Neuendorf **Standard Deviation**

A variable's standard deviation is a measure of dispersion of the scores in a sample. The more spread out the scores are, the larger the "sd." The more tightly the clustered the scores, the smaller the "sd." If there is <u>no</u> variation in the scores (e.g., if all subjects in the sample are 20 years old), the sd is zero. In order to calculate the sd for a variable, that variable must be measured at the interval/ratio level. The sd is, more or less, an estimate of the average deviation or distance (squared) from the mean for all the scores. Below is the formula for standard deviation and an example calculation. The imaginary situation for this example is that 12 subjects have reported their amount of TV viewing yesterday, in minutes.

NOTE: the Square of the standard deviation (sd^2) is called the variance.

$$sd = \sqrt{\frac{\sum x^2}{n-1}}$$

$\sum$ = "summation of"

x = each individual deviation score (score – mean)

n = sample size

Subject	x (Score)	*x* (Deviation score)	x^2 (Deviation score)^2
1	200	0	0
2	250	50	2500
3	220	20	400
4	150	-50	2500
5	100	-100	10000
6	180	-20	400
7	200	0	0
8	205	5	25
9	300	100	10000
10	210	10	100
11	185	-15	225
12	200	0	0
	sum = 2400		sum = 26150

$$\text{mean} = \frac{\sum x}{n} = \frac{2400}{12} = 200$$

$$\text{sd} = \sqrt{\frac{26150}{11}} = \sqrt{2237.27} = 48.76 = \text{approx. 49 min.}$$

Based on probability principles as applied to the normal curve, we would expect, given that the TV viewing scores are indeed normally distributed, the following:

> 68% of the scores will fall within 49 minutes either way of the mean, or between 151 and 249 minutes.
>
> 95% of the scores will fall within (1.96) x (49) minutes, or 96 minutes, either way of the mean – between 104 and 296 minutes.
>
> 99% of the scores will fall within (2.58) x (49) minutes, or 126.4 minutes, either way of the mean – between 73.6 and 326.4 minutes.

Remember that in this example we have looked at <u>individual scores in a single sample</u>. For estimating how well a sample represents a whole population, see the handout on standard error.

1/02