

I. Measures of Central Tendency:

-Allow us to summarize an entire data set with a single value (the midpoint).

1. Mode : The value (score) that occurs most often in a data set.

- Mo_x = Sample mode

- Mo = Population mode

2. Median : the point (score) which divides the data set in $\frac{1}{2}$: e.g. $\frac{1}{2}$ of the subjects are above the median and $\frac{1}{2}$ are below the median.

- Mdn_x = Sample Median

- Mdn = Population Median

3. Mean: the arithmetic average: Directly considers every score in a distribution.

$$- \bar{X} = \frac{\sum X}{n} = \text{Sample Mean}$$

$$- \mu = \frac{\sum X}{N} = \text{Population Mean}$$

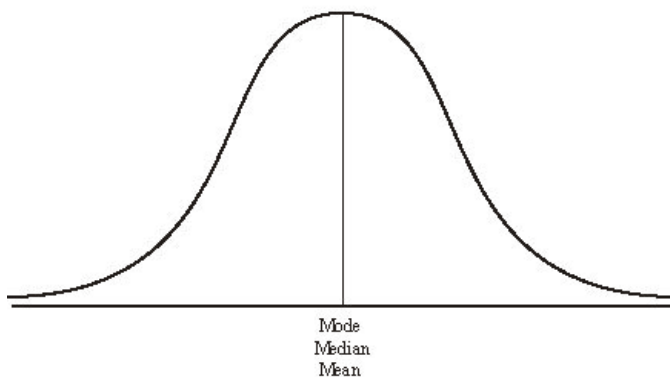
$$- n\bar{X} = \sum X$$

II. Skewed Distributions & the 3M's

-Skewness refers to the shape of the distribution which can be influenced by extreme scores.

- Skewness is also an estimate of the deviation of the Mean, Median, and Mode.

Figure 4.1 A Symmetrical Distribution

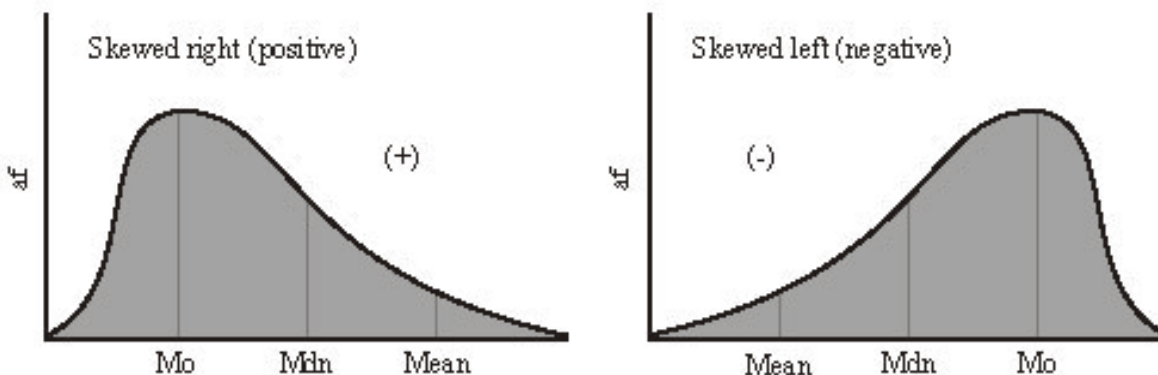


-Symmetrical Dist. = Mean, Median, Mode are all in the same location in the dist.

- Skewed Right (Positively Skewed) = Mode in peak of dist.(left of center), Median in center of distribution, Mean in right tail of distribution

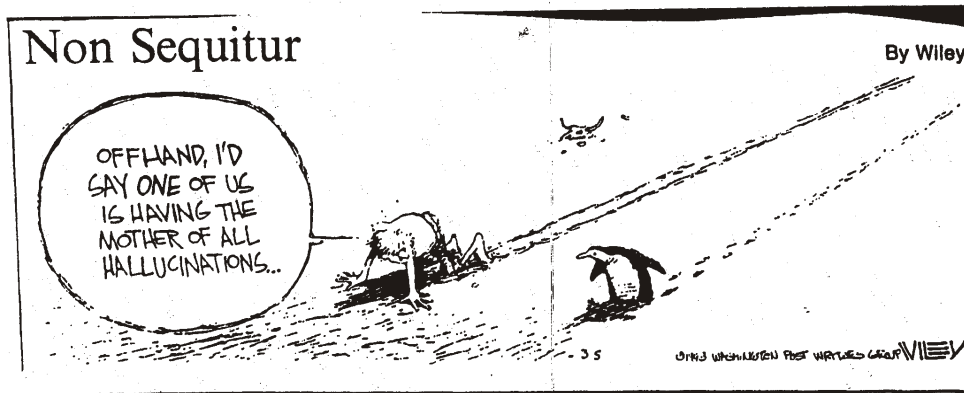
Skewed Left (Negatively Skewed) = Mode in peak of dist (right of center), Median in center of distribution, Mean in left tail of distribution

Figure 4.2 Positively and Negatively Skewed Distributions



I. Measures of Variability (Dispersion)

- Like measures of central tendency, measures of variability allow us to summarize our data set with a single value.
- When central tendency and variability are considered together we get a more accurate picture of our data set.
- The 3 main measures of variability: Range, Variance, and Standard Deviation.
 - these formulas are the root formulas for many of the statistical tests that will be covered later
- These measures tell us how much observations in a data set *vary* or how they are *dispersed* within the distribution.
- Although this information is indirectly contained within measures of central tendency, they don't tell us much about the variance within our data



Example.

Number of miles traveled before traveling companion appears human

n=8

Zoo Penguins

5
5
5
5
5
5
5
5

5
 $\Sigma X = 40$

South Pole Penguins

2
3
4
5
5
6
7
8

8
 $\Sigma X = 40$

Mean = 5, Mode = 5, Median = 5 for both data sets (They do not differ)

-all zoo penguins hallucinate after traveling 5 mile, while there is much more variability in the distances traveled by South Pole Penguins.

-In order to draw accurate conclusions about our data both central tendency and variability must be considered.

II. **Range** : The numerical distance between the largest (*X maximum*) and smallest values (*X minimum*), tells us something about the variation in scores we have in our data, or it tells us the width of our data set.

Range = X maximum - X minimum

- Range for Zoo penguins = 5-5 = 0;

- Range for South Pole P's = $8-2 = 6$

-Problems with Range, Again this is a summary measure that does not directly consider every value in the data set (here only the two extreme numbers; largest and smallest) therefore we do not know if most of the scores occur at the extremes of the distribution or toward the center. It is a very crude measure of variability.

For example:

Zoo Penguins	South Pole Penguins	North Pole Penguins
5	2	2
5	3	5
5	4	5
5	5	5
5	5	5
5	6	5
5	7	5
5	8	8
$\Sigma X = 40$	$\Sigma X = 40$	$\Sigma X = 40$
Range = 0	Range = 6	Range = 6

III. Variance = mathematically assess the total amount of variability in a data set by directly considering every observation.

-To do this requires a point from which each observation can be compared to assess the amount they fluctuate. The Mean can be used for this (since it considers every observation in its calculation).

Mean Deviation = $\Sigma(X - \bar{x})$

Zoo Penguins	X - Mean	South Pole Penguins	X - Mean	North Pole Penguins	X - Mean
5	0	2	-3	2	-3
5	0	3	-2	5	0
5	0	4	-1	5	0
5	0	5	0	5	0
5	0	5	0	5	0
5	0	6	1	5	0
5	0	7	2	5	0
5	0	8	3	8	3
$\Sigma X = 40$	$\Sigma(X-Mean) = 0$	$\Sigma X = 40$	$\Sigma(X-Mean) = 0$	$\Sigma X = 40$	$\Sigma(X-Mean) = 0$
Mean = 5		Mean = 5		Mean = 5	
Range = 0		Range = 6		Range = 6	

- The sum of the mean deviation for any data set is always 0. This limits the usefulness of the mean deviation for summarizing different data sets with a single point.

-If we square each deviation value then the negative values cancel out and we are left with a more

meaningful value.

Zoo Penguins	X - Mean	(X - Mean) ²	South Pole Penguins	X - Mean	(X - Mean) ²
5	0	0	2	-3	9
5	0	0	3	-2	4
5	0	0	4	-1	1
5	0	0	5	0	0
5	0	0	5	0	0
5	0	0	6	1	1
5	0	0	7	2	4
5	0	0	8	3	9
$\Sigma X = 40$	$\Sigma(X - \text{Mean}) = 0$	$\Sigma(X - \text{Mean})^2 = 0$	$\Sigma X = 40$	$\Sigma(X - \text{Mean}) = 0$	$\Sigma(X - \text{Mean})^2 = 28$
Mean = 5			Mean = 5		
Range = 0			Range = 6		

North Pole Penguins	X - Mean	(X - Mean) ²
2	-3	9
5	0	0
5	0	0
5	0	0
5	0	0
5	0	0
5	0	0
8	3	9
$\Sigma X = 40$	$\Sigma(X - \text{Mean}) = 0$	$\Sigma(X - \text{Mean})^2 = 18$
Mean = 5		
Range = 6		

-If we sum these values we no longer get 0, but a number that reflects the total variance for this data set, if we divide that number by N or n we get the average variance for this data set.

Definitional Population Formula $= \sigma^2 = \frac{\Sigma (X - \bar{x})^2}{N}$

Definitional Sample Formula $= s^2 = \frac{\Sigma (X - \bar{x})^2}{n-1}$

-Note sample variance uses n-1 rather than N because it is an estimate of the population variance, because of this reduced denominator the sample variance will always be slightly larger than the population variance.

Zoo Penguins = $\sigma^2 = \frac{\Sigma (X - \bar{X})^2}{N} = \frac{0}{8} = 0$

$s^2 = \frac{\Sigma (X - \bar{X})^2}{n-1} = \frac{0}{8-1} = \frac{0}{7} = 0$

$$\sigma^2 = \frac{\sum (X - \bar{X})^2}{N} = \frac{28}{8} = 3.5$$

South Pole Penguins =

$$s^2 = \frac{\sum (X - \bar{X})^2}{n - 1} = \frac{28}{8 - 1} = \frac{28}{7} = 4$$

$$\sigma^2 = \frac{\sum (X - \bar{X})^2}{N} = \frac{18}{8} = 2.25$$

North Pole Penguins =

$$s^2 = \frac{\sum (X - \bar{X})^2}{n - 1} = \frac{18}{8 - 1} = \frac{18}{7} = 2.5714$$

-gives a good idea of how we get variance but it is time consuming for large data sets, so we have developed mathematically identical (algebraic equivalents) formulas that are a little easier to calculate

Computational Formulas=

$$\text{Population Variance} = \sigma^2 = \frac{\sum X^2 - (\sum X)^2/N}{N}$$

$$\text{Sample Variance} = s^2 = \frac{\sum X^2 - (\sum X)^2/n}{n-1}$$

-Note sample variance uses n-1 rather than N because it is an estimate of the population variance, because of this reduced denominator the sample variance will always be slightly larger than the population variance.

	Zoo Penguins	X ²	South Pole Penguins	X ²	North Pole Penguins	X ²
	5	25	2	4	2	4
	5	25	3	9	5	25
	5	25	4	16	5	25
	5	25	5	25	5	25
	5	25	5	25	5	25
	5	25	6	36	5	25
	5	25	7	49	5	25
	5	25	8	64	8	64
$\sum X$	40		40		40	
$\sum X^2$		200		228		218

$$\sigma^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N} = \frac{200 - \frac{(40)^2}{8}}{8} = \frac{200 - \frac{1600}{8}}{8} = \frac{200 - 200}{8} = \frac{0}{8} = 0$$

Zoo Penguins:

$$s^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1} = \frac{200 - \frac{(40)^2}{8}}{8 - 1} = \frac{200 - \frac{1600}{8}}{7} = \frac{200 - 200}{7} = \frac{0}{7} = 0$$

$$\sigma^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N} = \frac{228 - \frac{(40)^2}{8}}{8} = \frac{228 - \frac{1600}{8}}{8} = \frac{228 - 200}{8} = \frac{28}{8} = 3.5$$

South Pole Penguins:

$$s^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1} = \frac{228 - \frac{(40)^2}{8}}{8-1} = \frac{228 - \frac{1600}{8}}{7} = \frac{228 - 200}{7} = \frac{28}{7} = 4$$

$$\sigma^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N} = \frac{218 - \frac{(40)^2}{8}}{8} = \frac{218 - \frac{1600}{8}}{8} = \frac{218 - 200}{8} = \frac{18}{8} = 2.25$$

North Pole Penguins:

$$s^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1} = \frac{218 - \frac{(40)^2}{8}}{8-1} = \frac{218 - \frac{1600}{8}}{7} = \frac{218 - 200}{7} = \frac{18}{7} = 2.5714$$

-Problems: This formula is the base for many other statistical formulas, however as a single summary measure it has little numerical meaning until it is converted to a standardized score.

- Right now it represents the average distance each penguin is from the mean, in squared mile units.

3. **Standard Deviation**= The square root of a variance. The standardized variance value. It provides us with a numerically meaningful measure of variance. This value (when combined with other stats methods) allow us to infer what percentage of our observations are a certain distance from the mean.

Standard Deviation (based on computational formula of variance)

Population St. Dev. :

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

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

Sample St. Dev. =

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

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

-The larger the value of variance or standard deviation, relative to the numerical values of the observations, the greater the amount of variability that is present in the data set.

$$\text{Zoo Penguins : } \sigma = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{0} = 0$$

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

$$s = \sqrt{s^2} = \sqrt{0} = 0$$

$$\text{South Pole Penguins : } \sigma = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{3.5} = 1.8708$$

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

$$s = \sqrt{s^2} = \sqrt{4} = 2$$

$$\text{North Pole Penguins : } \sigma = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.25} = 1.5$$

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

$$s = \sqrt{s^2} = \sqrt{2.5714} = 1.6935$$