Project Assignment 4: Upload to Canvas by, 26th April, 11:59 pm

Bivariate Analysis & Moderation

During this assignment you will:

1. Conduct tests of statistical inference (hypothesis tests) for 2 to 3 associations related to your research question.
2. Fully document the steps in each hypothesis test, conduct one or more post-hoc tests if needed, and state the type of error that *could* have been made.
3. (Optional) Consider generating a bivariate graph with a third moderating variable (such as gender or urban/rural) and conducting an inferential test for moderation.
4. Summarize the results of your research. How might you answer your original research question? Try to put the results in a broader context. What are the implications for policy or further research? State any limitations of the study.

The detailed instructions are below. After each step, save and backup your work! Knit frequently. And don't forget to add helpful comments to your code.

1. For the calculations needed for each test of statistical inference, use the Section 7 code chunk in the same RMarkdown file that you used for PA 3.
2. Below the Section 7 code chunk, document the steps in each hypothesis test plus a comment about the type of error that could have been made. Use the conventions of statistical communication (in other words, follow closely examples from class and OLI).
3. For each test:

    Step 1: State the claim

    - Formally state a null and alternate hypothesis for the test. [Remember, the form of the null and alternate hypothesis depends on the role-type classification and type of test, so check examples from OLI and class.]

    Step 2: Collect and summarize the sample

    - Summarize the sample using the appropriate summary statistics for the test.

        C to Q: List mean and standard for each category of the explanatory variable (i.e. group means from Section 5)

        C to C: List sample proportions for each category of the explanatory variable (i.e. look at the table of proportions from Section 6)

        Q to Q: Give the mean and standard deviation for the explanatory and response variable (from Section 5)

    - Sample size (counts) for each category of the explanatory variable (C to Q and C to C) or the overall sample size for Q to Q. Remember, sample sizes should not include NAs.
    - Are the conditions met for the test? Or, are there assumptions that must be made in order to meet the conditions of the test?

    Step 3: Assess the evidence

    - Report the test statistic and degrees of freedom (chi-square, ANOVA), or r and r^2 (Pearson)
    - Report the p-value for the test.

    Step 4: Make a conclusion

    - Apply a decision rule based on $\alpha = 0.05$
    - State the conclusion in context.

State the type of error that *could* have been made

Conduct a post-hoc test if necessary

- For C to Q: If the null hypothesis is rejected AND there are more than two categories in the explanatory variable, conduct a Post Hoc test to determine which pairs of categories are statistically significantly different
- For Q to Q: If the null hypothesis is rejected, calculate and report the slope of the regression line
- For C to C: (Optional) If the null hypothesis is rejected AND there are more than two categories in the explanatory variable, two by two chi-square tests can be done to determine which pairs of categories are statistically different.

5. Moderation (Optional). Generate a graph and test an association moderated by a third categorical variable. For example, do you think the association between financial well-being score and level of education could be different for men versus women? This can be done in Section 8.
6. At the end of your RMarkdown file, summarize the overall results of the hypothesis tests, post-hoc tests and moderation. What can you conclude about your original research question? If you conducted a post-hoc test, state which categories were significantly different. Mention the effect sizes, e.g. the difference in proportions (C to C) or means (C to Q) between significantly different categories. If a Pearson correlation test was significant, give an interpretation of the slope of the regression line.

   If the null hypothesis was not rejected and evidence of an association was not found, there are two possible conclusions: first, there genuinely is no association; or second, there are limitations to the study that make it difficult to identify such a relationship (small effect size, confounding or lurking variables, sample or measurement issues in the survey, etc..). These issues should be commented upon.

   Try to put the results in a broader context. What are possible implications for policy or further research? State any limitations of the study.

Upload the knitted MS Word document.