Stats 300 **Spring 2024**

Lab Assignment #2

This lab is due at 9:35 AM on Wednesday, 1/24 and is worth 6 points. This may be done individually, or in a group of 2 or 3 people.

Part A

Go to the website: https://en.wikipedia.org/wiki/List_of_chemical_elements . For this data set, answer the following questions.

1) What are the individuals in the data set?

2) What is the population size?

3) What might be an interesting question that could be answered using this data set, other than "what is the average number of letters in the names of all the elements?"

4) One variable is density. What are some values of this variable? List at least four.

5) What are some other variables in this data set? List at least four.

6) For density and each variable you list in Q5, classify as quantitative or qualitative.

7) Classify each quantitative variable from Q6 as discrete or continuous.

8a) Pretend that there is one more column in this table which is: number of letters in the name of the element. What type of variable would this be?

8b) Explain why the average number of letters in the names of all the elements is not a variable.

8c) What is it instead? Hint: it starts with "p".

Part B

Go to the website:
https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States .
Look at the first big table, the one that goes from Alabama to Wyoming. For this data set, answer the following questions.

9) What are the individuals in the data set?

10) What is the population size? (Careful: "population" still means the stats meaning, not the number of people in a particular state.)

11) What might be an interesting question that could be answered using this data set, other than "what fraction of US states have more than 10,000,000 people?"

12) One variable is water area, measured in square kilometers. What are some values of this variable? List at least four.

13) What are some other variables in this data set? List at least four.

14) Would you classify the variable "ratification or admission" as quantitative or qualitative? Explain your reasoning.

15) Would you classify the variable "population" (number of people in a particular state) as discrete or continuous? Explain your reasoning.

16) Would you classify "total area, square miles" as discrete or continuous? Explain your reasoning.

17a) Explain why the fraction of US states with more than 10,000,000 people is not a variable.

17b) What is it instead? Hint: it starts with "p".

Part C

Go to mlb.com. We will be looking at game scores from the 2019 season. The population for this question is: all games played in 2019. The interesting question is: what proportion of all games in 2019 were won by the home team?

Start by using this sample: all games played on May 2, June 17, Aug 1, and Sep 6. To find these results, go to mlb.com, click on scores, click on the calendar icon next to the dates, then find May 2019 and click on 2. Repeat for the other 3 dates.

For each game, determine whether or not the home team won the game. The number of runs (R) by each team determines the winning team. The home team is always on the bottom. So for example, the first score says the Reds had 0 and the Mets had 1, and the Mets are on the bottom, so the home team won that game. Meanwhile, in the Rays-Royals game, the road team (Rays) won 3-1.

18) What is the variable? Hint: the variable itself is a 2-option question.

19) What type of variable is this? (Qual/quant?)

20) How many games are in this sample (using all the games from these 4 selected dates)?

21) How many games in this sample were won by the home team?

22) What proportion of games in this sample were won by the home team? This is called the sample proportion. Write as a percent and write as a decimal. No fractions please.

23) What type of sampling method was used here? (Hint, not a simple random sample.)

24a) Explain why "the proportion of games in this sample that were won by the home team" is not a variable.

24b) What is it? (Starts with "s".)

Imagine the following data collection method. Look at each of the 30 teams. Randomly choose 3 of their games from the 2019 season. Determine whether or not the home team won each game. Don't actually do this data collection. Use your IMAGINATION.

25) What type of sampling method is this? (Hint, not the same as Q23, and not a simple random sample.)

Imagine the following data collection method. Look at all the games from the 2019 season, listed as they appear on mlb.com, from the first day of the season to the last. Use every 50th game on the list in your sample. Don't actually do this data collection. Use your IMAGINATION.

26) What type of sampling method is this? (Hint, not the same as Q23 or Q25, and not a simple random sample.)

New topic:

27) Describe the steps one would have to follow to create a simple random sample of 60 games from the 2019 season.

28) Is this something that you would want to do?

Finish part C:

To find the actual proportion of all 2019 MLB games won by the home team, go to mlb.com, click on standings, click on Regular Season, then select 2019 (make sure the date is September 29, that was the end of the season), and look under the "Home" heading. For example, the Yankees were 57-24 at home, so they won 57 home games. Add up these numbers for the 30 teams to get the total number of wins by home teams. Then divide by 2,429, which is the total number of games played. (Looks like there was one rainout that was not made up.)

29) How many games in 2019 were won by the home team?

30) What proportion of games in 2019 were won by the home team? Write as a percent and write as a decimal. No fractions please.

31) How far apart are your answers to Q22 and Q30? Write as a percent and write as a decimal.

32) We will learn later that generally we would expect the difference (written as a decimal) between Q22 and Q30 to be less than about $2 * \sqrt{\frac{p(1-p)}{n}}$, where *p* is the proportion of all games won by the home team, written as a decimal (Q30), and *n* is the sample size (Q20). Calculate $2 * \sqrt{\frac{p(1-p)}{n}}$. Write as a decimal. Write as a percent.

33) Is your answer to Q31 smaller than Q32? If yes, then there is nothing unusual about this sample. Even if the numbers are not exactly the same, or don't seem that "close."