **COVID 19 Awareness** of a sample and analyze your dataset with the statistical methodology we learnt in CENG 3516.

Objective: Measuring COVID 19 Awareness of subjects by asking them relevant questions.

1. You need to create **your own** COVID 19 Awareness quiz or test (at least 10 questions) and collect data from at least 30 people. Here is an example quiz: <https://www.srmist.edu.in/forms/sph/covid-19-awareness-champions-quiz>
Please feel free to search for other quizzes available to get inspired and make sure you have your signature questions. You may use Google forms <https://docs.google.com/forms> to collect data.
2. Give 10 points for each correct answer.
3. Calculate an awareness score by summing all correct answers for each respondent and call it "total.score".
4. You also need to ask 5 demographics questions (examples; gender, income, education etc.) before your quiz. You will use these questions to generate hypothesis.
5. The final project requires you to 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 total.score based on demographics.
 - b. Confidence intervals
Example: Can we compare confidence intervals of mean total.score for men and women?
 - 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 total.score of men and women?
 - d. ANOVA or Regression
Example: Are there significant differences between mean total.score of different income levels?

Example: can we create a model for $\text{total.score} \sim \text{gender} + \text{income} + \text{gpa}$? Can we create different models per gender?

Project grading guidelines

I will be looking for the following characteristics:

1. **Completeness:** Did you follow the guidelines above and provide the source code?
2. **Clarity:** Is it easy for your reader to understand what you did and the arguments you made?
3. **Relevancy:** Did you use statistical techniques wisely to address your question?
4. **Interest:** Did you tackle a challenging, interesting question (good), or did you just collect descriptive statistics (bad)? Or 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:

1. **Data collection:** Explain your quiz and how you collect data. Include questions you asked. Also, include response rates.
2. **Analyses:** Describe the analyses you did. Explain why you believe these methods are justified if needed.
3. **Results:** Present relevant results of the statistical methods listed above.
4. **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 eralp_dogu_hwfinal.html) through DYS.