

## Lesson 5 Measures of Dispersion

### Outline

Measures of Dispersion

- Range
- Interquartile Range
- Standard Deviation and Variance

### **Measures of Dispersion**

Measure of central tendency give us good information about the scores in our distribution. However, we can have very different shapes to our distribution, yet have the same central tendency. Measures of dispersion or variability will give us information about the spread of the scores in our distribution. Are the scores clustered close together over a small portion of the scale, or are the scores spread out over a large segment of the scale?

Range – The range is the difference between the high and low score in a distribution. Simply subtract the two numbers to find the range. So, in the distribution: 1, 3, 5, 9, 11 the range is  $11 - 1 = 10$ . Remember to subtract the two numbers to give one number for the final answer.

Interquartile Range (IQR) – The interquartile range is the range of the middle 50% of a distribution. Because any outliers in our distribution must be on the ends of the distribution, the range as a measure of dispersion can be strongly influenced by outliers. One solution to this problem is to eliminate the ends of the distribution and measure the range of scores in the middle. Thus, with the interquartile range we will eliminate the bottom 25% and top 25% of the distribution, and then measure the distance between the extremes of the middle 50% of the distribution that remains.

To actually compute some IQR's we would need to use calculus. Instead, of that possibility we will use a method that will yield a consistent, and somewhat accurate answer. Before we compute the value, let's learn some new definitions. A quartile is a quarter or 25% of the distribution. When we compute the IQR we will want to find each of the quartiles. The first quartile is the same as the 25<sup>th</sup> percentile because 25 percent of the distribution is at or below that point. The second quartile is the same thing as the 50<sup>th</sup> percentile and the median. The third quartile is the same as the 75<sup>th</sup> percentile. The IQR is the found by eliminating the values that lie between the bottom end and the first quartile (bottom 25%). We will also eliminate the values between the third quartile and the top of the distribution. We then subtract the new low and high score of the left over middle part of the distributions. So,  $IQR = \text{Quartile 3} - \text{Quartile 1}$  or  $IQR = 75^{\text{th}} \text{ percentile} - 25^{\text{th}} \text{ percentile}$ .

To compute the IQR first arrange your numbers from lowest to highest and 1) find the median. The median is the 50<sup>th</sup> percentile and second quartile. It's a starting point for us to find the other quartiles. 2) Next find the median of the bottom half of the distribution



To compute standard deviation and variance we first start by finding the deviation about the mean. Recall that we did the same thing when discussing properties of the mean. I'll use the same example with the simple distribution 1, 2, 3, 4, 5.

First we find the mean and the deviations about the mean. What we want to do is add up these deviations and find out how far on average the scores deviate from the mean. The problem we run into is that whenever we add the deviations (in order to find the average of the deviations) they will always sum to zero. How can we get an average if the sum is always zero?

X	X- $\mu$
1	1-3 = -2
2	2-3 = -1
3	3-3 = 0
4	4-3 = 1
5	5-3 = 2
$\mu = 3$	$\Sigma(X - \mu) = 0$

One solution is to square all of the deviations. When we square all the numbers the negative values will all become positive and we can then add the deviations without getting zero.

X	X- $\mu$	$(X - \mu)^2$
1	1-3 = -2	4
2	2-3 = -1	1
3	3-3 = 0	0
4	4-3 = 1	1
5	5-3 = 2	4
$\mu = 3$	$\Sigma(X - \mu) = 0$	$\Sigma(X - \mu)^2 = 10$

Once we add the squared deviations we have a measure of overall variability in the distribution. The sum of the average squared deviations is called the sums of squares, and will be used in almost everyone formula we learn this semester. Please refer back to this section if formulas give you problems later on in the course.

Once we divide these squared sums we will get the average squared deviation or variance. In this example it is  $10/5 = 2$ .

Since we are in squared units and not the same units as our scale we can take the square root of the variance in order to get the standard deviation. The standard deviation is the average deviation about the mean. For our example we take the square root of 2 and find 1.41 is the standard deviation.

The formula that contains all these operations is as follows. Note that  $\sigma^2$  is just the symbol we use for population variance and  $\sigma$  is the symbol we use to denote population standard deviation.

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

population variance

$$\sigma = \sqrt{\sigma^2}$$

population standard deviation

When dealing with a sample a minor change to the formula is made, and instead of subtracting the numerator by N, we divide by  $n - 1$ . Try the numbers in the above example to compute the sample variance and standard deviation (variance is 2.5, standard deviation is 1.58).

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

sample variance

$$s = \sqrt{s^2}$$

sample standard deviation

Please review the animated demonstration on variance and standard deviation for another example of how the population formula works. In addition an alternative formula for these same computations is presented. Although the formula detailed here is the best for understanding the concept, the one presented on the web page will be easier to use in the long run. Both appear in the homework packet formula section as well.

See <http://faculty.uncfsu.edu/dwallace/ssandrd1.html>