Problem Set 1: A Review of Basic Concepts
Due: Sep 06, 2024

This problem set is meant to refresh some basic ideas and concepts you might have learned previously in statistics. For every question, I have indicated the relevant section in the textbook for your reference. Some of these questions are open-ended and may not have one right/wrong answer.

Problem 1

A professor has a sample of 30 student scores with a mean of 10, a variance of 9, and a standard deviation of 3.

a. Realizing that the student scores are very low, the professor decides to multiply each student's score by 3 (how generous!). Explain whether (and why) the mean of the new scores would change. (see section 1.2)

b. Would the variance increase, decrease, or stay the same? Why? (see section 1.2.2)

c. Suppose the professor realizes that she had erroneously entered the highest score as 19 instead of 29 in the original scores (before multiplying by 3). basing on this new scenario, answer the following questions:

    i. Would the mean of the new scores change? Why? (see section 1.2)

    ii. Would the median of the new scores change? Why? (see section 1.2.1)

    iii. Would the variance increase, decrease, or stay the same? Why? (see section 1.2.2)

    iv. Would the range increase, decrease, or stay the same? Why? See (Defintion of range)

v. Would the interquartile range (IQR) of the new scores increase, decrease, or stay the same? Why? (see section 1.2.1)

Problem 2

The first step in conducting research is to identify topics or questions that are to be investigated. A clearly laid out research question is helpful in identifying what subjects or cases should be studied and what variables are important. It is also important to consider how data are collected so that the data are reliable and help achieve the research goals. Suppose you are conducting a study trying to answer the following research question: ***Does a new drug (call it drug X) reduce the number of deaths in patients with severe heart disease?***

    a. What type of study (experimental or observational) would you conduct? Explain why? (see section 1.4)

    b. Briefly describe how you would set up your study and the data you would collect to answer the above question. In your description describe the variables and their types (i.e., categorical or numerical). (see sections 1.2.1 - 1.2.2 and section 2.2)

    c. Suppose you wanted to conduct your study in Tompkins County. Explain what your population of interest? What would be your sample and what sampling strategies would you use to obtain it? (see section 2.1 and 2.2)

d. What are some potential ethical concerns you would need to consider before conducting your study?

Problem 3

Gamification is the application of game- design elements and game principles in non-game contexts. In educational settings, gamification is often implemented as educational activities to solve problems by using characteristics of game elements. Researchers investigating the effects of gamification on learning statistics conducted a study where they split college students in a statistics class into four groups: (1) no reading exercises and no gamification, (2) reading exercises but no gamification, (3) gamification but no reading exercises, and (4) gamification and reading exercises. Students in all groups also attended lectures. Students in the class were from two majors: Electrical and Computer Engineering (n = 279) and Business Administration (n = 86). After their assigned learning experience, each student took a final evaluation comprised of 30 multiple choice question and their score was measured as the number of questions they answered correctly. The researchers considered students' gender, level of studies (first through fourth year) and academic major. Other variables considered were expertise in the English language and use of personal computers and games, both of which were measured on a scale of 1 (beginner) to 5 (proficient). The study found that gamification had a positive effect on student learning compared to traditional teaching methods involving lectures and reading exercises. They also found that the effect was larger for females and Engineering students. (see Legaki et al., 2020).

    a. Assuming the data from this study were recorded in a single data frame, how many cases and how many variables would be in this data frame? (see section 1.2.3)

    b. List all variables in the study and identify their type by putting a check mark in the relevant column in the table below:

Variable Name	Categorical	Numerical

c. Was the study experimental or observational? Identify a suitable explanatory and response variable(s).

d. Name a variable for which we could use a histogram as a visualization tool.(see chapter 5)

e. Name a variable for which we could use a simple bar plot for visualization. (see chapter 5)

f. Name a variable set for which we could use a scatter plot for visualization.(see chapter 5)

Submission

- Write your answers in the spaces provided, then scan your document into a pdf (if you wrote on paper)
- Submit your scanned pdf on Canvas by the due date (Sep 06, 2024 at 11.59 pm).