Who is your TA (check one)?

🞏 De Andre 🞏 Corey 🞏 Jingkai 🞏 Neeraj

**SOC 3811/5811 - Problem Set #1**

[Note: SOC 5811 students should do ALL problems.

SOC 3811 students need not do the problems indicated as SOC 581 problems)

1. Read the article at <https://bit.ly/3kd6Y1Z> about a 2020 presidential election poll conducted this month in Pennsylvania by Monmouth University. (Be sure to read the “Methodology” section, too!) Then answer these questions:
   1. Opinion polls—like social surveys—are tools for studying some population. What population is being studied in this poll?

ANSWER: People registered to vote in Pennsylvania in the 2020 election

* 1. Was the entire population observed, or some sample of that population? If a sample, how was that sample selected? And how big was it?

ANSWER: A sample of that population was observed. Specifically, 400 registered voters were selected at random from a list of registered voters. The 400 “includes 139 contacted by a live interviewer on a landline telephone and 261 contacted by a live interviewer on a cell phone, in English.”

* 1. What key variables pertaining to how people will vote in the U.S. presidential election were measured about each individual?

ANSWER: Many questions were asked in the poll, but only 4 pertained to how people will vote in the presidential election: who the people will vote for (question 1); how sure they are that would will vote for that person (1b), how likely they are to vote for Trump (2), and how likely they are to vote for Biden (3).

1. Indicate whether the following variables are nominal, ordinal, or continuous.
   1. Sex: Male, female

ANSWER: Nominal

* 1. Height: Measured in inches

ANSWER: Continuous

* 1. Education: Measured as years of schooling completed

ANSWER: Continuous (or, arguably, ordinal)

* 1. Education: Measured as highest degree completed

ANSWER: Ordinal

* 1. Number of children: None, 1 or 2, 3 or 4, 5 or more

ANSWER: Ordinal (perhaps arguably continuous … but it’s a stretch)

1. Twenty people were surveyed and asked how many cars they own. Six people had zero cars, ten people had one car, two people had two cars, one person had three cars, nobody had four cars, and one person had five cars.

For the variable “number of cars owned,” generate a frequency distribution, a percentage distribution, a cumulative frequency distribution, and a cumulative percentage distribution.

ANSWER:

**Cum Cum**

**Value Freq Perc Freq Perc**

0 6 30% 6 30%

1 10 50% 16 80%

2 2 10% 18 90%

3 1 5% 19 95%

4 0 0% 19 95%

5 1 5% 20 100%

1. If you were to create a histogram of the variable “number of dogs owned” among a random subset of Americans today, do you think that histogram would be symmetric, right skewed, or left skewed? Approximately where do you think the center of the distribution would be located?

ANSWER: The distribution would be right skewed. It would be centered over 0 to 1 dogs, as most people have either 0 or 1 dogs.

1. Twenty people were surveyed and asked how many cars they own. Six people had zero cars, ten people had one car, two people had two cars, one person had three cars, nobody had four cars, and one person had five cars.

Compute the mean, median, mode, range, variance, standard deviation, and skewness of the distribution of the variable “number of cars owned.” Be sure to show your work.

**MEAN**



**MEDIAN**

If you arrange the observations from lowest to highest, you get:

0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 3 5

The median observation is the average of the 10th case (1) and the 11th case (1), or 1

**MODE**

The value appearing most frequently is 1.

**RANGE**

The difference between the maximum value and the minimum value equals 5 – 0 = 5.

**VARIANCE**

Y Yi – Mean(Y) [Yi – Mean(Y)]2 Y Yi – Mean(Y) [Yi – Mean(Y)]2

0 0–1.1=-1.1 -1.12=1.21 1 1–1.1=-0.1 -0.12=0.01

0 0–1.1=-1.1 -1.12=1.21 1 1–1.1=-0.1 -0.12=0.01

0 0–1.1=-1.1 -1.12=1.21 1 1–1.1=-0.1 -0.12=0.01

0 0–1.1=-1.1 -1.12=1.21 1 1–1.1=-0.1 -0.12=0.01

0 0–1.1=-1.1 -1.12=1.21 1 1–1.1=-0.1 -0.12=0.01

0 0–1.1=-1.1 -1.12=1.21 1 1–1.1=-0.1 -0.12=0.01

1 1–1.1=-0.1 -0.12=0.01 2 2–1.1=0.9 0.92=0.81

1 1–1.1=-0.1 -0.12=0.01 2 2–1.1=0.9 0.92=0.81

1 1–1.1=-0.1 -0.12=0.01 3 3–1.1=1.9 1.92=3.61

1 1–1.1=-0.1 -0.12=0.01 5 5–1.1=3.9 3.92=15.21



**STANDARD DEVIATION**



**SKEWNESS**



**THIS QUESTION IS JUST FOR SOC 5811 STUDENTS!**

1. Separately for men and women, compute the Index of Qualitative Variation for the variable “favorite kind of ice cream.” Be sure to show your work. Comparing the two values, would you say that women or men are distributed more evenly across categories of favorite ice cream?

**MEN**

pc=0.4, pv=0.2, ps=0.0, po=0.4

=0.64



**WOMEN**

pc=0.3, pv=0.3, ps=0.3, po=0.4

=0.72



Thus, we conclude that women are more evenly distributed across categories of favorite ice cream.

1. The distribution of the variable Y has a mean of 10 and a standard deviation of 4.

Compute the value of Y for cases that have the following Z scores

Z = 2 Y = (Y - 10)/4 = 2, so Y - 10 = 8, so Y = 18

Z = -1.5 Y = (Y - 10)/4 = -1.5, so Y - 10 = -6, so Y = 4

Z = 1.23 Y = (Y - 10)/4 = 1.23, so Y - 10 = 4.92, so Y = 14.92

Compute the Z score for cases that have the following values of Y

Y = 12 Z = (12-10)/4 = 0.5

Y = 7 Z = (7-10)/4 = -0.75

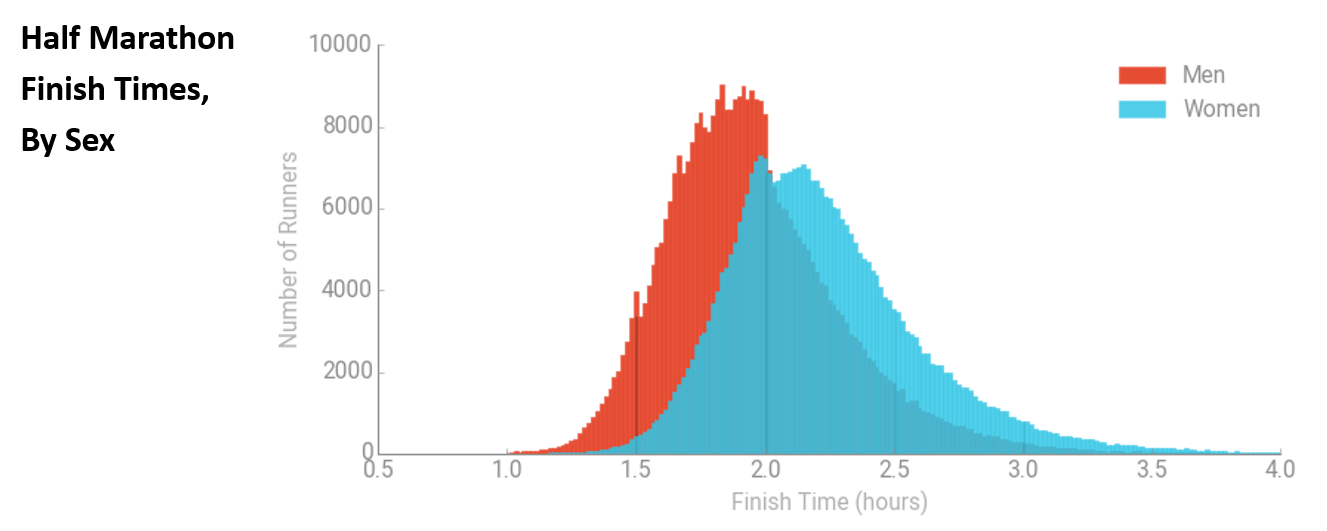
Y = 21 Z = (21-10)/4 = 2.75

1. If you learned that your score on an exam was 80 and the mean was 70, would you be more satisfied if the standard deviation was 5 or if it was 15? Explain.

ANSWER: If I scored 10 points above the mean and the standard deviation of exam scores equals 5, then I scored 2 standard deviation above the mean—because Z would equal (80-70)/5=2. That's pretty exceptional. However, if I scored 10 points above the mean and the standard deviation equals 15, then I only scored 2/3 of a standard deviation above the mean. Not so exceptional. Then, my score would be above average but not terrible high compared to many other students. So, I’d rather have the standard deviation equal 5.

**THIS QUESTION IS JUST FOR SOC 5811 STUDENTS!**

1. Below are distributions of the time it took men and women to run a half marathon. Evaluate the statement “Men run faster than women.” What does that statement mean, exactly? In what ways is that an accurate statement, and in what ways is it a misleading or incomplete statement?



ANSWER: In one sense, it is accurate to say that “men run faster than women” because the center of the distribution for men is at a lower time than the center of the distribution for women. The average man runs faster than the average woman. However, that does not mean that all men run faster than all women; there are many women who run fatser than many men. There are many men who run slower than the average women; there are also many women who run faster than the average man. Remermber that statements about distributions—where the center of a distribution is, for example—apply to the pattern of data depicted in the distribution. But distributions have variability, and statements that summarize a distribution may not apply to many individuals in the data.