QMS 101 Introductory Statistics

Topic III: Measures of Central Tendency

Dr. Anna Fome

Jordan University College

2026-05-16

Learning Outcomes

By the end of this topic, students will be able to:

- Explain the meaning and purpose of a measure of central tendency
- Calculate the **arithmetic mean** for ungrouped and grouped data
- Find the **median** and **mode** for ungrouped and grouped data
- Compute the **geometric mean** and **harmonic mean** and know when to use each
- Find **quartiles**, **deciles**, and **percentiles** from a dataset
- Choose the most appropriate measure for a given situation and **interpret** every result in plain language

Our Running Dataset — Used Throughout This Topic

The Study Context

A farm manager in a rural village wants to estimate the typical monthly expenditure on farm inputs (such as seeds, fertilizers, and pesticides) among small-scale farmers.

The raw data she collected **(TZS '000)** from 30 farmers were:

45	32	55	38	47	60	33	50	42	38
55	47	30	65	42	50	38	55	47	44
52	36	48	58	41	53	37	61	43	49

We will use the following frequency distribution:

Class (TZS '000)	Class boundary	f	Midpoint x	F
30 – 38	29.5 - 38.5	7	34	7
39 – 47	38.5 - 47.5	10	43	17
48 – 56	47.5 - 56.5	7	52	24
57 – 65	56.5 - 65.5	6	61	30
Total		**30**		

Tip

Every measure in this topic will be calculated from this same dataset.
You will see how each measure tells a different part of the same story.

3.1 What is a Measure of Central Tendency?

The Question Every Decision-Maker Asks

Real-Life Scenario

A farm manager in a rural village wants to estimate the typical monthly expenditure on farm inputs (such as seeds, fertilizers, and pesticides) among small-scale farmers.

- She has collected spending data from 30 farmers.
- However, she cannot present all 30 values to the finance officer.
- Instead, she needs one representative value that best describes a typical farmer's monthly spending.

  *This single representative value is called a **measure of central tendency**.*

The Question Every Decision-Maker Asks

Real-Life Scenario

A farm manager in a rural village wants to estimate the typical monthly expenditure on farm inputs (such as seeds, fertilizers, and pesticides) among small-scale farmers.

- She has collected spending data from 30 farmers.
- However, she cannot present all 30 values to the finance officer.
- Instead, she needs one representative value that best describes a typical farmer's monthly spending.
  This single representative value is called a ***measure of central tendency****.*

Definition

A **measure of central tendency** is a single numerical value that represents the *centre* or *typical value* of a dataset. It summarises an entire distribution into one meaningful number.

Five Measures — One Concept, Five Answers

The same question — *"What is the typical monthly spending?"* — can be answered five different ways:

Five Measures — One Concept, Five Answers

The same question — *"What is the typical monthly spending?"* — can be answered five different ways:

Measure	What It Finds	Best Situation
Arithmetic Mean	Balancing point of all values	Symmetric data, no extreme outliers
Median	The exact middle value	Skewed data or when outliers are present
Mode	The most common value	Categorical data; finding the most popular outcome
Geometric Mean	Average of growth rates or ratios	Growth rates, index numbers, ratios
Harmonic Mean	Average of rates	Speed, price per unit, productivity rates

Five Measures — One Concept, Five Answers

The same question — *"What is the typical monthly spending?"* — can be answered five different ways:

Measure	What It Finds	Best Situation
Arithmetic Mean	Balancing point of all values	Symmetric data, no extreme outliers
Median	The exact middle value	Skewed data or when outliers are present
Mode	The most common value	Categorical data; finding the most popular outcome
Geometric Mean	Average of growth rates or ratios	Growth rates, index numbers, ratios
Harmonic Mean	Average of rates	Speed, price per unit, productivity rates

Warning

No single measure is always correct. Part of learning statistics is knowing **which measure to choose** — and being able to **explain why**.

3.2 The Arithmetic Mean

Definition and Formula

The **arithmetic mean** ($\bar{x}$) is calculated by adding all values and dividing by the number of observations.

Definition and Formula

The **arithmetic mean** ($\bar{x}$) is calculated by adding all values and dividing by the number of observations.

Ungrouped data:

$$\bar{x} = \frac{\sum x}{n}$$

$\sum x$ = sum of all values
n = number of observations

Grouped data (frequency table):

$$\bar{x} = \frac{\sum fx}{\sum f}$$

f = class frequency
x = class **midpoint** (not limits!)
$\sum f = n$

Definition and Formula

The **arithmetic mean** ($\bar{x}$) is calculated by adding all values and dividing by the number of observations.

Ungrouped data:

$$\bar{x} = \frac{\sum x}{n}$$

$\sum x$ = sum of all values
n = number of observations

Grouped data (frequency table):

$$\bar{x} = \frac{\sum fx}{\sum f}$$

f = class frequency
x = class **midpoint** (not limits!)
$\sum f = n$

Warning

Critical rule: For grouped data, always use the class **midpoint** as x.

The midpoint $= \dfrac{\text{lower limit} + \text{upper limit}}{2}$

Never use the class limits themselves.

Arithmetic Mean — Ungrouped (from raw data)

Using the 30 raw values of monthly farm input expenditure (TZS '000):

$$\sum x = 45 + 32 + 55 + 38 + 47 + 60 + 33 + 50 + 42 + 38+$$

$$55 + 47 + 30 + 65 + 42 + 50 + 38 + 55 + 47 + 44 + 52 + 36+$$

$$48 + 58 + 41 + 53 + 37 + 61 + 43 + 49$$

. . .

$$\bar{x} = \frac{\sum x}{x} = \frac{1{,}391}{30} = \textbf{TZS 46,367}$$

Arithmetic Mean — Ungrouped (from raw data)

Using the 30 raw values of monthly farm input expenditure (TZS '000):

$$\sum x = 45 + 32 + 55 + 38 + 47 + 60 + 33 + 50 + 42 + 38 +$$

$$55 + 47 + 30 + 65 + 42 + 50 + 38 + 55 + 47 + 44 + 52 + 36 +$$

$$48 + 58 + 41 + 53 + 37 + 61 + 43 + 49$$

. . .

$$\bar{x} = \frac{\sum x}{x} = \frac{1{,}391}{30} = \textbf{TZS 46,367}$$

Interpretation

- The average monthly expenditure on farm inputs among the 30 farmers is approximately **TZS 46,367**.
- This means that a typical small-scale farmer spends about this amount per day on agricultural inputs such as seeds, fertilizers, and pesticides.
- A farm manager or policy planner could use this value as a benchmark for planning input subsidies or estimating average production costs in the area.

Arithmetic Mean — Grouped Data

Using the frequency distribution:

Class	f	Midpoint x	fx
30 – 38	7	34	238
39 – 47	10	43	430
48 – 56	7	52	364
57 – 65	6	61	366
Total	**30**		$\sum fx = 1{,}398$

Arithmetic Mean — Grouped Data

Using the frequency distribution:

Class	f	Midpoint x	fx
30 – 38	7	34	238
39 – 47	10	43	430
48 – 56	7	52	364
57 – 65	6	61	366
Total	**30**		$\sum fx = 1{,}398$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{1{,}398}{30} = \textbf{TZS 46,600}$$

Arithmetic Mean — Grouped Data

Using the frequency distribution:

Class	f	Midpoint x	fx
30 – 38	7	34	238
39 – 47	10	43	430
48 – 56	7	52	364
57 – 65	6	61	366
Total	**30**		$\sum fx = 1{,}398$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{1{,}398}{30} = \textbf{TZS 46,600}$$

Why the slight difference? (The ungrouped mean = TZS 46,367. The grouped mean = TZS 46,600.)

Because we assumed every observation in a class sits at the **midpoint** — an approximation. The two values are very close, confirming our class width was appropriate.

Practice Question 3.1 — Arithmetic Mean

Try This

Q1A student recorded the number of **hours slept per night** for 10 classmates:

$$7, \quad 6, \quad 8, \quad 5, \quad 7, \quad 9, \quad 6, \quad 7, \quad 8, \quad 6$$

(a) Calculate the arithmetic mean number of hours slept.
(b) The recommended sleep for a university student is 8 hours per night.
By how much does the average student in this group fall short of the recommendation?
Q2 Using the grouped data on small-scale farmers' monthly farm input expenditure, the mean is about TZS 46.6 thousand.
(a) A farmer spends TZS 30,000 per day. How does this compare to the average, and what does it suggest?
(b) A farmer spends TZS 60,000 per day. How does this compare to the average, and what does it suggest?

3.3 The Median

Definition

Definition

The **median** is the *middle value* when data is arranged in ascending order.

- Exactly **50%** of values lie below the median
- Exactly **50%** of values lie above the median

Definition

Definition

The **median** is the *middle value* when data is arranged in ascending order.

- Exactly **50%** of values lie below the median
- Exactly **50%** of values lie above the median

For ungrouped data:

Situation	Position of median
n is **odd**	$\dfrac{n+1}{2}$th value
n is **even**	Average of $\dfrac{n}{2}$th and $\left(\dfrac{n}{2}+1\right)$th values

Definition

Definition

The **median** is the *middle value* when data is arranged in ascending order.

- Exactly **50%** of values lie below the median
- Exactly **50%** of values lie above the median

For ungrouped data:

Situation	Position of median
n is **odd**	$\frac{n+1}{2}$th value
n is **even**	Average of $\frac{n}{2}$th and $\left(\frac{n}{2}+1\right)$th values

Warning

Always sort the data into ascending order before applying

Why the Median Matters — Outlier Resistance

A Quick Illustration

Eight farmers spend (TZS '000/month): **200, 250, 280, 300, 320, 350, 400, 3,500**
Mean = 5,600 / 8 = **TZS 700,000**
Median = average of 4th and 5th values = (300 + 320)/2 = **TZS 310,000**
Interpretation - Most farmers spend between TZS 200,000 and 400,000. One extreme value (TZS 3,500,000) is an outlier. This outlier pulls the *mean upward (TZS 700,000)* making it unrepresentative of a typical farmer.

- The median (TZS 310,000) better represents the typical spending level, since it is not affected by extreme values.

Why the Median Matters — Outlier Resistance

A Quick Illustration

Eight farmers spend (TZS '000/month): **200, 250, 280, 300, 320, 350, 400, 3,500**

Mean = 5,600 / 8 = **TZS 700,000**

Median = average of 4th and 5th values = (300 + 320)/2 = **TZS 310,000**

Interpretation - Most farmers spend between TZS 200,000 and 400,000. One extreme value (TZS 3,500,000) is an outlier. This outlier pulls the *mean upward (TZS 700,000)* making it unrepresentative of a typical farmer.

- The median (TZS 310,000) better represents the typical spending level, since it is not affected by extreme values.

This is why **income data, house prices, and waiting times** are almost always reported using the **median** — not the mean.

Median from Ungrouped Data — Our Dataset

Sorting all 30 spending values in ascending order:

30	32	33	36	37	38	38	38	41	42
42	43	44	45	47	47	47	48	49	50
50	52	53	55	55	55	58	60	61	65

Median from Ungrouped Data — Our Dataset

Sorting all 30 spending values in ascending order:

30	32	33	36	37	38	38	38	41	42
42	43	44	45	47	47	47	48	49	50
50	52	53	55	55	55	58	60	61	65

$n = 30$ (even) $\rightarrow$ average of **15th** and **16th** values

$$\text{15th value} = 47, \qquad \text{16th value} = 47$$

$$M_d = \frac{47 + 47}{2} = \mathbf{TZS\ 47{,}000}$$

Median from Ungrouped Data — Our Dataset

Sorting all 30 spending values in ascending order:

30	32	33	36	37	38	38	38	41	42
42	43	44	45	47	47	47	48	49	50
50	52	53	55	55	55	58	60	61	65

$n = 30$ (even) $\rightarrow$ average of **15th** and **16th** values

$$\text{15th value} = 47, \qquad \text{16th value} = 47$$

$$M_d = \frac{47 + 47}{2} = \textbf{TZS 47,000}$$

Interpretation

Half of the small-scale farmers spend less than TZS 47,000 per day on farm inputs, and the **other half** spend more than TZS 47,000.

Median from Grouped Data

When data is in a frequency table, we estimate the median using:

$$\boxed{M_d = L + \left[\frac{\frac{n}{2} - F}{f}\right] \times h}$$

Median from Grouped Data

When data is in a frequency table, we estimate the median using:

$$\boxed{M_d = L + \left[\frac{\frac{n}{2} - F}{f}\right] \times h}$$

Symbol	Meaning
L	Lower **boundary** of the median class
n	Total observations
F	Cumulative frequency of all classes **before** the median class
f	Frequency of the median class
h	Class width
Median class	The class where cumulative frequency **first reaches** $n/2$

Grouped Median — Step by Step

Class (TZS '000)	Class boundary	f	F
30 – 38	29.5 - 38.5	7	7
39 – 47	**38.5 - 47.5**	**10**	**17**
48 – 56	47.5 - 56.5	7	24
57 – 65	56.5 - 65.5	6	30
Total		**30**	

Grouped Median — Step by Step

Class (TZS '000)	Class boundary	f	F
30 – 38	29.5 - 38.5	7	7
39 – 47	**38.5 - 47.5**	**10**	**17**
48 – 56	47.5 - 56.5	7	24
57 – 65	56.5 - 65.5	6	30
Total		**30**	

Step 1: $n/2 = 30/2 = 15 \rightarrow$ median class is **39–47**
Step 2: Identify components: $L = 38.5$, $F = 7$, $f = 10$, $h = 9$

. . .

$$M_d = L + \left[\frac{\frac{n}{2} - F}{f}\right] \times h = 38.5 + \left[\frac{15 - 7}{10}\right] \times 9 = \textbf{TZS 45,700}$$

. . .

$$M_d = L + \left[\frac{\frac{n}{2} - F}{f}\right] \times h = 38.5 + \left[\frac{15 - 7}{10}\right] \times 9 = \textbf{TZS 45,700}$$

Note on the Difference

Ungrouped median = TZS 47,000 | Grouped median = TZS 45,700

The grouped formula is an **estimate** based on the assumption that data is spread evenly within each class.

Practice Question 3.2 — Median

Try This

The table below shows the **monthly mobile data usage (GB)** of 25 students:

Class (GB)	Frequency f
1 – 5	4
6 – 10	7
11 – 15	8
16 – 20	4
21 – 25	2

(a) Build the Class boundary and cumulative frequency columns.
(b) Identify the median class. Show how you found it.
(c) Apply the interpolation formula to estimate the median data usage.
(d) Write one sentence interpreting the median in the context

3.4 The Mode

Definition and Types

Definition

The **mode** is the value (or class) that appears **most frequently** in a dataset.

Definition and Types

Definition

The **mode** is the value (or class) that appears **most frequently** in a dataset.

Type	Meaning
Unimodal	One mode
Bimodal	Two modes
Multimodal	More than two
No mode	All values equal frequency

Definition and Types

Definition

The **mode** is the value (or class) that appears **most frequently** in a dataset.

Type	Meaning
Unimodal	One mode
Bimodal	Two modes
Multimodal	More than two
No mode	All values equal frequency

Unique Advantage

The mode is the **only** central tendency measure that works for **qualitative (categorical) data**.

Example: *Most popular mobile network among students = M-Pesa (22 out of 50)*

You cannot compute a mean or median of network names — but you can find the most f

Mode from Ungrouped Data — Our Dataset

Sorted data:

30	32	33	36	37	38	38	38	41	42
42	43	44	45	47	47	47	48	49	50
50	52	53	55	55	55	58	60	61	65

Mode from Ungrouped Data — Our Dataset

Sorted data:

30	32	33	36	37	38	38	38	41	42
42	43	44	45	47	47	47	48	49	50
50	52	53	55	55	55	58	60	61	65

Frequency count of repeated values:

Value	38	42	47	50	55
Count	**3**	2	**3**	2	**3**

Mode from Ungrouped Data — Our Dataset

Sorted data:

30	32	33	36	37	38	38	38	41	42
42	43	44	45	47	47	47	48	49	50
50	52	53	55	55	55	58	60	61	65

Frequency count of repeated values:

Value	38	42	47	50	55
Count	**3**	2	**3**	2	**3**

Values 38, 47, and 55 each appear **3 times** $\rightarrow$ **Multi-modal:** $M_o =$ **TZS 38,000, 47,000, and 55,000**

Mode from Ungrouped Data — Our Dataset

Sorted data:

30	32	33	36	37	38	38	38	41	42
42	43	44	45	47	47	47	48	49	50
50	52	53	55	55	55	58	60	61	65

Frequency count of repeated values:

Value	38	42	47	50	55
Count	**3**	2	**3**	2	**3**

Values 38, 47, and 55 each appear **3 times** → **Multi-modal:** $M_o =$ **TZS 38,000, 47,000, and 55,000**

Interpretation

The modes are TZS 38,000, 47,000, and 55,000 because these spending values occur most frequently among the farmers. This means these are the most common monthly spending levels on farm inputs in the study.

Mode from Grouped Data — Formula

For grouped data, the mode is **estimated** from the modal class:

$$\boxed{M_o = L + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times h}$$

Mode from Grouped Data — Formula

For grouped data, the mode is **estimated** from the modal class:

$$M_o = L + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times h$$

Symbol	Meaning
L	Lower **boundary** of the modal class (class with **highest** f)
f_1	Frequency of the modal class
f_0	Frequency of the class **immediately before** the modal class
f_2	Frequency of the class **immediately after** the modal class
h	Class width

Grouped Mode — Step by Step

Class (TZS '000)	Class boundary	f
30 – 38	29.5 - 38.5	7
39 – 47	**38.5 - 47.5**	**10**
48 – 56	47.5 - 56.5	7
57 – 65	56.5 - 65.5	6
Total		**30**

Grouped Mode — Step by Step

Class (TZS '000)	Class boundary	f
30 – 38	29.5 - 38.5	7
39 – 47	**38.5 - 47.5**	**10**
48 – 56	47.5 - 56.5	7
57 – 65	56.5 - 65.5	6
Total		**30**

Identify components:

$L = 38.5$, $f_0 = 7$, $f_1 = 10$, $f_2 = 7$, $h = 9$

Grouped Mode — Step by Step

Class (TZS '000)	Class boundary	f
30 – 38	29.5 - 38.5	7
39 – 47	**38.5 - 47.5**	**10**
48 – 56	47.5 - 56.5	7
57 – 65	56.5 - 65.5	6
Total		**30**

Identify components:

$L = 38.5$, $f_0 = 7$, $f_1 = 10$, $f_2 = 7$, $h = 9$

$$M_o = 38.5 + \left[\frac{10-7}{2(10)-7-7}\right] \times 9 = 38.5 + \left[\frac{3}{6}\right] \times 9 = \textbf{TZS 43,000}$$

Grouped Mode — Step by Step

Class (TZS '000)	Class boundary	f
30 – 38	29.5 - 38.5	7
39 – 47	**38.5 - 47.5**	**10**
48 – 56	47.5 - 56.5	7
57 – 65	56.5 - 65.5	6
Total		**30**

Identify components:
$L = 38.5,\ f_0 = 7,\ f_1 = 10,\ f_2 = 7,\ h = 9$

$$M_o = 38.5 + \left[\frac{10-7}{2(10)-7-7}\right] \times 9 = 38.5 + \left[\frac{3}{6}\right] \times 9 = \textbf{TZS 43,000}$$

Interpretation

The most common daily farm input expenditure among the small-scale farmers is approximately TZS 43,000.

Skewness: What the Order of Measures Tells You

Symmetric (No skew)

Mean = Median = Mode

Right-skewed

→

Mean > Median > Mode

Our dataset fall here

Left-skewed

→

Mean < Median < Mode

Summary: from the Example

Mean = TZS 46,600 | Median = TZS 45,700 | Mode = TZS 43,000

The order Mean > Median > Mode indicates a **right skew** — a few high spenders are pulling the mean upward.

Practice Question 3.3 — Mode

Try This

The following frequency table shows the **daily number of customer complaints** received by a shop over 30 days:

Class (complaints)	Frequency f
0 – 2	5
3 – 5	12
6 – 8	8
9 – 11	4
12 – 14	1

(a) Identify the modal class.
(b) Calculate the estimated mode using the interpolation formula.
(c) Given that the mean number of monthly complaints is 5.3, what does the relationship between mean and mode tell you about the shape of this distribution?

3.5 The Geometric Mean

Why Do We Need Another Type of Mean?

Note

In our study of small-scale farmers' daily expenditure on farm inputs, we used:

- Arithmetic mean $\rightarrow$ average spending
- Median $\rightarrow$ middle spending
- Mode $\rightarrow$ most common spending level

- These measures are appropriate because expenditure is additive data (values are summed).
- Why the Arithmetic Mean Fails for Growth Data

Example

A student invests TZS 1,000,000:

- Year 1: grows **+100%** $\rightarrow$ TZS 2,000,000
- Year 2: falls **−50%** $\rightarrow$ back to TZS 1,000,000
- **Arithmetic mean return** $= \frac{100 + (-50)}{2} = \mathbf{25\%\ per\ year}$
- In reality: But the student's money did **not grow at all**! The arithmetic mean **overstates** the return because growth rates *compound* — each year's gain or loss acts on the previous year's total, not on the original amount.

Example

A student invests TZS 1,000,000:

- Year 1: grows **+100%** → TZS 2,000,000
- Year 2: falls **−50%** → back to TZS 1,000,000
- **Arithmetic mean return** $= \frac{100 + (-50)}{2} = \mathbf{25\%\ per\ year}$
- In reality: But the student's money did **not grow at all**! The arithmetic mean **overstates** the return because growth rates *compound* — each year's gain or loss acts on the previous year's total, not on the original amount.

The **geometric mean** is the correct tool for averaging growth rates, ratios, and any multiplicative process.

Definition and Formula

Definition

The **geometric mean** is the nth root of the product of n positive values.

Definition and Formula

Definition

The **geometric mean** is the nth root of the product of n positive values.

Direct formula:

$$GM = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n}$$

Log formula (easier for large n):

$$GM = \text{Antilog}\left(\frac{\sum \log x}{n}\right)$$

Here "*Antilog*" is 10^y where $y = \left(\frac{\sum \log x}{n}\right)$

Definition and Formula

Definition

The **geometric mean** is the nth root of the product of n positive values.

Direct formula:

$$GM = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n}$$

Log formula (easier for large n):

$$GM = \text{Antilog}\left(\frac{\sum \log x}{n}\right)$$

Here "*Antilog*" is 10^y where $y = \left(\frac{\sum \log x}{n}\right)$

! Important

All values must be positive. The GM cannot be computed if any value is zero or negative.
This is why we convert percentage rates like -5% to a **growth factor** of 0.95 before computing the GM.

When to Use the Geometric Mean

Use GM when data involves:

- **Growth rates** (GDP growth, investment returns, population growth)
- **Percentage changes** (price indices, inflation rates)
- **Ratios** (exchange rates, financial ratios)
- **Data spanning very wide ranges** (earthquake magnitudes, population sizes)

When to Use the Geometric Mean

Use GM when data involves:

- **Growth rates** (GDP growth, investment returns, population growth)
- **Percentage changes** (price indices, inflation rates)
- **Ratios** (exchange rates, financial ratios)
- **Data spanning very wide ranges** (earthquake magnitudes, population sizes)

Tip

Simple test:
If the values *multiply together* to give a meaningful result $\rightarrow$ use GM.
If the values *add together* to give a meaningful result $\rightarrow$ use arithmetic mean.

Connecting GM to Our Dataset

Connecting GM to Our Dataset

Warning

Our dataset measures **monthly spending amounts in TZS** — not growth rates or ratios. So GM here is not appropriate. We introduce GM now with a separate example

Connecting GM to Our Dataset

Warning

Our dataset measures **monthly spending amounts in TZS** — not growth rates or ratios. So GM here is not appropriate. We introduce GM now with a separate example

Appropriate GM Example — Investment Returns

A finance student's portfolio grew at these annual rates over 5 years:

$$8\%, \quad 12\%, \quad -4\%, \quad 15\%, \quad 6\%$$

Convert to growth factors: $1.08,\ 1.12,\ 0.96,\ 1.15,\ 1.06$

Formula

Growth Factor $(x) = 1 +$ (rate of change) is derived f

Geometric Mean — Solved Step by Step

Data: $1.08,\ 1.12,\ 0.96,\ 1.15,\ 1.06$

Geometric Mean — Solved Step by Step

Data: $1.08,\ 1.12,\ 0.96,\ 1.15,\ 1.06$

Using the log formula:

x (growth factor)	$\log_{10} x$
1.08	0.03342
1.12	0.04922
0.96	−0.01773
1.15	0.06070
1.06	0.02531
Sum	$\sum \log x = 0.15092$

Geometric Mean — Solved Step by Step

Data: $1.08,\ 1.12,\ 0.96,\ 1.15,\ 1.06$

Using the log formula:

x (growth factor)	$\log_{10} x$
1.08	0.03342
1.12	0.04922
0.96	−0.01773
1.15	0.06070
1.06	0.02531
Sum	$\sum \log x = 0.15092$

$$\log(GM) = \frac{0.15092}{5} = 0.030184$$

$$GM = \text{Antilog}(0.030184) = 10^{0.030184} = \mathbf{1.0718}$$

Average annual growth rate = 7.18% per year

Verifying the Geometric Mean

Verify!

If the portfolio grows at a **constant** 7.18% per year for 5 years:

TZS $1{,}000{,}000 \times (1.0718)^5 =$ TZS $1{,}000{,}000 \times 1.4148 =$ TZS $\mathbf{1{,}414{,}8}$

Now check using the actual rates:

$1{,}000{,}000 \times 1.08 \times 1.12 \times 0.96 \times 1.15 \times 1.06 =$ TZS $\mathbf{1{,}414{,}800}$ ✓

Arithmetic mean would give $(8 + 12 - 4 + 15 + 6)/5 = 7.4\%$ per year
$\rightarrow$ TZS $1{,}000{,}000 \times 1.074^5 =$ TZS $1{,}429{,}000$ — **an overestimate**. The GM gives the exact correct answer.

Q1

The Consumer Price Index (CPI) in a country changed by the following rates over four years:

Year	1	2	3	4
Rate	+6%	+9%	+3%	+12%

(a) Convert each rate to a growth factor.
(b) Calculate the geometric mean annual inflation rate using the log formula.

(c) If a basket of goods cost TZS 50,000 at the start of Year 1, what would it cost at the end of Year 4?
(d) Why would using the arithmetic mean give an incorrect answer here?

Q2

A university library is tracking how many students are using their new "E-Book Portal." The usage grew by the following annual rates over 3 years:

- Year 1: $+20\%$ growth (many early adopters)
- Year 2: $+40\%$ growth (marketing campaign success)
- Year 3: $+5\%$ growth (market saturation)

Calculate the Geometric mean using direct formula

3.6 The Harmonic Mean

Motivation

The Trip Scenario (When Equal Distance Gives Unequal Time)

Dodoma Mbeya (400 km each way) **Outward:** 80 km/h **Return:** 60 km/h **Common Guess:** (80 + 60) ÷ 2 = **70 km/h**

The Reality Check

Leg	Speed	Distance	Time Spent
Outward	80 km/h	400 km	**5.00 hours**
Return	60 km/h	400 km	**6.67 hours**
Total		**800 km**	**11.67 hours**

True Average Speed: 800 km $\div$ 11.67 hrs $=$ $\mathbf{68.57}$ km/h

The “Why”

Definition and Formula

Definition

The **harmonic mean** is defined as the reciprocal of the arithmetic mean of the reciprocals of the values in a data set.

For a set of n numbers $(x_1, x_2, \cdots, x_n)$:

$$\boxed{HM = \frac{n}{\displaystyle\sum_{i=1}^{n} \frac{1}{x_i}}}$$

Zero Restriction: **HM** be calculated if any of the values in the data set are zero (as you cannot divide by zero)
Special case For two values, x_1 and x_2, use this simplified formula:

$$\text{HM} = \frac{2x_1x_2}{x_1 + x_2}$$

Harmonic Mean — Solved Example

Example 1 : Transport Example

Data: speeds of 80 km/h and 60 km/h over **equal distances**

$$HM = \frac{2}{\frac{1}{80} + \frac{1}{60}} = \frac{2}{0.01250 + 0.01667} = \frac{2}{0.02917} = \mathbf{68.57}\ \text{km/h}\ \checkmark$$

Harmonic Mean — Solved Example

Example 1 : Transport Example

Data: speeds of 80 km/h and 60 km/h over **equal distances**

$$HM = \frac{2}{\frac{1}{80} + \frac{1}{60}} = \frac{2}{0.01250 + 0.01667} = \frac{2}{0.02917} = \mathbf{68.57} \text{ km/h} \checkmark$$

This matches the true average speed we calculated from total time and distance.

Example 2 : Processing & Shelving rate Example

Use the Harmonic Mean when calculating average speeds for a fixed workload.
The Scenario: A library has two interns shelving a cart of 100 books.

- Intern A shelves at 20 books/hour.
- Intern B shelves at 10 books/hour.

The Question: What is the average shelving rate for the library?

Practice Question 3.5 — Harmonic Mean

Try This

A market researcher travels to three villages to conduct surveys. Each village is the **same distance** from the research office.

Village	Travel speed (km/h)
A	45
B	60

Choosing the Right Average — Summary

Measure	Formula	Use When	**Not** When
Arithmetic Mean	$\sum x/n$	Symmetric quantitative data	Skewed data; extreme outliers
Median	Middle value	Skewed data; outliers; income	Categorical data
Mode	Most frequent	Categorical; most common value	Precise numerical summaries
Geometric Mean	$\sqrt[n]{\prod x_i}$	Growth rates; ratios; indices	Absolute values; zeros present
Harmonic Mean	$n/\sum(1/x_i)$	Rates; equal quantity at diff. rates	Equal time at different rates

3.7 Quartiles

Dividing Data into Four Equal Parts

The median divides data into **two** equal halves. Quartiles go further:

$$\underbrace{25\%}_{\text{below } Q_1} \quad \Big| \quad Q_1 \quad \underbrace{25\%} \quad \Big| \quad Q_2 \quad \underbrace{25\%} \quad \Big| \quad Q_3 \quad \underbrace{25\%}_{\text{above } Q_3}$$

3.7 Quartiles

Dividing Data into Four Equal Parts

The median divides data into **two** equal halves. Quartiles go further:

$$\underbrace{25\%}_{\text{below } Q_1} \quad \Big| \quad Q_1 \quad \underbrace{25\%} \quad \Big| \quad Q_2 \quad \underbrace{25\%} \quad \Big| \quad Q_3 \quad \underbrace{25\%}_{\text{above } Q_3}$$

Quartile	Also known as	Meaning
Q_1	Lower quartile / 25th percentile	25% of data lies below here
Q_2	Median / 50th percentile	50% of data lies below here
Q_3	Upper quartile / 75th percentile	75% of data lies below here

3.7 Quartiles

Dividing Data into Four Equal Parts

The median divides data into **two** equal halves. Quartiles go further:

$$\underbrace{25\%}_{\text{below } Q_1} \quad \Big| \quad Q_1 \quad \underbrace{25\%} \quad \Big| \quad Q_2 \quad \underbrace{25\%} \quad \Big| \quad Q_3 \quad \underbrace{25\%}_{\text{above } Q_3}$$

Quartile	Also known as	Meaning
Q_1	Lower quartile / 25th percentile	25% of data lies below here
Q_2	Median / 50th percentile	50% of data lies below here
Q_3	Upper quartile / 75th percentile	75% of data lies below here

$$\text{Interquartile Range (IQR)} = Q_3 - Q_1$$

Quartiles from Grouped Data — Formula

$$\boxed{Q_k = L + \left[\frac{\frac{kn}{4} - F}{f}\right] \times h} \qquad (k = 1, 2, 3)$$

Symbol	Meaning
L	Lower **boundary** of the median class
n	Total observations
k	The Quarter. Which quartile $(k = 1(25\%), 2(50\%), 3(75\%))$
F	Cumulative frequency of all classes **before** the quartile class
f	Frequency of the class quartile
h	Class width

Quartiles — Worked on Our Dataset

Frequency table with cumulative frequency:

Class	f	F
30 – 38	7	7
39 – 47	10	17
48 – 56	7	24
57 – 65	6	30

Finding the quartile class:

The Q_k class is the class where the cumulative frequency **first reaches** $\dfrac{kn}{4}$.

Calculating Q1

- Find the quartile class: $\dfrac{1 \times 30}{4} = 7.5$
- then find first class where $F \geq 7.5 \rightarrow$ **39–47** (where $F = 17$)
- Calculate Q1

$$Q_1 = 38.5 + \left[\frac{7.5 - 7}{10}\right] \times 9$$

$$Q_1 = 38.5 + \frac{0.5}{10} \times 9 = \textbf{TZS 38,950}$$

Calculating Q2 (Median Check)

$\frac{2 \times 30}{4} = 15 \rightarrow$ **39–47** (where $F = 17$, first ≥ 15)

$$Q_2 = 38.5 + \left[\frac{15 - 7}{10}\right] \times 9$$

$$Q_2 = 38.5 + \frac{8}{10} \times 9 = 38.5 + 7.2$$

$$Q_2 = \textbf{TZS 45,700}$$

This matches our grouped median

Calculating Q3

$\frac{3 \times 30}{4} = 22.5 \rightarrow$ **48–56** (where $F = 24$, first ≥ 22.5)

$$Q_3 = 47.5 + \left[\frac{22.5 - 17}{7}\right] \times 9 = 47.5 + \frac{5.5}{7} \times 9 = 47.5 + 7.07 = \textbf{TZS 54,5}$$

Calculating Q3

$$\frac{3 \times 30}{4} = 22.5 \rightarrow$$ **48–56** (where $F = 24$, first ≥ 22.5)

$$Q_3 = 47.5 + \left[\frac{22.5 - 17}{7}\right] \times 9 = 47.5 + \frac{5.5}{7} \times 9 = 47.5 + 7.07 = \textbf{TZS 54,5}$$

$$\text{IQR} = Q_3 - Q_1 = 54{,}570 - 38{,}950 = \textbf{TZS 15,620}$$

Calculating Q3

$\frac{3 \times 30}{4} = 22.5 \rightarrow$ **48–56** (where $F = 24$, first ≥ 22.5)

$$Q_3 = 47.5 + \left[\frac{22.5 - 17}{7}\right] \times 9 = 47.5 + \frac{5.5}{7} \times 9 = 47.5 + 7.07 = \textbf{TZS 54,5}$$

$$\text{IQR} = Q_3 - Q_1 = 54{,}570 - 38{,}950 = \textbf{TZS 15,620}$$

Interpretation

- **25%** of farmers spend less than **TZS 38,950** per month — the low spenders
- **50%** of farmers spend less than **TZS 45,700** per month
- **75%** of farmers spend less than **TZS 54,570** per month
- The **middle 50%** of farmers' monthly spending spans a range of **TZS 15,620**
- Any farmer spending more than $Q_3 + 1.5 \times IQR = 54{,}570 + 23{,}430 =$ **TZS 78,000** is a statistical outlier

Practice Question 3.6 — Quartiles

Try This

The frequency table below shows the **weekly library visits** of 40 students:

Class (visits)	Frequency f
0 – 1	6
2 – 3	12
4 – 5	14
6 – 7	5
8 – 9	3

(a) Construct the cumulative frequency column.
(b) Calculate Q_1, Q_2, and Q_3 using the interpolation formula.
(c) Calculate the IQR. What does it tell you about the spread of library visits among students?
(d) A student visits the library 8 times in one week. Using the IQR rule ($> Q_3 + 1.5 \times IQR$), would this student be classified

3.8 Deciles and Percentiles

The Same Idea, More Divisions

Quartiles, deciles, and percentiles all use the **same interpolation formula** — only the denominator changes.

3.8 Deciles and Percentiles

The Same Idea, More Divisions

Quartiles, deciles, and percentiles all use the **same interpolation formula** — only the denominator changes.

Deciles — divide into **10** equal parts

$$D_k = L + \left[\frac{\frac{kn}{10} - F}{f}\right] \times h$$

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

D5 = Q2 = P50 = Median

Percentiles — divide into **100** equal parts

$$P_k = L + \left[\frac{\frac{kn}{100} - F}{f}\right] \times h$$

$k = 1, 2, \ldots, 99$

$P_{25} = Q_1 \mid P_{50} = Q_2 \mid P_{75} = Q_3$

3.8 Deciles and Percentiles

The Same Idea, More Divisions

Quartiles, deciles, and percentiles all use the **same interpolation formula** — only the denominator changes.

Deciles — divide into **10** equal parts

$$D_k = L + \left[\frac{\frac{kn}{10} - F}{f}\right] \times h$$

$k = 1, 2, \dots, 9$

D5 = Q2 = P50 = Median

Percentiles — divide into **100** equal parts

$$P_k = L + \left[\frac{\frac{kn}{100} - F}{f}\right] \times h$$

$k = 1, 2, \dots, 99$

$P_{25} = Q_1 \mid P_{50} = Q_2 \mid P_{75} = Q_3$

Tip

Memory aid: Quartile denominator = **4**, Decile = **10**, Percentile = **100**

Deciles and Percentiles — Our Dataset

Using the same frequency table ($n = 30$):

Class	f	F
30 – 38	7	7
39 – 47	10	17
48 – 56	7	24
57 – 65	6	30

Calculating D3 (3rd Decile)

$\dfrac{3 \times 30}{10} = 9 \rightarrow$ first class where $F \geq 9 \rightarrow$ **39–47** $(F = 17)$

$$D_3 = 38.5 + \left[\frac{9-7}{10}\right] \times 9 = 38.5 + \frac{2}{10} \times 9 = 38.5 + 1.8 = \textbf{TZS 40,300}$$

30% of students spend less than TZS 40,300 per day.

Calculating P80 (80th Percentile)

$\dfrac{80 \times 30}{100} = 24 \rightarrow$ first class where $F \geq 24 \rightarrow$ **48–56** ($F = 24$, exactly)

$$P_{80} = 47.5 + \left[\frac{24 - 17}{7}\right] \times 9 = 47.5 + \frac{7}{7} \times 9 = 47.5 + 9 = \textbf{TZS 56,500}$$

80% of students spend less than TZS 56,500 per day. The top 20% spend more than this.

Calculating P90 (90th Percentile)

$\frac{90 \times 30}{100} = 27 \rightarrow$ first class where $F \geq 27 \rightarrow$ **57–65** $(F = 30)$

$$P_{90} = 56.5 + \left[\frac{27 - 24}{6}\right] \times 9 = 56.5 + \frac{3}{6} \times 9 = 56.5 + 4.5 = \textbf{TZS 61,000}$$

Only the top **10%** of students spend more than TZS 61,000 per day on food.

Reading Positional Measures from the Ogive

Key Connection to Topic II

Deciles and percentiles can also be **read directly from the ogive** (cumulative frequency curve).

- Draw a **horizontal line** from the target cumulative frequency on the y-axis
- Where the line hits the ogive curve, drop a **vertical line** to the x-axis
- The x-value you read is your decile or percentile estimate

The formula gives the same result as reading from the ogive — the formula is simply more precise.

Practice Question 3.7 — Deciles and Percentiles

Try This

Using the **same library visits data** from Practice Question 3.6:

Class	0–1	2–3	4–5	6–7	8–9
f	6	12	14	5	3

$(n = 40)$
(a) Calculate D_4 (the 4th decile). Interpret your answer.
(b) Calculate P_{60}. Interpret your answer.
(c) Verify that $D_5 = Q_2$ (the 5th decile equals the median). Show both calculations and confirm they give the same result.
(d) A lecturer says: *"60% of my students visit the library at most 4.5 times per week."* Based on your P_{60}, is this statement correct?

3.9 Topic Summary

What We Covered in Topic III

1. **Arithmetic mean** — most common average; requires midpoints for grouped data; sensitive to outliers

3.9 Topic Summary

What We Covered in Topic III

1. **Arithmetic mean** — most common average; requires midpoints for grouped data; sensitive to outliers
2. **Median** — middle value; resistant to outliers; use interpolation formula for grouped data

3.9 Topic Summary

What We Covered in Topic III

1. **Arithmetic mean** — most common average; requires midpoints for grouped data; sensitive to outliers
2. **Median** — middle value; resistant to outliers; use interpolation formula for grouped data
3. **Mode** — most frequent; the only measure valid for categorical data; multimodal data is common

3.9 Topic Summary

What We Covered in Topic III

1. **Arithmetic mean** — most common average; requires midpoints for grouped data; sensitive to outliers
2. **Median** — middle value; resistant to outliers; use interpolation formula for grouped data
3. **Mode** — most frequent; the only measure valid for categorical data; multimodal data is common
4. **Geometric mean** — for growth rates and ratios; use log method for large n; requires all positive values

3.9 Topic Summary

What We Covered in Topic III

1. **Arithmetic mean** — most common average; requires midpoints for grouped data; sensitive to outliers
2. **Median** — middle value; resistant to outliers; use interpolation formula for grouped data
3. **Mode** — most frequent; the only measure valid for categorical data; multimodal data is common
4. **Geometric mean** — for growth rates and ratios; use log method for large n; requires all positive values
5. **Harmonic mean** — for rates where the same quantity is at different rates; not for absolute amounts

3.9 Topic Summary

What We Covered in Topic III

1. **Arithmetic mean** — most common average; requires midpoints for grouped data; sensitive to outliers
2. **Median** — middle value; resistant to outliers; use interpolation formula for grouped data
3. **Mode** — most frequent; the only measure valid for categorical data; multimodal data is common
4. **Geometric mean** — for growth rates and ratios; use log method for large n; requires all positive values
5. **Harmonic mean** — for rates where the same quantity is at different rates; not for absolute amounts
6. **Quartiles** (Q1, Q2, Q3) — divide data into 4 equal parts; IQR measures middle-50% spread; outlier detection

3.9 Topic Summary

What We Covered in Topic III

1. **Arithmetic mean** — most common average; requires midpoints for grouped data; sensitive to outliers
2. **Median** — middle value; resistant to outliers; use interpolation formula for grouped data
3. **Mode** — most frequent; the only measure valid for categorical data; multimodal data is common
4. **Geometric mean** — for growth rates and ratios; use log method for large n; requires all positive values
5. **Harmonic mean** — for rates where the same quantity is at different rates; not for absolute amounts
6. **Quartiles** (Q1, Q2, Q3) — divide data into 4 equal parts; IQR measures middle-50% spread; outlier detection
7. **Deciles and Percentiles** — same formula as quartiles, denominators 10 and 100; can be read from ogive

Next Topic

Topic IV: Measures of Dispersion

Range · Mean Deviation · Variance

Standard Deviation · Coefficient of Variation

Next Topic

Topic IV: Measures of Dispersion

Range · Mean Deviation · Variance
Standard Deviation · Coefficient of Variation

Tip

Preparation for next class:
The mean we calculated today — **TZS 46,600** — will be our reference point in Topic IV.
Every measure of dispersion asks the same question:
"How far do individual values deviate from this mean — and on average, by how much?"

References

- Francis, A. (2006). *Business Mathematics and Statistics* (10th ed.). Great Britain.
- Gupta, S.P. (2003). *Statistical Methods*. Sultan Chand and Sons.
- Anderson, D.R., Sweeney, D.J., & Williams, T.A. (2014). *Statistics for Business and Economics* (12th ed.). West Publishing.
- Miller, C., Salzmann, S., & Clendenin, G. (2005). *Business Mathematics* (10th ed.). Addison Wesley.
- Aczel, A.D. & Sounderpandian, J. (2006). *Complete Business Statistics* (6th ed.). Tata McGraw Hill.

Thank You

Questions?

Dr. Anna Fome
Department of Economics, Mathematics and Statistics
Jordan University College

"*The average tells you where the centre is.*
The spread — which we study next — tells you how much to trust it."