PROJECT FLOW AND REQUIREMENTS

- *Please prepare your project in R Markdown in the following format which is described in this document.*
- *"Clear comments before and after every little step" and "clear data visualizations with desriptions" will make good points.*
- *Don't ever put the whole data in the project report!! If you want to show your data, please just show the first rows (6-10); or use summaries- data visualization techniques instead of data itself.*
- *Write your names in the author part.*
- *Put your final docs (Project report as an html and a short video of no more than 10 minutes) in dys.*
- *You will be graded as: (82p following steps; 10p report clarity and quality, 10p video presentation) ☺*

1. **(3p) Data:**
   Please find your original dataset or datasets; and describe your data in the first step.

2. **(3p) Exploratory and descriptive data analysis:**
   Use "Exploratory and descriptive data analysis". Talk about your categorical and quantitative data or your ordinal variables etc. Write down your comments.

3. **(3p) Data Visualization:**
   Use at least 4 diffrerent "data visualization techniques" and talk about your data further (distribution, outliers, variability, etc). Use at least 2 of the visualizations to compare two groups (like female/male; smoker/non-smoker etc).

4. **(3p) Central Limit Theorem:**
   Choose one of your variables. Write down your own function which proves the central limit theorem (CLT): show that whatever the distribution is, the sampling data will be distribured normally. (have a look at the example in Statistics_Lecture_5, page9)

5. **(3p) Confidence Intervals:**
   Build 2 confidence intervals and make "clear comments" about your findings.

6. **(3p) Transformation:**
   Try to make one transformation (log transformation, Box-Cok transformation, etc) for one of your quantitative variables, which is not normally distributed; but will be normal or more normal, after the transformation.

7. **(2p every item) Single t-test (Welch t-test or Wilcoxon rank-sum test)**
   Implement a single t-test for one of your "normally or not-normally distributed" variable:

   a. **Aim**
      In words, what is your objective here?

b. **Hypothesis and level of significance:**
   Write your hypothesis in scientific form and determine the level of singnificance.
c. **Assumption Check:**
   Check the required assumptions and "comment on each of them is a must!".
d. **Indicate "which test you choose" "for what reason"**
e. **Result:**
   Give the output of the test and write down the result (ex: since p value is less than alpha, I reject the null hypothesis).
f. **Conclusion:**
   You got your result in item e. Write down the conclusion of your result, in such a way that, the reader who doesn't know any statistics can understand your findings.
g. **What can be Type-1 and Type-2 error here?**

8. **(2p every item) Paired t-test:**

   a. **Aim**
      In words, what is your objective here?
   b. **Hypothesis and level of significance:**
      Write your hypothesis in scientific form and determine the level of singnificance.
   c. **Assumption Check:**
      Tell why you use the paired t test, and Check the required assumptions. "comment on each of them is a must!".
   d. **Result:**
      Give the output of the test and write down the result (ex: since p value is less than alpha, I reject the null hypothesis).
   e. **Conclusion:**
      You got your result in item d. Write down the conclusion of your result, in such a way that, the reader who doesn't know any statistics can understand your findings.

9. **(2p every item) Fisher's exact test for count data**
   a. **Aim**
      In words, what is your objective here?
   b. **Hypothesis and level of significance:**
      Write your hypothesis in scientific form and determine the level of singnificance.
   c. **Result:**
      Give the output of the test and write down the result (since p value is less than alpha, I reject the null hypothesis).
   d. **Conclusion:**
      You got your result in item c. Write down the conclusion of your result, in such a way that, the reader who doesn't know any statistics can understand your findings.

e. **Odds Ratio:**
   Comment about the odds ratio, what does it indicate?

10. (2p every item) ANOVA and Tukey Test

a. **Aim**
   In words, what is your objective here?
b. **Hypothesis and level of significance:**
   Write your hypothesis in scientific form and determine the level of singnificance.
c. **Assumption Check:**
   Check the required assumptions. “comment on each of them is a must!”.
d. **Result of ANOVA:**
   Give the output of the test and write down the result (ex:since p value is less than alpha, I reject the null hypothesis)
e. **Conclusion of ANOVA:**
   You got your result in item d. Write down the conclusion of your result, in such a way that, the reader who doesn’t know any statistics can understand your findings.
f. **Result of Tukey:**
   Give the output of the test and write down the result (ex:since p value is less than alpha, I reject the null hypothesis)
g. **Conclusion of Tukey:**
   You got your result in item f. Write down the conclusion of your result, in such a way that, the reader who doesn’t know any statistics can understand your findings.

11. (2p every item) Multiple Linear Regression

a. **Aim**
   In words, what is your objective here?
b. **Regression Equation:**
   Multiple linear regression (MLR) is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. Which ones are your explanatory variables and which one is your response variable? Write down the equation of your regression using those variables.
c. **Hypothesis and level of significance:**
   Write your hypothesis in scientific form and determine the level of singnificance.
d. **Find the Best Model:**
   Use step function and find the best model, describe the reason which makes it the best one.
e. **Assumption Check:**
   Check the required assumptions, “comment on each of them is a must!”.
f. **Result:**
   Give the output of the best model and write down the result.

g. **Conclusion:**

You got your result in item f. Write down the conclusion of your result, in such a way that, the reader who doesn't know any statistics can understand your findings.

h. Prediction:

Think of a new X variable, or variables (just make it up). Use your best model in order to predict the new Y variable for those X variables.

ZEYNEP FILIZ EREN