Probability & Statistics Workbook Solutions

Analyzing data

krista king MATH

CENTRAL TENDENCY: MEAN, MEDIAN, AND MODE

■ 1. What is the mean of the data set?

$$105,\ 250,\ 358,\ 422$$

Solution:

To find the mean, add all the numbers in the data set, and then divide by the number of data points. This data set has 4 numbers so we get:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\mu = \frac{105 + 250 + 358 + 422}{4}$$

$$\mu \approx 283.75$$

The mean is 283.75.

■ 2. What is the median of the data set?

$$62,\ 64,\ 69,\ 70,\ 70,\ 71,\ 73,\ 74,\ 75,\ 77$$

Solution:

This data set has 10 values, which means to find the median we need to find the two middle numbers and take their mean.

$$\cancel{62, 64, 69, 70}, 70, 71, \cancel{73, 74, 75, 77}$$

Now we need to find the mean of 70 and 71.

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\mu = \frac{70 + 71}{2}$$

$$\mu = 70.5$$

The median of the data set is 70.5.

■ 3. What is the mode of the data set?

1	3 7 8
2	1 4 6
3	5 5
4	
5	2 6

$$1 | 3 = 13$$

Solution:

The mode of a data set is the number that repeats the most often. In the stem plot, the mode is 35.

■ 4. What number could you add to the data set that would give you a median of 15?

$$1, 2, 8, 13, 20, 30, 31$$

Solution:

The median of a data set is the middle number. In this data set, without changing anything, 13 is the middle number, so it's the median.

$$\cancel{1, 2, 8}, 13, \cancel{20, 30, 31}$$

If we were to add one more number to the data set, then to find the median we would take the average of the two middle numbers. We want the median to be 15, which means we'll need to insert a number larger than 13.

$$\cancel{1, 2, 8}, 13, \_\_, \cancel{20, 30, 31}$$

Let's call the missing number m. Then we can set up this equation:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$15 = \frac{13 + m}{2}$$

$$15(2) = 13 + m$$

$$30 = 13 + m$$

$$17 = m$$

This means 17 is the number we can add to the data set to force the median to be 15.

■ 5. A teacher lost Samantha's test after it was graded, but she knows the statistics for the rest of the class.

Class mean (including Samantha's test): $\mu = 85$

Total number of students who took the test: 18

Class test scores for everyone but Samantha were:

$$75, 75, 75, 80, 80, 80, 80, 80, 82, 82, 82, 82, 95, 95, 95, 95, 98$$

What did Samantha score on her test?

Solution:

To find the mean, add the test scores, then divide by the number of test scores. We know the mean is 85, and that there were 18 students who took the test. Let's call Samantha's missing test score s. Then we can set up this equation:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$85 = \frac{s + 3(75) + 5(80) + 4(82) + 4(95) + 1(98)}{18}$$

When we solve for s, we get

$$85(18) = s + 3(75) + 5(80) + 4(82) + 4(95) + 1(98)$$

$$1{,}530 = s + 3(75) + 5(80) + 4(82) + 4(95) + 1(98)$$

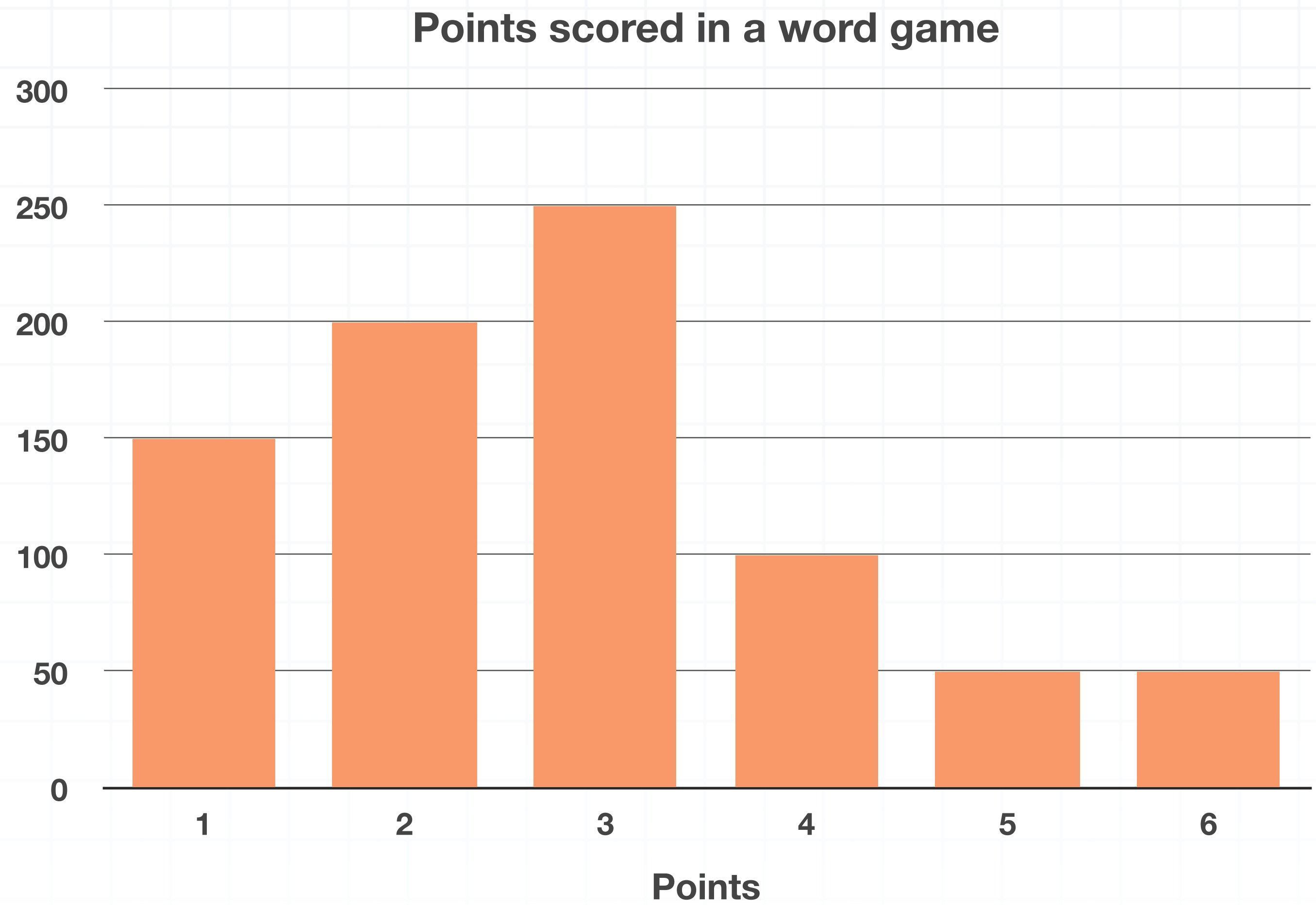
$$1{,}530 = s + 225 + 400 + 328 + 380 + 98$$

$$s = 1{,}530 - 225 - 400 - 328 - 380 - 98$$

$$s = 99$$

Samantha scored a 99 on her test.

6. What is the mode of the data set?



Solution:

This is a frequency histogram. More often than any other value, 3 points were scored in the word game. Therefore, the mode is 3.

SPREAD: RANGE AND IQR

1. Sarah is visiting dairy farms as part of a research project and counting the number of red cows at each farm she visits. Here is her data:

$$0,\ 1,\ 1,\ 1,\ 2,\ 5,\ 5,\ 7,\ 7,\ 18,\ 24,\ 24$$

Calculate the IQR and range of the data set.

Solution:

The range of the data is the largest number minus the smallest number. The smallest number in the data set is 0, and the largest number in the data set is 24, so the range is $24 - 0 = 24$.

To find the IQR, we need to find the upper median (called the upper quartile) and the lower median (called the lower quartile). To do this, we divide the data into four equal parts.

This data set has 12 data items, so we can find the median by crossing out the first five numbers and the last five numbers, and then take the average of the middle two numbers.

$$\cancel{0,\ 1,\ 1,\ 1,\ 2},\ 5,\ 5,\ \cancel{7,\ 7,\ 18,\ 24,\ 24}$$

The median is

$$\frac{5+5}{2} = \frac{10}{2} = 5$$

The lower half of the data set is $0, 1, 1, 1, 2, 5$, and its median is

$$\frac{1+1}{2}=\frac{2}{2}=1$$

The upper half of the data set is $5, 7, 7, 18, 24, 24$, and its median is

$$\frac{7+18}{2}=\frac{25}{2}=12.5$$

Therefore, the IQR is $12.5-1=11.5$.

2. A dog boarding company kept track of the number of dogs staying overnight and the frequency. What is the range of the data?

Number of dogs	Frequency
20	2
25	3
32	1
38	1
39	2
40	3
43	2

Solution:

The largest number in the data set is 43, and the smallest number is 20, so the range is $40 - 20 = 23$.

If you are wondering why we didn't really need the frequency side of the table, consider that, since this is a frequency table, the frequency tells us how many times each number appears. The list of data is actually

$$20, 20, 25, 25, 25, 32, 38, 39, 39, 40, 40, 40, 43, 43$$

So the range is still 23.

3. Catherine counted the number of lizards she saw in her garden each week and recorded the data in a table. What is the interquartile range of the data?

Number of lizards	Frequency
2	5
5	2
8	1
12	2
13	2
15	3
21	1

Solution:

Let's create a list from the table. The data set is

$$2, 2, 2, 2, 2, 5, 5, 8, 12, 12, 13, 13, 15, 15, 15, 21$$

There are 16 items in the data set, so we can cross off the first seven and last seven, and then find the average of the middle two numbers to get the median.

$$\cancel{2, 2, 2, 2, 2, 5, 5}, 8, 12, \cancel{12, 13, 13, 15, 15, 15, 21}$$

The median is

$$\frac{8+12}{2} = \frac{20}{2} = 10$$

The lower half of the data is $2, 2, 2, 2, 2, 5, 5, 8$, so the median of the lower half is

$$\frac{2+2}{2} = \frac{4}{2} = 2$$

The upper half of the data is $12, 12, 13, 13, 15, 15, 15, 21$, so the median of the upper half is

$$\frac{13+15}{2} = \frac{28}{2} = 14$$

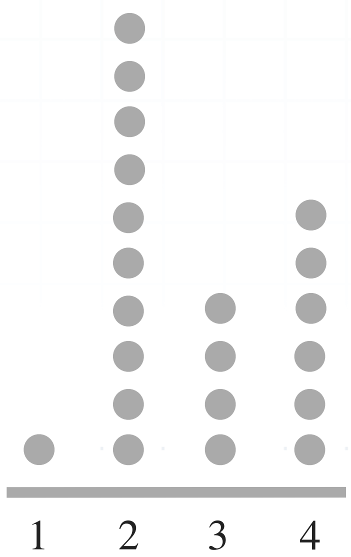
Therefore, the interquartile range is $14 - 2 = 12$.

■ 4. The median of the lower-half of a data set is 98. The interquartile range is 2. If the data set has 9 numbers, what can you say about the median of the entire data set?

Solution:

Since the median of the lower half of the data is 98 and the interquartile range is 2, you can find the median of the upper half of the data as $98 + 2 = 100$. This means the median of the data set is any number between or including 98 and 100.

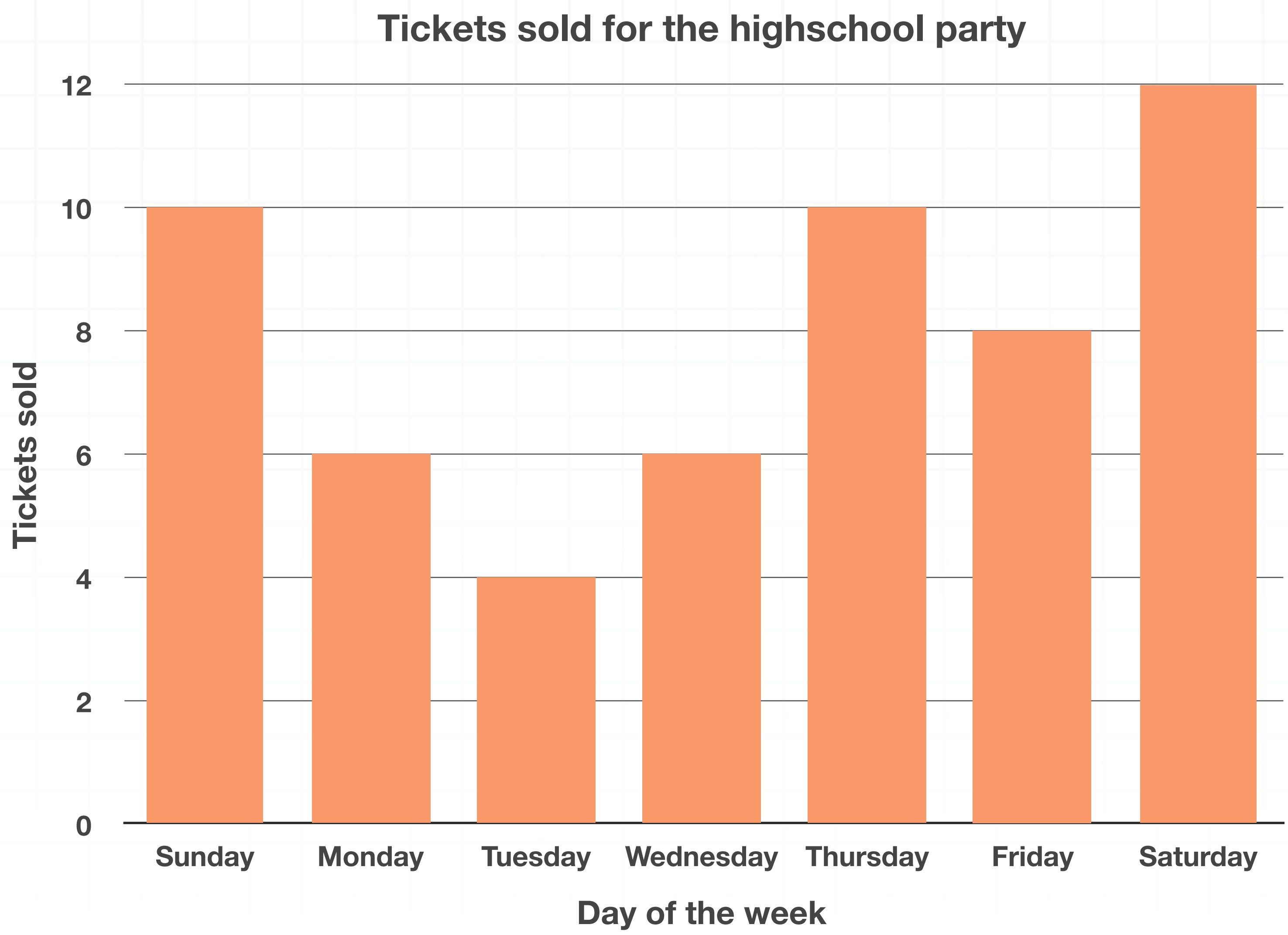
■ 5. The dot plot shows the number of trips to the science museum for a class of 4th graders. What is the range of the data set?



Solution:

The range is the largest number in the data set minus the smallest number. The largest number is 4, and the smallest number is 1, so the range is $4 - 1 = 3$. The range is 3.

6. The bar graph shows the number of tickets sold for the high school party each day. What is the interquartile range of the data set?



Solution:

We could list the data from the bar graph as

$$10, 6, 4, 6, 10, 8, 12$$

Put the data in order so we can find the median.

$$4, 6, 6, 8, 10, 10, 12$$

The median is 8. The lower half of the data is 4, 6, 6, so the median of the lower half is 6. The upper half of the data is 10, 10, 12, so the median of the upper half is 10.

Therefore, the IQR of the data set is $10 - 6 = 4$.

CHANGING THE DATA AND OUTLIERS

1. The students in an English class ended up with a mean score on their recent exam of 65 points. The range of exam scores was 25 points. If each score is increased by $10\,\%$, what are the new mean and range?

Solution:

Increasing the scores by $10\,\%$ is the same as multiplying the data set by 1.10. This multiplication both increases the scores and spreads out the data. This means that both the mean and the range will be multiplied by 1.10.

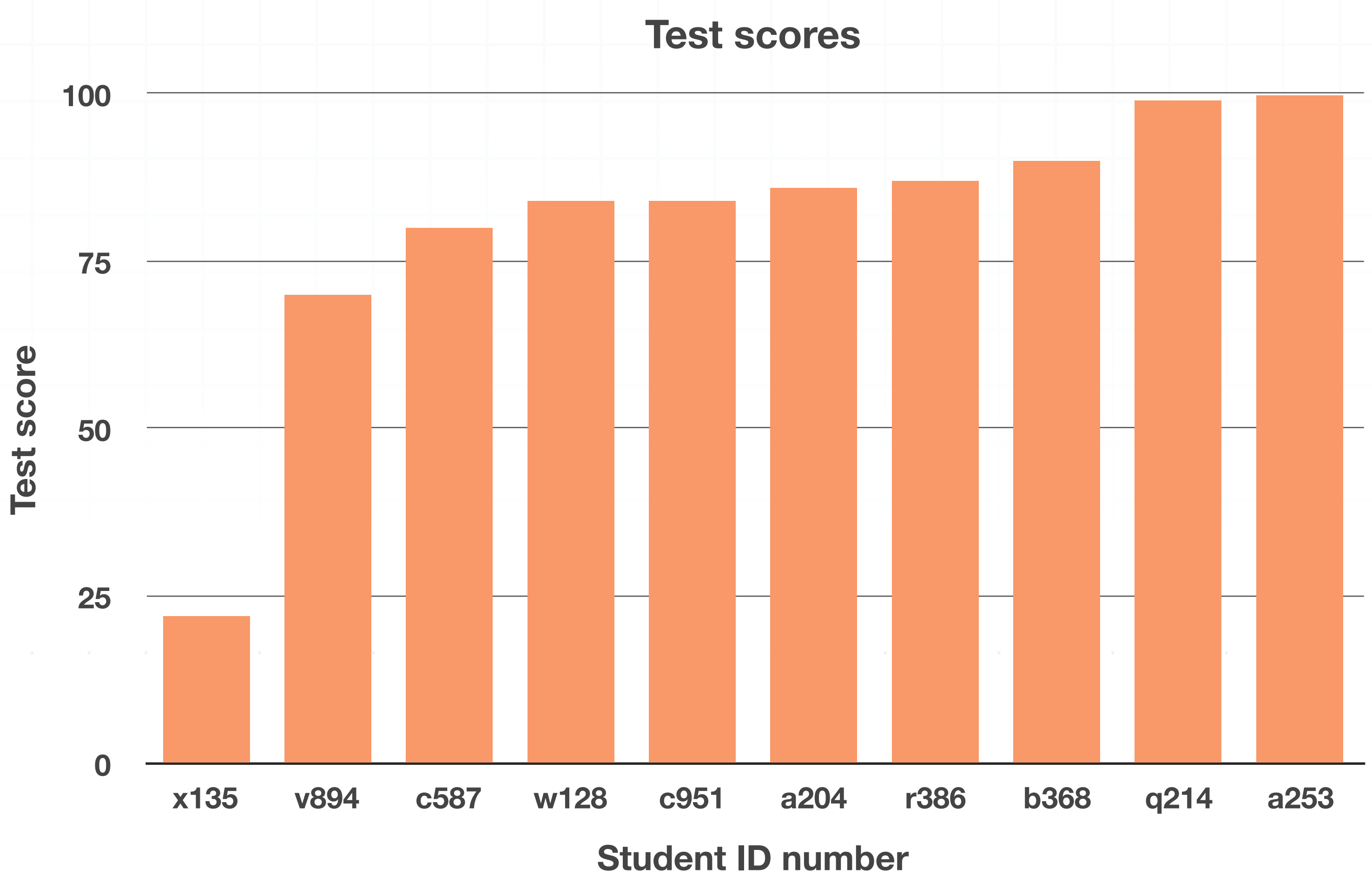
The original mean is 65 and the new mean is $65(1.10) = 71.5$. The original range is 25, and the new range is $25(1.10) = 27.5$.

2. Spencer asked students at his high school what percentage of the school budget they thought was spent on extracurricular activities. The mean response was $8\,\%$ and the median response was $5\,\%$. There was one outlier in the responses. What do the mean and median tell you about the outlier?

Solution:

The outlier was greater than the rest of the data because the mean is greater than the median. In other words, the outlier is pulling the mean toward the larger value. The median is more resistant to outliers, which is why it's much lower.

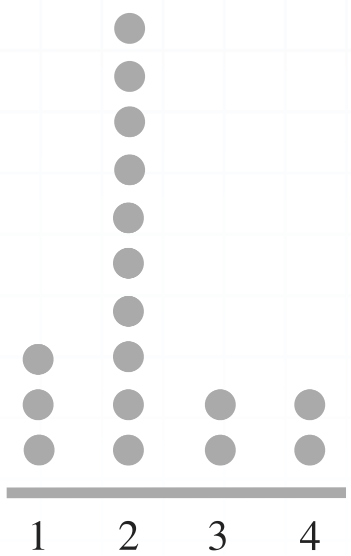
3. How does the mean compare to the median in the data from the bar graph?



Solution:

In this bar graph, there's one score from student $x135$ that's significantly lower than the rest. This means that the score is likely an outlier. This will make the mean smaller than the median because it'll pull the mean score down.

4. The dot plot shows the number of trips to the science museum for a class of 4th graders. How does the mean compare to the median in the data set below, and what does it tell you about the potential outliers in the data set?



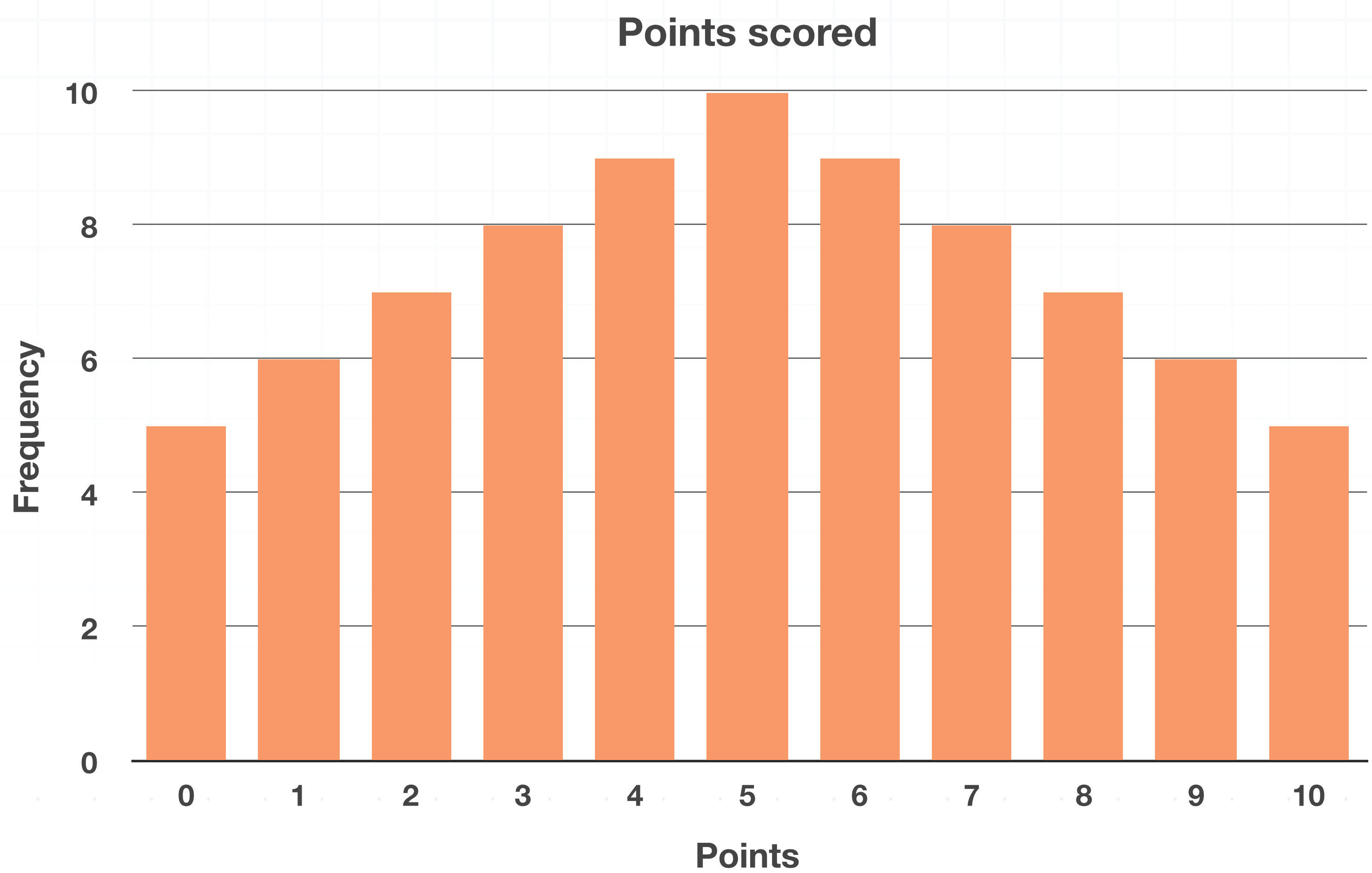
Solution:

You can calculate the mean and median from the dot plot. The median is 2 and the mean is

$$\mu = \frac{3(1)+10(2)+2(3)+2(4)}{3+10+2+2} \approx 2.18$$

The mean and median are close together, so this data set doesn't have outliers. The mean is being pulled upward from the median a bit because there are more students who visited the museum more than 2 times than there are students who visited the museum less than 2 times.

■ 5. What does the shape of this histogram tell you about the mean and median of the data?



Solution:

This data set is symmetric. The mean and median are equal to one another, and there are no outliers in the data.

6. An experiment is done in degrees Celsius. The original data had the following:

Mean: 102° Celsius

Median: 101° Celsius

Mode: 99° Celsius

Range: 7° Celsius

IQR: 4° Celsius

The formula to convert to degrees Fahrenheit is $F = (9/5)C + 32$. After the conversion to Fahrenheit, what are the new reported measures of the data set?

Solution:

Multiplying the data set by a constant value of $9/5$ will multiply all of these measures of center and spread as well.

Mean: $102^\circ(9/5) = 183.6^\circ$ Celsius

Median: $101^\circ(9/5) = 181.8^\circ$ Celsius

Mode: $99^\circ(9/5) = 178.2^\circ$ Celsius

Range: $7^\circ(9/5) = 12.5^\circ$ Celsius

IQR: $4^\circ(9/5) = 7.2^\circ$ Celsius

Shifting the data set by adding 32, will add 32 to the new mean, median and mode. The range and IQR will stay the same.

Mean: $183.6^\circ + 32^\circ = 215.6^\circ$ Celsius

Median: $181.8^\circ + 32^\circ = 213.8^\circ$ Celsius

Mode: $178.2^\circ + 32^\circ = 210.2^\circ$ Celsius

Range: 12.5° Celsius

IQR: 7.2° Celsius

BOX-AND-WHISKER PLOTS

■ 1. What is the range and interquartile range of the data set?

Median: $617{,}594$

Minimum: $216{,}290$

Maximum: $845{,}300$

First quartile: $324{,}528$

Third quartile: $790{,}390$

Solution:

The range is

$$845{,}300 - 216{,}290 = 629{,}010$$

The interquartile range is

$$790{,}390 - 324{,}528 = 465{,}862$$

■ 2. These are average lifespans in years of various mammals:

$$35,\ 10,\ 40,\ 40,\ 20,\ 10,\ 15,\ 14,\ 18,\ 35$$

Find the five-number summary for the data.

Solution:

The five-number summary is the list of the minimum, first quartile, median, third quartile and maximum values. We need to divide the data set into four parts to find the five-number summary, which we can do by arranging the numbers from least to greatest.

$$10,\ 10,\ 14,\ 15,\ 18,\ 20,\ 35,\ 35,\ 40,\ 40$$

Now we can see that the minimum is 10, that the maximum is 40, and that the median is

$$\frac{18+20}{2}=19$$

The lower half of the data set is $10,\ 10,\ 14,\ 15,\ 18$, so the median of the lower half is 14. The upper half of the data set is $20,\ 35,\ 35,\ 40,\ 40$, so the median of the upper half is 35. So we can summarize the five-number summary as

Min	Q1	Median	Q3	Max
10	14	19	35	40

■ 3. Create a box plot based on the following information about a data set.

Mode: 300

Minimum: 100

First Quartile: 300

Median: $2,000$

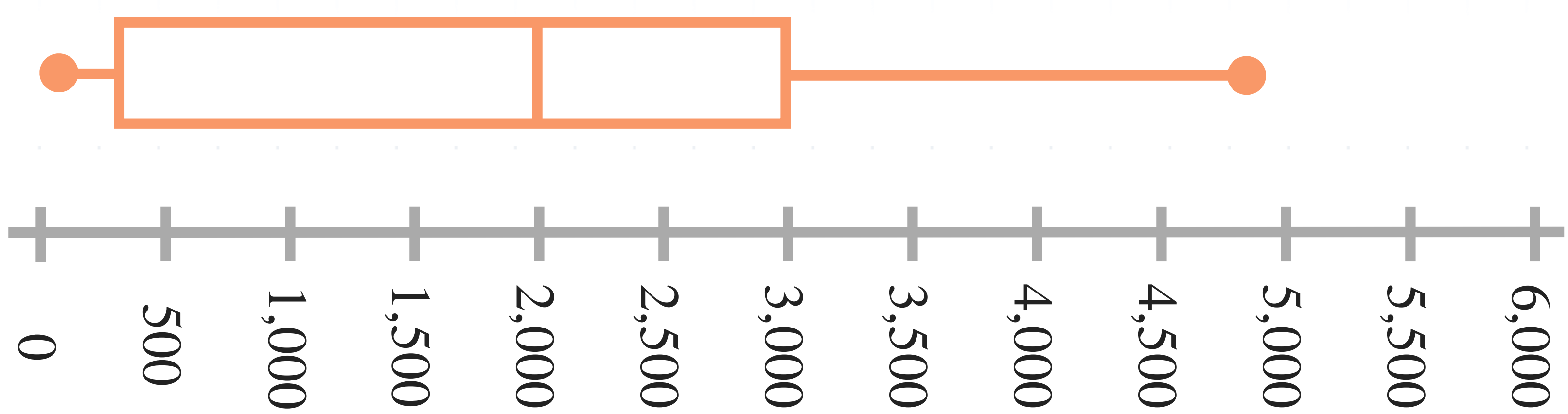
Mean: $1,887.5$

Third Quartile: $3,050$

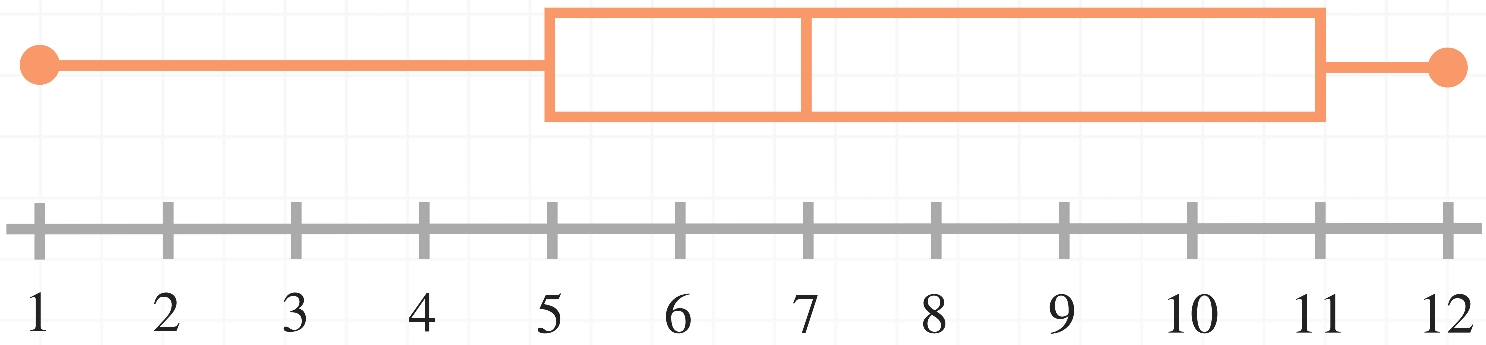
Maximum: $4,800$

Solution:

To create a box plot, you just need to use the five-number summary. The five-number summary is the list of the minimum, first quartile, median, third quartile, and maximum values. If we take those values from the question, then we can create the box plot.



■ 4. How does the amount of data between 1 and 5 compare to the amount of data between 11 and 12?

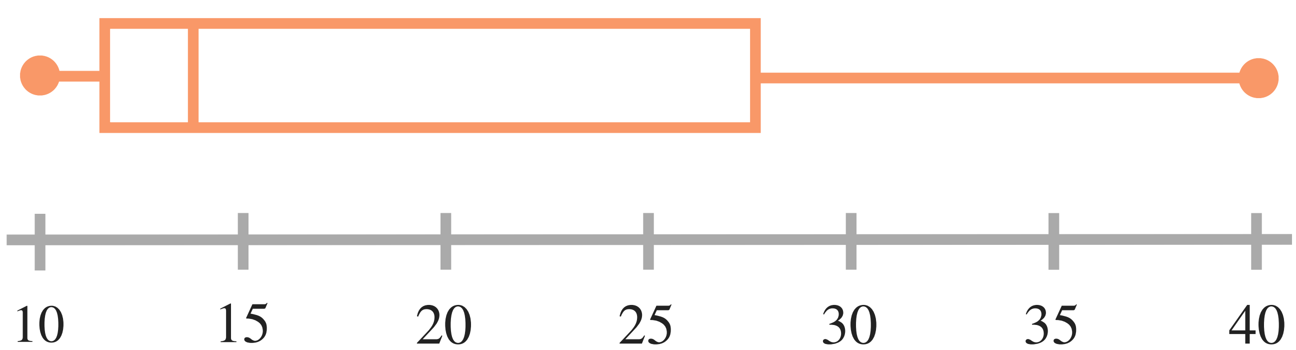


Solution:

The line from 1 to 5 is longer because the numbers in that quartile are more spread. In comparison the numbers in the quartile between 11 and 12 are less spread out.

For example, if the data set had 24 values, 6 of the values would fall between 1 and 5, and 6 of the values would fall between 11 and 12.

■ 5. In which quartile of the data is the number 23 located?



Solution:

The number 23 lies between the median and the right edge of the box, which is the third quartile of the data set.

6. Create the box-and-whisker plot for the book ratings given in the stem and leaf plot.

Stem	Leaf
1	3 7 8
2	1 4 6
3	5 5
4	
5	2 6

Key: $1\,|\,3 = 13$

Solution:

To create the box-and-whisker plot, we first need to create the five-number summary. The five-number summary is the list of the minimum, first quartile, median, third quartile, and maximum values. We need to divide the data set into four parts to find the five-number summary, so let's start by writing out the numbers in the data set.

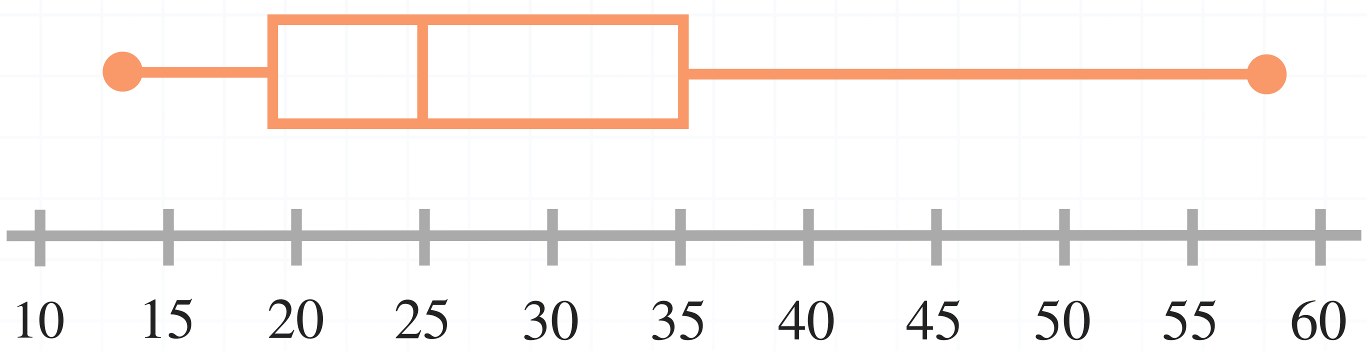
$$13,\ 17,\ 18,\ 21,\ 24,\ 26,\ 35,\ 35,\ 52,\ 56$$

Now we can see that the minimum of the data set is 13, and the maximum of the data set is 56. The median is

$$\frac{24+26}{2} = 25$$

The lower half of the data is $13, 17, 18, 21, 24$, so the median of the lower half is 18. The upper half of the data is $26, 35, 35, 52, 56$, so the median of the upper half is 35.

Now that we have the five-number summary given by the minimum, first quartile, median, third quartile, and maximum, we can sketch the box plot.



W W W . K R I S T A K I N G M A T H . C O M