Measures of Variability

The **standard deviation** of a data set is simply the

square root of the variance

Calculate the **Sum**, **Mean** and **SD** for these two sets of Ages

Set 1

Ages
66
50
57
68
55
70
70
68
55
59
58
65
66
67
68
69
72
73
88
87

Set 2

Ages
12
13
33
18
7
9
14
18
10
12
7
13
14
19
4
16
20
6
19
11

What can you say about the spread of these two sets?

	Ages		Ages
	66		12
	50		13
	57		33
	68		18
	55		7
	70		9
	70		14
	68		18
	55		10
	59		12
	58		7
	65		13
	66		14
	67		19
	68		4
	69		16
	72		20
	73		6
	88		19
	87		11
Sum	1331	Sum	275
Mean	66.55	Mean	13.75
SD	9.36	SD	6.34

Because the means are so different, difficult to compare the sets - use **Coefficient of Variation**

Measures of Variability

Coefficient of Variation

If we want to compare the variability of two data sets where the mean values are significantly different then the coefficient of variation is a good measure.

It is simply the size of the standard deviation relative to the mean, presented as a percentage. That is,

divide the standard deviation by the mean and multiply by 100

Calculate the Coefficient of Variation for these two

	Ages		Ages
	66		12
	50		13
	57		33
	68		18
	55		7
	70		9
	70		14
	68		18
	55		10
	59		12
	58		7
	65		13
	66		14
	67		19
	68		4
	69		16
	72		20
	73		6
	88		19
	87		11
Sum	1331	Sum	275
Mean	66.55	Mean	13.75
SD	9.36	SD	6.34

Coefficient of Variation for these two

	Ages		Ages
	66		12
	50		13
	57		33
	68		18
	55		7
	70		9
	70		14
	68		18
	55		10
	59		12
	58		7
	65		13
	66		14
	67		19
	68		4
	69		16
	72		20
	73		6
	88		19
	87		11
Sum	1331	Sum	275
Mean	66.55	Mean	13.75
SD	9.36	SD	6.34
Coefficient of variability	14%	Coefficient of variability	46%

Note: measured as a %

Using the Standard Deviation

The standard deviation of a data set can be used to determine how far a value is away from the mean.

This distance is referred to as the z-score.

One simply, subtracts the mean from the data value and divides by the standard deviation

This gives us how many standard deviations a value is from the mean.

NOTE: AN OUTLIER IS A VALUE THAT IS AT LEAST 3 STANDARD DEVIATIONS AWAY FROM THE MEAN

Calculate the z-scores

	Ages		Ages
	66		12
	50		13
	57		33
	68		18
	55		7
	70		9
	70		14
	68		18
	55		10
	59		12
	58		7
	65		13
	66		14
	67		19
	68		4
	69		16
	72		20
	73		6
	88		19
	87		11
Sum	1331	Sum	275
Mean	66.55	Mean	13.75
SD	9.36	SD	6.34
Coefficient of variability	14%	Coefficient of variability	46%

Calculate the z-scores

	Ages			Ages	
	66	-0.06		12	-0.28
	50	-1.77		13	-0.12
	57	-1.02		33	3.04
	68	0.15		18	0.67
	55	-1.23		7	-1.06
	70	0.37		9	-0.75
	70	0.37		14	0.04
	68	0.15		18	0.67
	55	-1.23		10	-0.59
	59	-0.81		12	-0.28
	58	-0.91		7	-1.06
	65	-0.17		13	-0.12
	66	-0.06		14	0.04
	67	0.05		19	0.83
	68	0.15		4	-1.54
	69	0.26		16	0.35
	72	0.58		20	0.99
	73	0.69		6	-1.22
	88	2.29		19	0.83
	87	2.19		11	-0.43
Sum	1331		Sum	275	
Mean	66.55		Mean	13.75	
SD	9.36		SD	6.34	
Coefficient of variability	14%		Coefficient of variability	46%	

Spot the outlier?

Calculate the z-scores

	Ages	
	66	-0.06
	50	-1.77
	57	-1.02
	68	0.15
	55	-1.23
	70	0.37
	70	0.37
	68	0.15
	55	-1.23
	59	-0.81
	58	-0.91
	65	-0.17
	66	-0.06
	67	0.05
	68	0.15
	69	0.26
	72	0.58
	73	0.69
	88	2.29
	87	2.19
Sum	1331	
Mean	66.55	
SD	9.36	
Coefficient of variability	14%	

Negative - below the mean

Positive - above the mean

We say "a value is x SD away from the mean"

So ... '87' is 2.19 SDs above the mean

Try this

A set of math test scores has a mean of 70 and a standard deviation of 8.

A set of English test scores has a mean of 74 and a standard deviation of 16.

For which test would a score of 78 be better?

To solve: Find the z-score for each test.

$$\text{math } z\text{-score} = \frac{78-70}{8} = 1$$

$$\text{English } z\text{-score} = \frac{78-74}{16} = .25$$

The math score would have the highest standing since it is 1 standard deviation above the mean while the English score is only .25 standard deviation above the mean.

Try this

A group of data with normal distribution has a mean of 45. If one element of the data is 60, will the z-score be positive or negative?

Try this

A group of data with normal distribution has a mean of 45. If one element of the data is 60, will the z-score be positive or negative?

The z-score must be positive since the element of the data set is above the mean.

Try this

Example: With a height of 70 inches, Roger was the shortest person in a particular club. The club has a mean height of 75.1 inches and a standard deviation of 2.4 inches.

a) What is the difference between Roger's height and the mean?

b) How many standard deviations is the value found in part a?

c) Convert Roger's height into a z score.

d) Is Roger's height usual or not?

Try this

Solution:

a) the difference is calculated as subtracting the smaller value from the larger value = 75.1 - 70 = 5.1

b) the number of standard deviations is calculated as the difference found divided by the standard deviation = 5.1 / 2.4 = 2.13

c) z = (70 - 75.1) / 2.4 = -2.13

d) The z score of -2.13 is unusual.

Chebychev's Theorem

At least $(1-1/k^2)$ of the items of any data set must be within k standard deviations of the mean where $k > 1$

At least 0.75 of the items must be within 2 standard deviations of the mean.

At least 0.89 of the items must be within 3 standard deviations of the mean.