8-4 Data Variability

What You'll Learn

CONTENT STANDARDS
7.SP.3, 7.SP.4

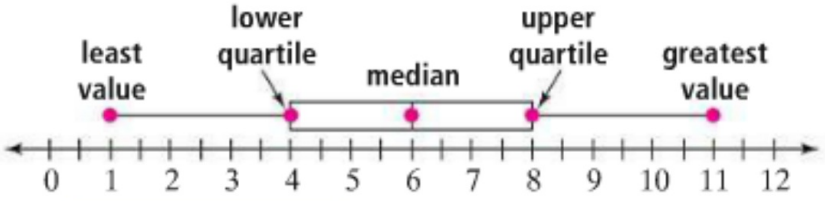
To compare data about two populations by using measures of center and measures of variability

New Vocabulary box plot, interquartile range (IQR), variability, mean absolute deviation (MAD)

Why Learn This?

You can use random samples to compare two populations.

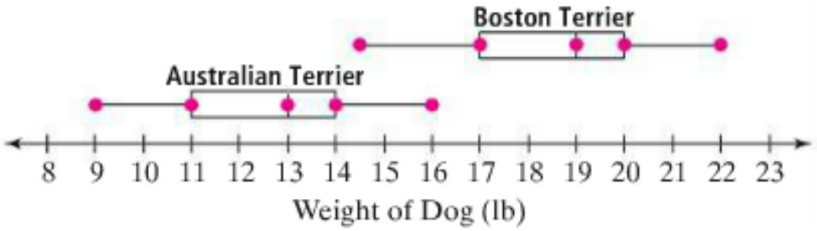
A **box plot** uses 5 points on a number line to summarize a data set. The box shows the middle 50% of the data. With a box plot you can observe the visual overlap of two data sets.



The **interquartile range** (IQR) of a data set is the difference between the upper and lower quartiles. The IQR is one measure of **variability**, which tells how much a data set is spread out.

EXAMPLE 1 Comparing Two Populations

A veterinarian collects data about the weights of the dogs she treats. Compare the IQRs of the data sets to draw an inference about the dogs.



IQR for Australian terriers:

$14 - 11 = 3$

IQR for Boston terriers:

$20 - 17 = 3$

The IQRs of the data sets are the same. So, you can infer that the weights of Boston terriers vary about as much as the weights of Australian terriers.

Example

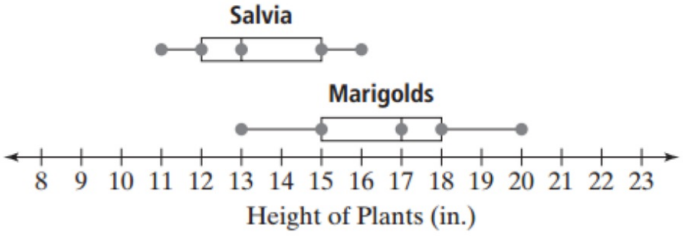
1 Comparing Two Populations A gardener collects information about the heights of the flowers she grows. Compare the IQRs of the data sets, and use the comparison to make an inference about the plants.

IQR for Salvia:

$15 - 12 = 3$

IQR for Marigolds:

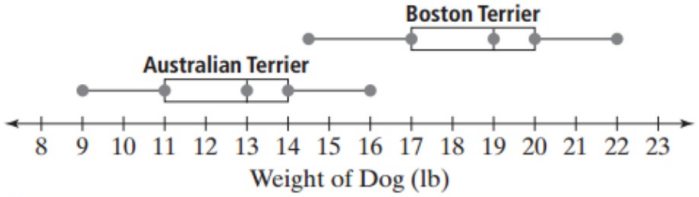
18 − 15 = 3



The IQRs of the data sets are the same. So, you can infer that the heights of Salvia vary about as much as the heights of Marigolds.

Quick Check

1. A veterinarian collects data about the weights of the dogs she treats. Compare the medians of the data sets 1, and use the comparison to make an inference about the plants.



A second measure of variability is **mean absolute deviation** (MAD). The MAD of a data set is the average distance between the mean and each data value.

You can determine the amount of overlap of two data sets by expressing the difference between their centers as a multiple of the MAD or IQR. A multiple less than 1 indicates a large amount of overlap. A multiple greater than 1 indicates a small or no amount of overlap.

Points Scored		
Eagles		Vikings
	0	✗✗
	1	✗
	2	✗✗
	3	✗
✗	4	✗✗
	5	
✗✗	6	
✗✗	7	
✗	8	
✗✗	9	

EXAMPLE 2 Determining Overlap of Data Sets

Sports The data table at the left shows the number of points scored by each player on two basketball teams during a game. Use measures of center and variability to express the amount of overlap between the data sets.

a. Calculate the mean of each data set.

Eagles: Mean $= \frac{56}{8} = 7$ **Vikings:** Mean $= \frac{16}{8} = 2$

b. Determine the MAD of each data set.

Eagles: Start by finding the MAD for points scored by the Eagles.

Points	4	6	6	7	7	8	9	9
Mean	7	7	7	7	7	7	7	7
Distance	3	1	1	0	0	1	2	2

$\text{MAD} = \frac{\text{total of the distances}}{\text{number of data values}} = \frac{10}{8} = \frac{5}{4} = 1.25$

Vikings: Follow the same steps. $\text{MAD} = \frac{10}{8} = \frac{5}{4} = 1.25$

The data sets have the same MAD, so their variability is similar.

c. What multiple n of the MAD equals the difference between the means?

$\text{MAD} \cdot n = \text{difference of means}$ ← **Write an equation.**

$1.25n = 7 - 2$ ← **Substitute.**

$1.25n = 5$ ← **Simplify.**

$\frac{1.25n}{1.25} = \frac{5}{1.25}$ ← **Divide each side by 1.25.**

$n = 4$ ← **Simplify.**

d. What does this number tell you about the overlap of the data sets?

The difference between the means is 4 times the MAD. So, the distance between the centers of the data sets is greater than the variability of either data set. There is little overlap in the data sets, which the data table confirms.

Example

2 Determining Overlap of Data Sets The line plot at the right shows the ages of teens who participated in a junior archery competition.

a. Calculate the mean of each data set.

Girls: Mean $= \frac{150}{10} = 15$ **Boys:** Mean $= \frac{\boxed{170}}{\boxed{10}} = \boxed{17}$

Participant Ages		
Girls		Boys
✗✗	13	✗
✗✗	14	
✗✗✗	15	
✗	16	✗✗
✗	17	✗✗✗
✗	18	✗✗
	19	✗✗

b. Determine the MAD for ages of girls in the competition.

Ages	13	13	14	14	15	15	15	16	17	18
Mean	15	**15**	**15**	**15**	**15**	**15**	15	**15**	**15**	**15**
Distance	2	2	1	1	0	0	0	1	2	3

$\text{MAD} = \frac{\text{total distance}}{\text{number of data items}} = \frac{\boxed{12}}{10} = \frac{\boxed{6}}{5} = \boxed{1.2}$

c. What number n multiplied by the MAD equals the difference between the means? What does this number tell you about the overlap of the data sets?

$\text{MAD} \cdot n = \text{difference of means}$ ← **Write an equation.**

$1.2n = 17 - \boxed{15}$ ← **Substitute.**

$1.2n = \boxed{2}$ ← **Simplify.**

$\frac{1.2n}{\boxed{1.2}} = \frac{2}{\boxed{1.2}}$ ← **Divide each side by 1.2.**

$n = \boxed{1.67}$ ← **Simplify.**

The difference between the means is $\boxed{1.67}$ times the MAD.
The multiple $\boxed{1.67}$ is greater than 1, which indicates a $\boxed{\text{small}}$ amount of overlap in the data sets.

Quick Check

2. The data table at the right shows the number of commercials during a random sample of hour-long shows on two television stations.

a. Calculate the mean of each data set.
Station A: 36 Station B: 35

b. Determine the MAD for Station A commercials. 1.2

c. What number n multiplied by the MAD equals the difference between the means? $n \approx$ 0.83.

What does this number tell you about the overlap of the data sets?

Sample answer: The difference between the means is 0.83 times the MAD.

Since 0.83 is less than 1, there is a large amount of overlap in the data sets.

Commercials During Hour-Long Shows		
Station A		Station B
	32	✗
✗	33	
	34	✗ ✗ ✗
✗ ✗ ✗	35	✗ ✗
✗ ✗ ✗	36	✗ ✗
✗	37	✗ ✗
✗	38	
✗	39	

We will now do a lab together. The problems on this lab are definitely like problems on Thursday's test.

You will need your Desmos calculator app open, and you will want to refer to your Evernote notes.

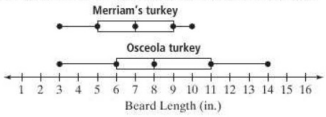
Thursday's test is open Evernote and open calculator.

Name _____ Class _____ Date _____

Practice 8-4

Data Variability

1. A biologist collects data about beard length on wild turkeys. Compare the IQRs of the data sets, and use the comparison to make an inference.



IQR for Merriam's wild turkey: ☐ IQR for Osceola wild turkey: ☐

What can you infer? Explain your reasoning.

_____

_____

_____

The line plot at the right shows the number of homeruns hit by players in a homerun derby.

Homeruns		
Teens		Adults
✗✗	1	✗
✗✗	2	
✗	3	✗✗
✗	4	✗✗✗
✗✗	5	
	6	
✗	7	✗✗
	8	✗
	9	✗
✗	10	

2. Calculate the mean of each data set.

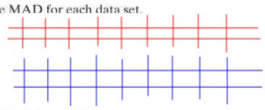
Teens: ☐

Adults: ☐

3. Determine the MAD for each data set.

Teens: ☐

Adults: ☐



4. What number n multiplied by the MAD equals the difference between the means?

MAD: ☐