Mean Absolute Deviation measures the dispersion within a data set.

example: 1, 2, 3, 2, 4, 5, 2, 3 (just some data) No drastically bigger or smaller values, so average!

find the mean: $\frac{1+2+3+2+4+5+2+3}{8} = \frac{22}{8} = 2.75$

Sum of deviations is always 0.

devations:

$1-2.75=-1.75$

$2-2.75=-0.75$

$3-2.75=0.25$

$2-2.75=-0.75$

$4-2.75=1.25$

$5-2.75=2.25$

$2-2.75=-0.75$

$3-2.75=0.25$

Say we want to find the average deviation:

$\frac{-1.75-0.75+0.25-0.75+1.25+2.25-0.75+0.25}{8} = \frac{0}{8} = 0$ (useless value)

Since sum of deviations for ANY data set is 0, this would always give us an average deviation of 0.

take the deviations in absolute value:

$$\frac{|1-2.75|+|2-2.275|+|3-2.75|+|2-2.75|+|4-2.75|+|5-2.75|+|2-2.75|+|3-2.75|}{8}$$

≈ 0.566 On average, the values deviate by about .566 units.

Sample variance:

example: $s^2 = \frac{(x_1-x_{mean})^2+(x_2-x_{mean})^2+(x_3-x_{mean})^2+\ldots+(x_n-x_{mean})^2}{n-1}$ (not n in bottom, n-1)

$(x_1-x_{mean})^2 \Leftarrow$ squared deviation from the average

$x_{mean} = \frac{x_1+x_2+\ldots+x_n}{n}$

A sample is obtained by reaching into a population and pulling some values out.

$s_1 = 4, 6, 8, 10, 12, 14, 16,$ mean of $s_1 = \frac{4+6+8+10+12+14+16}{7} = \frac{70}{7} = 10$

$$s_1^2 = \frac{(4-10)^2+(6-10)^2+(8-10)^2+(10-10)^2+(12-10)^2+(14-10)^2+(16-10)^2}{7-1} = 18.67 \text{ (measure of dispersion)}$$

Notice 18.67 is bigger than each data value. Reason is we're squaring, so things like $6^2 = 36$.

$s_2 = 6, 7, 9, 10, 11, 13, 14,$ *mean* of $s_2 = \frac{6+7+9+10+11+13+14}{7} = 10$ (same average as above)

$$s_2^2 = \frac{(6-10)^2+(7-10)^2+(9-10)^2+(10-10)^2+(11-10)^2+(13-10)^2+(14-10)^2}{7-1}$$

$$= \frac{(-4)^2+(-3)^2+(-1)^2+0^2+(1)^2+3^2+4^2}{6} = \frac{16+9+1+9+1+9+16}{6} = 8.7$$

Notice $s_2^1 > s_2^2$, so first data set has a bigger dispersion than the second data set. The data in s_1 is more spread than the data in s_2.

Sample Standard Deviation: Just take the square root of the sample variance.

$$s = \sqrt{\frac{(x_1-x_{mean})^2+(x_2-x_{mean})^2+\ldots+(x_n-x_{mean})^2}{n-1}}$$

example: $s = 1, 2, 3, 2, 4, 5, 2, 6,$ *mean* $\bar{s} = \frac{1+2+3+2+4+5+2+6}{8} = 3.1$

$$s = \sqrt{\frac{(1-3.1)^2+(2-3.1)^2+(3-3.1)^2+(2-3.1)^2+(4-3.1)^2+(5-3.1)^2+(2-3.1)^2+(6-3.1)^2}{8-1}} = 1.72 \text{ (measures dispersion)}$$

$b = 1, 3, 3, 5, 5, 6, 7, 1,$ *mean* of b$= \frac{1+3+3+5+5+6+7+1}{8} = 3.9$ (bigger than 3.1)

$$s = \sqrt{\frac{(1-3.9)^2+(3-3.9)^2+(3-3.9)^2+(5-3.9)^2+(5-3.9)^2+(6-3.9)^2+(7-3.9)^2+(1-3.9)^2}{8-1}} = 2.23$$

2.23>1.72, so second data set is more spread.

summary:

measures of central tendency: mean, median, mode, (use median with ouliers)

measures of dispersion: range, mean absolute deviation, variance, standard deviation (samples)

sample is obtained by reaching into a population and pulling a set of values out.

A good sample is one that's very representative of the largest possible set of opinions.

example: imagine we have a population of values: $\underbrace{1, 2, 3, 4, 1, 2, 4, 3, 5, 6, 7, 9, 6, 6, 5}_{\text{entire population}}$

reach in and pull a sample out of size , say, 4: 1, 4, 6, 5 (sample of size 4)

$$\left.\begin{array}{r} \text{sample average} = \dfrac{1+4+6+5}{4} = 4 \\ \text{population average} = \dfrac{1+2+3+4+1+2+4+3+5+6+7+9+6+6+5}{15} = 4.3 \end{array}\right\} \text{different values}$$

grab another sample:

3, 4, 1, 5 : $sample$ average$= \dfrac{3+4+1+5}{4} = 3.25$ (different from 4 and 4.3)

average of sample averages: $\dfrac{4+3.25}{2} = 3.625$

This is rigged because we know the population and the average. In real life, a population is like 320,000,000 people, so we DON 'T KNOW the population average. In this case, we have to design our samples well to make sure they are as representative of the population as possible.

sample 3: 1, 1, 2, 2, average$= \dfrac{1+1+2+2}{4} = 1.5$ (pop. average is 4)

In this case, the sample average is far from the pop. average of 4.