Applied Statistics
Course No: CSE 3107

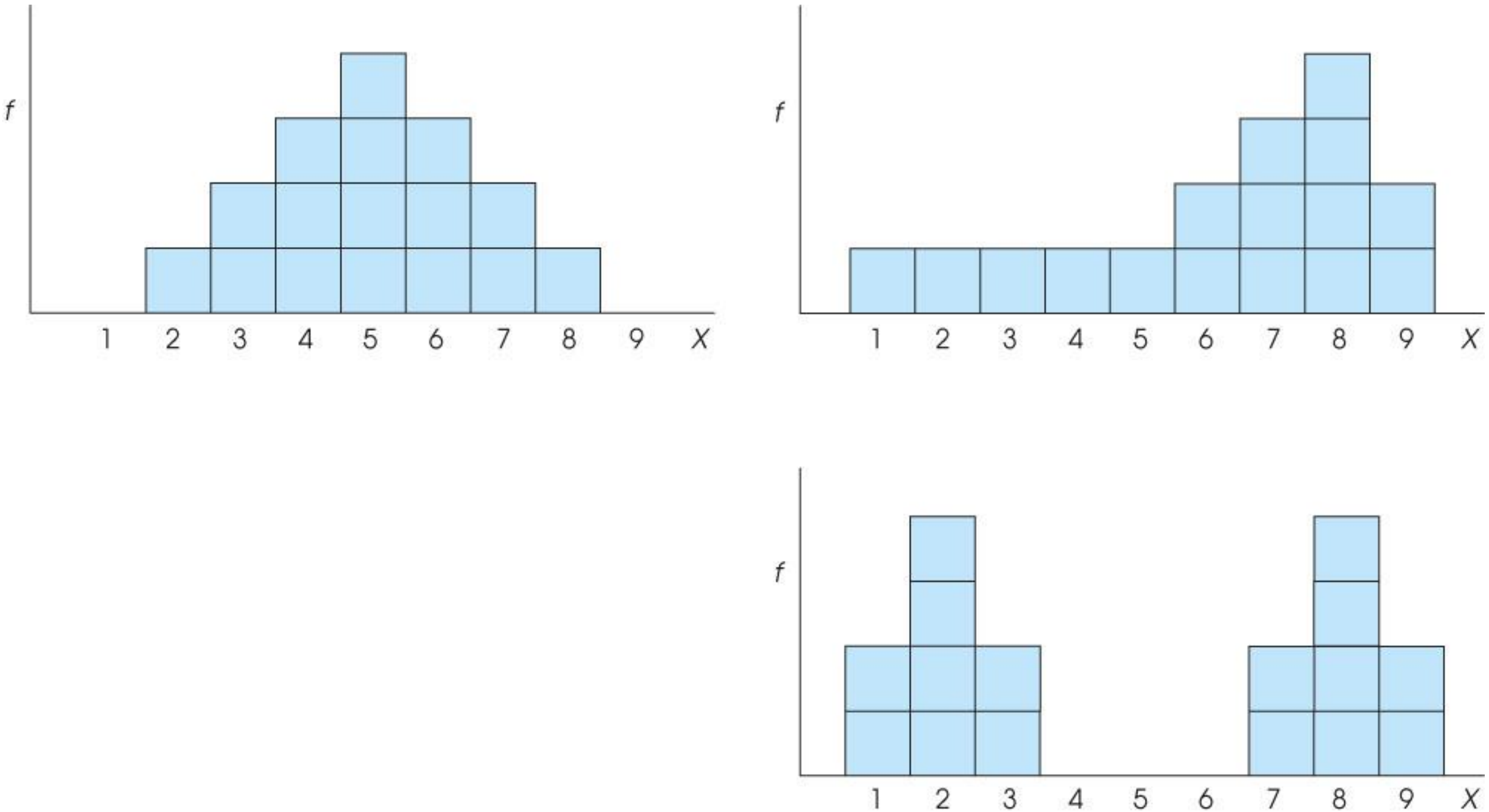
Topic 2: Measures of Central Tendency

Julia Rahman

Central Tendency

- In statistics, a **central tendency** is a central value or a typical value for a probability distribution.
- Central tendency determines a single value that accurately describes the center of the distribution and represents the entire distribution of scores.
- The goal of central tendency is to identify the single value that is the best representative for the entire set of data.
- By identifying the "average score," central tendency allows researchers to summarize or condense a large set of data into a single value.
- This single value describes the PERFORMANCE of the group.

Central Tendency



Central Tendency

Three commonly used measures of central tendency

1. Mean
2. Median
3. Mode

Arithmetic Mean

- Mean – arithmetic average
- It is the most commonly used measure of the center of data
- If we get the mean of the sample, we call it the sample mean and it is denoted $\bar{x}$ by (read "x bar").
- The MEAN of a set of values or measurements is the sum of all the measurements divided by the number of measurements in the set.
- Computation of Sample Mean

$$\bar{X} = \frac{\sum X}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

- If we compute the mean of the population, we call it the parametric or population mean denoted by μ

Arithmetic Mean

- Example 1: 9 students' exam scores: 75; 69; 88; 93; 95; 54; 87; 88; 27

  Mean $\bar{x} = \frac{75+69+88+93+95+54+87+88+27}{9}$ = 75:11

- Example 2: Scores of 15 students in Mathematics I quiz consist of 25 items. The highest score is 25 and the lowest score is 10. Here are the scores: 25, 20, 18, 18, 17, 15, 15, 15, 14, 14, 13, 12, 12, 10, 10. Find the mean.

$$\bar{X} = \frac{\sum X}{N} = \frac{228}{15} = 15.2$$

Geometric Mean

- The geometric mean is well defined only for sets of positive real numbers.
- This is calculated by multiplying all the numbers (call the number of numbers n), and taking the nth root of the total.
- A common example of where the geometric mean is the correct choice is when averaging growth rates.
- The geometric mean is NOT the arithmetic mean and it is NOT a simple average.
- Mathematical definition: The nth root of the product of n numbers.

Geometric Mean

- For Ungrouped Data

Geometric mean of X = G = Anti$\left(\frac{\sum logx}{n}\right)$

- For grouped Data

Geometric mean of X = G = Anti $\left(\frac{\sum flogx}{n}\right)$

Geometric Mean

- Example: Find the geometric mean of the following values: 15, 12, 13, 19, 10

Solution:

n = 5 ; $\sum logx$ = 5.648

G = Anti$(\frac{\sum logx}{n})$

= Anti$(\frac{5.648}{5})$

= Anti (1.129)

= 13.48

x	*Log x*
15	1.1761
12	1.0792
13	1.1139
19	1.2788
10	1.0000
Total	5.648

Harmonic Mean

- Harmonic mean is quotient of “number of the given values” and “sum of the reciprocals of the given values”.
- For Ungrouped Data

Harmonic mean of X = $\bar{X} = \frac{n}{\sum(\frac{1}{x})}$

- For grouped Data

Harmonic mean of X = $\bar{X} = \frac{\sum f}{\sum(\frac{f}{x})}$

Harmonic Mean

- Example: Calculate the harmonic mean of the numbers: 13.2, 14.2, 14.8, 15.2 and 16.1

Solution:

$$\bar{X} = \frac{n}{\sum(\frac{1}{x})} = \frac{5}{.3147} = 14.63$$

X	$\frac{1}{X}$
13.2	0.0758
14.2	0.0704
14.8	0.0676
15.2	0.0658
16.1	0.0621
Total	$\sum\frac{1}{X}$ =0.3147

Harmonic Mean

- Example: Calculate the harmonic mean for the given below

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
F	2	3	11	20	32	25	7

Solution:

$$\bar{X} = \frac{\sum f}{\sum(\frac{f}{x})}$$

$$= \frac{100}{1.4368} = 69.60$$

Marks	x	f	$\frac{f}{x}$
30-39	34.5	2	0.0580
40-49	44.5	3	0.0674
50-59	54.5	11	0.2018
60-69	64.5	20	0.3101
70-79	74.5	32	0.4295
80-89	84.5	25	0.2959
90-99	94.5	7	0.0741
Total		100	1.4368

Mean

Relation Between arithmetic mean, geometric mean and harmonic mean:

- Geometric mean of a set of positive numbers is less than or equal to their arithmetic mean but is greater than or equal to their harmonic mean.

$$H \le G \le \bar{X}$$

Weighted Mean

- Weighted mean is the mean of a set of values wherein each value or measurement has a different weight or degree of importance.
- The weighted mean of observations $x_1; x_2; \ldots ; x_n$ using weights $w_1; w_2; \ldots ; w_n$ is given by:

$$\text{weighted mean} = \frac{w_1 x_1 + w_2 x_2 + \ldots + w_n x_n}{w_1 + w_2 + \cdots + w_n} = \frac{\sum W_i X_i}{\sum W_i}$$

- Simple mean is a specific case of weighted mean.

Weighted Mean

- Example: Find the Grade Point Average (GPA) of Paolo Adade for the first semester of the school year 2013-2014. Use the table below:

$$\bar{X} = \frac{\sum w_i x_i}{\sum w_i}$$

$$= \frac{32}{26}$$

$$= 1.23$$

Subjects	Grade (X_i)	Units (w_i)	(X_i) (w_i)
BM 112	1.25	3	3.75
BM 101	1.00	3	3.00
AC 103	1.25	6	7.50
EC 111	1.00	3	3.00
MG 101	1.50	3	4.50
MK 101	1.25	3	3.75
FM 111	1.50	3	4.50
PE 2	1.00	2	2.00
		$\Sigma(w_i) = 26$	$\Sigma(X_i)w_i)=32.00$

Mean

Properties of the Mean

- It measures stability. Mean is the most stable among other measures of central tendency because every score contributes to the value of the mean.
- Mean can be calculated for any set of numerical data, so it always exists.
- A set of numerical data has one and only one mean.
- It may easily affected by the extreme scores.
- It can be applied to interval level of measurement.
- It may not be an actual score in the distribution.
- It is very easy to compute.

Mean

When to use the Mean

- Sampling stability is desired.
- Other measures are to be computed such as standard deviation, coefficient of variation and skewness.

Advantages of Mean

- Sensitive to any change in the value of any observation

Disadvantages of Mean

- Very sensitive to outliers

Median

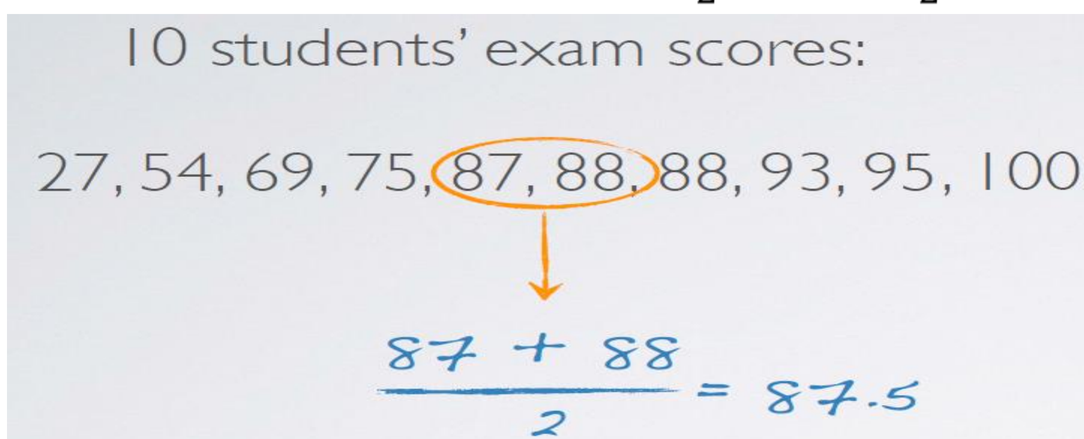
- Median – midpoint of the distribution
- Median is the middle value of the sample when the data are ranked in order according to size.
- Median is what divides the scores in the distribution into two equal parts.
- Fifty percent (50%) lies below the median value and 50% lies above the median value.
- It is also known as the middle score or the 50th percentile.

Median

Calculation of Median of Ungrouped Data:

- Sort the data
- Take the mid point i.e. x[ceil($\frac{N}{2}$)] of the ordered data as the median. Here, N is the sample size.

- When N is even:
  - Sort the data and take the mean of x[$\frac{N}{2}$] and x[$\frac{N}{2}$ + 1] as the median.



Median

- Find the median score of 8 students in an English class. 30, 19, 17, 16, 15, 10, 5, 2

Arrange them: 2, 5, 10, 15, 16, 17, 19, 30

$$\tilde{x} = \frac{15+16}{2} = 15.5$$

Median

- **Calculation of Median of Grouped Data:**
- Formula:

$$\tilde{x} = L_B + \frac{\frac{n}{2} - cfp}{f_m} * \text{ci}$$

$\tilde{x}$ = median value

L_B = lower boundary of the median class (MC)

MC = median class is a category containing the $\frac{n}{2}$

cfp = cumulative frequency before the median class if the scores are arranged from lowest to highest value

f_m = frequency of the median class

ci = size of the class interval

Median

Steps in Solving Median for Grouped Data

1. Complete the table for cf<
2. Get $\frac{n}{2}$ of the scores in the distribution so that you can identify MC.
3. Determine L_B, cfp, f_m, and ci.
4. Solve the median using the formula

Median

- Example: Scores of 40 students in a science class consist of 60 items and they are tabulated below. The highest score is 54 and the lowest score is 10.

X	f	cf<
10 - 14	5	5
15 - 19	2	7
20 - 24	3	10
25 - 29	5	15
30 - 34	2	17 (cfp)
35 - 39	9 (fm)	26
40 - 44	6	32
45 - 49	3	35
50 - 54	5	40
	n = 40	

Median

- Example: Scores of 40 students in a science class consist of 60 items and they are tabulated below. The highest score is 54 and the lowest score is 10. (cont.)

Solution: $\frac{n}{2} = \frac{40}{2}$ = 20

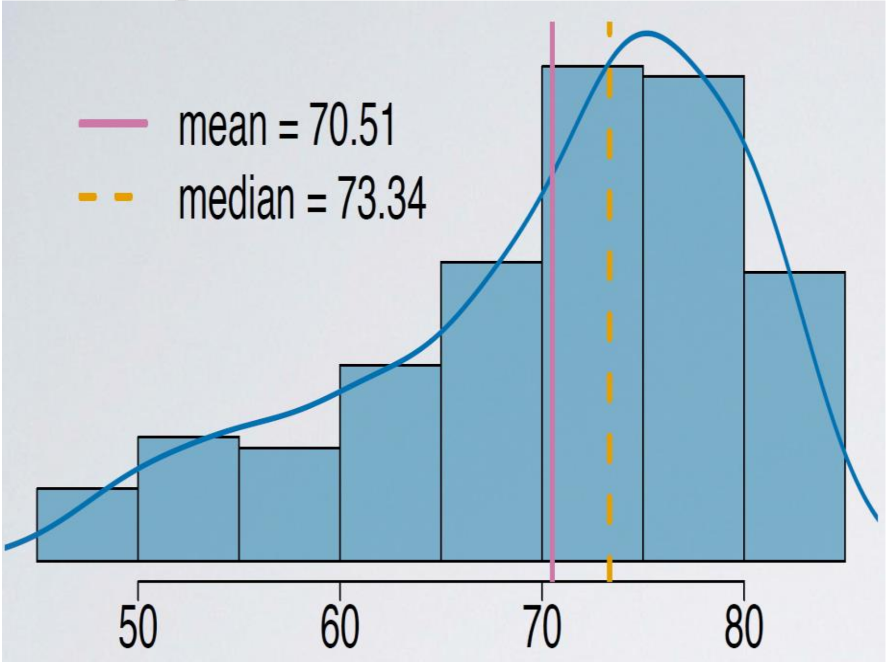
The category containing $\frac{n}{2}$ is 35 – 39.

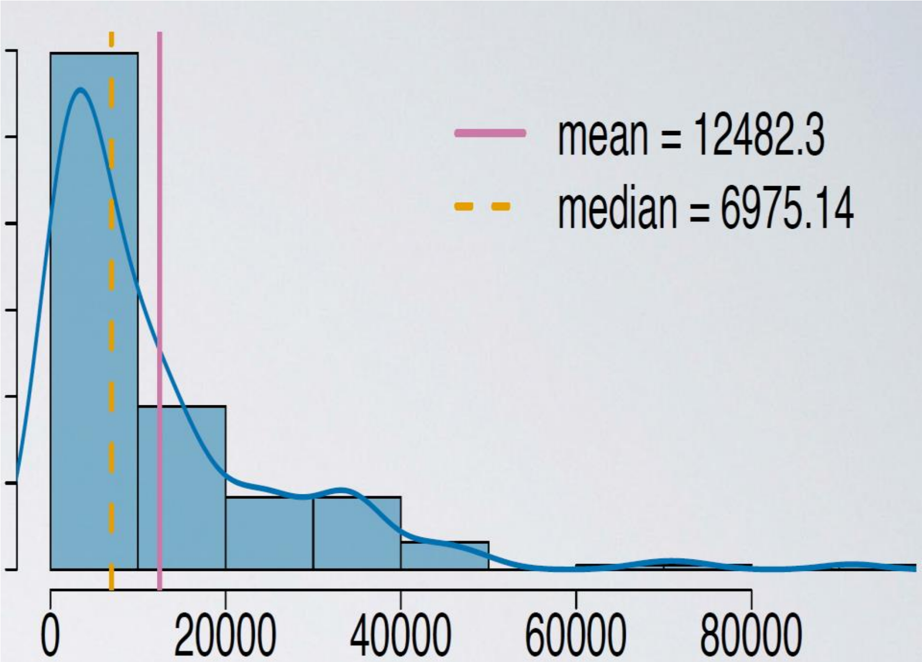
LL of the MC = 35; L_B = 34.5

cfp = 17; f_m = 9; ci = 5

$\tilde{x}$ = $L_B + \frac{\frac{n}{2} - cfp}{f_m}$ * ci = 34.5+ $\frac{20 - 17}{9}$ * 5 = 34.5+ $\frac{15}{9}$ = 36.17

Median





Average life expectancy histogram

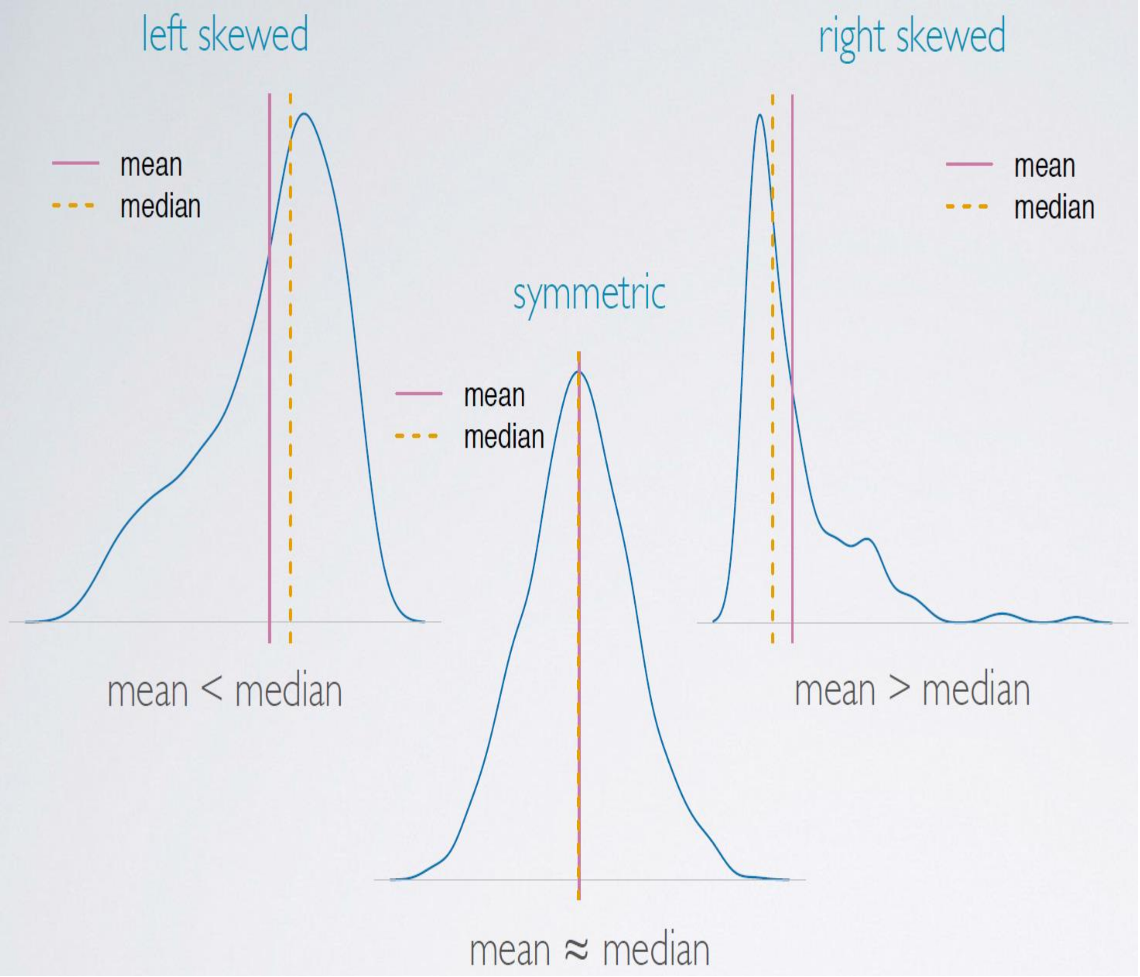
- few values in the left i.e. few low values make the mean smaller.
- few values in the left pull the mean towards them i.e. to the left.

Income per person histogram

- few values in the right i.e. few high values make the mean larger.
- few values in the right pull the mean towards them i.e. to the right.

Median

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



Median

Properties of the Median

- It may not be an actual observation in the data set.
- It can be applied in ordinal level.
- It is not affected by extreme values because median is a positional measure.
- Median exists in both quantitative or qualitative data.

When to Use the Median

- The exact midpoint of the score distribution is desired.
- There are extreme scores in the distribution.

Median

Advantages of the Median

- Median can be calculated in all distributions.
- Median can be understood even by common people.
- Median can be ascertained even with the extreme items.
- It can be located graphically
- It is most useful dealing with qualitative data.

Median

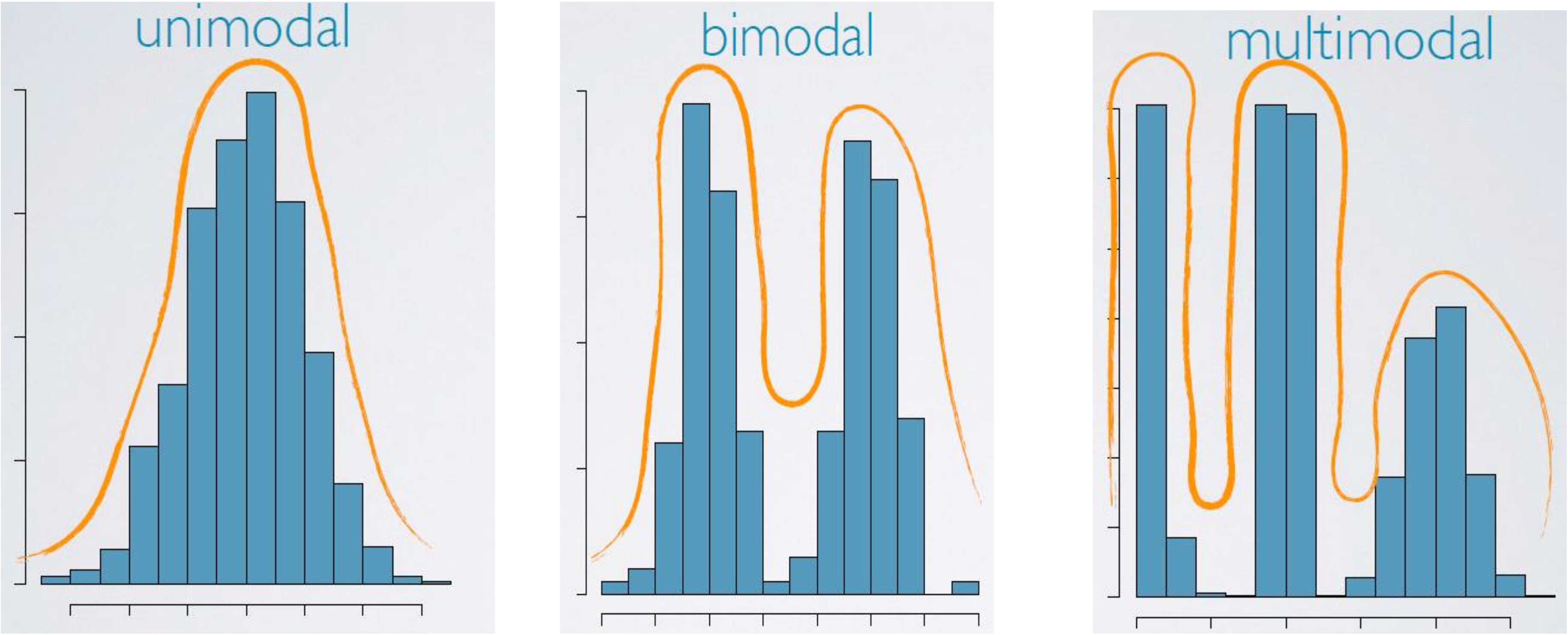
Disadvantages of the Median

- It is not based on all the values.
- It is not capable of further mathematical treatment.
- It is affected fluctuation of sampling.
- In case of even no. of values it may not the value from the data.

Mode

- Mode – most frequent observation
- Mode is the value which occurs most frequently in a set of measurements or values
- The mode or the modal score is a score or scores that occurred most in the distribution.
- It is classified as unimodal, bimodal, trimodal or mulitimodal.
  - Unimodal is a distribution of scores that consists of only one mode.
  - Bimodal is a distribution of scores that consists of two modes.
  - Trimodal is a distribution of scores that consists of three modes or multimodal is a distribution of scores that consists of more than two modes.

Mode



Mode

- Example: 9 students' exam scores: 75; 69; 88; 93; 95; 54; 87; 88; 27

Mode: most frequent observation = 88

Mode

- Example: Scores of 10 students in Section A, Section B and Section C.

- In Section A, more frequent is 20, hence, the mode of Section A is 20. There is only one mode, so it is called **unimodal**.
- In Section B more frequents are 18 and 24, since both 18 and 24 appeared twice. There are two modes in Section B, so it is a **bimodal** distribution.
- In Section C more frequents are 18, 21, and 25. There are three modes for Section C, therefore, it is called a **trimodal or multimodal distribution**

Scores of Section A	Scores of Section B	Scores of Section C
25	25	25
24	24	25
24	24	25
20	20	22
20	18	21
20	18	21
16	17	21
12	10	18
10	9	18
7	7	18

Mode

Properties of the Mode

- It is a quick approximation of the average.
- It may not be unique.
- It is affected by extreme values.
- It is the most unreliable among the three measures of central tendency because its value is undefined in some observations

When to Use the Median

- When the “typical” value is desired.
- When the data set is measured on a nominal scale.

Mode

Advantages of the Mode

- Mode is readily comprehensible and easily calculated
- It is the best representative of data
- It is not at all affected by extreme value.
- The value of mode can also be determined graphically.
- It is usually an actual value of an important part of the series.

Mode

Disadvantages of the Mode

- It is not based on all observations.
- It is not capable of further mathematical manipulation.
- Mode is affected to a great extent by sampling fluctuations.
- Choice of grouping has great influence on the value of mode.

Conclusion

Which one is better: mean, median, or mode?

- Most often, the mean is selected by default
- The mean's key advantage is that it is sensitive to any change in the value of any observation
- The mean's disadvantage is that it is very sensitive to outliers
- We really must consider the nature of the data, the distribution, and our goals to choose properly

Conclusion

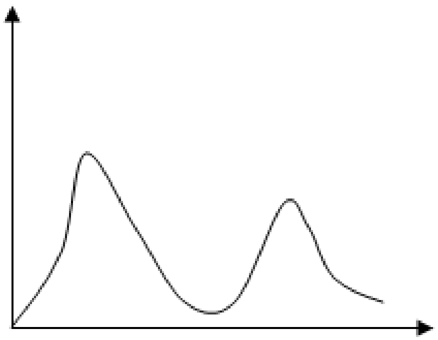
Which one is better: mean, median, or mode? (cont.)

- The **mean** is valid only for **interval** data or **ratio** data.
- The **median** can be determined for **ordinal** data as well as **interval** and **ratio** data.
- The **mode** can be used with **nominal**, **ordinal**, **interval**, and **ratio** data
- **Mode** is the **only** measure of central tendency that can be used with **nominal** data

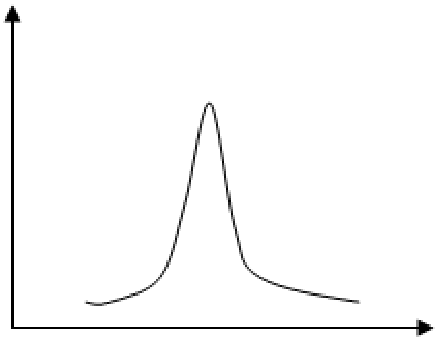
Conclusion

Which one is better: mean, median, or mode? (cont.)

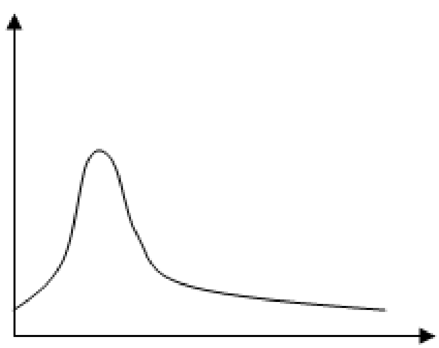
- It also depends on the nature of the distribution



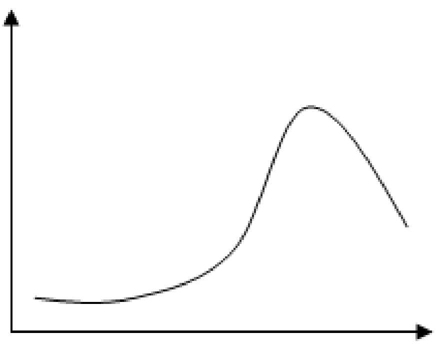
Multi-modal distribution



Unimodal symmetric



Unimodal skewed



Unimodal skewed

Conclusion

Which one is better: mean, median, or mode? (cont.)

- It also depends on your goals
- Consider a company that has nine employees with salaries of 35,000 a year, and their supervisor makes 150,000 a year.
- If you want to describe the typical salary in the company, which statistics will you use?
- I will use mode or median (35,000), because it tells what salary most people get

Conclusion

Which one is better: mean, median, or mode? (cont.)

- It also depends on your goals
- Consider a company that has nine employees with salaries of 35,000 a year, and their supervisor makes 150,000 a year
- What if you are a recruiting officer for the company that wants to make a good impression on a prospective employee?
- The mean is (35,000*9 + 150,000)/10 = 46,500 I would probably say: "The average salary in our company is 46,500" using mean