Peer-graded Assignment: Data Analysis Project

Try again
You earned 26 points, but you need to earn at least 35 / 50 points to pass. Review your feedback below, improve your submission, and resubmit when you're ready.

Edit to resubmit

It looks like this is your first peer-graded assignment. Learn more

Instructions | My submission | Discussions

Inferential Statistics Lab

Submitted on September 16, 2018

Shareable Link

PROMPT

Upload the HTML file of your project.

stat_inf_project.html

RUBRIC

Did the author use the R Markdown to complete their project and upload the resulting HTML file?

- **1 point**
  Yes
- 0 points

No

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

Did the author describe the sampling method, mentioning that a random sample was used?

- **1 point**
  Yes
- 0 points
  No

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

Did the author decide that the results can be generalized to the US population?

- 1 point
  Yes
- **0 points**
  No

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

If the author discussed potential sources of bias, comment on this discussion.

Jeremy Ledger

Bias is mentioned, but you do well to point out that the survey is ultimately voluntary.

Norman Cho

I anticipate a bias in scores toward people who have more free time such as older Americans without young kids. You can do a distribution on the age to see whether this is true.

Charles Crawford

Yes, the author says bias may occur.

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

Did the author decide that no random assignment was used, and hence causality cannot be inferred?

1 point
Yes

0 points

No

RESEARCH QUESTION: Is the research question phrased in a non-causal way?

- (•) **1 point**
  Yes
- ( ) 0 points
  No

RESEARCH QUESTION: 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.

- ( ) 1 point
  Yes
- (•) **0 points**
  No

RESEARCH QUESTION: Is it clear why this research question is of interest to the author and/or the reader?

- (•) **1 point**
  Yes
- ( ) 0 points
  No

EDA (Exploratory Data Analysis)

PLOTS: Do the plots address the research questions?

Note: There is no requirement on 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.

- **1 point**
  Yes
- 0 points
  No

EDA (Exploratory Data Analysis)

PLOTS: Are the plots constructed correctly?

- **1 point**
  Yes
- 0 points
  No

EDA (Exploratory Data Analysis)

PLOTS: Are the plots formatted well? (Size not too large, not too small, etc.)

- **1 point**
  Yes
- 0 points
  No

EDA (Exploratory Data Analysis)

SUMMARY STATISTICS: Do the summary statistics address the research question?

- **1 point**

Yes

0 points
No

EDA (Exploratory Data Analysis)

SUMMARY STATISTICS: Are the summary statistics calculated correctly?

1 point
Yes

0 points
No

EDA (Exploratory Data Analysis)

SUMMARY STATISTICS: Are the summary statistics formatted well? (Not taking up pages and pages, etc.)

1 point
Yes

0 points
No

EDA (Exploratory Data Analysis)

NARRATIVE: Is each plot and R output accompanied by a narrative?

1 point
Yes

0 points
No

EDA (Exploratory Data Analysis)

NARRATIVE: Does the narrative interpret the visuals and summary statistics correctly?

- **1 point**
  Yes
- 0 points
  No

EDA (Exploratory Data Analysis)

NARRATIVE: Does the narrative address the research question?

- **2 points**
  Yes
- 0 points
  No

INFERENCE: *Statistical inference via hypothesis testing and/or confidence interval.*

Are the hypothesis stated clearly and matches the research question?

- 4 points
  Yes hypotheses are stated clearly, and they match the research question
- **2 points**
  Yes hypotheses are stated clearly, but they do not match research question
- 0 points
  Not present or not clear

INFERENCE: *Statistical inference via hypothesis testing and/or confidence interval.*

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)?

- 4 points
  All conditions checked correctly and in context of the data
- 2 points
  Correct conditions are listed as a generic list, but not checked in context of the data
- **0 points**
  Not present or incorrect

INFERENCE: *Statistical inference via hypothesis testing and/or confidence interval.*

Are the appropriate method(s) the writer will be using stated? Did the author provide a discussion of why they chose these methods, and described how they work? Note that in this part the author 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 they were teaching the method to someone with no background in statistics.

- 4 points

Present and correct

- **0 points**
  Not present or incorrect

INFERENCE: *Statistical inference via hypothesis testing and/or confidence interval.*

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)

- 6 points
  All required code and output present and correct
- **3 points**
  Only some of the required techniques are presented, but others are missing
- 0 points
  Not present or incorrect

INFERENCE: *Statistical inference via hypothesis testing and/or confidence interval.*

Are correct interpretations and conclusions for all output provided? Note that 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.

- 7 points
  All conclusions/interpretations present, correct, and in context of the data and the research question
- 4 points
  All conclusions/interpretations present and correct, but not put in context of the data and the research question
- **2 points**
  Conclusions/interpretations are provided and correct for only some of the required techniques
- 0 points
  Not present or incorrect

INFERENCE: *Statistical inference via hypothesis testing and/or confidence interval.*

Is whether or not results from hypothesis test and confidence interval agree stated? Or, if doing ANOVA or chi-square testing, did the author state that no other methods were applicable and hence there's nothing to compare?

- 3 points
  Present and correct
- **0 points**
  Not present or incorrect

OVERALL: *Organization and readability.*

The document follows the organization of parts outlined in the template.

- **1 point**
  Yes
- 0 points
  No

OVERALL: *Organization and readability.*

The narrative uses correct grammar and clearly and succinctly addresses the research question.

- **2 points**
  Yes
- 0 points
  No

OVERALL: *Organization and readability.*

The code is clear, readable, well organized, and uses syntax and packages taught in the course.

- **2 points**
  Yes
- 0 points
  No

OVERALL ASSIGNMENT RUBRIC

Please make any general constructive comments on this project that the student would find helpful.

Jeremy Ledger

First, I'm very sorry to hear about your mother's cancer. The question you pose is indeed an important one.
As for the assignment, you can gain more points on this assignment by mentioning causality and generalizability in the 'data' section. You can also make your research question more clear and gain more points there. It is not quite clear what you are asking and from your question it is also unclear what variables you will be looking at. Some of the wording in the question section needs to be clarified as well. You should point out that you're only looking at 2012, for instance.
You also need to state your null and alternative hypotheses and probably do an inference analysis or confidence interval, as the assignment requested. Or, if you think inference/confidence intervals are not applicable to your questions, you need to say why.
Overall this is an excellent project. I tried to be generous with the points allocation, but I had to take some points off. If you follow the assignment guidelines a little closer, you'll easily get all of the points.

Norman Cho

Would recommend removing the NAs.

It jumps from Americans and their thoughts on assisted suicide to a chi-squared test on religion. Please state the differences by religion in the beginning

Charles Crawford

A little more work on the meaning of the p-value would be in order.

Is there anything in the author's writing that is distracting in their writing style? Please provide comments that might be helpful for the author in the subsequent phase of the project.

Jeremy Ledger

Some typos? A few unclear sentences too. Could easily be fixed by proofreading.

Norman Cho

As a native English speaker, I could understand what was written.

Charles Crawford

Writing is clear, but a little more study is needed.

Based on having evaluated this author's project, what have you learned that might help you with your own project?

Jeremy Ledger

Maybe.

Norman Cho

n/a

Charles Crawford

I need to describe conditions for the tests more

Edit submission

Comments

Comments left for the learner are visible only to that learner and the person who left the comment.

share your thoughts...