Variation

Range: The range of a set of data is the difference between the maximum value and the minimum value. The range can be found for both population data and sample data.

$$Range = (\text{Maximum Value}) - (\text{Minimum Value})$$

In this Example: Suppose we want to discuss the adults who moved to Phoenix in 2012 from outside of the state of Arizona to become residents. There were 50 people who moved to Phoenix from outside AZ in 2012. There ages are as follows.

25	20	32	31	45	69	70	76	27	45
46	47	37	38	22	31	33	51	55	46
63	41	80	35	36	36	65	25	25	34
37	59	59	45	44	34	29	27	29	32
72	64	63	36	48	37	31	21	21	22

To find the range in the ages, first you must find the minimum value and the maximum value of the data set. The maximum value is the largest age = 80 and the minimum value is the smallest age = 20.

$$Range = (\text{Maximum Value}) - (\text{Minimum Value})$$
$$Range = (80) - (20)$$
$$Range = 60 \text{ years}$$

Standard Deviation

The standard deviation of a set of sample values is measured of variation of values about the mean. It is a type of average deviation of values from the mean. There are two types of data, population data and sample data. The standard deviation is calculated differently for each.

Sample Standard Deviation

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

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

Example: Suppose we want to discuss the adults who moved to Phoenix in 2011 from outside of the state of Arizona to become residents. We want to find the sample standard deviation of the age of those people. There were 50 adults who moved to Phoenix from outside AZ in 2011. We look at a random sample of 20 adults.

25	37	72	47	72	41	59	32	59	37
80	63	80	31	38	31	45	36	45	22

Either of the formulas above can be used to do this.

Formula 1:

Make a table. Column one is x = the sample data values. Column two is $x - \bar{x}$, the sample data minus the sample mean (found in a previous example to be 47.6 years). Column three is the square of column two.

x	x- $\bar{x}$	$(x - \bar{x})^2$
25	-22.6	510.76
37	-10.6	112.36
72	24.4	595.36
47	-0.6	0.36
72	24.4	595.36
41	-6.6	43.56
59	11.4	129.96
32	-15.6	243.36
59	11.4	129.96
37	-10.6	112.36

80	32.4	1049.76
63	15.4	237.16
80	32.4	1049.76
31	-16.6	275.56
38	-9.6	92.16
31	-16.6	275.56
45	-2.6	6.76
36	-11.6	134.56
45	-2.6	6.76
22	-25.6	655.36

Once you have the table find the sum of column three.

$$\text{sum of } \left(x-\bar{x}\right)^2 = \sum\left(x-\bar{x}\right)^2 = 6256.8$$

Plug this into the formula

$$s = \sqrt{\frac{\sum\left(x-\bar{x}\right)^2}{n-1}} = \sqrt{\frac{6256.8}{20-1}} = \sqrt{\frac{6256.8}{19}}$$

$$s = \sqrt{329.147} = 18.14$$

The sample standard deviation is 18.14 years

Formula 2:

To use the second formula also make a table. Column one is the data values but the second column is the square of the data values.

X	x^2
25	625
37	1369
72	5184
47	2209
72	5184
41	1681
59	3481
32	1024
59	3481
37	1369
80	6400
63	3969

80	6400
31	961
38	1444
31	961
45	2025
36	1296
45	2025
22	484

Find the sum of both columns

$$\text{sum of data values} = \sum x = 952$$

$$\text{sum of data values squared} = \sum x^2 = 51572$$

Plug these values into the formula

$$s = \sqrt{\frac{n\left(\sum x^2\right) - \left(\sum x\right)^2}{n(n-1)}} = \sqrt{\frac{20(51572) - (952)^2}{20(20-1)}}$$

$$s = \sqrt{\frac{1031440 - 906304}{20(19)}} = \sqrt{\frac{125136}{380}} = \sqrt{329.305}$$

$$s = 18.14$$

Notice the sample standard deviation is 18.14 years which is the same as it was using the first formula.

POPULATION STANDARD DEVIATION

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$$

Example: Suppose we want to discuss the adults who moved to Phoenix in 2012 from outside of the state of Arizona to become residents. We want to find the standard deviation of the age of those people. There were 50 adults who moved to Phoenix from outside AZ in 2012.

25	20	32	31	45	69	70	76	27	45
46	47	37	38	22	31	33	51	55	46
63	41	80	35	36	36	65	25	25	34
37	59	59	45	44	34	29	27	29	32
72	64	63	36	48	37	31	21	21	22

To solve for population standard deviation make a table. The first column is the data values. Column two is $x-\mu$, the sample data minus the population mean (found in a Measure of Center example video to be 41.92 years. Column three is the square of column two.

x	$(x-\mu)$	$(x-\mu)^2$
25	-16.92	286.2864
46	4.08	16.6464
63	21.08	444.3664
37	-4.92	24.2064
72	30.08	904.8064
20	-21.92	480.4864
47	5.08	25.8064
41	-0.92	0.8464
59	17.08	291.7264
64	22.08	487.5264
32	-9.92	98.4064
37	-4.92	24.2064
80	38.08	1450.086
59	17.08	291.7264
63	21.08	444.3664
31	-10.92	119.2464
38	-3.92	15.3664
35	-6.92	47.8864

45	3.08	9.4864
36	-5.92	35.0464
45	3.08	9.4864
22	-19.92	396.8064
36	-5.92	35.0464
44	2.08	4.3264
48	6.08	36.9664
69	27.08	733.3264
31	-10.92	119.2464
36	-5.92	35.0464
34	-7.92	62.7264
37	-4.92	24.2064
70	28.08	788.4864
33	-8.92	79.5664
65	23.08	532.6864
29	-12.92	166.9264
31	-10.92	119.2464
76	34.08	1161.446
51	9.08	82.4464
25	-16.92	286.2864
27	-14.92	222.6064
21	-20.92	437.6464
27	-14.92	222.6064
55	13.08	171.0864
25	-16.92	286.2864
29	-12.92	166.9264
21	-20.92	437.6464
45	3.08	9.4864
46	4.08	16.6464
34	-7.92	62.7264
32	-9.92	98.4064
22	-19.92	396.8064

Then find the sum of the third column and plug it into the formula.

$$\text{sum of the third column} = \sum (x-\mu)^2 = 12701.68$$

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}} = \sqrt{\frac{12701.68}{50}} = \sqrt{254.034}$$

$$\sigma = 15.9384$$

The standard deviation is 15.9384 years.

Variance

The variance of a set of values is a measure of variation equal to the square of the standard deviation.

Sample Variance

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

Population Variance

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

To calculate the variance(sample and population), follow the steps to find the standard deviation and then square the answer.

For Sample: Suppose we want to discuss the adults who moved to Phoenix in 2012 from outside of the state of Arizona to become residents. We want to find the sample variance of the age of those people. There were 50 adults who moved to Phoenix from outside AZ in 2012. We look at a random sample of 20 adults.

25	37	72	47	72	41	59	32	59	37
80	63	80	31	38	31	45	36	45	22

$$s = 18.14 \text{ years}$$
$$s^2 = 18.14^2 = 338.928$$

For Population: Using the same example of adults who moved to Phoenix in 2012 from outside of the state of Arizona to become residents. We want to find the variance of the age of those people. Instead of taking a sample of the population we will use the full population of 50 people.

25	20	32	31	45	69	70	76	27	45
46	47	37	38	22	31	33	51	55	46
63	41	80	35	36	36	65	25	25	34
37	59	59	45	44	34	29	27	29	32
72	64	63	36	48	37	31	21	21	22

$$\sigma = 15.9384 \text{ years}$$
$$\sigma^2 = 254.033$$

Coefficient of Variation

Standard deviation has the same units as the data. This means that comparing standard deviation between data sets can be complicated. If the data are in different units then a direct comparison is not possible. The coefficient of variation for a set of non-negative sample or population data, expressed as a percent, describes the standard deviation relative to the mean.

Sample

$$CV = \frac{s}{\overline{x}} \cdot 100\%$$

Population

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Example: Suppose we want to compare the adults who moved to Phoenix in 2012 from outside of the state of Arizona to become residents to the adults who moved to LA in 2012 from outside the state of California to become residents. We want to compare the standard deviation of the ages between the two populations but in California they gather the data in months rather than years so a direct comparison is not possible.

Phoenix Ages

25	20	32	31	45	69	70	76	27	45
46	47	37	38	22	31	33	51	55	46
63	41	80	35	36	36	65	25	25	34
37	59	59	45	44	34	29	27	29	32
72	64	63	36	48	37	31	21	21	22

LA Ages

264	216	228	288	300	300	780	540	384	804
864	252	252	804	888	456	528	504	288	324
360	372	360	696	492	276	336	348	396	504
708	228	240	252	252	288	264	276	336	456
432	444	408							

Notice that there are 50 adults who moved to Phoenix in 2012 from outside of the state of Arizona to become residents and 43 adults who moved to LA in 2012 from outside

the state of California to become residents. This means $N = 50$ for the Arizona data and $N = 43$ for the California data. The population size does not have to be the same for the two data sets in order to compare the coefficient of variations.

We found the population standard deviation for the Arizona data previously. The population standard deviation is $\sigma = 15.9384$ years. The same method is used to find the population standard deviation for the California data. Make a table …

x	$x-\mu$	$(x-\mu)^2$
264	-154.3	23808.49
216	-202.3	40925.29
228	-190.3	36214.09
288	-130.3	16978.09
300	-118.3	13994.89
300	-118.3	13994.89
780	361.7	130826.9
540	121.7	14810.89
384	-34.3	1176.49
804	385.7	148764.5
864	445.7	198648.5
252	-166.3	27655.69
252	-166.3	27655.69
804	385.7	148764.5
888	469.7	220618.1
456	37.7	1421.29
528	109.7	12034.09
504	85.7	7344.49
288	-130.3	16978.09
324	-94.3	8892.49
360	-58.3	3398.89
372	-46.3	2143.69
360	-58.3	3398.89
696	277.7	77117.29
492	73.7	5431.69
276	-142.3	20249.29
336	-82.3	6773.29
348	-70.3	4942.09
396	-22.3	497.29
504	85.7	7344.49
708	289.7	83926.09
228	-190.3	36214.09

	240	-178.3	31790.89
	252	-166.3	27655.69
	252	-166.3	27655.69
	288	-130.3	16978.09
	264	-154.3	23808.49
	276	-142.3	20249.29
	336	-82.3	6773.29
	456	37.7	1421.29
	432	13.7	187.69
	444	25.7	660.49
	408	-10.3	106.09
sum	17988	1.1	1520231
Mean	418.3256		

$$\text{sum of the third column} = \sum (x-\mu)^2 = 1520231$$

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}} = \sqrt{\frac{1520231}{43}} = \sqrt{35354.2093}$$

$$\sigma = 188.0271504$$

The population standard deviation for the California is 188.0271504 months.

To compare the standard deviation of the ages of the adults who moved to Phoenix with the standard deviation of the ages of the adults who moved to LA find the coefficient of variation for each.

Arizona Data

$$\textit{Population}$$

$$CV = \frac{\sigma}{\mu} \cdot 100\% = \frac{15.9384}{41.92} \cdot 100\%$$

$$CV = (.000388438) \cdot 100\% = .0388438292\%$$

California Data

$$\textit{Population}$$

$$CV = \frac{\sigma}{\mu} \cdot 100\% = \frac{188.0271504}{418.3256} \cdot 100\%$$

$$CV = (.4494756) \cdot 100\% = 44.94756\%$$

Which population appears to have more variability?

The ages of the adults who moved to LA has a much larger coefficient of variation than the ages to the adults who moved to Phoenix.