Midterm – EDU 7043

Threat this midterm like an exam. That means, **work on it by yourself and not with partners or in a group**. This midterm is divided into two parts. The first part is definitional, the second will examine your knowledge of using statistical methods to answer everyday questions. You are required to turn in all your code along with a write-up of the questions asked. For part two, do not just answer the questions, but give me an essay about your results. Imagine that you are reporting findings to the agency mentioned. Include appropriate graphs, tables, charts, and whatever you feel would be helpful to the shareholders of this report.

Turn in all code and assignments to my e-mail: matthew_martinez@brown.edu by Sunday, October 14th, 2018. *Do not post them to GitHub*.

Part I – Definitions – 20 Points (2 each)

In your own words define the following:

1. Mean
2. Mode
3. Central Tendency
4. Histogram
5. Probability
6. Conditional Probability
7. Z-Score
8. Normal Distribution
9. Interquartile Range
10. Confidence Interval

Part II – Data Analyses – 80 Points

You have been approached by a marketing company called A&A Consulting to complete some analysis work. They are beginning a marketing campaign and want to know more about potential clients. They have decided to use the General Social Survey – 2016 to figure out basic demographics. There was an internal debate about using the GSS as opposed to going to a Research Data Center and using 100% count of the 2010 Census. The debate centered around cost. The GSS is free, but access to the RDC would cost the company $500,000 in man hours, accreditation, and travel to the RDC. While the GSS is free, some question the validity of the results because it is a survey (think sample). Briefly weigh in on this discussion as to the merits of using the GSS versus the 100% Census data.

The company wants to know the following assuming the GSS can be representative of potential clients:

- What is the average age?
    - For everyone in the sample
    - For whites, blacks, and Other Race separately
    - For males and females separately
- What is the average level of education?
    - For everyone in the sample
    - For whites, blacks, and Other Race separately
    - For males and females separately

- Political affiliation (See Hint)
    - For everyone in the sample
    - For whites, blacks, and Other Race separately
    - For males and females separately

Hint: There are two questions in the survey that ask about political affiliation. Both range from Extremely One Way to Extremely the Other with categories that fall in between. It is up to you on how to present this information. You may want to lump the responses into different categories or you may wish to present them as is.

Along with the basic demographics, A&A Consulting also wants to get a sense of how the nation feels about government spending. In the data you will find a series of questions that ask respondents about their views on government spending. Choose three of these items and provide descriptive statistics for the 1) the entire sample; 2) Whites, blacks, and Other Race separately, and 3) by age. Because age is continuous, it may not be very informative to provide means for each year, but rather for age categories. Which age categories you choose is up to you.

When providing means estimates it may be helpful to also give some since about the variance of these variables. It is up to you the analyst to decide what is important to report and how to report it. You will find the dataset for labeled GSS_2016_AA in the data folder on Github. Along with the dataset, you will find documentation of the variables in the data.

Additionally, the CEO is afraid that one of the data managers is forging some data. They became aware of the potential problem, but do not have evidence to say one way or another. They have provided you with the data ("Accounts_Data.csv" in Github) they believe to be forged. Develop an analysis of the data and make a recommendation on weather you believe there is impropriety concerning the data providing your rational for the recommendation.