Inferential Statistics with R

Data Analysis Project Checklist

This checklist is intended to help you critically evaluate your Data Analysis Project. Remember that the more thoroughly you evaluate your project, the more useful this exercise will be to you and will set you up for success in future research projects.

Format

- [ ] Did you use the R Markdown to complete your project?

Data

Generalizability: Describe how the observations in the sample are collected, and the implications of this data collection method on generalizability.

- [ ] Did you describe the sampling method, mentioning that a random sample was used?
- [ ] Did you decide that the results can be generalized to the US population?
- [ ] If you discussed potential sources of bias, tell us more about how these affected your conclusions.

Causality: Describe how the observations in the sample are collected, and the implications of this data collection method on causality.

- [ ] Did you decide that no random assignment was used, and hence causality cannot be inferred?

Research Question

- [ ] Is the research question phrased in a non-causal way?
- [ ] Is the research question well defined / not vague?

Note: "Well defined" means it is obvious from the research questions which variables will be involved in the analysis.

- [ ] Is it clear why this research question is of interest to you and/or the reader?

Exploratory Data Analysis

Plots

- [ ] Do the plots address the research questions?

Note: There is no requirement on the minimum number of plots to be provided. A single plot can be sufficient, as long as it addresses the research question, or multiple plots may be needed.

- [ ] Are the plots properly constructed?
- [ ] Are the plots formatted well? (Size not too large, not too small, etc.)

Summary Statistics

- [ ] Do the summary statistics address the research question?
- [ ] Are the summary statistics calculated correctly?
- [ ] Are the summary statistics formatted well? (Not taking up pages and pages, etc.)

Narrative

- [ ] Is each plot and R output accompanied by a narrative?
- [ ] Does the narrative interpret the visuals and summary statistics correctly?
- [ ] Does the narrative address the research question?

Inference

Statistical inference via hypothesis testing and/or confidence interval.

- [ ] Is the hypothesis stated clearly and matches the research question?
- [ ] Are the conditions checked in context of the data (not just a generic bullet point list of the conditions, but reasoning through them for the given dataset)?
- [ ] Did you state the appropriate method(s) to use for your project? Did you provide a discussion of

why you chose these methods, and described how they work?

Note: in this part you should display a thorough and conceptual understanding of how the methodology works, however the write-up does not need to be as detailed as if you were teaching the method to someone with no background in statistics.

Was the correct code used and output provided for all required techniques? See below for which situation requires which technique: ☐

One numerical and one categorical variable (with only 2 levels): hypothesis test + confidence interval

- parameter of interest = difference between two means (theoretical or simulation)
- parameter of interest = difference between two medians (simulation only)

One numerical and one categorical variable (with more than 2 levels): hypothesis test only, compare means across several groups, no defined parameter of interest, ANOVA and pairwise tests (theoretical only)

Two categorical variables (each with only 2 levels): hypothesis test + confidence interval

- parameter of interest = difference between two proportions (theoretical if success-failure condition met, simulation if not)

Two categorical variables (either one or both with more than 2 levels): hypothesis test only, compare proportions across several groups, no defined parameter of interest, Chi-square test only (theoretical if expected sample size condition met, simulation if not)

Are correct interpretations and conclusions for all output provided? ☐

Note: this portion should be evaluated based on criteria stated earlier about which technique is required when). This includes some or all of

- conclusions of hypothesis tests,
- interpretations of p-values as conditional probabilities, and
- interpretations of confidence intervals

depending on the methods used. All interpretations must be in context of the data and the research question.

Is whether or not results from hypothesis test and confidence interval agree stated? ☐

Or, if doing ANOVA or chi-square testing, did you state that no other methods were applicable and hence there's nothing to compare?

Overall Project Details: Organization/Readability

- [ ] The document follows the organization of parts outlined in the template.
- [ ] The narrative uses correct grammar and clearly and succinctly addresses the research question.
- [ ] The code is clear, readable, well organized, and uses syntax and packages taught in the course.