**Final Project Assignment (55 points)**

This final project draws upon the material covered throughout the course. I recommend working backward, to some extent. In other words, I recommend working through your complete analysis in R and then responding to the assignment questions.

**Part I: Background Information**

1. (2 points) State the research question:

**What is the relationship between viewers of a particular news source and perceived media bias? Rather, does the perception of bias in the news significantly differ between Fox News viewers and CNN viewers?**

1. (2 points) Write a short paragraph that provides some context for the research question. Why is this an interesting relationship to study? Why do you expect there to be a relationship between your independent and dependent variable?

**It will be interesting to determine whether CNN viewers identify media bias as more or less of an issue than Fox News viewers. I suspect that Fox viewers to perceive there to be more bias in the media because a major talking point for right-leaning political commentators is bias in the media.**

1. (1 point) Cite the dataset you’ll be using (Chicago style):

Pew Research Center for the People & the Press, Pew Research Center: Factual and Opinion News Statements March 2018 Survey, GfK Group, (Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, 2018), Dataset, DOI: {doi:10.25940/ROPER-31115098}.

1. (1 point) What is the unit of analysis?

**Individuals. Each observation represents a different respondent.**

1. (2 points) Complete Table 1. Only include those variables you’ll be using in your analysis. Adjust the number of rows as necessary. Use a meaningful variable name (not a label from your dataset).

**Table 1: Variable Definitions**

|  |  |
| --- | --- |
| **Variable Name** | **Definition** |
| **Perceived.bias** | **Perceived bias of the media (1=favors one side, 0=otherwise/unbiased)** |
| **freq\_cons** | Frequency of news media consumption (Most of or all the time=1, Only when something interesting is happening=0) |
| **accuracy** | Expected accuracy of news stories (largely accuracte=1, largely inaccurate=0 |
| **trust** | Respondent’s level of trust for the media (1=Not at all, 2=Not too much, 3=some, 4=A lot) |
| **medium** | Whether news was (1=Read, 2=Watched, or 3=Listened to) |
| **inform** | How informed does the national news media leave the respondent (1=Not at all, 2=Not well, 3=Some, 4=A Lot) |
| **news.source** | Fox News viewer or CNN viewer |

**Part 2: Descriptive Information**

1. (4 points) Complete the table of summary statistics. Add or delete rows as necessary.

NOTE: If you are examining unordered categorical variables (e.g., race, favorite color) please feel free to modify the table. For example, for an unordered categorical variable, you could include the mode and number of observations, or you could include the percentage of observations in each category.

**Table 2: Summary Statistics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Observations** | **Maximum** | **Mean** | **Standard Deviation** | ***N*** |
| **freq\_cons** |  |  |  |  |  |
| **accuracy** |  |  |  |  |  |
| **trust** |  |  |  |  |  |
| **medium** |  |  |  |  |  |
| **inform** |  |  |  |  |  |
| **news.source** |  |  |  |  |  |

Bias

1. (4 points) Create one visualization of the dependent variable (e.g. bar graph, line graph, boxplot, histogram). Be sure the visualization includes all the critical components, including a title, meaningful axes labels and any other useful labels. Chart, bar chart

   Description automatically generated
2. (2 points) Write a 1-2 sentence summary of the visualization above. What is something interesting about the variable that the visualization highlights?

**The visualization above shows that a clear majority of the people who favor CNN or FOX News prefer to consume their news media by watching it, either on television or online. It would be interesting to discover whether that has any relationship to their perception of its accuracy.**

1. (4 points) Create one visualization of the independent variable of interest (e.g. bar graph, line graph, histogram). Be sure the visualization includes all the critical components, including a title, meaningful axes labels and any other useful labels. If you have more than one independent variable of interest, select one for this question.

Bar chart

Description automatically generated with medium confidence

1. (2 points) Write a 1-2 sentence summary of the visualization above. What is something interesting about the variable that the visualization highlights?

**The visualization above shows that the majority of people who favor CNN or FOX News believe the news is mostly accurate. Although, it also reveals that a substantial portion of people believe the news is mostly inaccurate. I wonder who they are?**

1. (4 points) Create one bivariate visualization that shows the relationship between the two variables used above (e.g. scatter plot, side-by-side bar plot, line graph).
2. (2 points) Write a 1-2 sentence summary of the visualization above. What is something interesting about the relationship that the visualization highlights?

**Part 3: Regression Analysis**

1. (6 points) Run a bivariate regression model that includes your dependent variable and one independent variable of interest.

Run a multivariate regression model that includes your dependent variable, one or more independent variables of interest and at least one control variable.

If you are using categorical variables, you can choose either to enter them linearly (in the case of ordered categorical variables) or use a dummy set. Just be clear, in a note below the table, about how the variables are coded.

Complete Table 3 with your regression results. List the independent variables (using meaningful variable names) in the first column. Include your bivariate model results in the second column and the multivariate model results in the third column. *For each variable, list the estimated coefficient and the estimated standard error beneath in parentheses. Include only two digits beyond the decimal point.* Add or delete rows as necessary.

**Table 3: Regression Results**

**Dependent variable:**

|  |  |  |
| --- | --- | --- |
| **Independent Variable** | **Bivariate Model** | **Multivariate Model** |
| *Example row: variable name* | *(standard error)* | *(standard error)* |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| *N (number of observations)* |  |  |

1. (3 points) Explain your justification for one of the control variables that you chose to include in the multivariate model.
2. (3 points) Interpret the coefficient on the independent variable of interest in the bivariate model.
3. (3 points) Interpret the coefficient on the independent variable of interest in the multivariate model.
4. (3 points) Conduct a hypothesis test on the independent variable of interest in the multivariate model. State the null hypothesis, alternative hypothesis, *t*-score and result of the test.

* *H0*:
* *HA*:
* *t*-score:
* Result of the test:

1. (2 points) In a few sentences, describe your overall findings from your analysis and their real-world importance.
2. (2 points) If you were to continue to study this research question, what would be some useful next steps? In a few sentences describe what would be valuable to do next, such as collecting data on a new variable, conducting an experiment, operationalizing a variable in a different way, etc. Be specific in your recommendations (one paragraph).
3. (3 points) Paste your R code (not the output, just the code) here:

Bias\_Data12 <- read\_excel("Downloads/Bias\_Data12.xlsx")

FOX <- subset(Bias\_Data12, (Bias\_Data12$...8 == 'Fox News'))

CNN <- subset(Bias\_Data12, (Bias\_Data12$...8 == 'CNN'))

Bias.Table <- rbind(FOX,CNN)

### Summary Table Code When finished

medium<- sort(table(Bias.Table$...7), decreasing = TRUE)

barplot(medium, xlab="Medium", ylab="Number of Respondents",

col= TRUE, ylim = c(10, 1000), main = "Preferred News Medium")

accuracy<- sort(table(Bias.Table$...3), decreasing = TRUE)

barplot(accuracy, xlab = "Responses", ylab="Number of Respondents",

col= TRUE, ylim = c(0, 1000), main ="Perceived Accuracy of the News")

trust<- sort(table(Bias.Table$...6), decreasing = TRUE)

accuracy<- sort(table(Bias.Table$...3), decreasing = TRUE)

The Code in Yellow is where I am stuck on question 11.

I can’t make a side by side bar plot with the relevant variables and I’m not sure what I’m doing wrong.

barplot(accuracy, xlab = "Responses", ylab="Number of Respondents",

col= TRUE, ylim = c(0, 1000), main ="Perceived Accuracy of the News"

p.accurate<-(Bias.Table$...3 == 'Largely be accurate')

p.inaccurate<-(Bias.Table$...3 == 'Largely be inaccurate')

sort(table(Bias.Table$...3))

true<-data.matrix(rbind(p.accurate, p.inaccurate))

Watchers<-c(Bias.Table$...7=='Watching it')

Readers<-c(Bias.Table$...7=='Reading it')

Listeners<-c(Bias.Table$...7=='Listening to it')

media.type<-data.matrix(rbind(Watchers, Readers, Listeners))

entities<-c("Watch", "Read", "Listen")

barplot(true, beside=T, ylab = "Number of Respondents", names.arg=c("Watchers","Readers","Listeners"),

col = c("darkblue", "maroon"), ylim = c(0, 1000), main = "News Media Type and Perceived Accuracy")