STA13 Homework 1

September 25, 2021

1. Please reflect and explain the following key concepts for Chapter 1.

- What is the relationship between Sample and Population?
- How to describe univariate categorical data graphically?
- How to describe univariate numerical data graphically?
- If you were given a set of samples for a univariate numerical data, how would you summarize its key characteristics in terms of center and variability? Please try to explain these terms both in exact formula and intuition.

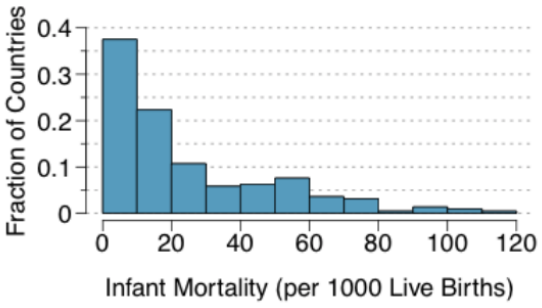
---

2. Make-up exam. In a class of 25 students, 24 of them took an exam in class and 1 student took a make-up exam the following day. The professor graded the first batch of 24 exams and found an average score of 74 points with a standard deviation of 8.9 points. The student who took the make-up the following day scored 64 points on the exam.

1. Does the new student's score increase or decrease the average score?
2. What is the new average?
3. Does the new student's score increase or decrease the standard deviation of the scores?

---

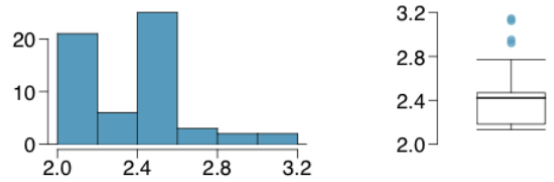
3. Infant mortality. The infant mortality rate is defined as the number of infant deaths per 1,000 live births. This rate is often used as an indicator of the level of health in a country. The relative frequency histogram below shows the distribution of estimated infant death rates for 224 countries for which such data were available in 2014.



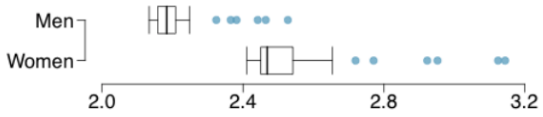
Estimate Q1, the median, and Q3 from the histogram.

---

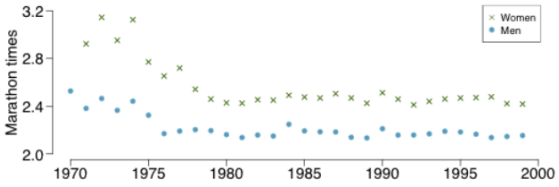
4. Marathon winners. The histogram and box plots below show the distribution of finishing times for male and female winners of the New York Marathon between 1970 and 1999.



1. What features of the distribution are apparent in the histogram and not the box plot? What features are apparent in the box plot but not in the histogram?
2. What may be the reason for the bimodal distribution? Explain.
3. Compare the distribution of marathon times for men and women based on the box plot shown below.



time series plot shown below is another way to look at these data. Describe what is visible in this plot but not in the others.



---

5. Sleeping in college. A recent article in a college newspaper stated that college students get an average of 5.5 hrs of sleep each night. A student who was skeptical about this value decided to conduct a survey by randomly sampling 25 students. On average, the sampled students slept 6.25 hours per night. Identify which value represents the sample mean and which value represents the claimed population mean.