Lab-2: Suggested guidelines

The operationalization

- Begin with an introduction to the topic and specify the general interest
- Identify a few concepts that are essential for studying that interest
- Define those concepts
- Spell out precisely how you will measure those concepts
- Discuss possible discrepancies between the operational definition and the intended concept

Explanatory data analysis

Basic Data Understanding, Data Cleaning & Data Manipulation

- Check the total number of entries and the variable types
- Check and handle any null values
- Check and handle duplicate entries
- Check for any other possible issues in data such as outliers or obvious wrong data entries.
- Comment on them in the text

Visualization

- Create the simplest plot, graph, Table, etc. for each concept you try to explore considering:
  - ✓ Variable type
  - ✓ Attribute used to create a plot. (It is a variation within one variable or covariation between different variables)
- Use understandable title and axis labels
- Select meaningful axis ranges
- Use layering, facet, and different colors or shape to compare different categories
- Try to explain the most common values, rare values, any unusual pattern, etc.
- **Never** leave any plot, graphs or table without commenting on it in the text

Test selection

- Determine the population parameter of interest
- State your null and alternate hypothesis
  - ✓ Better to be conservative and use two-tailed tests, unless there was strong ground for the one-tailed test

- Choose the proper test considering:
  - ✓ Variable type (Cardinal V.S. Ordinal)
  - ✓ Paired vs. Unpaired sample
- Check the test assumption
  - ✓ For the parametric tests, when you invoke the CLT, check the conditions of CLT such as iid, and talk about the relationship between the skewness of the underlying random variable and the number of observations
  - ✓ **Never** discuss CLT when variables are ordinal
- Run the test and compute the test statistic and P-value
  - ✓ **Never** change the null hypothesis and run more than one test

Conduction test and conclusion

- Specify desired significance level.
- Reach the decision and interpret your result
  - ✓ **Never** accept Null hypothesis
- Draw conclusions
  - ✓ Use practical significance to draw a conclusion about the size and importance of your result, especially when the sample size is large
  - ✓ Use Cohen's d and correlation (r), when there is no more an understandable measure such as the difference between the means
  - ✓ Use proper correlation for ordinal variables