Psych 610, HW7

Due Wednesday, October 25, 2023, 1:30pm on Canvas

* Submit **both** an HTML file and a .Rmd file
* Email Ben, LiChen, or Nick with any questions

# Conceptual Questions

## 1.

A researcher presents the following equation representing the relationship between the number of books a child reads and whether or not that child has a diagnosis of dyslexia, where dyslexia is coded 1 (does not have dyslexia) and -1 (has dyslexia):

Books = 1.52 + 0.5 \* dyslexia.

If the researcher recoded dyslexia as -.5 (does not have dyslexia) and .5 (has dyslexia), what would be the new regression equation representing the predicted number of books a child reads?

## 2.

In a mediation analysis, what is meant by the “direct effect”? Why is this term potentially misleading?

## 3.

What is our measurement of the total prediction error associated with a given model? In the context of testing statistical models, what does it mean to “reduce error?”

## 4.

What does it mean for an estimate to be “unbiased”? What is our estimate for the population mean? Is this estimate biased or unbiased? What is our estimate for the population SD?

**5.**

**A.**

Imagine you are at a conference when you see someone present a mediation model. Their model indicates that the relationship between extraversion and the likelihood of agreeing to attend an unknown event (measured continuously) is mediated by mood. Assume that participants answered all questions (including whether they would attend the event) in the same survey. Is this an appropriate design to test for mediation?

**B.**

The researcher in part A reports that extraversion significantly predicts the likelihood of attending the unknown event, b1 = 3.5. Further the researcher reports that extraversion predicts having a more positive mood, b1 = 0.6. Finally, when controlling for the effects of extraversion, mood significantly predicts the likelihood that someone attends the event, b1 = 2.50. Oops! The researcher forgot to report one estimate from their mediation model. What did they forget to report? Calcuate the value of that effect.

# Data Analysis

Researchers at UW-Madison conduct a study about laughing. They have a series of hypotheses about the social function of laughter, but for this homework you will focus on only one of these hypotheses: that alcohol consumption has a positive relationship with amount of laughing.

After a home football game, researchers recruited both Badger fans and the rival team’s fans in pairs as they were exiting Camp Randall Stadium, and brought them to a lab in Brogden. Half of the subjects were paired with a fan of their own team (whom they had never met before), and half were paired with a fan of the rival team. Since gender differences exist in average laughter frequency, all participants were paired with same-gender partners (Smoski & Bachorowski, 2003). Participants were not given an opportunity to talk to each other before the study began. The subject pairs were seated together in front of a TV and watched 7 video clips from *America’s Funniest Home Videos*. The researchers filmed the session so they could code the total duration of participants’ laughter that occurred during each video clip. After the video viewing stage, the participants separately completed scales measuring how much they liked and how similar they felt to their partner and reported how many drinks they had had that day (since the researchers expected more intoxicated participants to laugh more). The researchers didn’t know how to handle the non-independence of the subject pairs, so they just computed average scores for each pair on all the variables. The codebook for the data (“laughter\_hw\_06.csv”) is given below.

|  |  |  |
| --- | --- | --- |
| **Var. name** | **Description** | **Range/values** |
| team | Whether participant pairs were fans of the same team (“same”) or rival teams (“different”) | “same”  “different” |
| laughter1 to laughter7 | The total duration (in seconds) of laughter observed during each of the 7 video clips | 0-23.75 |
| liking | Mean/composite score of a scale measuring how much the participant pairs liked each other at the end of the study (0 = strongly dislike, 7 = strongly like) | 0-7 |
| similarity | Mean/composite score of a scale measuring how similar the participants found their partners to be (averaged for each pair, 0 = very dissimilar, 7 = very similar) | 0-7 |
| drinks | Number of drinks consumed that day (averaged for each pair) | 0-18 |

## 1.

Read the data into R and inspect them. Based on descriptive statistics and univariate plots, are there any hints that some cases might have been miscoded and should be excluded?

## 2.

Observations in this dataframe should have unique identifiers. Create a variable called pair\_id (an id for each observation in the dataset) that ranges from “p1001” to “p1209”. Reorder your variables so this id is in the first column. Google it or use ChatGPT if you need help with any of this data wrangling. Str\_c(), paste(), and select() are three functions that might help you get the job done.

## 3.

Exclude any cases that have clearly been miscoded. While in your own data you might look further into this miscoding to see if you can fix it, for the purposes of this assignment, simply remove the entire pair when you find evidence of clear miscoding. Do a quick check for any additional miscoded values.

## 4.

Focus on the variables measuring laughter during each video clip (i.e., laughter1-7).

### A.

Determine the reliability of the clips. Assume for this question only that these clips were used in a previous study and using them to elicit laughter is a standardized procedure. Are there any clips that should not be included in the composite score? Write a short comment explaining your reasoning.

### B.

After doing some digging, you found out that your researchers did not use the standardized version of the laughter procedure and accidentally showed pairs a (very sad) clip from the opening of the Pixar movie “Up,” which is why one of the “laughter” variables looked different than the others. Create a composite measure of laughter, laughter\_m, excluding this anomolous clip.

## 5.

Explore the dataframe by obtaining the bivariate correlations and scatterplot matrix on some of the variables of interest: laughter\_m, liking, similarity, and drinks. Note any strong relationships between these variables.

#### **6.**

Fit a model in which you predict the average laughter from drinking. In a sentence, describe your findings (report F, p values, and partial eta squared).

#### **7.**

What is the interpretation of the intercept in this model? Is this value meaningful?

#### **8.**

Make a publication quality plot of the relationship between drinks and laughter. Below the plot, describe what the error bands in your plot represent.

## 9.

Fit a model in which you predict the average laughter from drinking when controlling for the effect of similarity. In a sentence, describe your findings (report F, p values, and partial eta squared).

## 10.

Provide an interpretation for each of the parameter estimates for the model you fit in question 9. You answer must include a b value and a description that can be understood by someone who is not a statistician.

## 11.

What are 2 topics or concepts that you are still unsure about that you would like to go over before the in-class exam?

## 12.

How many hours did you spend on this HW assignment?