

Chapters covered: Chapter 1, 2, 4

Show your work to receive full credit.

- **Problem 1** Statistics department at a large university wants to determine what proportion of students are interested in taking intro statistics courses during summer. The department asks 20 randomly selected students from Stat Club meetings and found none of them are interested in taking intro statistics during the summer.
Identify the following :
 - a) Population of interested :
 - b) Sample :
 - c) Is it a good sample? Why/why not?
 - d) Parameter :
 - e) Statistic :

- **Problem 2** Identify each of the following variables as continuous, discrete or categorical.
 - a) The exact length of time to run a marathon
 - b) The number of courses for which a student has received credit
 - c) The amount of money in your pocket (Hint: You could regard a number such as \$12.75 as 1275 in terms of “the number of cents”)
 - d) The amount of fuel in a car when you measure it precisely with values such as 4.573 gallons, etc.

- **Problem 3** For each of the following variables, indicate whether you would expect its histogram to be symmetric, skewed to the right, or skewed to the left. Explain why.
 - a) Assessed value of houses in a large city (Hint: There are only a few homes with extremely high assessed value.)
 - b) Time spent on a difficult exam (where most students used up the maximum time allowed (2 hours) while a few finished early or gave up)
 - c) Number of siblings, for each student in your class.
 - d) The height of female college students

- **Problem 4** The distribution of sales of a fast-food restaurant has a symmetric unimodal bell-shaped with a mean of \$1100 and a standard deviation \$250. Today the sale equaled \$2000. Is today an unusually good sale day? Use the general rule of standard deviation (three-standard deviation rule from the course lecture notes page 30) to justify your answer.

- **Problem 5** During the 1980s approximately 22,000 physicians over the age of 40 agreed to participate in a long-term study called the Physicians’ Health Study. One question investigated was whether aspirin helps lower the rate of heart attacks. The physicians were randomly assigned to take aspirin or take a placebo.
 - a) Identify the response variable and the explanatory variable.

- b) What is the type of this study? Is this an experiment or an observational study? Explain why.
- c) There are other explanatory variables, such as the amount of exercise a physician got, that we would expect to be associated with the response variable. Explain how such a variable is dealt with by the randomized nature of the experiment.

R Problems

For R Problems 1 - 3, we will use the Class survey data from STAT 3011 Fall 2022 consists of 391 respondents and 24 variables.

R Problem 1:

Use the following R command to import the data set.

```
survey_f22<-read.csv("http://users.stat.umn.edu/~parky/Fall2022Survey.csv", header=TRUE)
```

- a) Explore the imported data and identify at least one variable that is i) categorical, and ii) quantitative. Remember R is case-sensitive.
- b) Construct a histogram of a variable that represents the number of hours spent on social media. i) Submit the plot. ii) Describe the overall shape of the distribution iii) State if there are any outliers.

R Problem 2

In this question, we will explore the distribution of students' sleep hours during the week.

- a) What is the name of the corresponding variable?
- b) What is the 5-number summary of students' weekday sleep hours? (Hint: Use `summary()` command).
- c) Use R to calculate the mean and standard deviation of weekday sleep hours. Suppose that you slept 9 hours and 30 minutes. How many standard deviations is your sleep hours away from the mean?

R Problem 3

Suppose this time, you want to compare sleep hour distribution for students who live on campus v.s. off campus. i) Construct a side-by-side boxplot. ii) Submit your plot with an appropriate title. iii) State at least one similarity and one difference between these two groups.