Measures of Central Tendency

Measure of central tendency and its types

- Arithmetic mean for ungroup and group data. Its uses and applications
- Median for ungroup and group data. Its uses and applications
- Mode for ungroup and group data. Its uses and applications
- Empirical relationship between mean, median and mode.

Measures of Central Tendency

- A *measure of central tendency* is a descriptive statistic that describes the average, or typical value of a set of scores.

Types of Averages

- There are five common measures of central tendency.
  - Arithmetic Mean
  - Median
  - Mode
  - Geometric Mean
  - Harmonic Mean
- First three known as primary and last two known as secondary.

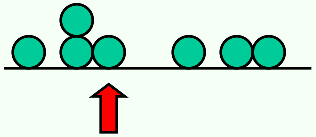
Measures of Central Tendency

Central Tendency

- Mean

$$\sum_{i=1}^{n} x_i \Big/ n$$

- Median



- Mode
- Geometeric Mean
- Harmonic Mean

The Mean

$\bar{x}$

- The *mean* is sum of all observations divide by no of observations.
- The mean of a population is represented by the Greek letter μ; the mean of a sample is represented by $\bar{X}$
  - the arithmetic average of all the scores (ΣX)/N

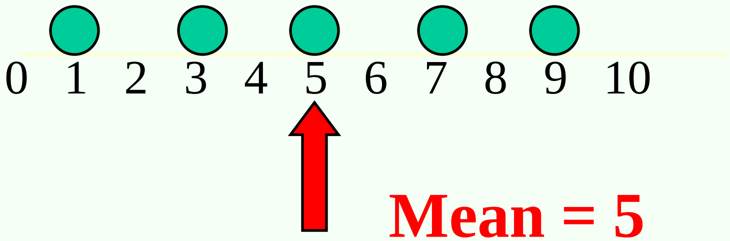
The Mean (Arithmetic Mean)

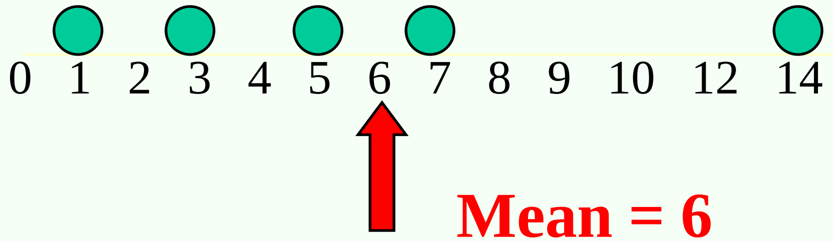
- It is the Arithmetic Average of data values:

$$\bar{x} = \sum_{i=1}^{n} x_i \Big/ n = \frac{x_i + x_2 + \bullet \bullet \bullet + x_n}{n}$$

Sample Mean

- The Most Common Measure of Central Tendency





Calculating the Mean

- Calculate the mean of the following data:
  1   5   4   3   2
- Sum the scores (ΣX):
  $1 + 5 + 4 + 3 + 2 = 15$
- Divide the sum ($\Sigma X = 15$) by the number of scores ($N = 5$):
  $15 / 5 = 3$
- Mean = $\bar{X}$ = 3

Mean for group data

Hourly wages	No of workers		xi	fx
50-54	4		52	208
55-59	8		57	456
60-64	12		62	744
65-69	20		67	1340
70-74	16		72	1152
75-79	10		77	770
.80-84	5		82	410
Total	75			5080

Where x is mid point and calculte by using formula

mid point is equal to Lower limit +upper limit and divide by 2.

$$\bar{X} = \frac{\sum fx}{\sum f} = \frac{5080}{75} = 67.733$$

Properties of A.M

- the number, m, that makes $\Sigma(X - m)$ equal to 0
- the number, m, that makes $\Sigma(X - m)^2$ a minimum

Draw back

- Affected by Extreme Values (Outliers)

When To Use the Mean

- You should use the mean when
  - the data are interval or ratio scaled
    - Many people will use the mean with ordinally scaled data too
  - and the data are not skewed
- The mean is preferred because it is sensitive to every score
  - If you change one score in the data set, the mean will change

The Median

- The *median* is simply another name for the 50th percentile
  - It is the score in the middle; half of the scores are larger than the median and half of the scores are smaller than the median

How To Calculate the Median

- Conceptually, it is easy to calculate the median
  - There are many minor problems that can occur; it is best to let a computer do it
- Sort the data from highest to lowest or lowest to highest.
- Find the score in the middle
  - middle = (n+ 1) / 2 th value

- If n is odd, the median is the middle number.
- If n is even, the median is the average of the 2 middle numbers.

Median Example

- What is the median of the following scores:
  10 8 14 15 7 3 3 8 12 10 9
- Sort the scores:
  15 14 12 10 10 9 8 8 7 3 3
- Determine the middle score:
  middle = (n + 1) / 2 = (11 + 1) / 2 = 6
- Middle score = median = 9

Median Example

- What is the median of the following scores:
  24 18 19 42 16 12
- Sort the scores:
  42 24 19 18 16 12
- Determine the middle score:
  middle = (N + 1) / 2 = (6 + 1) / 2 = 3.5
- Median = average of 3rd and 4th scores:
  (19 + 18) / 2 = 18.5

Median for group data

$$Median = l + \frac{h}{f}(\frac{n}{2} - c)$$

- L lower class limit of selected class
- h class interval
- f frequency of selected class
- n total frequency
- C preceding cummulative frequency

When To Use the Median

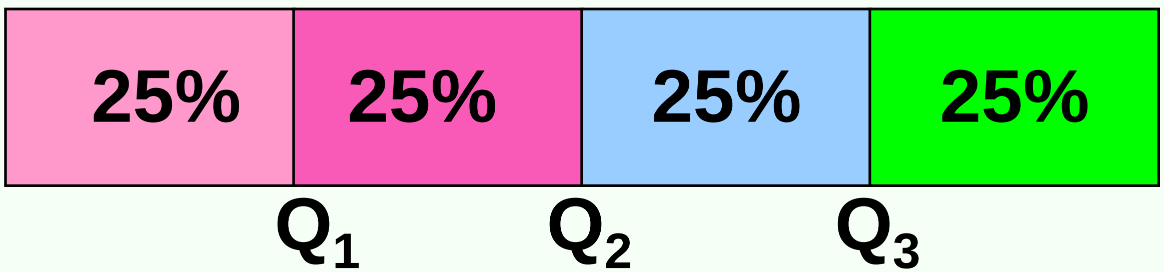
- The median is often used when the distribution of scores is either positively or negatively skewed
  - The few really large scores (positively skewed) or really small scores (negatively skewed) will not overly influence the median.
  - Median is not Affected by Extreme Values

Quantiles

- Quartiles
- Deciles
- Percentiles

Quartiles

- **Not a Measure of Central Tendency**
- **Split Ordered Data into 4 Quarters**



- **Position of i-th Quartile:** **position of point** $Q_i = \frac{i(n+1)}{4}$

Data in Ordered Array: 11 12 13 16 16 17 18 21 22

Position of Q_1 = $\frac{1 \cdot (9 + 1)}{4}$ = 2.50 Q_1 =12.5

For Group data

$$Q_1 = L + \frac{h}{f}(\frac{N}{4} - CF)$$

$$Q_2 = L + \frac{h}{f}(\frac{2N}{4} - CF)$$

$$Q_3 = L + \frac{h}{f}(\frac{3N}{4} - CF)$$

Decile

$$D_i = l + \frac{h}{f}\left(\frac{in}{10} - c\right)$$

$$D_1 = l + \frac{h}{f}\left(\frac{n}{10} - c\right)$$

$$i = 1, 2, \ldots, 9$$

Percentile

$$Pi = l + \frac{h}{f}\left(\frac{iN}{100} - c\right);\ i = 1, 2, 3 \ldots.., 99$$

Where:

l = lower boundary of Percentile group

h = Width of Percentile group

f = Frequency of Percentile group

N = Total number of observations i.e. sum of the frequencies

c = Cumulative frequency preceding Percentile group

Relationship between quantiles

$First\ Quartile\ Q_1 = P_{25}$ $\quad$ $First\ Decile\ D_1 = P_{10}$

$Second\ Quartile\ Q_2 = P_{50}$ $\quad$ $Second\ Decile\ D_2 = P_{20}$

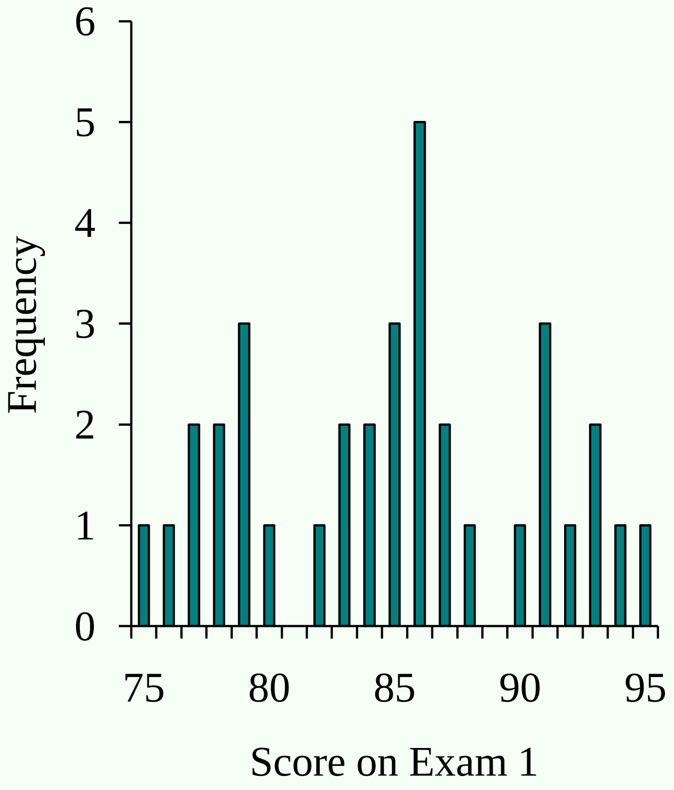
$Third\ Quartile\ Q_3 = P_{75}$ $\quad$ $Fifth\ Decile\ D_5 = P_{50}$ and so on

$$Second\ Quartile = Fifth\ Decile = 50th\ Percentile = Median$$

$$Q_2 = D_5 = P_{50} = Median$$

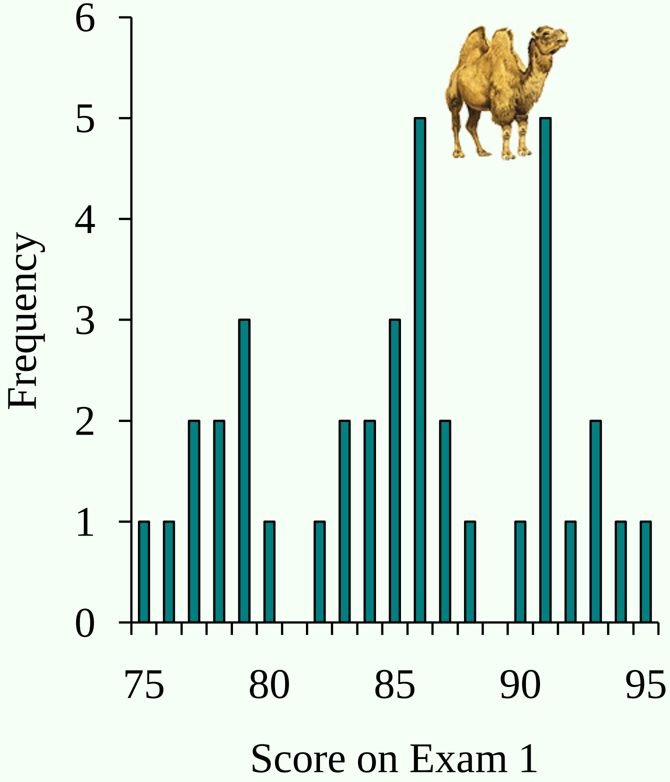
The Mode

- The *mode* is the score that occurs most frequently in a set of data



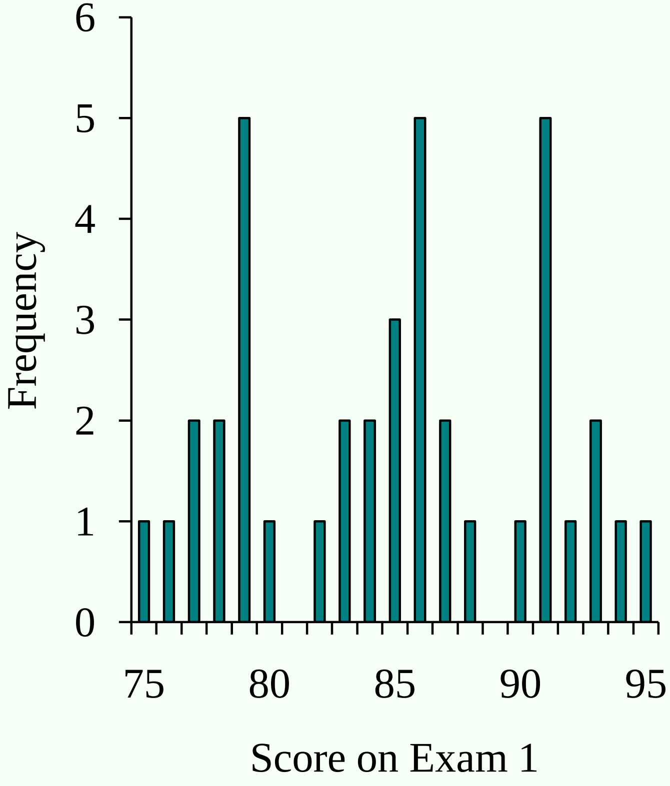
Bimodal Distributions

- When a distribution has two "modes," it is called *bimodal*



Multimodal Distributions

- If a distribution has more than 2 "modes," it is called *multimodal*



Mode for group data

$$\text{mode} = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$$

- L is lower class limit of modal class
- fm maximum frequency
- f1 preceeding frequency of modal class
- f2 preceeding frequency of modal class
- h class interval

For group Data

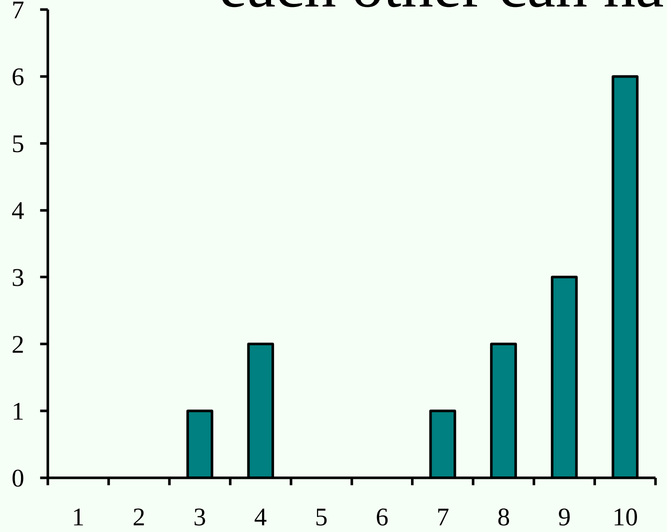
Hourly wages	No of workers	
50-54	4	
55-59	8	
60-64	12	f1
65-69	20	fm
70-74	16	f2
75-79	10	
.80-84	5	

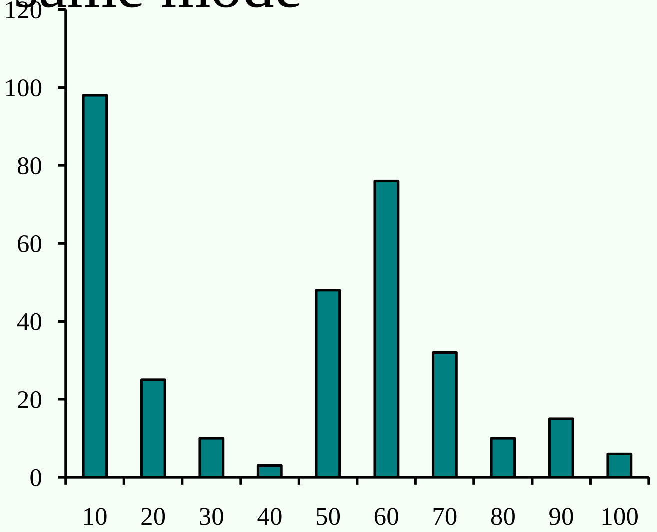
Calculte the mode by putting the value in formula

- Mode=65+$\frac{(20-12)}{(20-12)+(20-16)}\times 5$

When To Use the Mode

- The mode is not a very useful measure of central tendency
  - It is insensitive to large changes in the data set
    - That is, two data sets that are very different from each other can have the same mode



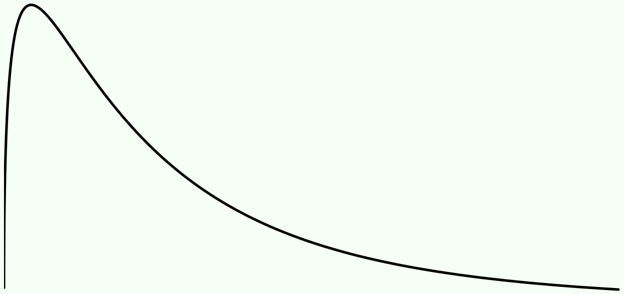


When To Use the Mode

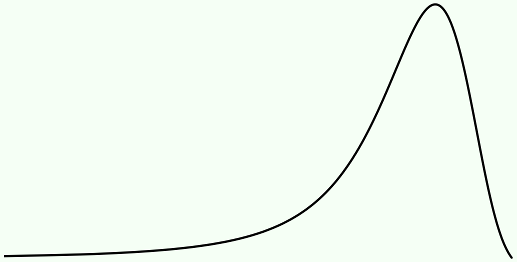
- The mode is primarily used with nominally scaled data
  - It is the only measure of central tendency that is appropriate for nominally scaled data

Relations Between the Measures of Central Tendency

- In symmetrical distributions, the median and mean are equal
  - For normal distributions, mean = median = mode



- In positively skewed distributions, the mean is greater than the median and mode



- In negatively skewed distributions, the mean is smaller than the median &

Shapes

- **Describes How Data Are Distributed**

Measures of Shape:

- Symmetric or skewed