Biostatistics

Lecture 3

Measure of Central Tendency

There are several types of descriptive measures that can be computed from a set of data

Ungrouped

or Grouped

The three most commonly used are:

the Mean

the Median

and the Mode

1. Ungrouped Data

Mean:

-Mean is the first and the most commonly measure of central tendency.

- It is determined by adding all the values in population or sample and dividing by the total number of values that are adding.

The population mean is defined by μ

$$\mu = \frac{X_1 + X_2 + \cdots + X_N}{N} = \frac{\sum_{i=1}^{N} X_i}{N}$$

The sample Mean is defined by $\bar{X}$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

The properties of Mean are:

1- For a given set of data there is only one Mean.

2- Simplicity: is easily understood and easy to compute.

3- Mean is affected by extreme value.

Example: A set data (5, 7, 9, 5, 4).

$$\overline{X} = \frac{5+7+9+5+4}{5} = 6$$

Other set of data with extreme value (5, 7, 9, 5, 24).

$$\overline{X} = \frac{5+7+9+5+24}{5} = 10$$

Median:

The median is that value which located in middle of observations, if these observations are ordered from smallest to largest.

If the number of observations is **even**, the median is the average of two middle values.

$$\text{Median} = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$$

Example:

A set data (10, 54, 10, 33, 21, 53)

Ordered set (10, 10, 21, 33, 53,54)

Rank order (1, 2, 3, 4, 5, 6)

then n=6

$$\text{Median} = \frac{X_{\frac{6}{2}} + X_{\frac{6}{2}+1}}{2} = \frac{X_3 + X_4}{2} = \frac{21+33}{2} = 27$$

If the number of observations is **odd**, the median is

$$Median = X_{\frac{n+1}{2}}$$

Example:

A set data (10, 10, 33, 21, 53)

Ordered set (10, 10, 21, 33, 53)

Rank order (1, 2, 3, 4, 5)

then n=5

$$\text{Median} = X_{\frac{5+1}{2}} = X_3 = 21$$

Mode:

The mode is the value, which occurs most frequency.

Example:

(2, 6, 3, 7, 0, 10, 4)

No Mode

(5, 6, 10, 12, 6, 7, 6)

One Mode, 6

(0, 2, 5, 4, 2, 1, 2, 4, 4)

Two Mode, 2, 4

Comparison of the mean, median and mode

Exp: Data set (2, 2, 4, 5, 7, 9)

Mean= 4.833 Median= 4.5 Mode= 2

If we change the last data point from 9 to 28, the mean equal 8, but the median and the mode remain the same.

Discussion:

1. The mean is sensitive to extremes value but the median is not.

2. For some data set, the mean can give a misleading picture of the observation.

Example:

(2, 2, 2, 2, 17)

The mean is not accurately representative of the data set.

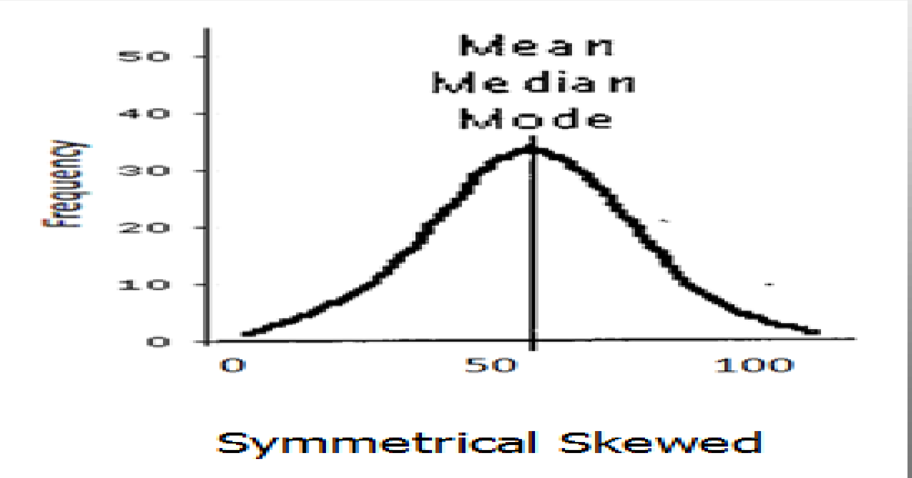
3. **The median sometimes ignores potentially useful information because only the middle value (or two middle values) affects the median. The median is the best measure of central tendency when the distributions are small or extremely skewed.**

4. **The mode can sometimes be useful, but it tends to characterize individuals more than groups.**

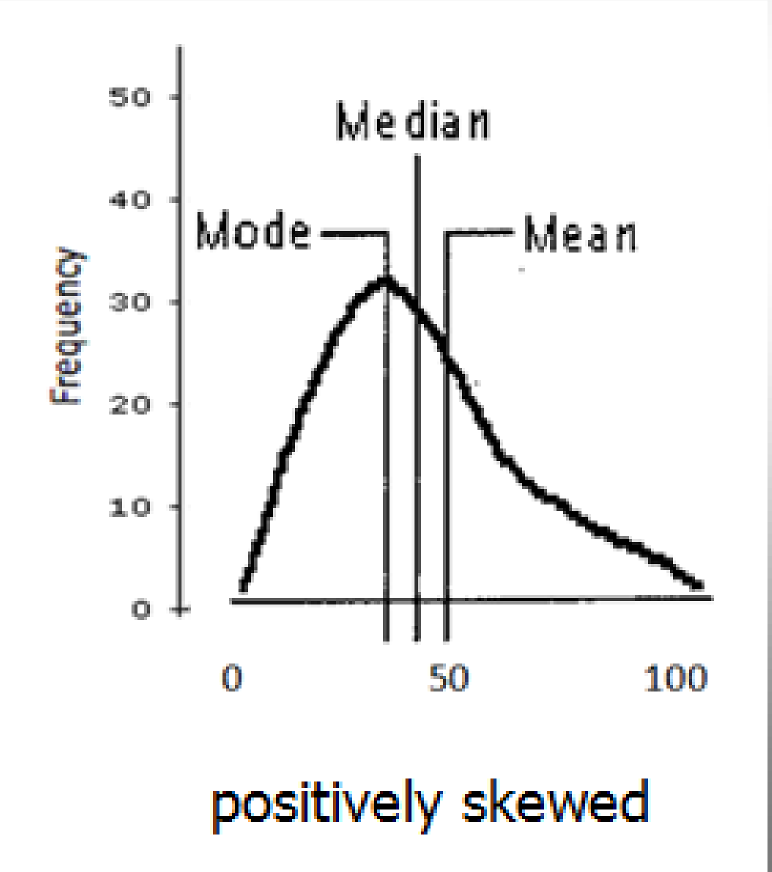
Relationship of the Mean, Median and Mode

The relationship of the mean, median and mode to each other can provide some information about the relative shape of the data distribution.

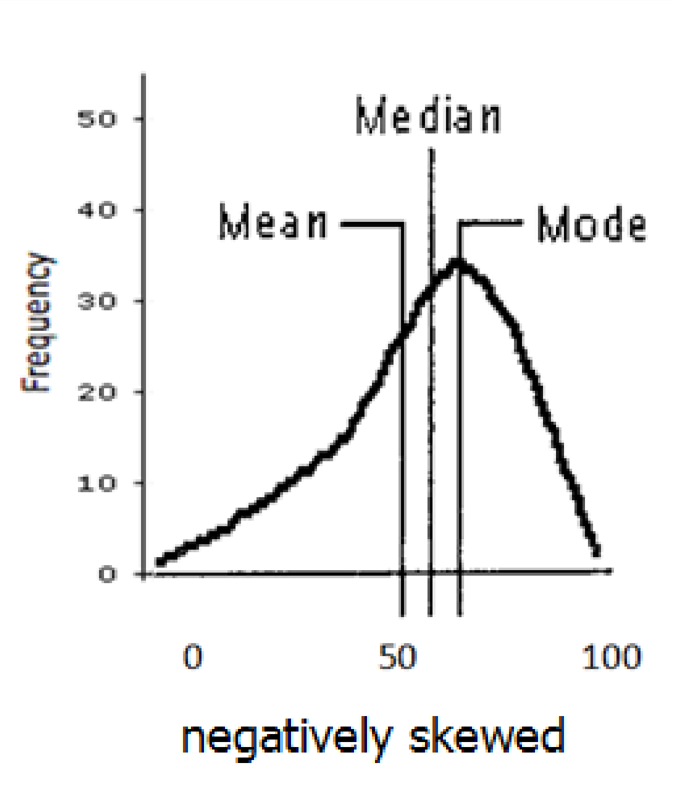
1. If the mean, median, and mode are approximately equal to each other, the distribution can be assumed to be approximately symmetrical.



2. If the mean > median > mode, the distribution will be skewed to the left or positively skewed.



3. If the mean < median < mode, the distribution will be skewed to the right or negatively skewed.



2. Grouped data

Mean:

The mean of grouped data equal the sum of multiplied mid point by corresponding frequencies and divided by the sum of frequencies.

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{k} mifi}{\sum_{i=1}^{k} fi}$$

<u>Example:</u> The following are hemoglobin values (g/dl) of 30 children receiving treatment for hemolytic anemia.

Classes (Hemoglobin)	Midpoint (mi)	Frequency (f)	Cumulative Frequency	Relative Frequency
6.5 – 7.5	7	1	1	0.033
7.5 – 8.5	8	5	6	0.167
* 8.5 – 9.5	9	11	17	0.367
9.5 – 10.5	10	9	26	0.300
10.5 – 11.5	11	3	29	0.100
11.5 – 12.5	12	1	30	0.033
Total	---	n=30	---	1.000

Classes (Hemoglobin)	Midpoint (mi)	Frequency (f)	Cumulative Frequency	Relative Frequency
6.5 – 7.5	7	1	1	0.033
7.5 – 8.5	8	5	6	0.167
* 8.5 – 9.5	9	11	17	0.367
9.5 – 10.5	10	9	26	0.300
10.5 – 11.5	11	3	29	0.100
11.5 – 12.5	12	1	30	0.033
Total	---	n=30	---	1.000

$$\frac{7\times1+8\times5+9\times11+10\times9+11\times3+12\times1}{1+5+11+9+3+1} = 9.3$$

Median:

The first step in computing the median from grouped data is to locate the median in which class is located.

$$\left(\sum_{i=1}^{k} \frac{fi}{2} \quad \text{or} \quad \frac{n}{2} \right) = \frac{30}{2} = 15 \text{ .... (The median location)}$$

$$\text{Median} = a + \left[\frac{\frac{n}{2} - n_1}{f_m}\right] \times \Delta$$

Where:

a= lower value of class containing median.

n= total frequency.

n1= cumulative frequency of previous class containing median.

fm= No. of observation in class containing median.

∆= value of class interval.

Classes (Hemoglobin)	Midpoint (mi)	Frequency (f)	Cumulative Frequency	Relative Frequency
6.5 – 7.5	7	1	1	0.033
7.5 – 8.5	8	5	6	0.167
* 8.5 – 9.5	9	11	17	0.367
9.5 – 10.5	10	9	26	0.300
10.5 – 11.5	11	3	29	0.100
11.5 – 12.5	12	1	30	0.033
Total	---	n=30	---	1.000

$$\text{Median} = a + \left[\frac{\frac{n}{2} - n_1}{f_m}\right] \times \Delta = 8.5 + \left[\frac{\frac{30}{2} - 6}{11}\right] \times 1 = 9.367$$

Mode:

The mode of grouped data, we usually refer to the modal class, where the modal class is the class interval with highest frequency. Refer to the previous example; the mode is taken as the midpoint of the modal class.

The modal class is (8.5 – 9.5)

The mode = 9

Which Measure Do I Use?

When the distribution is symmetrical the three measures are about the same (mean, med & mod).

- The mean is favored because it is closely related to the variance and standard deviation.

- The median is used when there are:

1. extreme scores.
2. the distribution is skewed.
3. when there are open ended data (undetermined values).

- The mode is used as an extra descriptive and when a mean and/or median can’t be calculated, such as with a nominal variable, and when discrete values are used (number of children).

GOOD LUCK