

MEASURES OF SPREAD IN DATA STANDARD DEVIATION

What can you infer, justify and conclude about John and Tara's tests scores (seen below)?
 (Hint: Calculate the mean, median and mode for each. What do they tell you?)

John's Tests: 76, 45, 83, 68, 64

Tara's Tests: 67, 70, 70, 62, 62

John's Mean = 67.2

Tara's Mean = 66.2

45, 64, 68, 76, 83
 Median = 68
 Mode = NONE

Median = 67
 Mode = 62, 70

These results tell us:

- John has a higher average BUT
- Tara's marks are more consistent

MEASURES OF SPREAD

MEAN, MEDIAN & MODE are all good ways to find the centre of your data.

This information is most useful when the sets of data being compared are similar.

It is also important to find out how much your data is spread out. This gives a lot more insight to data sets that vary from each other.

Example 1

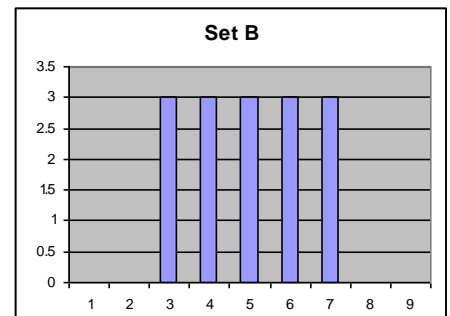
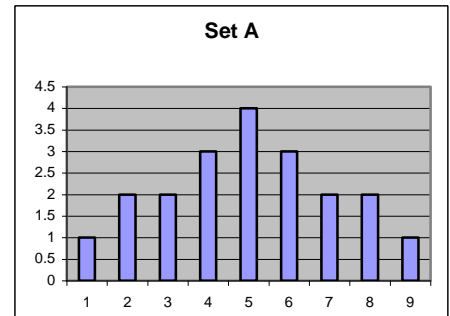
Consider the following two data sets with identical mean and median values.
 Why is this information misleading?

Set A: 0, 2, 2, 4, 4, 6, 6, 6, 8, 8, 8, 8, 10, 10, 10, 12, 12, 14, 14, 16

Mean = 8 Median = 8

Set B: 4, 4, 4, 6, 6, 6, 8, 8, 8, 10, 10, 10, 12, 12, 12

Mean = 8 Median = 8



Because, when graphed one is BELL shaped, the other is uniform. Calculations make them APPEAR similar. In graph, they SPREAD OUT quite differently

What is something that can be done to further compare these graphs?
 LOOK AT THE **RANGE** IN THE DATA SETS

Range: is the difference between the highest and lowest numbers.

A Range = $\frac{16-0}{16}$ B Range = $\frac{12-4}{8}$

SET B is more consistent since it has a smaller range.

Example 2

Twins, Toby and Moby, both work at a local pizza shop. Their manager has decided to give a raise to her best employee. She looks at their data.

Number of Pizzas Made per Shift								
Toby	54	152	180	12	72	126	104	132
Moby	132	104	102	120	86	12	180	96

Who is more deserving of the raise?

Solution: She starts by finding the mean number of pizzas made by each employee and their range.

TOBY
 MEAN (μ) $\Rightarrow \frac{832}{8}$
 $= 104$

MOBY
 MEAN $= \frac{832}{8}$
 $= 104$

RANGE $= 180 - 12$
 $= 168$

RANGE $= 180 - 12$
 $= 168$

These statistics leave both employees equal.

Who do you think is more consistent? _____

She decides to calculate the standard deviation for each.

Standard Deviation (σ) - BEST choice for measuring spread of DATA

Steps for calculating Standard Deviation (σ):

1. Find the DIFFERENCE between each value and the mean (average)
2. Square each difference
3. Add up all your answers from step 2
4. Divide this sum by the number of values in the data set
5. Take the square root to find your answer

Mathematical Formula:
$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

where σ = standard deviation
 \bar{x} = mean (μ - also a symbol uses for mean)
 n = number of entries
 Σ = sum

Standard deviation for Toby:

Number of Pizzas x	$x - \bar{x}$ ↑ mean	$(x - \bar{x})^2$
54	$54 - 140 = -86$	7396
152	$152 - 140 = 12$	$12^2 = 144$
180	$180 - 140 = 40$	$40^2 = 1600$
12	$12 - 140 = -128$	16384
72	$72 - 140 = -68$	4624
126	$126 - 140 = -14$	196
104	$104 - 140 = -36$	1296
132	$132 - 140 = -8$	64
	Total=	31704

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

$$\begin{aligned} \text{s.d} = \sigma &= \sqrt{\frac{31704}{8}} \\ &= \sqrt{3963} \\ &= 62.95 \end{aligned}$$

In order for this standard deviation to be significant, you must compare it to another data set.

Standard deviation for Moby

mean = 140

Number of Pizzas x	$x - \bar{x}$	$(x - \bar{x})^2$
132	$132 - 140 = -8$	$(-8)^2 = 64$
104	$104 - 140 = -36$	$(-36)^2 = 1296$
102	$102 - 140 = -38$	$(-38)^2 = 1444$
120	$120 - 140 = -20$	$(-20)^2 = 400$
86	$86 - 140 = -54$	$(-54)^2 = 2916$
12	$12 - 140 = -128$	$(-128)^2 = 16384$
180	$180 - 140 = 40$	1600
96	$96 - 140 = -44$	1936
	Total=	26040

$$\begin{aligned} \sigma &= \sqrt{\frac{26040}{8}} \\ &= \sqrt{3255} \\ &= 57.05 \end{aligned}$$

Since the s.d. for Moby is less, his data (PIZZA PRODUCTION) is less spread out - he is more consistent.

∴ He is more deserving of the raise.

MEASURES OF SPREAD – PRACTICE
(please answer on a separate sheet of paper)

1. True or False? The standard deviation cannot be a negative.
2. Calculate the range, variance and the standard deviation of the following data:
4, 8, 6, 3, 12, 9, 7, 6
3. The machine packaging cookies has been considered defective. The packages are labelled as containing 150g. A sample of 15 packages was selected and the masses are given.
145, 151, 152, 150, 147, 152, 149, 148, 153, 150, 146, 152, 148, 149, 151
 - a) Calculate the mean.
 - b) If any packages are deviate than 2.2g from the mean, it is defective. How many are defective?
 - c) Should the machine be fixed?
4. A group of student landscapers are to keep track of their own weekly hours. They are as follows: 44, 52, 43, 39, 42, 41, 38, 43, 46, 45, 44, 39, 40, 42, 45
 - a) Find the range.
 - b) Find the mean.
 - c) Find the standard deviation.
 - d) What can be said about the entry of 52 hours/week?
 - e) Calculate the standard deviation again without the 52 hours/week entry.
5. The sale prices of the last 10 homes sold in **1985** were: \$198 000, \$185 000, \$205 200, \$225 300, \$206 700, \$201 850, \$200 000, \$189 000, \$192 100, \$200 400.
 - a) What is the average sale price?
 - b) What is the range of sale prices?
 - c) What is the standard deviation?
 - d) Do you think that a price of \$240 000 would be considered unusual? Why or why not?

Some Solutions

2. a) range = 9; s.d. = 2.85
3. a) 149.5g b) 7
4. a) 14hrs b) 42.9hrs c) 3.50hrs e) 2.52 hrs
5. a) \$200 355.00 b) \$40 300 c) \$11 189.04

