POPULATION
MEAN: μ (mu)
VARIANCE: σ^2 (sigma squared)
STANDARD DEVIATION: σ (sigma)

FST NOTES 1-6
TOPIC: Measures of Spread: Variance and Standard Deviation

GOAL
Introduce two of the most common measures of spread, variance and its square root, standard deviation.

SPUR Objectives

A Calculate measures of center and spread for data sets.
D Describe relations between measures of center and spread.
E Use statistics to draw conclusions about data.

Vocabulary

range MAX − MIN (USES ONLY 2 DATA POINTS)
deviations DIFFERENCE EACH DATA VALUE IS FROM THE MEAN
population variance σ^2 MEAN OF SQUARED DEVIATIONS − DIVIDE BY "n"
population standard deviation σ $\sqrt{\text{POP VARIANCE}}$
sample variance s^2 MEAN OF SQUARED DEVIATIONS − DIVIDE BY "n − 1"
sample standard deviation s $\sqrt{\text{SAMPLE VARIANCE}}$

Warm-Up

Suppose $a_1 = 50$, $a_2 = 70$, $a_3 = 100$, and $\overline{a}$ is the mean of a_1, a_2, and a_3.
Calculate:

1. $\underbrace{\sum_{i=1}^{3} a_i}_{\text{SUM}}$ $50 + 70 + 100 = \boxed{220}$

2. MEAN $\dfrac{\sum_{i=1}^{3} a_i}{3} - \overline{a}$ MEAN $\dfrac{220}{3} - \dfrac{220}{3} = \boxed{0}$ $73.\overline{3}$

3. SAMPLE VARIANCE $\dfrac{\overbrace{\sum_{i=1}^{3}(a_i - \overline{a})^2}^{\text{SUM OF SQUARED DEVIATIONS}}}{2 \ (n-1)}$ $(50 - 73.\overline{3})^2 + (70 - 73.\overline{3})^2 + (100 - 73.\overline{3})^2 = \dfrac{1266.\overline{6}}{2} = \boxed{633.\overline{3}}$

4. SAMPLE S.D. $\sqrt{\dfrac{\sum_{i=1}^{3}(a_i - \overline{a})^2}{2 \ (n-1)}}$ $\sqrt{633.\overline{3}} = \boxed{25.2}$

Three measures of center are: MEAN, MEDIAN, and MODE

Four measures of spread are: RANGE, IQR, VARIANCE, and STANDARD DEVIATION.

VARIANCE and STANDARD DEVIATION are measures of spread based on the **mean**.

IQR is a measure of spread related to the **median**.

RANGE is the simplest measure of the spread of distribution. It is the difference of the maximum and minimum values.

☆ VARIABILITY, SPREAD + DISPERSION ARE SYNONYMS. THEY LOOK AT HOW "SPREAD OUT" DATA IS.

SEC 1-4
p.2

Definition of Variance and Standard Deviation of a Population

Let μ be the mean of the population data set $x_1, x_2, \ldots, x_n$. Then the variance σ^2 and **standard deviation σ of the population are**

$$\sigma^2 = \frac{\text{sum of squared deviations}}{n} = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

and

$$\sigma = \sqrt{\text{variance}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}.$$

STANDARD DEVIATION IN CALC. = σx

Definition of Variance and Standard Deviation of a Sample

Let $\bar{x}$ be the mean of the sample data set $x_1, x_2, \ldots x_n$.

Then the **variance** s^2 and **standard deviation** s **of the sample are**

$$s^2 = \frac{\text{sum of squared deviations}}{n-1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

and

$$s = \sqrt{\text{variance}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}.$$

STANDARD DEVIATION IN CALC. = Sx

FST 1-6

The number of miles in **thousands**, obtained in five tests of two different tires is listed in the table below.

* SAMPLES *

Tire A	66	43	37	50	54	mean = 50
Tire B	54	49	47	48	52	mean = 50

A

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
66	66 - 50 = 16	$16^2 = 256$
43	43 - 50 = -7	$(-7)^2 = 49$
37	37 - 50 = -13	$(-13)^2 = 169$
50	50 - 50 = 0	$(0)^0 = 0$
54	54 - 50 = 4	$4^2 = 16$
Total 250	SUM OF DEVIATIONS 0	490

* ALWAYS "0" *

Variance = $\frac{490}{5-1} = 122.5$

Standard Deviation = $\sqrt{122.5}$

$= 11.068$

$= 11,068$ miles

B

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
54	54 - 50 = 4	$4^2 = 16$
49	49 - 50 = -1	$(-1)^2 = 1$
47	47 - 50 = -3	$(-3)^2 = 9$
48	48 - 50 = -2	$(-2)^2 = 4$
52	52 - 50 = 2	$(2)^2 = 4$
Total 250	SUM OF DEVIATIONS 0	34

* ALWAYS "0" *

Variance = $\frac{34}{5-1} = 8.5$

Standard Deviation = $\sqrt{8.5}$

$= 2.92$

$= 2920$ miles

Tire A has a greater variability than Tire B because it has a larger Standard deviation. Tire B is more consistent.

SEC 1-6
p. 4

1-6 Measures of Spread: Variance and Standard Deviation

Activity 1: Given the two data sets below, fill in the table with each statistic.

Data Set 1: 70, 71, 71, 73, 74, 75, 78, 79, 79, 80

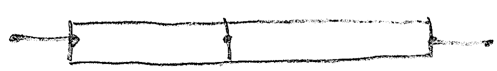
Data Set 2: 70, 74, 75, 75, 75, 75, 75, 75, 76, 80

Statistic	Data Set 1	Data Set 2
Number of elements, n	10	10
Minimum	70	70
Q1	71	75
Median	74.5	75
Q3	79	75
Maximum	80	80
Range Max - Min	10	10
IQR $Q_3 - Q_1$	8	0
Mean	75	75

Review: Make a box plot of each data set, using the same number line. (There are no outliers)

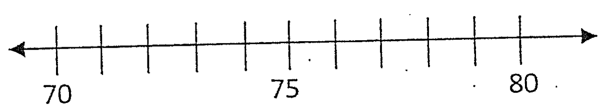
min: 70 Q_1: 71 Q_2: 74.5 Q_3: 79 Max: 80

Data Set 1



min: 70 Q_1: 75 Q_2: 75 Q_3: 75 Max: 80

Data Set 2



Compare the two values for range. Using range alone, are you able to determine which set is more spread out than the other?

No, both have a range of 10.

What about using the IQR?

Yes, Data 1 IQR of 8, Data 2 IQR of 0.
Data 1 more spread than Data 2.

1-6 Measures of Spread: Variance and Standard Deviation

Variance & Standard Deviation – their calculations depend on whether the data set is from a population or a sample.

Symbols:

	Mean	Variance	Standard Deviation
Population	μ (mu)	σ^2	σ (sigma)
Sample	$\bar{x}$ (x-bar)	s^2	s

Equations:

	Variance	Standard Deviation
Population	$\sigma^2 = \dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$	$\sigma = \sqrt{\sigma^2}$
Sample	$s^2 = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$	$s = \sqrt{s^2}$

Activity 2: Calculate the population and $\boxed{\text{sample}}$ standard deviation and variance of the data sets from Activity 1.

Data Set 1: 70, 71, 71, 73, 74, 75, 78, 79, 79, 80 $\quad \bar{x} = 750/10 = 75$

Data Set 2: 70, 74, 75, 75, 75, 75, 75, 75, 76, 80 $\quad \bar{x} = 750/10 = 75$

Set 1

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
70	-5	25
71	-4	16
71	-4	16
73	-2	4
74	-1	1
75	0	0
78	3	9
79	4	16
79	4	16
80	5	25
		128

$VAR = \dfrac{128}{10-1}$

$\boxed{VAR = 14.22}$

$SD = \sqrt{14.22}$

$\boxed{SD = 3.77}$

Set 2

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
70	-5	25
74	-1	1
75	0	0
75	0	0
75	0	0
75	0	0
75	0	0
75	0	0
76	1	1
80	5	25
		52

$VAR = \dfrac{52}{10-1}$

$\boxed{VAR = 5.78}$

$SD = \sqrt{5.78}$

$\boxed{SD = 2.40}$

☆ Set 1 has a greater variability than Set 2 due to higher S.D.

☆ Set 2 more consistent.