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
deviations
population variance σ^2
population standard deviation σ
sample variance s^2
sample standard deviation s

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

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

3. $\frac{\sum_{i=1}^{3}(a_i - \overline{a})^2}{2}$ $(50-73.3)^2 + (70-73.3)^2 + (100-73.3)^2$ $\frac{1266.67}{2}$ $\boxed{633.3}$

4. $\sqrt{\frac{\sum_{i=1}^{3}(a_i - \overline{a})^2}{2}}$ $\sqrt{633.3} = \boxed{25.2}$

Three measures of center are: mean, median, and mode

Four measures of spread are: range, IQR, variance, and Standard deviation

affected by extreme values

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" a group of scores are.

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

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

$(a_i - \bar{a})^2$

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

calculator $= \sigma x$

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

calculator $= Sx$

Khan Academy

Sample

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

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

FST 1-6

The number of miles in **thousands** (,000), obtained in five tests of two different tires is listed in the table below.

$\bar{x}$ = mean

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

$\bar{x} = 50$

$\bar{x} = 50$

50,000

Tire 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	$(0)^2 = 0$
54	54 − 50 = 4	$(4)^2 = 16$
Total Sum	~~X~~	490

B

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
54	4	16
49	−1	1
47	−3	9
48	−2	4
52	2	4
Total	~~X~~	34

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

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

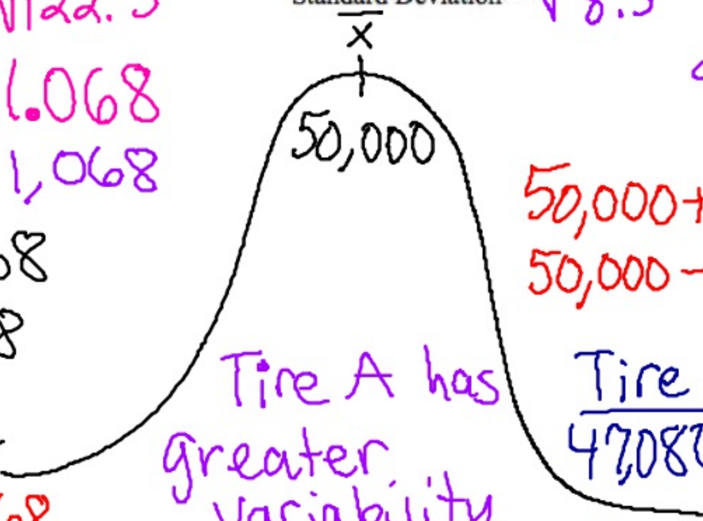
Standard Deviation = $\sqrt{122.5}$

11.068

11,068

Standard Deviation = $\sqrt{8.5} = 2.92$

2,920



50,000 + 11,068

50,000 − 11,068

Tire Range A

38,932 - 61,068

50,000 + 2920

50,000 − 2920

Tire Range B

47,080 − 52,920

Tire A has greater variability because the Standard deviation is larger than Tire B

Tire B is more consistent

STAT #1 enter data in L1   STAT → Calc #1

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

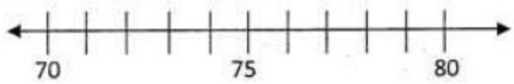
used to make box plot

Statistic	Data Set 1	Data Set 2
Number of elements, n		
Minimum		
Q1		
Median		
Q3		
Maximum		
Range Max − Min		
IQR Q3 − Q1		
Mean $\bar{x}$		

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

Data Set 1

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?

What about using the IQR?

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			
Sample			

Equations:	Variance	Standard Deviation
Population		
Sample		

Activity 2: Calculate the ~~population and~~ 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 $\bar{x} = 75$

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

Data Set 1

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$

Variance =

Standard deviation =

Data Set 2

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$

Variance =

Standard deviation =

STAT #1 enter data in L1 STAT → Calc #1

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

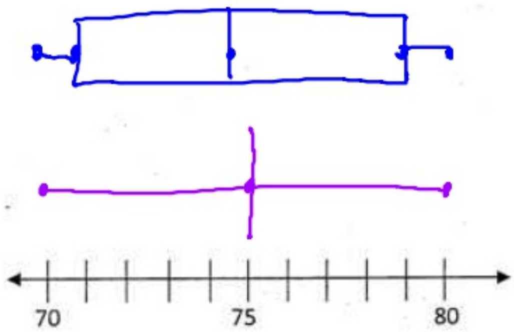
Data Set 2: 70, 74, 75, 75, 75, 75, 75, 75, 76, 80

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 Q3-Q1	8	0
Mean $\bar{x}$	75	75

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

Data Set 1

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 the same range

What about using the IQR?

Yes Data Set 1 has an IQR of 8 and Data Set 2 has an IQR of 0. Data Set 1 is more spread out than Data Set 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			
Sample			

Equations:

	Variance	Standard Deviation
Population		
Sample		

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 $\bar{x} = 75$

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

Data 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

Variance $= \frac{128}{9} = 14.22$

Standard deviation $= \sqrt{14.22}$
$= 3.77$

Data 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

Variance $= \frac{52}{9} = 5.78$

Standard deviation $= \sqrt{5.78}$
$= 2.40$