DATA 3441: Assignment 1

Due Time/Date: 23:59 Aug. 30, 2025

1. (IPS10-1.2) Coffee ratings. A website ranks 50 different varieties of coffee. The data include the following variables: name of the coffee, price for a 12-ounce serving, overall rating (0 to 100), roast (light, medium, or dark), flavor, aroma, and body ratings (0 to 10). (10 Points)
   (a) What are the cases?
   (b) Identify the variables and their possible values.
   (c) Classify each variable as categorical or quantitative.
   (d) Was a label used? Explain your answer.
   (e) Summarize the key characteristics of your data set.
2. (IPS10-1.4) An experiment on haptic feedback. A group of technology students is interested in whether haptic feedback (forces and vibrations applied through a joystick) is helpful in navigating a simulated game environment they created. To investigate this, they randomly assign 20 students to each of three joystick controller types and record the time it takes to complete a navigation mission. The joystick types are (1) a standard video game joystick, (2) a game joystick with force feedback, and (3) a game joystick with vibration feedback. The data collected included an ID variable that uniquely identifies each student, which of the three types of joystick was used, the time taken to complete the navigation mission, the age of the student, and the student's satisfaction with the navigation, rated on a scale of 1 to 5 with 5 being the highest satisfaction. (10 Points)
   (a) What are the cases?
   (b) Identify the variables and their possible values.
   (c) Classify each variable as categorical or quantitative.
   (d) Was a label used? Explain your answer.
   (e) Summarize the key characteristics of your data set.
3. (IPS10-1.8) Attending college in your state or in another state. The U.S. Census Bureau collects a large amount of information concerning higher education.2 For example, the bureau provides a table that includes the following variables: state, number of students from the state who attend college, and number of students who attend college in their home state. (10 Points)
   (a) What are the cases for this set of data?
   (b) Is there a label variable? If yes, what is it?
   (c) Identify each variable as categorical or quantitative.
   (d) Explain how you might use each of the quantitative variables to explain something about the states.
   (e) Consider a variable computed as the number of students in each state who attend college in the state divided by the total number of students from the state who attend college. Explain how you would use this variable to explain something about the states.
4. (IPS10-1.12) Frequent users of social media. A recent survey by the Pew Research Center asked social media users about how often they visited various sites. Pew defined a frequent user to be someone who visited a site several times a day. Here are the percents of users who are frequent users for several popular sites:
   Use a bar graph to describe the percents of frequent users of these sites and write a short summary of the data based on your graph. (10 Points)

Social media	Frequent users (%)
Facebook	51
Snapchat	46
Instagram	42
YouTube	32
Twitter	25

5. (IPS10-1.22) Vehicle colors. Vehicle colors differ among regions of the world. Here are data on the most popular colors for vehicles in North America:(10 Points)

Color	Percent
White	24
Black	19
Silver	16
Gray	15
Red	10
Blue	7
Brown	5
Other	4

   (a) Describe these data with a bar graph.

   (b) Describe these data with a pie chart.

   (c) Which graphical summary do you prefer? Give reasons for your answer.

6. (IPS10-1.25) Sketch a skewed distribution. Sketch a histogram for a distribution that is skewed to the left. Suppose that you and your friends emptied your pockets of coins and recorded the year marked on each coin. The distribution of dates would be skewed to the left. Explain why. (10 Points)

7. (IPS10-1.26) Describe the self-concept scores. Based on a suitable graph, briefly describe the distribution of self-concept scores for the students in TABLE 1.2 (downloadlink attached). Be sure to identify any suspected outliers. (10 Points)
ex01-026sevengr.csv

8. (IPS10-1.34) Potassium from a supplement. Refer to Exercise 1.16 (page 22), where you examined the potassium absorption of a group of 29 adults who ate a controlled diet that included 40 mEq of potassium from a supplement for five days. In Exercise 1.16, you used a stemplot to examine the distribution of the potassium absorption. (10 Points)
ex01-034ksup40.csv

   (a) Make a histogram and use it to describe the distribution of potassium absorption.

   (b) Make a boxplot and use it to describe the distribution of potassium absorption.

   (c) Make a stemplot and use it to describe the distribution of potassium absorption (Hint: round data point to the nearest integer, and divide them into 3100, 3200, etc.).

   (d) Compare the stemplot, the histogram, and the boxplot as graphical summaries of this distribution. Which do you prefer? Give reasons for your answer.

9. (IPS10-1.37) Gosset's data on double stout sales. William Sealy Gosset worked at the Guinness Brewery in Dublin and made substantial contributions to the practice of statistics.22 In his work at the brewery, he collected and analyzed a great deal of data. Archives with Gosset's handwritten tables, graphs, and notes have been preserved at the Guinness Storehouse in Dublin.23 In one study, Gosset examined the change in the double stout market before and after World War I (1914–1918). For various regions in England and Scotland, he calculated the ratio of sales in 1925, after the war, as a percent of sales in 1913, before the war. Here are the data: (10 Points)

   (a) Compute the mean for these data.

   (b) Compute the median for these data.

Bristol	94	Glasgow	66
Cardiff	112	Liverpool	140
English Agents	78	London	428
English O	68	Manchester	190
English P	46	Newcastle-on-Tyne	118
English R	111	Scottish	24

(c) Which measure do you prefer for describing the center of this distribution? Explain your answer. (You may include a graphical summary as part of your explanation.)

10. (IPS10-1.38) Measures of spread for the double stout data. Refer to the previous exercise.(10 Points)

(a) Compute the standard deviation for these data.

(b) Compute the quartiles for these data.

(c) Which measure do you prefer for describing the spread of this distribution? Explain your answer. (You may include a graphical summary as part of your explanation.)