

Day 8b Homework Answers

43. 10, 16, 18, 20, 22, 28

a) mean = 19 std.dev. = 6.0

b) 110, 116, 118, 120, 122, 128

The std. dev. will be the same because the spread of the numbers didn't change. They are still the same distance from the average, it's just that the average changed.

44. 10, 14, 15, 16, 20

a) mean = 15 std.dev. = 3.6

b) 10, 11, 15, 19, 20

The std. dev, will now be higher because 11 and 19 are farther from the mean than 14 and 16 were.

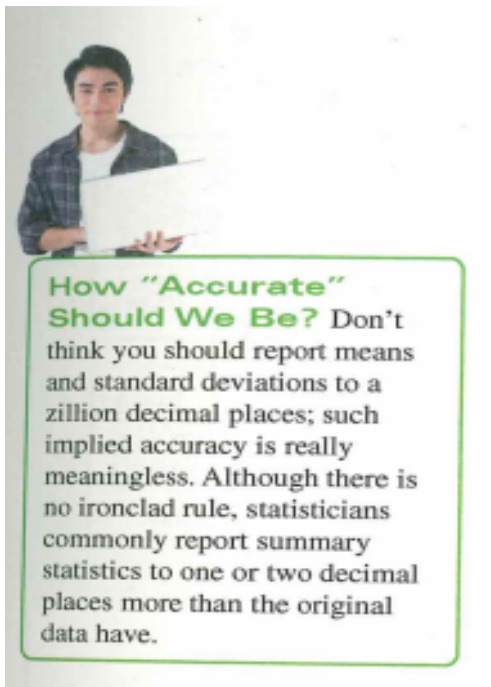
Think, Show, Tell

What should you *Tell* about a quantitative variable?

Textbook pg. 67

- Start by making a histogram, dotplot, or stem-and-leaf display, and describe the shape of the distribution.
- Next, discuss the center *and* spread.
 - (1) Always report *both* center and spread.
 - (2) Always pair the median with the IQR or the mean with the standard deviation.
 - (3) If the shape is skewed, report the median and IQR. You may want to include the mean and standard deviation as well, but you should point out why the mean and median differ.
 - (4) If the shape is symmetric, report the mean and standard deviation (and possibly the median and IQR as well). For unimodal symmetric data, the IQR is usually a bit larger than the standard deviation. If that's not true of your data set, look again to make sure that the distribution isn't skewed and there are no outliers.
- Also, discuss any unusual features:
 - (1) If there are multiple modes, try to understand why. If you can identify a reason for separate modes (for example, women and men typically have heart attacks at different ages), it may be a good idea to split the data into separate groups.
 - (2) If there are any clear outliers, point them out. If you are reporting the mean and standard deviation, you may want to report them with the outliers present and with the outliers removed. The differences may be revealing. (Of course, the median and IQR won't be affected very much by the outliers.)

symmetric
Skewed L/R
Uniform

A photograph of a man with dark hair, wearing a plaid shirt over a white t-shirt, holding a laptop. He is standing in front of a whiteboard. A green-bordered text box is overlaid on the bottom right of the image.

How "Accurate" Should We Be? Don't think you should report means and standard deviations to a zillion decimal places; such implied accuracy is really meaningless. Although there is no ironclad rule, statisticians commonly report summary statistics to one or two decimal places more than the original data have.

textbook pg. 67

pg. 67-68 from textbook...

Step-By-Step Example SUMMARIZING A DISTRIBUTION



One of the authors owned a 1989 Nissan Maxima for 8 years. Being a statistician, he recorded the car's fuel efficiency (in mpg) each time he filled the tank. He wanted to know what fuel efficiency to expect as "ordinary" for his car. (Hey, he's a statistician. What would you expect?¹²) Knowing this, he was able to predict when he'd need to fill the tank again and to notice if the fuel efficiency suddenly got worse, which could be a sign of trouble.

QUESTION: How would you describe the distribution of *Fuel efficiency* for this car?

THINK ➡ Plan State what you want to find out.

Variable Identify the variable and report the W's.

Be sure to check the appropriate condition.

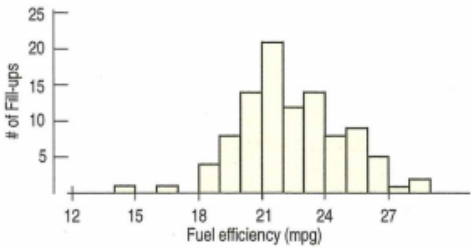
I want to summarize the distribution of Nissan Maxima fuel efficiency.

The data are the fuel efficiency values in miles per gallon for the first 100 fill-ups of a 1989 Nissan Maxima between 1989 and 1992.

✓ **Quantitative Data Condition:** The fuel efficiencies are quantitative with units of miles per gallon. Histograms and boxplots are appropriate displays for displaying the distribution. Numerical summaries are appropriate as well.

SHOW ➡ **Mechanics** Make a histogram. Based on the shape, choose appropriate numerical summaries.

REALITY CHECK A value of 22 mpg seems reasonable for such a car. The spread is reasonable, although the range looks a bit large.



A histogram of the data shows a fairly symmetric distribution with a low outlier.

Count	100
Mean	22.4 mpg
StdDev	2.45
Q1	20.8
Median	22.0
Q3	24.0
IQR	3.2

The mean and median are close, so the outlier doesn't seem to be a problem. I can use the mean and standard deviation.

TELL ➡ **Conclusion** Summarize and interpret your findings in context. Be sure to discuss the distribution's shape, center, spread, and unusual features (if any).

The distribution of mileage is unimodal and roughly symmetric with a mean of 22.4 mpg. There are one or two low outliers that should be investigated, but it does not influence the mean very much. The standard deviation suggests that from tankful to tankful, I can expect the car's fuel economy to differ from the mean by an average of about 2.45 mpg.

Note: Remember that we use mean and std. dev. to describe symmetric data, and we use median and IQR to describe skewed data.

Classwork: Pg. 77 #8, 9, 10 Pg. 78 #12, 13, 14

8. **Thinking about shape** Would you expect distributions of these variables to be uniform, unimodal, or bimodal? Symmetric or skewed? Explain why.

- a) The number of speeding tickets each student in the senior class of a college has ever had.
- b) Players' scores (number of strokes) at the U.S. Open golf tournament in a given year.

8. **Thinking about shape.**

- a) The distribution of the number of speeding tickets each student in the senior class of a college has ever had is likely to be unimodal and skewed to the right. Most students will have very few speeding tickets (maybe 0 or 1), but a small percentage of students will likely have comparatively many (3 or more?) tickets.
- b) The distribution of player's scores at the U.S. Open Golf Tournament would most likely be ~~unimodal~~ ^{bimodal (mis.w)} and slightly skewed to the right. The best golf players in the game will likely have around the same average score, but some golfers might be off their game and score 15 strokes above the mean. (Remember that high scores are undesirable in the game of golf!)

9. More shapes Would you expect distributions of these variables to be uniform, unimodal, or bimodal? Symmetric or skewed? Explain why.

- a) Ages of people at a Little League game.
- b) Number of siblings of people in your class.

9. More shapes.

- a) The distribution of the ages of people at a Little League game would likely be bimodal and skewed to the right. The average age of the players would be at one mode and the average age of the spectators (probably mostly parents) would be at the other mode. The distribution would be skewed to the right, since it is possible to have a greater variety of ages among the older people, while there is a natural left endpoint to the distribution at zero years of age.
- b) The distribution of the number of siblings of people in your class is likely to be unimodal and skewed to the right. Most people would have 0, 1, or 2 siblings, with some people having more siblings.

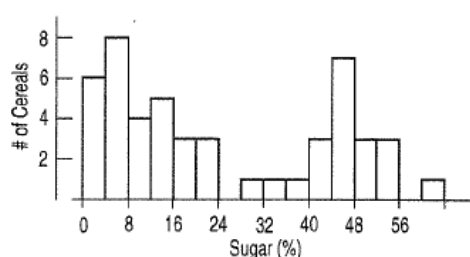
10. About shapes III Would you expect distributions of these variables to be uniform, unimodal, or bimodal? Symmetric or skewed? Explain why.

- a) Weights of female babies born in a particular hospital over the course of a year.
- b) The length of the average hair on the heads of students in a large class.

10. About shapes II.

- a) The weights of female babies in a particular hospital over the course of a year will likely have a distribution that is unimodal and symmetric. Most newborns have about the same weight, with some babies weighing more and less than this average. There may be slight skew to the left, since there seems to be a greater likelihood of premature birth (and low birth weight) than post-term birth (and high birth weight).
- b) The distribution of the length of the average hair on the heads of students in a large class would likely be bimodal and skewed to the right. The average hair length of the males would be at one mode, and the average hair length of the females would be at the other mode, since women typically have longer hair than men. The distribution would be skewed to the right, since it is not possible to have hair length less than zero, but it is possible to have a variety of lengths of longer hair.

- 12. Sugar in cereals** The histogram displays the sugar content (as a percent of weight) of 49 brands of breakfast cereals.

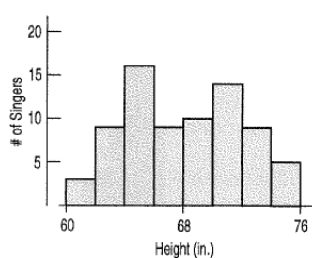


- Describe this distribution.
- What do you think might account for this shape?

12. Sugar in cereals.

- The distribution of the sugar content of breakfast cereals is bimodal, with a cluster of cereals with sugar content around 10% sugar and another cluster of cereals around 48% sugar. The lower cluster shows a bit of skew to the right. Most cereals in the lower cluster have between 0% and 10% sugar. The upper cluster is symmetric, with center around 45% sugar.
- There are two different types of breakfast cereals, those for children and those for adults. The children's cereals are likely to have higher sugar contents, to make them taste better (to kids, anyway!). Adult cereals often advertise low sugar content.

13. Singers The display shows the heights of some of the singers in a chorus, collected so that the singers could be positioned on stage with shorter ones in front and taller ones in back.

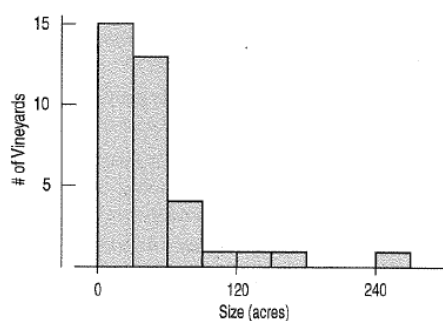


- Describe the distribution.
- Can you account for the features you see here?

13. Singers.

- The distribution of the heights of singers in the chorus is bimodal, with a mode at around 65 inches and another mode around 71 inches. No chorus member has height below 60 inches or above 76 inches.
- The two modes probably represent the mean heights of the male and female members of the chorus.

14. Vineyards The histogram shows the sizes (in acres) of 36 vineyards in the Finger Lakes region of New York. Write a brief description of this distribution (shape, center, spread, unusual features).



14. Vineyards.

The distribution of the size of 36 Finger Lakes vineyards is skewed to the right. Most vineyards are smaller than 75 acres, with a few larger ones, from 90 to 160 acres. One vineyard was larger than all the rest, over 240 acres. The mode of the distribution is between 0 and 30 acres.

Homework:

Pg. 78-79 #15 - 23

Book

Problems have been
copied into packet
PS. 66-67