

Central Tendency & Dispersion

S C AGARKAR

DR V N BEDEKAR INSTITUTE OF MANAGEMENT STUDIES

THANE 400 601

Measures of Central Tendencies

If we have to use data for decision making then we need to look at two characteristics: Central tendency and dispersion.

Central tendency is the middle point of distribution. Measures of central tendencies are also called measures of location.

Three central tendencies are usually studied for a given data. They are Mean, Median and Mode.

Dispersion is the spread of data in a distribution, that is, the extent to which the observations are scattered.



Mean - Average

- Most common measure of central tendency
- Best for making predictions
- Applicable under two conditions:
 1. scores are measured at the interval level, and
 2. distribution is more or less normal [symmetrical].
- Symbolized as: \bar{X}
 - \bar{x} for the mean of a sample
 - μ for the mean of a population



Finding the Mean

- $\bar{X} = (\Sigma X) / N$

- If $X = \{3, 5, 10, 4, 3\}$

$$\bar{X} = (3 + 5 + 10 + 4 + 3) / 5$$

$$= 25 / 5$$

$$= 5$$



Find the Mean

Generator: 01 02 03 04 05 06 07 08 09 10

Days out

of Service: 07 23 04 08 02 12 06 13 09 04

$$\text{Mean } X = (\Sigma X) / N$$

$$= 7+23+4+8+2+12+6+13+9+4 / 10$$

$$= 88/10$$

$$= 8.8$$

Advantages and Disadvantages of Mean

Advantages

1. Its concept is familiar to most people and intuitively clear
2. Every data set has a mean and it is unique for a particular data set.
3. It is useful for performing statistical procedures like comparing the means of several data.

Disadvantages

1. The mean gets affected by extreme values in the data
2. It is tedious to compute as every data point is to be taken into account.
3. We are unable to compute the mean for a data set that has open-ended classes.



Median

- It is the middle-most Value
- 50% of observations are above the Median, 50% are below it
- The difference in magnitude between the observations does not matter
- Therefore, it is not sensitive to outliers
- Formula Median = $n + 1 / 2$



To compute the median

- first you rank order the values of X from low to high: \rightarrow 85, 94, 94, 96, 96, 96, 96, 97, 97, 98
- then count number of observations = 10.
- add 1 = 11.
- divide by 2 to get the middle score \rightarrow the $5 \frac{1}{2}$ score
here 96 is the middle score score



Examples to Solve

■ Find the Median

4 5 6 6 7 8 9 10 12

■ Find the Median

4.2 5.0 9.0 5.1 4.3 4.7 4.8

■ Find the Median

86 31 11 30 49 52 35 43

Advantages and disadvantages of Median

Advantages

1. Extreme values do not affect the median as strongly as they affect mean.
2. The median can be calculated from any kind of data both grouped and ungrouped.
3. Median can be found out even for qualitative data

Disadvantages

1. to find out median we have to arrange data in ascending or descending order.
2. Certain statistical procedures that use the median are more complex in nature

The Mode

The mode is the value that is repeated most often in the data set.

Here is the data of number of delivery trips per day made by Redi-Mix Concrete Plant.

0 2 1 15 19 12 8 7 7 0 1 2 6 5 6 15 4 5 4 15

It is evident from the data that the mode value is 15.

If the data is in grouped format then we find out the modal class.

0-3 trips are 6, 4-7 trips are 8, 8-11 trips are 1 and 12-15 are 5

Hence the modal class is 4-7 trips

Advantages and Disadvantages of Mode

Advantages

1. Mode like median can be used as a central location for qualitative as well as quantitative data.
2. The mode is not unduly affected by extreme values.
3. it can be used even if one or more classes are open ended.

Disadvantages

1. Too often there is no modal value in the data.
2. Some time every value is a mode as it occurs the same number of times.
3. When the data set contain many modes they are difficult to compare.

Using Central Tendencies

The Choice of central tendency depends on the shape of the distribution.

In a single mode symmetrical distribution the values of mean, mode and median are the same.

In a positively skewed distribution the median is to the right of mode and the mean is to the right of both.

In a negatively skewed distribution the mean is to the left of mode and median is in between these two values.

The choice of central tendency also depends on the nature of variable.

For nominal variable Mode is the appropriate measure

For Ordinal variable Median is the appropriate measure

For interval variable Mean is the appropriate measure

Normally Distributed Curve

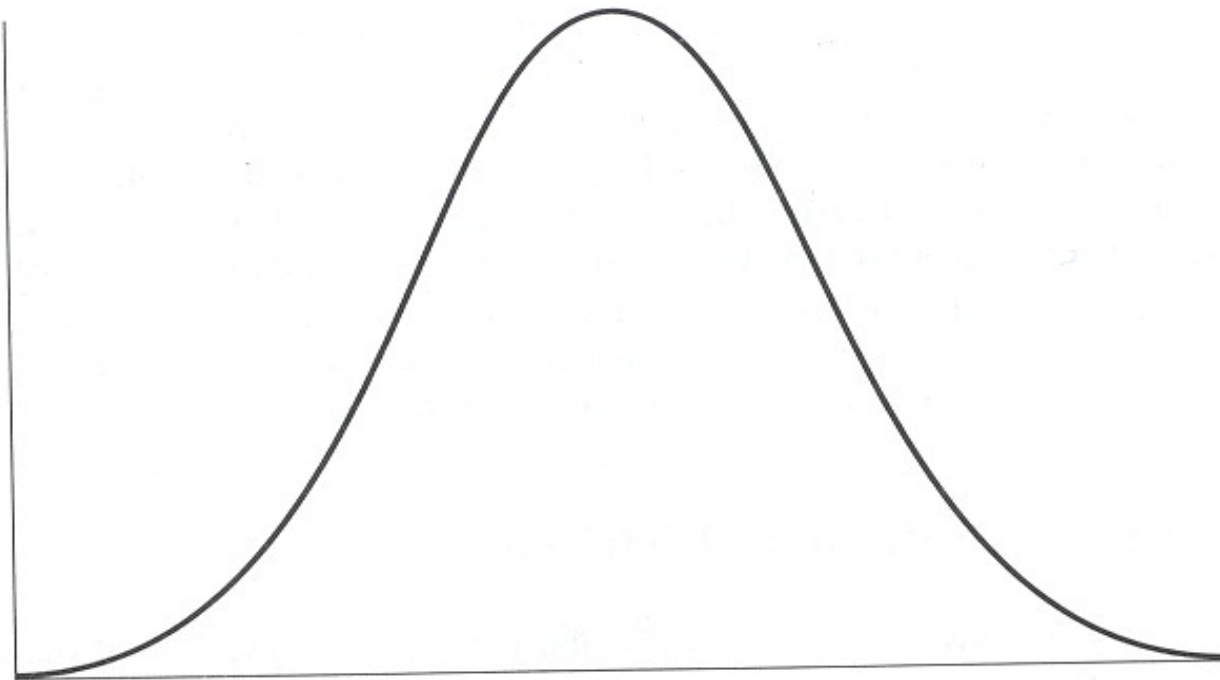
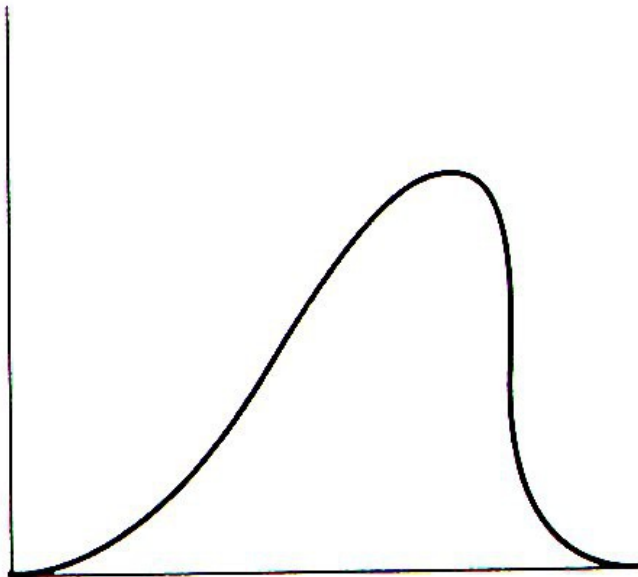
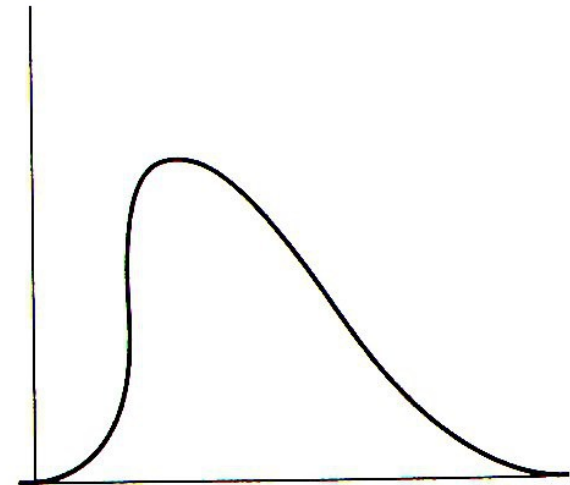


FIGURE 5.4 The Shape of the Normal Curve

Skewed Distributions



(a) Negatively skewed



(b) Positively skewed

Dispersion

Once you determine that the variable of interest is normally distributed, ideally by producing a histogram of the scores, the next question to be asked about the NDC is its dispersion: **how spread out are the scores around the mean.**

Dispersion is a key concept in statistical thinking.

The basic question being asked is how much do the scores deviate around the Mean? The more “bunched up” around the mean the better your ability to make accurate predictions.

Mean Values

Consider these means for weekly candy bar consumption.

$$X = \{7, 8, 6, 7, 7, 6, 8, 7\}$$

$$X = (7+8+6+7+7+6+8+7)/8$$

$$X = 7$$

$$X = \{12, 2, 0, 14, 10, 9, 5, 4\}$$

$$X = (12+2+0+14+10+9+5+4)/8$$

$$X = 7$$





Mean Deviation

The key concept for describing normal distributions and making predictions from them is called **deviation from the mean.**

We could just calculate the average distance between each observation and the mean.

- We must take the absolute value of the distance, otherwise they would just cancel out to zero!

Formula:

$$\sum \frac{|\bar{X} - X_i|}{n}$$



Mean Deviation

Data: 6 10 5 4 9 8

$$\bar{X} = 42 / 6 = 7$$

$X - X_i$	Abs. Dev
$7 - 6$	1
1. $7 - 10$	3
2. $7 - 5$	2
3. $7 - 4$	3
4. $7 - 9$	2
5. $7 - 8$	1
6. Total	12

1. Compute \bar{X} (Average)

1. Compute $X - \bar{X}$ and take the Absolute Value to get Absolute Deviations

1. Sum the Absolute Deviations

1. Divide the sum of the absolute deviations by N



Variance and Standard Deviation

- Instead of taking the absolute value, we square the deviations from the mean. This yields a positive value.
- This will result in measures we call the Variance and the Standard Deviation

Sample-

Population-

s: Standard Deviation

σ : Standard Deviation

s^2 : Variance

σ^2 : Variance



Formulae

Variance

$$s^2 = \frac{\sum (\bar{X} - X_i)^2}{N}$$

Standard Deviation

$$s = \sqrt{\frac{\sum (\bar{X} - X_i)^2}{N}}$$



Illustrative Example

Data: $X = \{6, 10, 5, 4, 9, 8\}$;

X	$X - \bar{X}$	$(X - \bar{X})^2$
-----	---------------	-------------------

$N = 6$
Mean:

$$\bar{X} = \frac{\sum X}{N} = \frac{42}{6} = 7$$

Variance:

$$s^2 = \frac{\sum (\bar{X} - X)^2}{N} = \frac{28}{6} = 4.67$$

Standard Deviation:

$$s = \sqrt{s^2} = \sqrt{4.67} = 2.16$$

Thank You

