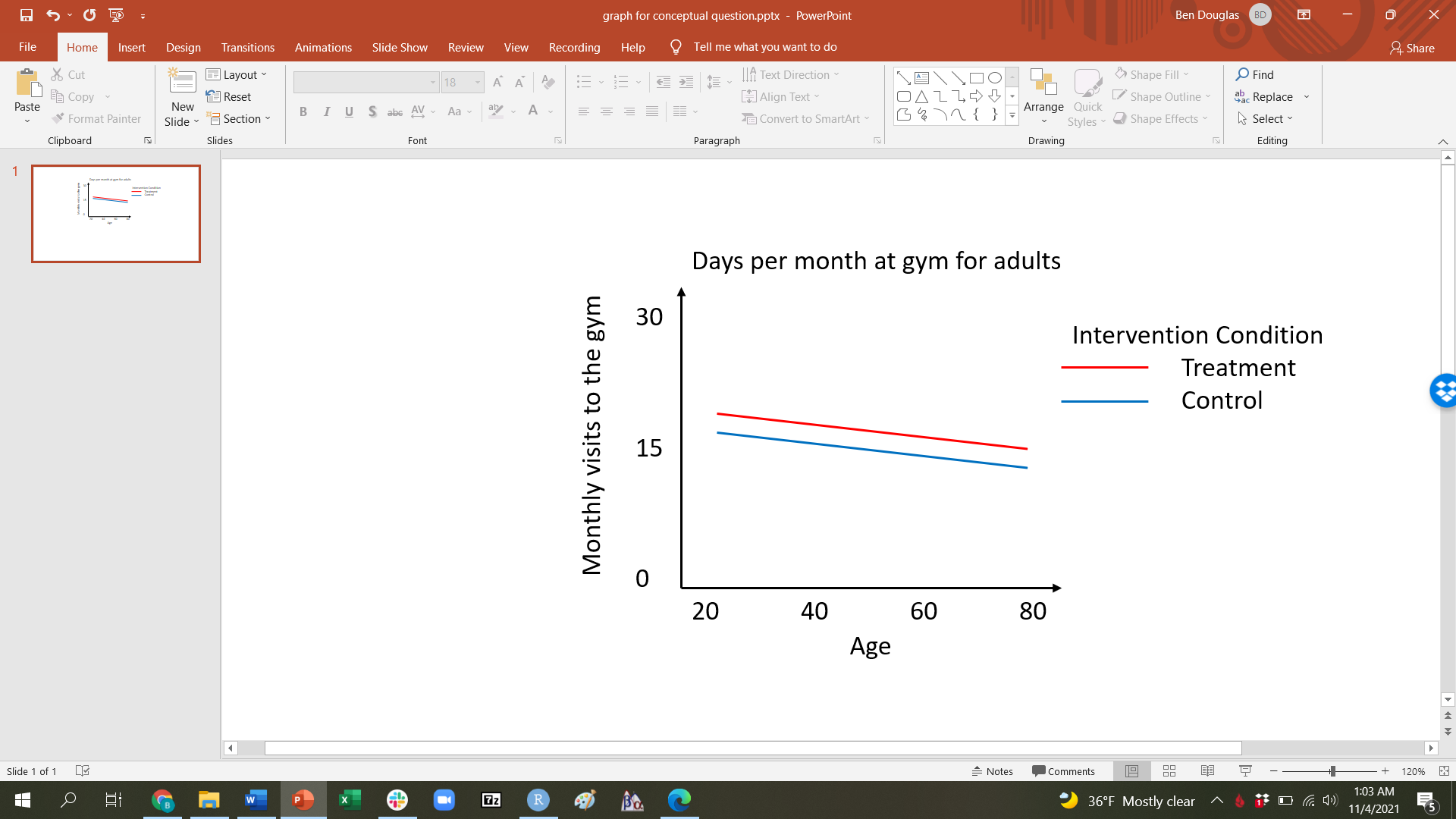
**Homework 6**

Due date: Wednesday, October 18th, 1:30pm

Submission instructions: Complete all questions in R Markdown. Submit both an R Markdown and a knitted HTML file to Canvas.

**Conceptual Questions**

1. Fill in the blanks:
   1. The goal of \_\_\_\_\_\_\_\_\_\_\_\_ is to understand the limiting conditions of the effect of a predictor on an outcome.
   2. The goal of \_\_\_\_\_\_\_\_\_\_\_\_ is to understand the mechanism of the effect of a predictor on an outcome.
2. How do c and c’ differ (in terms of absolute magnitude) in cases where a third variable is suppressing the effect of X on Y?
3. Jaquira is studying how a motivational intervention and age affect going to the gym. She plots the graph below for the model: gym attendance = b0 + b1 \* age + b2 \* intervention condition



1. Identify where on the graph we see b1
2. Identify where on the graph we see b2
3. Identify where on the graph we see b0 (Treatment is coded .5 and control -.5)

**Data analysis part A**

For this data analysis portion, you will be using the Prestige dataset.

Canadian researchers are interested in understanding the potential predictors of income for any given job. They have sampled 103 different jobs of varying types (professional, white collar, and blue collar). They also collected data on the average percentage of women in each profession, how prestigious each job is considered to be by the public, and the average amount of education required to work each job (measured in years). The researchers believe that jobs requiring more years of education must have a greater average monthly income. However, they know from their previous studies that prestige has a positive relationship with income and believe that they should statistically control for prestige in any models that they fit. Why don’t you give the research team a hand?

1. Read “prestige.csv” into R and inspect the dataframe.
2. Mean center both prestige and education. Note that these are both continuous predictors!
3. First fit a linear model in which income is regressed on education\_c. Save this model as m0.
4. Report your findings (F, p value, partial eta squared) and provide a conceptual interpretation of b1.
5. Now fit a linear model in which income is regressed on education\_c and prestige\_c. Call that model m1.
6. Report your findings (F, p values, partial eta squared) for any significant relationships and provide a conceptual interpretation of each parameter estimate. Start with the focal predictor.
7. Interpret partial eta squared and delta r squared for m1 for education\_c and prestige\_c.

**Data analysis part B**

For this data analysis portion, you will be using the Outgroup dataset. Below is the study description and codebook for this dataset.

Researchers have consistently found that people perceive their own group as more variable (i.e., heterogeneous) than other groups. For example, people from Madison might believe that “Badgers” are more diverse than “Gophers” (students at the University of Minnesota). The investigators were interested in elucidating the mechanism behind this effect. Participants all rated a target group on a variety of characteristics. Participants were randomly assigned to rate either the ingroup (a target group that they were part of) or an outgroup (a target group that they were not a member of). For this study of undergraduate students, ingroup was defined as the same major as the participant and outgroup was defined as another major. The investigators explored three different variables that they believed might mediate the relationship between the group membership (ingroup vs. outgroup) and how variable one believes the group is.

Codebook for the Outgroup dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Column | Variable | Description | Values |
| 1 | condition | Whether the participant was assigned to rate the ingroup or an outgroup | 1 = ingroup (participant’s major)  2 = outgroup (other major) |
| 2 | heterog | How variable the participants perceived the target group to be | 1-9\* |
| 3 | individ | Number of people in the target group that the participant knows personally | 4-19 |
| 4 | subgr | Number of subgroups the participants named within the target group | 1-10 |
| 5 | like | The participants' liking of the target group | 1-9\* |

\*higher numbers mean more variable/heterogenous, and more liking.

NOTE: Having your dependent variable be a count (such as individ and subgr) is problematic for GLM since counts are likely to be skewed. Count data are usually analyzed using poisson models. Given that this is fake data we are going to ignore this issue, but if you typically work with count data it would be worthwhile to look into some of these models.

1. Read the data into R. Explore the data set by obtaining the recommended univariate and bivariate statistics and plots.
2. Create a correlation matrix with density plots and scatter plots.
3. Run a model to test if participants' average liking score differs significantly from a value of 4.5.
4. Center the condition variable (around 0).
5. Run a model to test whether participants who rated their ingroup had higher heterog scores than participants who rated their outgroup. What is the name of this particular statistical test?
6. Write a sentence summary, including the appropriate statistics and interpretation of b1.
7. What does the intercept of this model represent?
8. Are the data consistent with a hypothesized mediational model according to which the number of group members that participants know (individ) mediates the relation between ingroup/outgroup and perceived heterogeneity? Add brief comments on whether the conditions for mediation are satisfied.
9. Are the data consistent with a hypothesized mediational model according to which the number of subgroups that participants know (subgr) mediates the relation between ingroup/outgroup and perceived heterogeneity? Add brief comments on whether the conditions for mediation are satisfied.
10. Are the data consistent with a hypothesized mediational model according to which liking (like) mediates the relation between ingroup/outgroup and perceived heterogeneity? Add brief comments on whether the conditions for mediation are satisfied.
11. Write a results section describing the findings of this experiment, focusing on the relation between ingroup/outgroup, heterogeneity, and the potential mediators. See the end of the lecture slides on mediation for an example write-up of a mediation analysis. Include all information you think is required to understand the relation between ingroup/outgroup and heterogeneity; don’t forget to include parameter estimates, F statistics (with appropriate dfs), *p* values, and variance-based effect size estimates when appropriate.
12. Create a mediation figure for one of the three mediators. Fill in names and values appropriately, and provide stars to indicate significance. There is a .pptx template for mediation figures on the course website (for this HW).
13. Optional challenge: find a code that allows you to insert an image with your mediation figure into your RMarkdown file.
14. How long did this assignment take you?

**Knitting Note**: The more complicated a function, the longer R takes to knit the file. For procedures like Bootstrapping R may take while to knit those particular lines of code. Be patient, it may take several minutes for R to knit the file.