
Statistics
organization
scientific
descriptive
collection
interpretation
data
statistical
Inference
mathematical
theory
experiments
science
population
prediction
probability
mathematics
parameters
design
pertaining
discipline
called
often
including
well
analysis
samples
wide
word
also
provides
concerned
inferential
consider
total
statistic
used
referring
branch
experience
application
deduce
part
problems
accounts
considered
social
working
useful
ways
improve
process
soundness
terms
empirical
summarize
successful
element
necessary
using
logically
modeled
expertise
companies
usually
sciences
applications
people
gained
models direct
thinking
planning
inferences
collecting
since
given
larger
roots
wrong
survey
one
confused
theoretical
prove
vital
draw
statistician
art
applied
difference
variety
studied
interpreting
randomness
guiding
based
natural
starts
versed
someone
help
others
moves
organizations
inductive
grouped
observations
provide
mean
disciplines
set
hypothesis
presentation
methods
consultants
particularly
increase
particular
sampling
forecasting
explanation
closely
research
seems
may
focus
probabilities
misleading
new
academic
way
opposite
together
communicating
uncertainty
calculated
quality
plural
rather
fields
method
patterns
true
results
government
direction
deals
number
singular
distinct
subject
predictions
holds
basis
roughly
whose
addition
relevant
predictive
quantity
tools
describe
tested
surveys
related
available
without
advance
leads
a.k.a
applicable
business
access
median

Lecture Notes on Measure of Central Tendency

Lecturer: Dr. Ajijola, L.A.

Learning Outcomes

- Describe the various central and non-central location measures
- Calculate and interpret each of these location measures
- Describe the appropriate central location measure for different data types

Introduction to Statistical Thinking for Decision Making

- Today's good decisions are driven by data.
- In all aspects of our lives, and importantly in the business context, an amazing diversity of data is available for inspection and analytical insight.
- Business managers and professionals are increasingly required to justify decisions on the basis of data.
- They need statistical model-based decision support systems.
- Statistical skills enable them to intelligently collect, analyze and interpret data relevant to their decision-making.
- Statistical concepts and statistical thinking enable them to:
  - solve problems in a diversity of contexts.
  - add substance to decisions.
  - reduce guesswork.

Statistics in Management

- A course in business statistics is part of every management education programme offered today by academic institutions, business schools and management colleges worldwide.

Why?

1. Management Decision Making

The reason lies in the term ‘management decision support systems’. Decision making is central to every manager’s job. Managers must decide, for example,

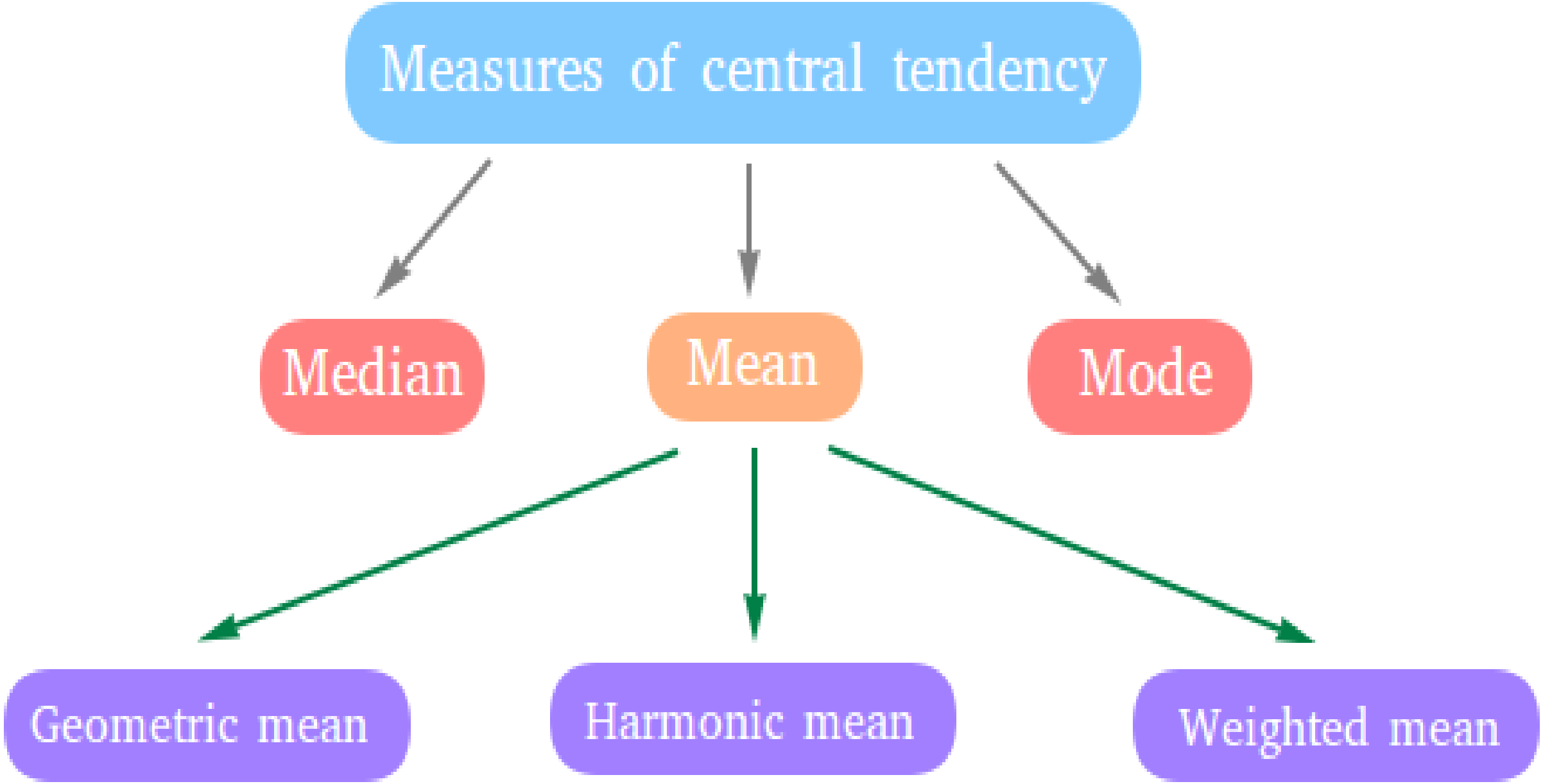
- which advertising media are the most effective;
- who the company’s high-value customers are;
- which machinery to buy;
- whether a consignment of goods is of acceptable quality;
- where to locate stores for maximum profitability;
- and whether females buy more of a particular product than males.

2. Information

To make sound business decisions, a manager needs high-quality information. Information must be timely, accurate, relevant, adequate and easily accessible. However, information to support decision making is seldom readily available in the format, quality and quantity required by the decision maker. More often than not, it needs to be generated from data.

3. Data

What is more readily available – from a variety of sources and of varying quality and quantity – is data. Data consists of individual values that each conveys little useful and usable information to management. Three examples of data are: the purchase value of a single transaction at a supermarket (e.g. ₦2500); the time it takes a worker to assemble a single point (e.g. 7.35 minutes); the brand of cornflakes that a particular consumer prefers (e.g. Kellogs).


Measures of central tendency
Median
Mean
Mode
Geometric mean
Harmonic mean
Weighted mean

Introduction

- The description of statistical data may be quite elaborate or quite brief depending on two factors: the nature of data and the purpose for which the same data have been collected.
- While describing data statistically or verbally, one must ensure that the description is neither too brief nor too lengthy.
- The measures of central tendency enable us to compare two or more distributions pertaining to the same time period or within the same distribution over time.
- For example, the average consumption of tea in two different territories for the same period or in a territory for two years, say, 2003 and 2004, can be attempted by means of an average.
- Often times, we are interested in a typical numerical value to help us describe a data set.
- This typical value is often called an average value or a measure of central tendency.
- We are looking for a single number that is in some sense representative of the complete data set.

- Summary tables and graphs are useful to communicate *broad overviews* of the profiles of random variables.
- However, managers also need specific *numerical measures* (called statistics) to convey more precise information about the behaviour of random variables.
- This is the purpose of numeric descriptive statistics.Three characteristics are commonly used to describe the data profile of a random variable. These are:
  - measures of location (both central and non-central)
  - measures of spread (or dispersion)
  - a measure of shape (skewness).
- **Location** refers to the where the data values are *concentrated*. Central location is a representative ‘middle’ value of concentration of the data, while non-central location measures identify relevant ‘off-centre’ reference points in the data set (such as quartiles).

- **Dispersion** refers to the extent to which the data values are *spread* about the central location value.
- Finally, **skewness** identifies the shape (or degree of *symmetry*) of the data values about the central location measure.
- To illustrate, an electronic goods company has recorded the daily sales (in Naira) over a 12-month trading period.
- The average daily sales are a measure of central location, while the extent to which daily sales vary around the average daily sales would be a measure of dispersion.
- Finally, a measure of skewness would identify whether any very large or very small daily sales values relative to the average daily sales have occurred over this period.

Central Location Measures

- A **central location statistic** is a single number that gives a sense of the 'centrality' of data values in a sample.
- Managers and management reports often make statements containing phrases such as:
  - the average salary per job grade
  - the most popular healthcare plan
  - the average quantity of milk purchased by a household of four persons
  - half of our employees spend less than ₦17000 per month commuting to work
  - the mean age of our employees.
- These statements all refer to a *typical* or *central data value* used to represent where the majority of data values lie. They are called central location statistics.

- Three commonly used central location statistics are:
  a) The Mean (Arithmetic, Geometric and Harmonic Mean)
  b) The Median (also called the second quartile, the middle quartile or the 50th percentile)
  c) The Mode (or modal value).
  d) Fractiles (Quartiles, Deciles, Percentiles and other Fractiles)
- All these measures (mean, median and mode, fractiles) can be used for numeric data, while only the mode is valid for categorical data.

Arithmetic Mean (Average)

- The **arithmetic mean** (or average), $\bar{x}$, lies at the *centre* of a set of numeric data values.
- It is found by adding up all the data values and then dividing the total by the sample size as shown in the following formula:

$$\bar{x} = \frac{\text{Sum of all the observations}}{\text{Number of observations}} = \frac{\sum x_i}{n}$$

Where: $\bar{x}$ = the sample arithmetic mean (average)

n = the number of data values in the sample

x_i = the value of the ith data value of random variable x

$\sum x_i$ = the sum of the n data values, i.e. $x_1 + x_2 + x_3 + x_4 + \ldots + x_n$

Categorization of Data

- Basically we are going to treat measures of central tendency under the following categories:
- Type I: Ungrouped data without frequency
- Type II: Ungrouped data with frequency
- Type III: Grouped data with frequency

Calculating Arithmetic Mean for Ungroup Data without Frequency

- The arithmetic mean of a set of N observations $x_1, x_2, x_3, x_4, \ldots, x_N$ is the denoted by $\bar{x}$ and is defined by:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \ldots + x_n}{N} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{\sum x}{N}$$

- **Examples 1:** Find the average of this data $2, 4, 5, 8, 5, 10, 5, 4$
- **Solution:**

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{2 + 4 + 5 + 8 + 5 + 10 + 5 + 4}{8} = \frac{43}{8} = 5.375$$

Geometric Mean

- The **geometric mean** is used to find the average of *percentage change data*, such as indexes, growth rates or rates of change. When each data value is calculated from a different base, then the appropriate measure of central location is the geometric mean.
- **<u>Calculating Geometric Mean for Ungroup Data without Frequency</u>**
- The geometric mean GM or $\bar{x}_G$ of a set of n observation $x_1 , x_2 , x_3 , x_4 , \dots , x_n$ is the nth root of the product of the numbers. The formula for the geometric mean (GM) is as follows:

$$GM = \bar{x}_G = \sqrt[n]{x_1 \times x_2 \times x_3 \times x_4 \times \ldots \times x_n}$$

$$= (x_1 \times x_2 \times x_3 \times x_4 \times \cdots \times x_n)^{1/n} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

- **Examples 2:** Find the Geometric Mean (GM) of this data $2, 4, 5, 8, 5, 10, 5, 4$
- **Solution**

$$\bar{x}_G = \sqrt[8]{2 \times 4 \times 5 \times 8 \times 5 \times 10 \times 5 \times 4} = \sqrt[8]{320{,}000} = 4.88$$

Harmonic Mean

- The harmonic mean H or $\bar{x}_H$ of a set of n numbers $x_1, x_2, x_3, x_4, \ldots, x_n$ is the reciprocal of the arithmetic mean of the reciprocals of the numbers. The formula is:

$$H = \bar{x}_G = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Calculating Harmonic Mean for Ungroup Data without Frequency

- **Examples 3:** Find the Harmonic Mean (H) of this data $2, 4, 5, 8, 5, 10, 5, 4$
- **Solution**

$$\bar{x}_G = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{8}{\frac{1}{2} + \frac{1}{4} + \frac{1}{5} + \frac{1}{8} + \frac{1}{5} + \frac{1}{10} + \frac{1}{5} + \frac{1}{4}} = 4.38$$

Extreme Value

- Extreme value are those value that stand out ***conspicuously*** away from other value in a given data. For example, if you are asked to calculate the arithmetic mean of the following sets of data 2,4,5,10,5,105. The AM is:

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{2+4+5+10+5+105}{6} = \frac{131}{6} = 21.8$$

- We can see that this value is not among the given that and this is due to extreme value 105
- If the extreme value is very high, Harmonic Mean will be very suitable.
- If the extreme value is moderately high, Geometric Mean will be suitable.
- If there is no extreme value, use Arithmetic Mean.

Calculating Arithmetic Mean for Ungroup Data with Frequency

- If the numbers $x_1\ ,\ x_2\ ,\ x_3\ ,\ x_4\ , \ldots , x_n$ occur $f_1\ ,\ f_2\ ,\ f_3\ ,\ f_4\ , \ldots , f_n$ times, respectively (i.e., occur with frequencies $f_1\ ,\ f_2\ ,\ f_3\ ,\ f_4\ , \ldots , f_n$) the arithmetic mean is

$$\bar{x} = \frac{f_1x_1 \ + \ f_2x_2 \ + \ f_3x_3 \ + \ f_4x_4 \ + \ \ldots \ + \ f_nx_n}{f_1 + \ f_2 + \ f_3 + \ f_4 + \ \ldots + \ f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum fx}{\sum f} = \frac{\sum fx}{N}$$

- **Examples 4:** Find the arithmetic mean of the data in the table below:

x	2	4	5	8	10
f	1	2	3	1	1

- **Solution**

x	f	fx
2	1	2
4	2	8
5	3	15
8	1	8
10	1	10
	$\sum f = 8$	$\sum fx = 43$

$$\therefore \bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{43}{8} = 5.375$$

Calculating Geometric Mean for Ungroup Data with Frequency

- The formula for Geometric mean is:

$$\bar{x}_G = G = \sqrt[\sum_{i=1}^{n} f_i]{x_1^{f_1} \times x_2^{f_2} \times x_3^{f_3} \times \cdots \times x_n^{f_n}}$$

$$= \left(x_1^{f_1} \times x_2^{f_2} \times x_3^{f_3} \times \cdots \times x_n^{f_i}\right)^{\frac{1}{\sum_{i=1}^{n} f_i}}$$

Example 5: Use the data in example 4 to find the Geometric mean

x	2	4	5	8	10
f	1	2	3	1	1

Solution

$$\bar{x}_G = (2^1 \times 4^2 \times 5^3 \times 8^1 \times 10^1)^{\frac{1}{8}} = 4.88$$

Calculating Harmonic Mean for Ungroup Data with Frequency

- The formula for Harmonic mean is:

$$H = \bar{x}_H = \frac{f_1 + f_2 + f_3 + \cdots + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \frac{f_3}{x_3} + \cdots + \frac{f_n}{x_n}} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$

Example 6: Use the data in example 4 to find the Harmonic mean

x	2	4	5	8	10
f	1	2	3	1	1

Solution

$$\bar{x}_G = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}} = \frac{1 + 2 + 3 + 1 + 1}{\frac{1}{2} + \frac{2}{4} + \frac{3}{5} + \frac{1}{8} + \frac{1}{10}} = 4.38$$

Calculating Arithmetic Mean for Grouped Data with Frequency

- The formula for calculating A.M. for group data with frequency is

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum fx}{\sum f}$$

Where $f =$ is the frequency

$x =$ the mid-point of the class limit or class boundary

$n =$ the upper limit of the summation.

Example 7:Find the arithmetic mean in the table below:

Class	$5-10$	$10-15$	$15-20$	$20-25$	$25-30$	$30-35$	$35-40$	$40-45$	$45-50$
f	3	5	8	10	12	7	6	2	1

- **Solution**

Class	f	$Mid\ Point\ (x)$	fx
$5-10$	3	7.5	22.5
$10-15$	5	12.5	62.5
$15-20$	8	17.5	140
$20-25$	10	22.5	225
$25-30$	12	27.5	330
$30-35$	7	32.5	227.5
$35-40$	6	37.5	225
$40-45$	2	42.5	85
$45-50$	1	47.5	47.5
	$\sum f = 54$		$\sum fx = 1365$

$$\therefore \bar{x} = \frac{\sum fx}{\sum f} = \frac{1365}{54} = 25.28$$

Advantages and Disadvantages of Arithmetic Mean

The arithmetic mean has the following two *advantages*.

- It uses all the data values in its calculation.
- It is an unbiased statistic (meaning that, on average, it represents the true mean).

These two properties make it the most widely used measure of central location.

The arithmetic mean, however, has two *drawbacks*.

- It is not appropriate for categorical (i.e. nominal or ordinal-scaled) data. For example, it is not meaningful to refer to the 'average' colour of cars or 'average' preferred brand or 'average' gender. The arithmetic mean can only be applied to numeric (i.e. interval and ratio-scaled) data.
- It is distorted by **outliers**. An outlier is an extreme value in a data set. For example, the mean of 3, 4, 6 and 7 is 5. However, the mean of 3, 4, 6 and 39 is 13, which is not representative of the majority of the data values.

These two drawbacks require that other measures of central location be considered. Two alternative central location measures are the median and the mode.

The Mode

- The **mode** (M_o) is defined as the *most frequently occurring value* in a set data. It can be calculated both for categorical data and numeric data. The mode of a set of numbers is that value which occurs with the greatest frequency (i.e. it is the most common value. The mode may not exist and even if it does, it may not be unique.
- The following are illustrative statements that refer to the mode as the central location measure:
  - Colgate is the brand of toothpaste most preferred by households.
  - The most common family size is four.
  - The supermarket frequented most often in Kimberley is Checkers.
  - The majority of machine breakdowns last between 25 and 30 minutes.

Steps

- Follow these steps to calculate the mode:
  - For small samples of ungrouped data, rank the data from lowest to highest, and identify, by inspection, the data value that occurs most frequently.
  - For large samples of discrete or categorical (nominal and ordinal-scaled) data:
    - construct a categorical frequency table
    - identify the modal value or modal category that occurs most frequently.

Calculating Mode for Ungroup Data without Frequency

- **Examples 8:** Find the mode of the following:
  1. The set 2,2,5,7,9,9,9,10,10,11,12 and 18 has 9 as it's mode.
  2. The set 3,5,8,10,12,15 and 16 has no mode.
  3. The set 2,3,4,4,4,5,5,7,7,7 and 9 has two modes, 4 and 7 and it is called "Bimodal". A distribution with only one mode is called "Unimodal".
- **Calculating Mode for Ungroup Data with Frequency**
- Find the mode in the table below

x	28	30	32	33	35	38	40
f	1	2	2	6	2	1	1

- The mode is 33

Calculating Mode for Grouped Data with Frequency

- The formula for calculating mode for grouped data with frequency is

$$Mode = M_o = X_L + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] c$$

Where X_L = Lower class boundary of the modal class

f_1 = frequency of the modal class

f_0 = frequency of the pre-modal class

f_2 = frequency of the post-modal class

c = class width or class size

Another formula is

$$Mode = M_o = X_L + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] c$$

Where $\Delta_1 = f_1 - f_0$

$\Delta_2 = f_1 - f_2$

Example 9

- Find the mode for the data in the table below:

Class	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50
f	3	5	8	10	12	7	6	2	1

- **Solution**

Class boundary	Frequency (f)
5 – 10	3
10 – 15	5
15 – 20	8
20 – 25	10
25 – 30	12
30 – 35	7
35 – 40	6
40 – 45	2
45 – 50	1

$$Mode = M_o = X_L + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] c$$

Here $X_L = 25;\quad f_1 = 12;\quad f_0 = 10;$

$f_2 = 7;\ c = 5$

$$\therefore M_o = 25 + \left[\frac{12-10}{2\times12-10-7}\right] \times 5 = \mathbf{26.43}$$

Advantages and Disadvantages of Mode

The mode has several *advantages:*

- It is a valid measure of central location for *all data types* (i.e. categorical and numeric). If the data type is categorical, the mode defines the most frequently occurring category. If the data type is numeric, the mode is the most frequently occurring data value (or the midpoint value of a modal interval, if the numeric data has been grouped into intervals).
- The mode is not influenced by outliers, as it represents the most frequently occurring data value (or response category).

The mode also has one main *disadvantage:*

- It is a representative measure of central location only if the histogram of the numeric random variable is unimodal (i.e. has one peak only). If the shape is bi-modal, there is more than one peak, meaning that two possible modes exist, in which case there is no single representative mode.

The Median

- The **median** (M_e) is the *middle number* of an ordered set of data. It divides an ordered set of data values into two equal halves (i.e. 50% of the data values lie below the median and 50% lie above it).

- ***Follow these steps to calculate the median for ungrouped (raw) numeric data:***
  - Arrange the n data values in ascending order.
  - Find the median by first identifying the middle position in the data set as follows:
    - If n is odd, the median value lies in the $\left(\frac{n+1}{2}\right)^{th}$ position in the data set.
    - If n is even, the median value is found by identifying the $\left(\frac{n}{2}\right)^{th}$ position and then averaging the data value in this position with the next consecutive data value.

Calculating Median for Ungroup Data without Frequency

- **Examples 10:** Find the median of the following sets of data

1. $2, 4, 5, 8, 5, 10, 5, 4$
2. $8,\ 3,\ 5,\ 12,$ and 10
3. $6, 4, 1, 9, 8, 3, 7, 10, 5$

Solutions

1. Rearrange the data $2, 4, 4, \mathbf{5}, \mathbf{5}, 5, 8, 10$. Therefore the median is $\frac{5+5}{2} = 5$
2. Rearrange the data $3, 5, \mathbf{8}, 10, 12$. Therefore the median is 8
3. Rearrange the data $1, 3, 4, 5, \mathbf{6}, 7, 8, 9, 10$. Therefore the median is 6

Calculating Median for Ungroup Data with Frequency

- **Example 11:** Find the median in the table below

$Age(x)$	3	5	7	9	10
$No\ of\ Children(f)$	2	3	4	5	6

- Solution: First and foremost, find the cumulative frequency (cf)

x	f	cf
3	2	2
5	3	5
7	4	9
9	5	**14**
10	6	20
	$n = \sum f = 20$	

Find the $\left(\frac{n+1}{2}\right)^{th}$ in the $cf = \left(\frac{20+1}{2}\right)^{th}$

$= 10.5^{th}$ position

Locate the place where 10.5 first fall in the cf against $x\ (Age)$

Calculating Median for Grouped Data with Frequency

- Use these methods when the data is already summarised into a numeric frequency distribution (or ogive).
- ***Graphical approach***

  Arithmetic approach: Steps to follow in using the Arithmetic/Analytical formula

  - Based on the sample size, n, calculate $\frac{n}{2}$ to find the median position (i.e. compute half of the frequency)
  - Compute the cumulative frequency
  - Use the cumulative frequency to identify the group where the median belongs (i.e. the median class
  - Compute the lower class boundary
  - Apply the formula
- ***Note: You must convert class limit to class boundary in order to calculate median.***

- **The Arithmetic/Analytical formula is:**
- $Median = M_e = X_L + \left[\frac{\frac{1}{2}\sum f - f_0}{f_{me}}\right] c$
- Where X_L = Lower class boundary of the median class

  f_0 = Cumulative frequency before the median class

  f_{me} = frequency of the median class

  c = class width or class size
- **Example 12**
- Find the Median for the data in the table below

Class	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50
f	3	5	8	10	12	7	6	2	1

- Solution

$Class$	f	cf
$5-10$	3	3
$10-15$	5	8
$15-20$	8	16
$20-25$	10	26
$25-30$	12	38
$30-35$	7	45
$35-40$	6	51
$40-45$	2	53
$45-50$	1	54
	$\sum f = 54$	

Calculate half of the frequency $\frac{1}{2}\sum fx = \frac{1}{2} \times 54 = 27$

The median class is where 27 first fall i.e. $25-30$

From the median class $X_L = 25; f_0 = 26; f_{me} = 12; and\ c = 5.$

$$\therefore Median = X_L + \left[\frac{\frac{1}{2}\sum f - f_0}{f_{me}}\right] c = 25 + \left[\frac{\frac{1}{2}\times 54 - 26}{12}\right] \times 5$$

$$= 25.42$$

Advantages and Disadvantages of Median

- **The median has one major *advantage* over the mean** – it is not affected by outliers. It is therefore a more representative measure of central location than the mean when significant outliers occur in a set of data.
- **A *drawback* of the median**, however, is that it cannot be calculated for categorical data. It makes no sense, for example, to refer to a ‘middle’ brand of fuel types. Thus a median, like the mean, can also only be applied to numeric data.

Which central location measure is best?

- A central location measure must be *representative* of its data values. The choice depends mainly on (a) the data type of the random variable being analysed and (b) whether outliers are present or not in the data set.
- ***Data type:*** If the data type is *categorical* (nominal or ordinal scaled), then the *mode* is the only valid and representative measure of central location. All three measures (mean, median and mode) can, however, be used for *numeric* (interval or ratio-scaled) data.
- ***Outliers:*** Outliers distort the mean but do not affect the median or the mode. Thus, if outliers are detected in a set of data, then the *median* (or mode) should be chosen. In such cases, the median is preferred to the mode as it can be used in further analysis. However, if there are good reasons to remove the outlier(s) from the data set (because they are errors or atypical data values), then the mean can again be used as the best central location measure.

Non-central Location Measures (Fractiles)

- **Quartiles: Quartiles** are non-central measures that divide an ordered data set into quarters (i.e. four equal parts).
  - The *lower quartile*, Q_1, is that data value that separates the lower (bottom) 25% of (ordered) data values from the top 75% of ordered data values.
  - The *middle quartile*, Q_2, is the median. It divides an ordered data set into two equal halves.
  - The *upper quartile*, Q_3, is that data value that separates the top (upper) 25% of (ordered) data values from the bottom 75% of ordered data values.
- Quartiles are calculated in a similar way to the median. The only difference lies in the identification of the quartile position, and consequently in the choice of the quartile interval.

- Follow these steps to calculate quartiles (lower, middle and upper) for *Grouped data with frequency:*
- Sort the data in ascending order.
- Each quartile position is determined as follows by the formula below (regardless of whether *n* is even or odd):

$$Q_1 = X_L + \left[\frac{\frac{1}{4}\sum f - f_0}{f_{Q_1}}\right]c$$

Where $X_L =$ Lower class boundary of the Q_1 class

$f_0 =$ Cumulative frequency before the Q_1 class

$f_{Q_1} =$ frequency of the Q_1 class

$c =$ class width or class size

$$Q_2 = Median = X_L + \left[\frac{\frac{1}{2}\sum f - f_0}{f_{Q_2}}\right]c$$

Where X_L = Lower class boundary of the Q_2 class

f_0 = Cumulative frequency before the Q_2 class

f_{Q_1} = frequency of the Q_2 class

c = class width or class size

$$Q_3 = X_L + \left[\frac{\frac{3}{4}\sum f - f_0}{f_{Q_3}}\right]c$$

Where X_L = Lower class boundary of the Q_3 class

f_0 = Cumulative frequency before the Q_3 class

f_{Q_1} = frequency of the Q_3 class

c = class width or class size

- **Example 13:** Use the information in example 12 to calculate the Q_1
- **Solution**

$Class$	f	Cf
$5-10$	3	3
$10-15$	5	8
$15-20$	8	16
$20-25$	10	26
$25-30$	12	38
$30-35$	7	45
$35-40$	6	51
$40-45$	2	53
$45-50$	1	54
	$\sum f = 54$	

Compute $\frac{1}{4}$ of the frequency i.e. $\frac{1}{4}\sum f = \frac{1}{4} \times 54 = 13.5$.

This mean the Q_1 class is $15-20$

Since $Q_1 = X_L + \left[\frac{\frac{1}{4}\sum f - f_0}{f_{Q_1}}\right] c$. It implies that $X_L = 15$;

$f_0 = 8; \quad f_{Q_1} = 8; and\ c = 5$

$\therefore Q_1 = 15 + \left[\frac{\frac{1}{4}\times 54 - 8}{8}\right] \times 5 = 18.44$

DECILES

- **Deciles** are non-central measures that divide an ordered data set into ten (10) equal parts.

  $D_1 =$ first deciles

  $D_2 =$ second deciles

  $D_3 =$ third deciles

  .

  .

  .

  $D_9 =$ ninth deciles

- Deciles are calculated in a similar way to the median. The only difference lies in the identification of the decile position, and consequently in the choice of the quartile interval.

- Follow these steps to calculate deciles for *Grouped data with frequency:*
  - Sort the data in ascending order.
  - Each quartile position is determined as follows by the formula below (regardless of whether *n* is even or odd):

$$D_1 = X_L + \left[\frac{\frac{1}{10}\sum f - f_0}{f_{D_1}}\right]c$$

Where X_L = Lower class boundary of the D_1 class

f_0 = Cumulative frequency before the D_1 class

f_{D_1} = frequency of the D_1 class

c = class width or class size

$$D_2 = X_L + \left[\frac{\frac{2}{10}\sum f - f_0}{f_{D_2}}\right]c$$

Where X_L = Lower class boundary of the D_2 class

f_0 = Cumulative frequency before the D_2 class

f_{D_2} = frequency of the D_2 class

c = class width or class size

$$D_5 = Median = X_L + \left[\frac{\frac{5}{10}\sum f - f_0}{f_{D_5}}\right]c$$

Where X_L = Lower class boundary of the D_5 class

f_0 = Cumulative frequency before the D_5 class

f_{D_5} = frequency of the D_5 class

c = class width or class size

$$D_9 = X_L + \left[\frac{\frac{9}{10}\sum f - f_0}{f_{D_9}}\right]c$$

Where X_L = Lower class boundary of the D_9 class

f_0 = Cumulative frequency before the D_9 class

f_{D_9} = frequency of the D_9 class

c = class width or class size

- **Example 14**: Find the value of D_7 by using the information in the table below:

Class	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50
f	3	5	8	10	12	7	6	2	1

- **Solution**

Compute $\frac{7}{10}$ of the frequency i.e. $\frac{7}{10}\sum f = \frac{7}{10} \times 54 = 37.8$. This mean the D_7 class is $35 - 40$

$Class$	f	Cf
$5 - 10$	3	3
$10 - 15$	5	8
$15 - 20$	8	16
$20 - 25$	10	26
$25 - 30$	12	38
$30 - 35$	7	45
$35 - 40$	6	51
$40 - 45$	2	53
$45 - 50$	1	54
	$\sum f = 54$	

Since $D_7 = X_L + \left[\frac{\frac{7}{10}\sum f - f_0}{D_7}\right] c$. It implies that $X_L = 35$;

$f_0 = 45; \quad f_{D_7} = 6; and\ c = 5$

$\therefore D_7 = 35 + \left[\frac{\frac{7}{10} \times 54 - 45}{6}\right] \times 5 = 29$

Percentile

- **Percentile** are non-central measures that divide an ordered data set into 100 equal parts.

  The $P_{10} = 10^{th}$ percentile

  The $P_{20} = 20^{th}$ percentile

  The $P_{50} = 50^{th}$ percentile = Median

  .

  .

  .

  The $P_{99} = 99^{th}$ percentile

- Percentiles are calculated in a similar way to the median. The only difference lies in the identification of the percentiles position, and consequently in the choice of the percentile interval.

- Follow these steps to calculate percentile for *ungrouped* (*raw*) *data:*
  - Sort the data in ascending order.
  - Each percentile position is determined as follows by the formula below (regardless of whether *n* is even or odd):

$$P_{10} = X_L + \left[\frac{\frac{10}{100}\sum f - f_0}{f_{P_{10}}}\right]c$$

Where $X_L =$ Lower class boundary of the P_{10} class

$f_0 =$ Cumulative frequency before the P_{10} class

$f_{P_{10}} =$ frequency of the P_{10} class

$c =$ class width or class size

$$P_{20} = X_L + \left[\frac{\frac{20}{100}\sum f - f_0}{f_{P_{20}}}\right] c$$

Where X_L = Lower class boundary of the P_{20} class

f_0 = Cumulative frequency before the P_{20} class

$f_{P_{20}}$ = frequency of the P_{20} class

c = class width or class size

$$P_{50} = Median = X_L + \left[\frac{\frac{50}{100}\sum f - f_0}{f_{P_{50}}}\right] c$$

Where X_L = Lower class boundary of the P_{50} class

f_0 = Cumulative frequency before the P_{50} class

$f_{P_{50}}$ = frequency of the P_{50} class

c = class width or class size

- **Example 15**: Use the data in example 14 to calculate 20th percentile.
- **Solution**

Compute $\frac{20}{100}$ of the frequency i.e. $\frac{20}{100}\sum f = \frac{20}{100} \times 54 = 10.8.$

$Class$	f	Cf
$5-10$	3	3
$10-15$	5	8
$15-20$	8	16
$20-25$	10	26
$25-30$	12	38
$30-35$	7	45
$35-40$	6	51
$40-45$	2	53
$45-50$	1	54
	$\sum f = 54$	

This mean the P_{20} class is $10-40$

Since $P_{20} = X_L + \left[\frac{\frac{20}{100}\sum f - f_0}{P_{20}}\right] c.$

It implies that $X_L = 10;\ f_0 = 3;\ \ f_{P_{20}} = 5; c = 5$

$$\therefore P_{20} = 10 + \left[\frac{\frac{20}{100} \times 54 - 3}{5}\right] \times 5 = 17.8$$

Exercises

1. Find the arithmetic mean of this data 8, 3, 5, 12, and 10
2. The grades of students on six examinations were 84, 91, 72, 68, 87 and 78. Find the arithmetic mean of the grades.
3. A student's final grades in mathematics, physics, English and Home Economics are, respectively, $82, 86, 90$ and 70. If the respective credits received for these courses are $3, 5, 3$ and 1, determine an appropriate average grade.
4. Find the arithmetic mean if $5, 8, 6$, and 2 occur with $3, 2, 4$ and 1, respectively.
5. The number of seminar training days attended last year by 20 financial advisors is shown in observations below:
   16, 20, 13, 19, 24, 22, 18, 18, 15, 20, 21, 21, 18, 20, 18, 20, 15, 20, 18, 20.
   What is the average number of training days attended by these financial advisors?

6. The electricity tariff has increased by 12%, 8% and 16% per annum over a three-year period. Find the average annual percentage increase in the electricity tariff.
7. The harmonic mean of the numbers $2, 4$ and 8 is?

More Exercises

8. The data below represent the monthly car sale of a company

(a)Find the median number of cars sold per month by a dealer over the past nine months based on the following monthly sales figures: 27, 38, 12, 34, 42, 40, 24, 40, 23

(b)Find the median number of cars sold per month by a dealer over the past ten months based on the following monthly sales figures: 27, 38, 12, 34, 42, 40, 24, 40, 23,18.

9. Find the median, first quartile and second quartile using data in the table below

Age	3	5	7	9	10
No of Children	2	3	4	5	6

10. A courier company recorded 30 delivery times (in minutes) to deliver parcels to their clients from its depot. The data is summarised in the numeric frequency distribution in Table below: Find the Arithmetic mean, median, mode, Q_1, D_6, and P_{70}.

Time Interval	$10-20$	$20-30$	$30-40$	$40-50$	$50-60$
Frequency count	3	5	9	7	6

More Exercises

11. The daily electricity consumption in kilowatt hours (kWh) by a sample of 20 households in Rustenburg is recorded in Table below.

(a)Find the lower quartile value of daily household electricity consumption.

(b)Find the upper quartile value of daily household electricity consumption.

(c)Find the third decile value of daily household electricity consumption.

(d)Find the eight decile value of daily household electricity consumption.

(e)Find the 25th percentile value of daily household electricity consumption.

(f) Find the 75th percentile value of daily household electricity consumption.

Table: Daily household electricity consumption (kWh)

58	50	33	51	38	43	60	55
51	47	40	37	43	48	61	55