Sunday, October 08, 2017
5:04 PM

Statistics Honors Chapter 1 Standard Deviation Day 1 Name KEY

DO NOW: Find the ***sample size*** and the ***mean*** for the given data.

Barry Bond's homerun counts for years 1986-2001:

16, 19, 24, 25, 25, 33, 33, 34, 34, 37, 37, 40 ,42, 46, 49, 73

n = 16 $\bar{x}$ = 35.4375 $\bar{x} = \frac{\sum x_i}{n}$

* sample size

CLASS NOTES:

The ***variance*** for a sample of data, called s^2, is a measure of spread.

$$s^2 = \frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2$$

n = # of observations
x_i= each piece of data
$\bar{x}$ = mean, $\sum$ = sum

- The ***variance*** is large if the observations are widely spread about their mean.
- The ***variance*** is small if the observations are all close to the mean.

The ***standard deviation***, called s, also measures spread by looking at how far the observations are from their mean. The standard deviation, s, is simply the square root of the variance, s^2.

$$s = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2}$$

- A deviation tells us how far, and in what direction, an observation is from the mean. The sum of the ***deviations*** will *always* be zero because some observations are above the mean (equals a positive deviation) and some observations are below (equals a negative deviation). When you add them up, the sum will be **zero**.

Variance and Standard Deviation are both measures of SPREAD.

Both tell us how far the data is spread from the mean.

Properties of Standard Deviation

- s measures spread about the mean and should *only* be used when the ***mean*** is chosen as the ***measure of center***.
- $s = 0$ only when there is no difference in ***spread*** => all data values are the same value = equal distance from the ***mean***. If data values are different and become more ***spread*** out about the ***mean***, s gets larger.
- Standard deviation, s, and the mean, $\bar{x}$, are impacted by outliers and skewedness. Both are **not** resistant. Therefore, when choosing a measure of center and spread, we only choose the ***mean*** and ***standard deviation*** when we have data that is Symmetric and no ***outliers***.

You will be required to know how to find the standard deviation by using a step-by-step methodology. We will use a table to help us in this process.

REVISITING THE DO NOW EXAMPLE1:

Barry Bond's homerun counts for years 1986-2001:

16, 19, 24, 25, 25, 33, 33, 34, 34, 37, 37, 40 ,42, 46, 49, 73

n = 16 $\bar{x}$ = 35.4375

L1 Observation x_i	$L2 = L1 - 35.4375$ Deviations $x_i - \bar{x}$	$L3 = L2 \wedge 2$ Squared Deviations $(x_i - \bar{x})^2$
16	16−35.4375 = −19.4375	377.8164
19	−16.4375	270.1914
24	−11.4375	130.8164
25	−10.4375	108.9414
25	−10.4375	108.9414
33	−2.4375	5.9414
33	−2.4375	5.9414
34	−1.4375	2.0664
34	−1.4375	2.0664
37	1.5625	2.4414
37	1.5625	2.4414
40	4.5625	20.8164
42	6.5625	43.0664
46	10.5625	111.5664
49	13.5625	183.9414
73	37.5625	1410.9414

on calc! 2nd stat math sum

Sum of deviations = 0 Sum of squared deviations = 2787.9375

* variance

$$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2 = \frac{1}{15}(2787.9375) = \boxed{185.8625}$$

* std dev.

$$s = \sqrt{s^2} = \sqrt{185.8625} = \boxed{13.6331}$$

Example2: Now we will remove the ***outlier***, 73, from the data from Example1 and recalculate the standard deviation s.

Hint: You must recalculate $\bar{x}$ and the new n to find the new sum of the squared deviations.

n = 15 $\bar{x}$ = 32.9333 $\bar{x} = \frac{\sum x_i}{n} = \frac{494}{15}$

Observation L1	Deviations L2 = L1 − 32.9333	Squared Deviations L3 = L2^2
X_i	$X_i - \overline{X}$	$\left(X_i - \overline{X}\right)^2$
16	−16.9333	286.7366
19	−13.9333	194.1368
24	−8.9333	79.8038
25	−7.9333	62.9372
25	−7.9333	62.9372
33	.0667	.0045
33	.0667	.0045
34	1.0667	1.1378
34	1.0667	1.1378
37	4.0667	16.5380
37	4.0667	16.5380
40	7.0667	49.9382
42	9.0667	82.2050
46	13.0667	170.7386
49	16.0667	258.1388

2nd STAT CALC MATH SUM

Sum of deviations = 0 Sum of squared deviations = 1282.9333

variance:

$$s^2 = \frac{1}{n-1}\sum\left(x_i - \overline{x}\right)^2 = \frac{1}{14}(1282.9333) = \boxed{91.6381}$$

Std. dev.: $s = \sqrt{91.6381} = \boxed{9.5728}$

What happens to the spread when the *outlier* is removed ?

The spread got smaller, so the std. deviation and the variance got smaller.

Important Information when describing data's distribution: *SOCS*!

The ***median and IQR*** should be used together and are the best measure of center and spread when dealing with data that **have *outliers* and/or are *strongly skewed***.
The ***mean and standard deviation*** go should be used together and are the best measure of center and spread when dealing with data that is ***symmetrical* and has no *outliers***. *(Recall: the **mean** and **standard deviation** are strongly effected / not resistant to **outliers**.)*

Partner Practice:

A person's metabolic rate is the rate at which the body consumes energy. Metabolic rate is important in studies of weight gain, dieting, and exercise. Below are the metabolic rates of 7 men who took part in a study of dieting. (The units are calories per 24 hours, which are the same calories used to describe the energy content of foods).

1362 1439 1460 1614 1666 1792 1867

What is the mean for this data? $\bar{x} =$ 1600

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{11{,}200}{7}$

Fill in the table below: n = 7

L1 | L2= L1 − 1600 | L3 = L2 ^2

Observations x_i	Deviations $x_i - \bar{x}$	Squared Deviations $(x_i - \bar{x})^2$
1362	−238	56644
1439	−161	25921
1460	−140	19600
1614	14	196
1666	66	4356
1792	192	36864
1867	267	71289

2nd STAT MATH sum L3

Sum of deviations = 0 Sum of squared deviations = 214870

$$s^2 = \frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2 = \frac{1}{6}(214870) = \boxed{35811.6667}$$

Standard Deviation (S) =

$S = \sqrt{35811.6667}$

$\boxed{S = 189.2397}$

SUMMARY: *YOU MUST KNOW THIS*

Measures of Center	Measures of spread	When to use	Advantage	Graph
mean	std. deviation	symmetric data	—	Histogram
median	IQR	skewed data or outliers	Resistant	Box Plot