**Exercise 7**

**Names:**

**OVERVIEW**

Each group is tasked with exploring the Netflix survey data to identify bivariate relationships.

Everyone downloads the "Descriptive_Table_Netflix" Word document.

At least one group member downloads this worksheet to document and save the work.

At least one group member downloads "Exercise_7" RScript" to modify and save the programming.

Note: "11_PPT_SOC303" and "RScript_3" should be helpful resources.

**TASKS**

A. Select two interval-ratio, continuous, variables to examine simultaneously.

1a. Continuous Variable: _____

1b. Continuous Variable: _____

2a. Generate a bivariate plot of Continuous Variable 1a by Continuous Variable 1b, or vice versa – whichever option makes more sense. Insert plot here:

2b. Interpret the bivariate plot shown in 2a.

B. Select a continuous variable to examine by a binary (i.e., 0,1) nominal variable of your choice.

3. Which variables did you select?

Continuous Variable: _____

Nominal Variable: _____

Nominal Variable Category (0): _____, n=_____ (____%)

Nominal Variable Category (1): _____, n=_____ (____%)

4a. Produce a histogram that depicts the within-group frequency distribution of the continuous variable for category 0 and another for category 1of the nominal variable. Insert histograms here:

4b. What does a comparison of the two histograms in 4a appear to show?

5a. Produce a histogram that depicts the within-group proportional distribution of the continuous variable for category 0 and another for category 1of the nominal variable. Insert histograms here:

5b. What does a comparison of the two histograms in 5a appear to show?

5c. Did you observe any differences in your interpretation of the histograms depicting frequencies versus the histograms depicting proportions? If so, what were they, and why might this be?

6. What is the within-group mean and standard deviation (sd) of the continuous variable?

Within Category (0): mean: _____, sd: _____

Within Category (1): mean: _____, sd: _____

7. Are the group means statistically different? Follow the steps below:

7a. State the null ($H_0$) and alternative ($H_a$) hypothesis by highlighting the option below that is most appropriate based on your group's reasoning.

***Two-tailed***

$$H_0: category\ (0)\ mean = \ category\ (1)\ mean, H_a: category\ (0)\ mean \neq \ category\ (1)\ mean$$

***One-tailed less than***

$$H_0: category\ (0)\ mean \geq \ category\ (1)\ mean, H_a: category\ (0)\ mean < \ category\ (1)\ mean$$

***One-tailed greater than***

$$H_0: category\ (0)\ mean \leq \ category\ (1)\ mean, H_a: category\ (0)\ mean > \ category\ (1)\ mean$$

7b. Briefly explain why you selected this hypothesis test.

7c. Choose an alpha, significance, level.

Alpha ($\alpha$) = _____

7d. Find the corresponding critical value in the t-table, available on Canvas or Google (maybe even ask ChatGPT). Note: degrees of freedom (df) = 22 – 1 = 21.

Critical value (*t**): _____

7e. Highlight the R code below that reflects your group's hypothesis test:

t.test(ContinVar , BinaryVar , conf.level=(1-alpha), alternative="two.sided")

t.test(ContinVar , BinaryVar , conf.level=(1-alpha), alternative="less")

t.test(ContinVar , BinaryVar , conf.level=(1-alpha), alternative="greater")

7f. Run the hypothesis test in R and report the results below:

Obtained *t*: _____

Obtained p-value: _____

7g. Interpret the results of your hypothesis test and include a comparison of the critical value (*t**) to the obtained *t* score as well as a comparison of the selected alpha ($\alpha$) to the obtained p-value.

C. Select a nominal or ordinal variable to be examined simultaneously by a nominal variable of your choice.

8. Which variables did you select?

Ordinal/Nominal Variable (Var1): _____

Nominal Variable (Var2): _____

9a. Create a bivariate contingency table that contains cell frequencies, as well as row (Var1) and column (Var2) summed frequencies. Use R to obtain the frequency counts but create the table in this Word document opposed to pasting the R output. Create this table below:

9b. Compute within-Var2 proportions and add those to a new version of the bivariate contingency table. Create this table below:

9c. Interpret the bivariate contingency table shown in 9b and include an explanation for why you considered each selected variable to be placed in the row versus column position in this table.

10. Estimate a chi-squared test of independence to determine whether there is a statistically significant association between the two selected categorical variables.

$H_0: Var1\ and\ Var2\ are\ indpendent$, $H_a: Var1\ and\ Var2\ are\ not\ independent$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$f_o$ = the cell frequencies observed in the bivariate table

$f_e$ = the cell frequencies that would be expected if the variables were independent

10a. Choose an alpha, significance, level.

Alpha ($\alpha$) = _____

10b. Copy and paste the bivariate contingency table from 9a below and modify this table based on the following steps to conduct a chi-squared test of independence. Insert table here:

10c. Compute expected counts ($f_e$) for each cell and add those to the table.

$$f_e = \frac{row\ marginal\ \times column\ marginal}{N}$$

10d. Compute discrepancies between observed ($f_o$) and expected counts ($f_e$) and add those to the table.

$$\frac{(f_o - f_e)^2}{f_e}$$

10e. Sum the discrepancies between observed ($f_o$) and expected counts ($f_e$) to obtain a chi-squared statistic.

10c. Determine the appropriate degrees of freedom (df) for your chi-squared test of independence.

df = (rows - 1) X (columns - 1)

df = _____

10d. Locate a chi-squared probability distribution table using Google or another search engine and use this to find the critical value that corresponds with your degrees of freedom and selected alpha.

Critical value ($\chi^2$ $*$) = _____

10e. Interpret the test results. Recall, if the obtained $\chi^2$ statistic is greater than the critical value then reject the null hypothesis ($H_0$). Conversely, if the obtained $\chi^2$ statistic is less than the critical value then fail reject the null hypothesis ($H_0$).

10f. Attempt to run this chi-squared test of independence in R. Note, the sample size is small (N=22), so cell sizes are unlikely to be large enough to meet the requirements of this test. Report the R output below.