

Chapters covered: Chapter 6

Show all of your work to receive full credit.

Textbook Problems : (Statistics, the art and science of learning from data, 4th)

- **Problem 1:** Let $X = \{0, 1, 2, 3, 4, 5, 6, 7\}$ be the number of days it rains in a week. Based on a monthly record, the estimate is $P(0)=0.28$, $P(1)= 0.28$, $P(2)= 0.15$, $P(3)=0.12$, $P(4)=0.09$, $P(5)=0.06$, $P(6)=0.02$, $P(7)=0$
 - a. Explain why it does not make sense to compute the average number of days that rains as $(0+1+2+3+4+5+6+7)/7=4$
 - b. Calculate the correct mean.

- **Problem 2:** Exercise 6.11 (a,b,c)
- **Problem 3:** Exercise 6.25
- **Problem 4:** Exercise 6.32
- **Problem 5:** Exercise 6.58
- **Problem 6:** Exercise 6.64
- **Problem 7:** Exercise 6.71

R Problem 1

We will look at the distribution of the number of tattoos for male vs. female students in STAT3011 using R. Let X be the random variable representing the number of tattoos for female students and Y be the random variable for males. Use the following codes to load the dataset and show relevant R codes when addressing the questions below.

```
data <- read.csv("http://users.stat.umn.edu/~parky/SurveyFall2019.csv", header = T)
attach(data)
```

- a. Create a frequency table of the number of tattoos that female students have.
- b. Find the number of female students in the dataset.
- c. Use the table from part a. to construct the probability distribution.
- d. Check that the probability distribution in part b. is legitimate using the rules for what constitutes a probability distribution.
- e. Find the probability that a randomly selected female student has at least one tattoo.

R Problem 2

Consider the Georgia student survey dataset from the textbook, *Statistics: The Art and Science of Learning from Data* by Agresti, Franklin, and Klingenberg, 4th edition. Use the `read.csv()` command to load the dataset in R and show relevant R codes when addressing the questions below.

- a. Plot the histogram and density curve of HSGPA. Comment on the shape of the histogram.
- b. Plot the normal Q-Q plot for height of the students. Is it heavy-tailed? Assess the normality.
- c. Find the probability that a randomly selected student's height is greater than or equal to 70 inches assuming the random variable follows a normal distribution.
- d. Find the probability that a randomly selected student's height is less than or equal to 60 inches assuming the random variable follows a normal distribution.

REFERENCE: Agresti, A. and Franklin, C. (2017). Statistics: the art and science of learning from data. Pearson, 4th ed.