

Homework 2

Measurement and Evaluation of HCC Systems

How to hand in this homework

- Please email the homework to me as a PDF.
- Late assignments get a penalty of 20% when submitted after the deadline, plus an additional 10% per hour late.
- Make sure you include the R input you used to get to your answer, but do not “dump” the resulting R output on the paper. Copy from the output selectively, and explain it in your own words.
- You may collaborate on this homework, but not copy from others... again, please write your answers in your own words.
- Please include a collaboration statement that says: “I collaborated on this homework with [name].” or “I worked alone on this homework”

Dataset

You are going to use the same dataset as in Homework 1. Please refer to Homework 1 for a reminder of how the data was collected, etc.

Question 1. Plotting

For this assignment we are going to focus on the effects of the inspectability manipulation (list view versus graph view) and the control manipulation (none vs. item-control vs. friend-control) on users’ perceived control and understandability of the system.

Let’s start with understandability:

- a. Create a double bar chart of the mean understandability in the list view and graph view conditions. Give your chart normal error bars.
- b. Just by looking at this graph, does it look like this effect will be significant? Why (not)?
- c. Split up the same bar chart by control, using the `facet_wrap()` function.
- d. Does it look like there is an interaction effect? Why (not)?

Now let’s do perceived control:

- e. Create a line chart of the mean perceived control at different levels of control.

- f. Just by looking at this graph, does it look like this effect will be significant? Why (not)?
- g. Split up the same line chart by inspectability. Put the two line charts in the same plot area, using different colors for list view and graph view.
- h. Does it look like there is an interaction effect? Why (not)?

Question 2. T-test

Let's look at the effect of the inspectability manipulation (list view versus graph view) on understandability. Our hypothesis is that understandability is higher in the graphview condition than in the listview condition (keep this in mind when reporting p -values!).

Let's start with a power analysis:

- a. Let's say we want to detect an increase of at least 1 point in understandability between the listview and the graphview condition. Use the formula $d = \text{difference} / \sqrt{(\text{var}_g + \text{var}_l)/2}$ to calculate the effect size that this difference represent. Note: var_g and var_l are the variances of understandability in the graphview and listview conditions.
- b. What sample size would you need to detect this increase with $\alpha = .05$ and 85% power? (Note: think about whether you are going to conduct a 1-sided or 2-sided test!)

Now let's test some assumptions:

- c. Use R to get the skewness, kurtosis and Shapiro-Wilk test for the understandability variable, split by inspectability. Also plot a double histogram (one for listview and one for graphview) with binwidth = 1 and a normal line overlaid on it.
- d. Are these values cause for concern? Why (not)?
- e. We are not testing for heteroscedasticity. Why is this not necessary?

Time for some t-tests!

- f. Disregard the assumptions, and run a regular t-test of the difference between list view and graph view on understandability. Interpret the results (including t and df , (one-sided or two-sided) p -value, and a description of the estimated difference between the conditions).
- g. Run a bootstrapped t-test with no trimming (!) and 10,000 bootstrap samples on the same variables.
- h. Run bootstrapped t-test with 20% trimming and 10,000 bootstrap samples on the same variables.
- i. Run bootstrapped t-test with M-estimators and 10,000 bootstrap samples on the same variables.
- j. Answer the following questions based on the results of the 4 t-tests you have run: Did bootstrapping change the model? Did trimming change the model? Is 20% trimming a good amount? Why (not)?

- k. Go back to your original t-test (question 2f). Calculate the effect size r for this t-test. How much of the variance in understandability is explained by the inspectability manipulation? (Hint: calculate R^2).
- l. Report your results for the original t-test. For reporting, follow the format of slide 27 of the t-test slides. Also, mention in your report that you also conducted robust tests, and briefly explain whether this changed the results (and if so, how).

Question 3. ANOVA

Let's look at the effect of the control manipulation (none vs. item vs. friend) on perceived control. We hypothesize two things: 1. any form of control leads to more perceived control than the "none" condition, and 2. there is a difference between item control and friend control (note that we don't hypothesize which one is larger).

Let's start with the assumptions:

- a. Use R to get the skewness, kurtosis and Shapiro-Wilk test for the perceived_control variable, split by the three control conditions. Also plot a three histograms (one for each control condition) with binwidth = 1 and a normal line overlaid on it.
- b. Are these values cause for concern? Why (not)?
- c. Test for heteroscedasticity with a boxplot, Levene's test, and the variance ratio test. Is there heteroscedasticity?

Time for an ANOVA!

- d. Write two contrasts "none_v_any" and "item_v_friend" based on the hypotheses outlined at the beginning of question 3. Load the contrasts into the control variable. Are these orthogonal contrasts? Explain your answer.
- e. Run the ANOVA, and request the summary of the model. Is there a significant overall effect?
- f. Request the lm summary of the model. What is the p-value of our first hypothesis? What is the p-value of our second hypothesis? Interpret these results.
- g. Run Welch's F-test. Are the results much different from the regular ANOVA? Does this surprise you? Why (not)?
- h. Conduct a Tukey post-hoc test. Interpret the results.

What about the effect sizes? Let's see:

- i. Get the overall effect of the ANOVA model, both in terms of r and omega.
- j. Get the effect sizes r of the two contrasts (hint: use the t -values to calculate r).
- k. Get the effect size d of the difference between the "none" and "friend" condition.

Let's do some reporting:

- l. Report on the results. Follow the format of slides 61-63 of the ANOVA slides. For the omnibus test, provide the f -statistic with its degrees-of-freedom parameters, p -value, and

omega effect size. For the planned contrasts, provide the t -statistics with their degrees of freedom, (one- or two-tailed) p -values, and effect sizes r . For the post-hoc tests, provide the t -statistics with their degrees of freedom, corrected p -values, and effect sizes d .

Let's say that we think that results of the ANOVA model are interesting enough to run another experiment. We want to do a power analysis to see how many participants are needed to get the current ANOVA significant with a power of 85%.

- m. Do a power analysis to see how many participants would be needed to get the current ANOVA effect significant at $\alpha = .05$ with 85% power. Note that you can convert the omega effect size of the current model into an f effect size (suitable for G*Power) by using the following formula: $f = \sqrt{\omega^2/(1-\omega^2)}$.

Question 4. Factorial ANOVA

Finally, we want to test if there is an interaction effect between inspectability and control on users' perception of understandability.

- Write contrasts for the independent variables and assign them to the variables. For control, use the same contrasts as before.
- Run the factorial ANOVA. We don't expect an interaction effect, so we should request Type II sums of squares.
- Interpret the results of the factorial ANOVA. Is there a significant overall effect? Is there an effect of inspectability on understandability? Is there an effect of control on understandability? Is there an interaction effect?
- Just to make sure, run the factorial ANOVA with 20% trimming. Do the results change much?

We will make an adjustment to the model by removing the interaction effect:

- Remove the interaction effect from the model (hint: use the formula `understandability ~ inspectability+control`), and run it again. Did the sums of squares of the two main effects change? Why (not)? Hint: your answer depends on the type of sum of squares used for the ANOVA.
- Get the `lm` summary of this model. Interpret the contrasts, both in terms of t -statistics, degrees of freedom, and (one- or two-sided) p -values, as well as what the estimates of the contrasts mean (what differences do they express?).

Let's test the simple effects:

- Use a simple effects analysis to test if the effect of graphview vs. listview is significant at every level of control. Hint: create a new variable using `tw$simple <- interaction(tw$control, tw$inspectability)`, create simple effect contrasts for this new variable, and run a regular ANOVA with these contrasts (see "Simple effects" slides). Make

sure to interpret the results; in particular, interpret the three contrasts that test the effect of inspectability in each of the three control conditions.

What about the effect sizes? Let's see:

- h. Get the overall effect size r of the ANOVA model from question 4f.
- i. Get the effect sizes ω for each manipulation (use my `omega_aov` function).
- j. Get the effect sizes r of the three contrasts of the ANOVA model from question 4f.
- k. Get the effect size d of the difference between the “listview” and “graphview” condition at each level of control.

Finally, let's do some reporting:

- l. Report on the results. Follow the format of Field section 12.9. Also, mention in your report that you also conducted a robust test, and briefly explain whether this changed the results (and if so, how).