STAT200 Introduction to Statistics

Assignment #1: Descriptive Statistics Data Analysis Plan

Assignment #1: Prepare Descriptive Statistics Data Analysis Plan

Before conducting any statistical analyses, researchers develop a plan for how they will analyze their data to answer their research questions. The purpose of this assignment is to provide an experience developing a descriptive statistics analysis plan. Note: This first assignment is a **plan only**; no statistics will be calculated or graphs created. The second assignment will involve carrying out the plan, after receiving feedback from your instructor.

Assignment Steps:

Step #1: Review the STAT200 data set file. (Note: This data set will be used for all three of this term's written assignments).

The data is a subsample from the US Department of Labor's Consumer Expenditure Surveys (CE) and provides information about the composition of households and their annual expenditures (https://www.bls.gov/cex/). Detailed information on the sample and variables is included with the data set file; please carefully review this information to familiarize yourself with the data (Note: This information will be used in Assignment #2 to describe the dataset).

Step #2: Develop descriptive statistics data analysis plan.

- **Task 1: Develop scenario.** Imagine that you are head of a household and have to determine a household budget plan based on the data available from the dataset. For instance, you are a 35 year old single parent with a high school diploma and one child.

- **Task 2: Select variables for analysis that match the scenario developed in Task 1.**The data set provides information on household consumption; there are socioeconomic variables and expenditures variables. The socioeconomic variable names start with "SE-" and the expenditure variable names start with a "USD;" all expenditures are in US dollars. All students must use income as one variable. Select **two additional socioeconomic variables** (one qualitative and one quantitative) and **two expenditures** for your analysis that match the scenario you developed for Task 1. For instance, using the example scenario of a 35 year old single parent with a high

school diploma and one child, you could select “income,” “education,” and “number of children” as socioeconomic variables and then pick two household expenditure items to show the distribution of costs and compare that with your income. When selecting variables, think about the following three questions:

- Why am I choosing these variables?
- What interests me about these variables?
- What do I think will be the outcome?

➢ **Task 3: Determine appropriate measures of central tendency and dispersion for the selected variables.** For each **quantitative** variable, select at least **one** measure of central tendency **and** at least **one** measure of dispersion (Please see below table for list of measures). For the **qualitative** variable, select **one** measure of central tendency. When determining the measures of central tendency and dispersion, think about what is appropriate given the level of measurement and type of variable. Recommend referring to the text and information posted in our LEO classroom to help with this task (Note: you will use this information to provide a rationale for your choice of measures).

Measures of Central Tendency	Measures of Dispersion
● Mean ● Mode ● Median	● Range ● Sample Standard Deviation ● Variance

➢ **Task 4: Determine appropriate graph and/or table for each of the selected variables.** Select **one** graph **or** table for each variable (Please see below table for list of graphs and tables). When determining the graphs and tables, think about what is appropriate given the level of measurement and type of variable. Recommend referring to the text and information posted in our LEO classroom to help with this task (Note: you will use this information to provide a rationale for your choice of graphs and/or tables).

Types of Graphs	Types of Tables
● Pie Chart ● Bar Chart ● Histogram ● Box Plots (also known as Box-and-Whiskers Plot)	● Frequency Table ● Relative Frequency Table ● Grouped Frequency Table

Step #3: Complete the "Assignment #1: Descriptive Statistics Data Analysis Plan Template."
Remember, you will not be conducting any statistical analysis, drawing any graphs, or compiling any tables for the first assignment. Rather, you need to wait for feedback from your instructor on this assignment and use that feedback to complete Assignment #2.

Here are the main sections for this assignment (i.e., completing the plan template):

- ✓ **Identifying Information**. Fill in information on name, class, instructor, and date.
- ✓ **Scenario.** In this section, **briefly** (2-3 sentences) describe the scenario you developed in Step #2, Task 1.
- ✓ **Complete Table 1: Variables Selected for the Analysis.** Enter information the variables selected for analysis in Step #2, Task 2. For each selected variable be sure to include its: name as listed in the data set, description, and variable type.
- ✓ **Reason(s) for Selecting the Variables and Expected Outcome(s):** In this section, for **each** selected variable, please answer the following questions:
    - ✓ Why did I choose this variable?
    - ✓ What interests me about this variable?
    - ✓ What do I think will be the outcome?
- ✓ **Complete Table 2. Numerical Summaries of the Selected Variables.** Enter information on selected measures of central tendency and dispersion for **each** selected variable. Be sure to **briefly** explain why you choose those measurements. Note: The information for the required variable, "Income," has already been completed and can be used as a guide for completing information on the remaining variables.
- ✓ **Complete Table 3. Type of Graphs and/or Tables for Selected Variables.** Enter information on selected graph and/or table for each selected variable. Be sure to **briefly** explain why you choose those measurements. Note: The information for the required variable, "Income," has already been completed and can be used as a guide for completing information on the remaining variables.

Assignment Submission: Name the file that contains your **completed "Assignment #1: Descriptive Statistics Data Analysis Plan Template"** using the following format: "Assignment1-StudentLastName."

Then, submit the file via the Assignments area in the LEO classroom in the "Assignment #1: Descriptive Statistics Data Analysis Plan" folder and wait for your instructor's feedback.

Grading Rubric for Written Assignment #1

Scenario and Selection of Related Variables ● Clear description of scenario ● Selected variables and reasons are appropriate for the scenario.	20%
Selection of Measures of Central Tendency and Dispersion For each variable: ● Appropriate measures selected. ● Rationale is provided and appropriate.	30%
Selection of Graphs and/or Tables For each variable: ● Appropriate measures selected. ● Rationale is provided and appropriate.	30%
Writing Quality: Completes all sections of template. Writes clearly, concisely, and with few errors.	20%