

Biostat 120A Discussion 2 Exercises

Fundamentals of Biostatistics (Lectures 1-4)

Problem 1. Summary Statistics and Data Visualization (Lecture 2)

The data in `serum_cholesterol.xlsx` are a sample of cholesterol levels taken from 24 hospital employees who were on a standard American diet and who agreed to adopt a vegetarian diet for 1 month. Serum-cholesterol measurements were made before adopting the diet and 1 month after. See Table 1 for the first 10 observations.

(Problem adopted from Exercises 2.13 - 2.18 Bernard Rosner, Fundamentals of Biostatistics, 2nd Edition).

Serum-cholesterol levels (mg/dL) before and after adopting a vegetarian diet

Subject	Before	After
1	195	146
2	145	155
3	205	178
4	159	146
5	244	208
6	166	147
7	250	202
8	236	215
9	192	184
10	224	208

(a) With your favorite statistical software, create a new column named `difference`, which is the difference between `Before` and `After` variables such that `difference = Before - After`.

(b) Generate the following summary statistics: mean, median, standard deviation, range, and IQR and comment on the distribution of the data. If you had to pick summary statistics to an investigator, which summary statistics would you choose and why?

(c) Construct a stem-and-leaf plot of the cholesterol changes.

(d) Construct a box-plot of the cholesterol changes. Are there any outliers in your boxplot? Explain.

(e) Some investigators believe that the effects of diet on cholesterol are more evident in people with high rather than low cholesterol levels. If you split the data in according to whether baseline cholesterol is above or below the median, can you comment descriptively on this issue?

Hint: the best way to do this would be to create an indicator variable that indicates whether baseline cholesterol is above or below the median and then run the analysis.

Define variables: `AbsDiff = abs(Difference)` and `CholesterolStatus = ifelse(Before >= median(Before), "High Cholesterol", "Low Cholesterol")`.

Problem 2. Dice Race (Lecture 3).

You and your friend are in a Dice Race where roll a 6-sided dice with numbers 1 through 6. You win the dice race if your number is greater than your friend's number. For example, if I roll a 6 and my friend rolls a 2, I win the dice race.

(a) What is the probability that you win the dice race in one round?

(b) What is the probability that you win at least one round out of five rounds?

Problem 3. The Freshman Flu? or Norovirus? (Lecture 3).

You are a doctor at the Arthur Ashe Student Center and it is allegedly freshman flu season. You have been observing a high percentage of students coming to the clinic with a high fever, a symptom of both the Flu and Norovirus.

In the general student population, it has been believed that the prevalence of students with the flu is 40% and the prevalence of students with Norovirus is 60%.

If a student has the Flu, the probability of the student developing a fever is 85% and if a student has Norovirus, the probability of the student developing a fever is 25%.

(a) For a randomly selected student, what is the probability that this patient has a Fever?

(b) If a student has a fever, what is the probability that the student has the Flu? What is the probability that the student has Norovirus?