

QMS 101 Introductory Statistics

Topic IV: Measures of Dispersion

Dr. Anna Fome

Jordan University College

2026-05-28

Learning Outcomes

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

- ▶ Explain the meaning and purpose of **measures of dispersion**
- ▶ **Derive** the formulas for range, mean deviation, variance, standard deviation, and coefficient of variation
- ▶ Calculate all five measures for ungrouped and grouped data
- ▶ Use the **coefficient of variation** to compare variability across different datasets
- ▶ Interpret every result in plain language and link dispersion to the Topic III mean

Our Running Dataset

The Study Context

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

She collected the following data (**TZS '000**) from **30 farmers**:

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

Frequency table (from Topic III):

Class (TZS '000)	f	Midpoint x	F
30 – 38	7	34	7
39 – 47	10	43	17
48 – 56	7	52	24
57 – 65	6	61	30
Total	30		

4.1 What is Dispersion — and Why Does It Matter?

The Limitation of Central Tendency Alone

Two Villages — Same Average, Very Different Reality

A ministry official reports that the **average** monthly farm input expenditure in two villages is the same: **TZS 46,600**.

Village A (TZS '000)	Village B (TZS '000)
44, 45, 47, 46, 48, 47, 46	10, 20, 65, 80, 15, 72, 64
Mean = 46.7	Mean = 46.6

Are the two villages really the same?

The Limitation of Central Tendency Alone

Two Villages — Same Average, Very Different Reality

A ministry official reports that the **average** monthly farm input expenditure in two villages is the same: **TZS 46,600**.

Village A (TZS '000)	Village B (TZS '000)
44, 45, 47, 46, 48, 47, 46	10, 20, 65, 80, 15, 72, 64
Mean = 46.7	Mean = 46.6

Are the two villages really the same?

No. Village A farmers all spend similarly — the mean is a reliable guide.

Village B farmers have wildly different spending — the mean misleads completely.

Definition

Measures of dispersion (also called measures of variability or spread) quantify how much the values in a dataset **differ from each other** and from the central value.

Visualising the Problem

Same Mean – Very Different Spread

Both villages have mean . TZS 46,600. But Village B values are far more scattered.

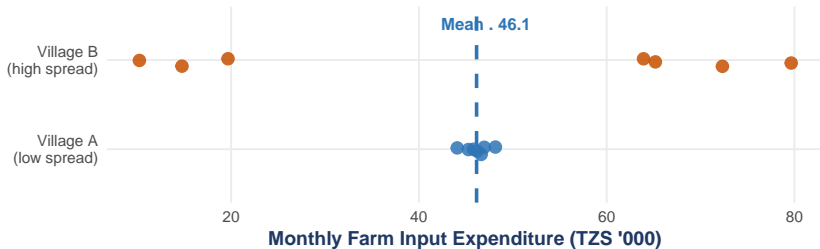


Figure 1

Visualising the Problem

Same Mean – Very Different Spread

Both villages have mean . TZS 46,600. But Village B values are far more scattered.

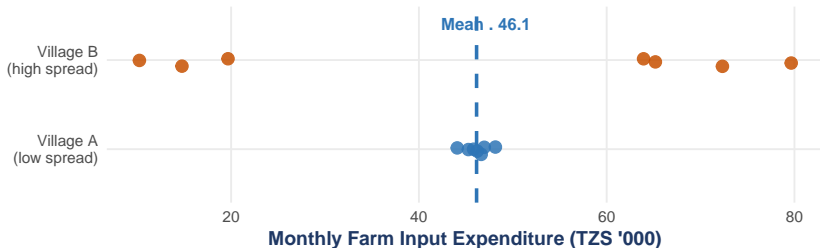


Figure 1

Central tendency answers: *Where is the centre?*

Dispersion answers: *How reliable is that centre as a summary?*

Five Measures of Dispersion

Measure (Symbol)	What It Measures
Range (R)	Distance from smallest to largest value
Mean Deviation (MD)	Average of absolute distances from the mean
Variance (σ^2 or s^2)	Average of squared distances from the mean
Standard Deviation (σ or s)	Typical distance from the mean (same units as data)
Coefficient of Variation (CV)	Relative spread as a percentage (unitless)

Tip

Each measure builds on the previous one.

By the end of this topic you will see that **standard deviation** is the most useful —

4.2 The Range

Deriving the Range

The simplest possible question: How far apart are the most extreme values?

Deriving the Range

The simplest possible question: How far apart are the most extreme values?

Sort the data and look at the two ends:

$$\underbrace{x_{min}}_{\text{smallest}} \quad \cdots \quad \underbrace{x_{max}}_{\text{largest}}$$

$$R = x_{max} - x_{min}$$

Deriving the Range

The simplest possible question: How far apart are the most extreme values?

Sort the data and look at the two ends:

$$\underbrace{x_{min}}_{\text{smallest}} \quad \cdots \quad \underbrace{x_{max}}_{\text{largest}}$$

$$R = x_{max} - x_{min}$$

That is all. No formula to derive — just the difference between the two extremes.

Range — Our Dataset

From raw data:

Min = 30, Max = 65

$$R = 65 - 30 = \mathbf{TZS\ 35,000}$$

Range — Our Dataset

From raw data:

Min = 30, Max = 65

$$R = 65 - 30 = \textbf{TZS 35,000}$$

Interpretation

The farm input expenditure among these 30 farmers spans a range of **TZS 35,000**.

The most frugal farmer spends TZS 30,000; the highest spender pays TZS 65,000 per month.

Range — Advantages and Limitations

Same Range = 35 – But Very Different Distributions

Range uses only 2 values and ignores everything in between.

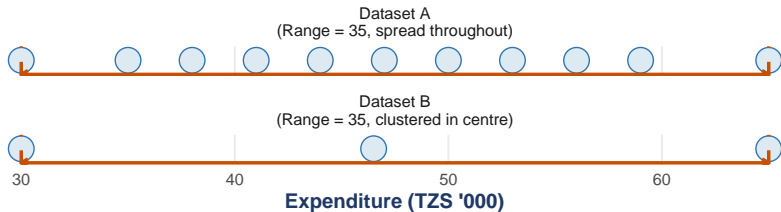


Figure 2

Range — Advantages and Limitations

Same Range = 35 – But Very Different Distributions

Range uses only 2 values and ignores everything in between.

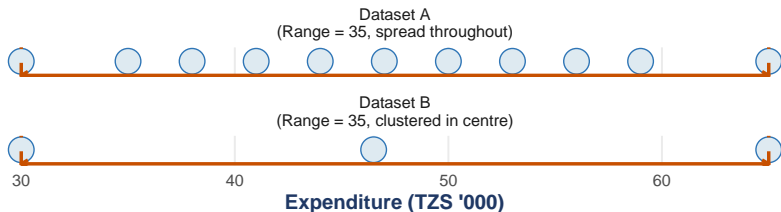


Figure 2

Advantage: Instantly simple to compute and understand.

Limitation: Uses only 2 values — one extreme value completely changes it.

Practice 4.1 — Range

Your Turn

An agricultural extension officer recorded the **number of bags of maize harvested** by 8 smallholder farmers:

12, 28, 15, 34, 9, 22, 31, 18

- (a) Calculate the range of the harvest data.
- (b) A ninth farmer joins the record with a harvest of **52 bags** — an unusually good season.

Recalculate the range. What does this show about the range's weakness?

- (c) The following year, all 9 farmers increase their harvest by 5 bags each.

Without recalculating, what is the new range? Explain your reasoning.

4.3 Mean Deviation

The Core Idea — Measuring from the Centre

The range ignores all middle values. A better measure would use every observation.

The Core Idea — Measuring from the Centre

The range ignores all middle values. A better measure would use every observation.

Natural idea: Measure how far each value is from the mean, then average those distances.

The Core Idea — Measuring from the Centre

The range ignores all middle values. A better measure would use every observation.

Natural idea: Measure how far each value is from the mean, then average those distances.

For each observation x_i , the **deviation from the mean** is:

$$d_i = x_i - \bar{x}$$

The Core Idea — Measuring from the Centre

The range ignores all middle values. A better measure would use every observation.

Natural idea: Measure how far each value is from the mean, then average those distances.

For each observation x_i , the **deviation from the mean** is:

$$d_i = x_i - \bar{x}$$

Problem: If we simply average the deviations:

$$\frac{\sum(x_i - \bar{x})}{n} = \frac{0}{n} = 0 \quad \text{always!}$$

Positive and negative deviations cancel out — giving zero every time.

The Core Idea — Measuring from the Centre

The range ignores all middle values. A better measure would use every observation.

Natural idea: Measure how far each value is from the mean, then average those distances.

For each observation x_i , the **deviation from the mean** is:

$$d_i = x_i - \bar{x}$$

Problem: If we simply average the deviations:

$$\frac{\sum(x_i - \bar{x})}{n} = \frac{0}{n} = 0 \quad \text{always!}$$

Positive and negative deviations cancel out — giving zero every time.

Solution: Ignore the sign. Take the **absolute value** $|d_i| = |x_i - \bar{x}|$.

Deriving the Mean Deviation Formula

Step 1: Compute each deviation from the mean: $x_i - \bar{x}$

Step 2: Remove the sign (take absolute value): $|x_i - \bar{x}|$

Step 3: Average these absolute deviations:

$$MD = \frac{\sum |x_i - \bar{x}|}{n}$$

Deriving the Mean Deviation Formula

Step 1: Compute each deviation from the mean: $x_i - \bar{x}$

Step 2: Remove the sign (take absolute value): $|x_i - \bar{x}|$

Step 3: Average these absolute deviations:

$$MD = \frac{\sum |x_i - \bar{x}|}{n}$$

For grouped data — weight each class deviation by its frequency:

$$MD = \frac{\sum f|x - \bar{x}|}{\sum f}$$

where x is the class midpoint and f is the class frequency.

Mean Deviation — Our Dataset (Ungrouped)

$\bar{x} = 46.37$ (ungrouped mean)

x	$x - \bar{x}$	$ x - \bar{x} $
30	-16.37	16.37
32	-14.37	14.37
33	-13.37	13.37
61	+14.63	14.63
65	+18.63	18.63
Total	0.00	$\sum x - \bar{x} = 302.93$

Mean Deviation — Our Dataset (Ungrouped)

$$\bar{x} = 46.37 \text{ (ungrouped mean)}$$

x	$x - \bar{x}$	$ x - \bar{x} $
30	-16.37	16.37
32	-14.37	14.37
33	-13.37	13.37
61	+14.63	14.63
65	+18.63	18.63
Total	0.00	$\sum x - \bar{x} = 302.93$

$$MD = \frac{302.93}{30} = \mathbf{TZS\ 10,098}$$

Mean Deviation — Grouped Data

Using $\bar{x} = 46.6$ (grouped mean):

Class	f	x	$x - \bar{x}$	$ x - \bar{x} $	$f x - \bar{x} $
30–38	7	34	-12.6	12.6	88.2
39–47	10	43	-3.6	3.6	36.0
48–56	7	52	+5.4	5.4	37.8
57–65	6	61	+14.4	14.4	86.4
Total	30				248.4

$$MD = \frac{\sum f|x - \bar{x}|}{\sum f} = \frac{248.4}{30} = \text{TZS } 8,280$$

Interpretation

On average, a farmer's monthly expenditure deviates from the group mean of TZS 46,600 by about **TZS 8,280**. A smaller MD would mean farmers spend more uniformly; a larger MD

Practice 4.2 — Mean Deviation

Your Turn

A veterinary officer recorded the **weight of 6 goats (kg)** at a livestock market:

18, 24, 31, 22, 27, 16

- (a) Calculate the arithmetic mean weight.
(b) Complete the deviation table:

x	$x - \bar{x}$	$ x - \bar{x} $
18		
24		
31		
22		
27		
16		

4.4 Variance

Why Not Just Use Mean Deviation?

The mean deviation is intuitive and useful. But there is a mathematical problem:

Why Not Just Use Mean Deviation?

The mean deviation is intuitive and useful. But there is a mathematical problem:

Absolute values ($| \cdot |$) are difficult to work with algebraically. They cannot be differentiated smoothly, which complicates more advanced statistics.

Why Not Just Use Mean Deviation?

The mean deviation is intuitive and useful. But there is a mathematical problem:

Absolute values ($| \ |$) are difficult to work with algebraically. They cannot be differentiated smoothly, which complicates more advanced statistics.

A better approach — square the deviations instead:

$$|x - \bar{x}| \longrightarrow (x - \bar{x})^2$$

Why Not Just Use Mean Deviation?

The mean deviation is intuitive and useful. But there is a mathematical problem:

Absolute values ($| \ |$) are difficult to work with algebraically. They cannot be differentiated smoothly, which complicates more advanced statistics.

A better approach — square the deviations instead:

$$|x - \bar{x}| \longrightarrow (x - \bar{x})^2$$

Squaring achieves two things:

1. Removes the sign (negative squared = positive):
 $(-5)^2 = 25 = (+5)^2$
2. **Penalises large deviations more heavily** — a deviation of 10 gives $10^2 = 100$, while a deviation of 2 gives only $2^2 = 4$.

Deriving the Variance Formula

Start with the same idea as MD — average the “distances” from the mean.

Replace absolute values with squares:

Population variance:
$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Deriving the Variance Formula

Start with the same idea as MD — average the “distances” from the mean.

Replace absolute values with squares:

Population variance:
$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

For a sample (which is our case — 30 farmers, not all farmers):

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

Deriving the Variance Formula

Start with the same idea as MD — average the “distances” from the mean.

Replace absolute values with squares:

Population variance:
$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

For a sample (which is our case — 30 farmers, not all farmers):

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

Why $n - 1$ and not n ?

When we estimate the population variance from a **sample**, dividing by n systematically **underestimates** the true population variance. Dividing by $n - 1$ corrects for this.

This is sometimes called “Bessel’s correction”.

Visualising Variance — Squaring the Deviations

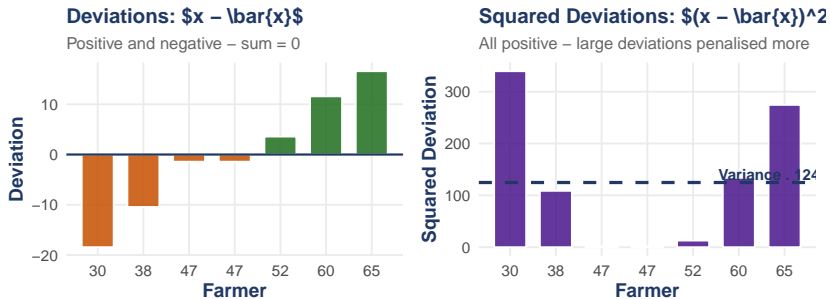


Figure 3

Deriving the Computational Shortcut

Working out $(x - \bar{x})^2$ for every observation is tedious for large datasets.

We can derive a faster equivalent formula algebraically in four steps.

Step 1 — Expand $(x - \bar{x})^2$

$$\sum (x - \bar{x})^2 = \sum (x^2 - 2x\bar{x} + \bar{x}^2)$$

...

Step 2 — Separate** the three sums (since \bar{x} is a constant):

$$\sum (x - \bar{x})^2 = \sum x^2 - 2\bar{x} \underbrace{\sum x}_{=n\bar{x}} + n\bar{x}^2$$

Step 3 — Substitute $\sum x = n\bar{x}$:

$$= \sum x^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum x^2 - n\bar{x}^2$$

Step 4 — Divide by n :

$$\sigma^2 = \frac{\sum x^2}{n} - \bar{x}^2 \quad (\text{ungrouped})$$

$$\sigma^2 = \frac{\sum fx^2}{\sum f} - \bar{x}^2 \quad (\text{grouped})$$

Shortcut vs. Definition — Quick Numerical Check

Verify with three values: 2, 4, 6 — $\bar{x} = 4$ (treat these data as the entire population)

Definition formula:

x	$x - \bar{x}$	$(x - \bar{x})^2$
2	-2	4
4	0	0
6	+2	4
Total	0	8

$$\sigma^2 = \frac{8}{3} = 2.67$$

Shortcut vs. Definition — Quick Numerical Check

Verify with three values: 2, 4, 6 — $\bar{x} = 4$ (treat these data as the entire population)

Definition formula:

x	$x - \bar{x}$	$(x - \bar{x})^2$
2	-2	4
4	0	0
6	+2	4
Total	0	8

$$\sigma^2 = \frac{8}{3} = \mathbf{2.67}$$

Shortcut formula:

$$\begin{aligned}\sum x^2 &= 4 + 16 + 36 = 56 \\ \bar{x}^2 &= 4^2 = 16\end{aligned}$$

$$\sigma^2 = \frac{56}{3} - 16 = 18.67 - 16 = \mathbf{2.67} \checkmark$$

Both give exactly the same answer.

The shortcut eliminates computing every $(x - \bar{x})$ individually.

Variance — Our Dataset (Ungrouped)

Using the 30 raw values with $\bar{x} = 46.37$: (treat the data as the entire population)

Step	Calculation	Result
$\sum x$	$45 + 32 + 55 + \dots + 49$	1,391
\bar{x}	$1,391 \div 30$	46.37
$\sum x^2$	$45^2 + 32^2 + 55^2 + \dots + 49^2$	66,731
$\frac{\sum x^2}{n}$	$66,731 \div 30$	2,224.37
\bar{x}^2	$(46.37)^2$	2,150.17
σ^2	$2,224.37 - 2,150.17$	74.20

Variance — Our Dataset (Ungrouped)

Using the 30 raw values with $\bar{x} = 46.37$: (treat the data as the entire population)

Step	Calculation	Result
$\sum x$	$45 + 32 + 55 + \dots + 49$	1,391
\bar{x}	$1,391 \div 30$	46.37
$\sum x^2$	$45^2 + 32^2 + 55^2 + \dots + 49^2$	66,731
$\frac{\sum x^2}{n}$	$66,731 \div 30$	2,224.37
\bar{x}^2	$(46.37)^2$	2,150.17
σ^2	$2,224.37 - 2,150.17$	74.20

Ungrouped vs. Grouped

Ungrouped Variance = **TZS 74,200** (exact, from raw data)

Variance — Our Dataset (Grouped)

Using $\bar{x} = 46.6$:

Class	f	x	x^2	fx^2
30–38	7	34	1,156	8,092
39–47	10	43	1,849	18,490
48–56	7	52	2,704	18,928
57–65	6	61	3,721	22,326
Total	30			$\sum fx^2 = 67,836$

Variance — Our Dataset (Grouped)

Using $\bar{x} = 46.6$:

Class	f	x	x^2	fx^2
30–38	7	34	1,156	8,092
39–47	10	43	1,849	18,490
48–56	7	52	2,704	18,928
57–65	6	61	3,721	22,326
Total	30			$\sum fx^2 = 67,836$

$$\sigma^2 = \frac{\sum fx^2}{\sum f} - \bar{x}^2 = \frac{67,836}{30} - (46.6)^2 = 2,261.2 - 2,171.56 = \mathbf{89.64} \text{ (TZ)}$$

Variance — Our Dataset (Grouped)

Using $\bar{x} = 46.6$:

Class	f	x	x^2	fx^2
30–38	7	34	1,156	8,092
39–47	10	43	1,849	18,490
48–56	7	52	2,704	18,928
57–65	6	61	3,721	22,326
Total	30			$\sum fx^2 = 67,836$

$$\sigma^2 = \frac{\sum fx^2}{\sum f} - \bar{x}^2 = \frac{67,836}{30} - (46.6)^2 = 2,261.2 - 2,171.56 = \mathbf{89.64} \text{ (TZ)}$$

Warning

The variance has **squared units** — (TZS '000)² — which makes it hard to interpret directly.

Practice 4.3 — Variance

Your Turn

The following table shows the **daily water consumption (litres)** recorded at 5 households in a village:

20, 35, 28, 42, 15

- (a) Calculate the arithmetic mean.
(b) Complete the squared-deviation table:

x	$x - \bar{x}$	$(x - \bar{x})^2$
20		
35		
28		
42		
15		

- (c) Calculate the variance (divide by n here, since this is the

4.5 Standard Deviation

From Variance to Standard Deviation

The problem with variance: Its units are *squared* — e.g., (TZS '000)² — making interpretation difficult.

From Variance to Standard Deviation

The problem with variance: Its units are *squared* — e.g., (TZS '000)² — making interpretation difficult.

The fix: Take the square root to return to the **original units** of the data.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

Grouped:
$$\sigma = \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2}$$

From Variance to Standard Deviation

The problem with variance: Its units are *squared* — e.g., (TZS '000)² — making interpretation difficult.

The fix: Take the square root to return to the **original units** of the data.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

Grouped:
$$\sigma = \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2}$$

The standard deviation answers: *“By how much does a typical observation differ from the mean?”*

It is the **most widely used** measure of dispersion in statistics, science, business, and medicine.

Standard Deviation — Our Dataset

From the variance already computed: $\sigma^2 = 89.64$

$$\sigma = \sqrt{89.64} = \mathbf{9.47 \text{ TZS '000}}$$

Standard Deviation — Our Dataset

From the variance already computed: $\sigma^2 = 89.64$

$$\sigma = \sqrt{89.64} = \mathbf{9.47 \text{ TZS '000}}$$

Interpretation

A typical farmer's monthly expenditure on farm inputs deviates from the group mean of **TZS 46,600** by about **TZS 9,470**.

This is the “typical spread” around the average. Most farmers (roughly 68%, if the data is normally distributed) spend within the range:

$$\bar{x} \pm \sigma = 46.6 \pm 9.47 \implies \text{between TZS } \mathbf{37,130} \text{ and } \mathbf{56,070}$$

Visualising Standard Deviation

Standard Deviation – Typical Distance from the Mean

Green band = Mean \pm 1 SD. Most observations fall within this band.

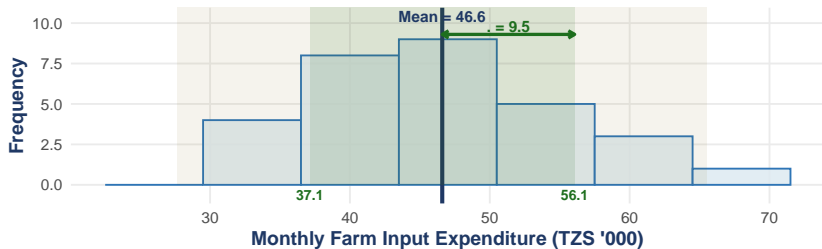


Figure 4

Properties of Standard Deviation

Understanding these properties helps you work efficiently and spot errors:

1. **Always non-negative:** $\sigma \geq 0$. SD equals zero only when all values are identical.

Properties of Standard Deviation

Understanding these properties helps you work efficiently and spot errors:

1. **Always non-negative:** $\sigma \geq 0$. SD equals zero only when all values are identical.
2. **Same units as the data:** If x is in TZS, then σ is also in TZS — unlike variance which is in TZS².

Properties of Standard Deviation

Understanding these properties helps you work efficiently and spot errors:

1. **Always non-negative:** $\sigma \geq 0$. SD equals zero only when all values are identical.
2. **Same units as the data:** If x is in TZS, then σ is also in TZS — unlike variance which is in TZS².
3. **Sensitive to outliers:** Squaring amplifies the effect of extreme values. One outlier can substantially increase SD.

Properties of Standard Deviation

Understanding these properties helps you work efficiently and spot errors:

1. **Always non-negative:** $\sigma \geq 0$. SD equals zero only when all values are identical.
2. **Same units as the data:** If x is in TZS, then σ is also in TZS — unlike variance which is in TZS².
3. **Sensitive to outliers:** Squaring amplifies the effect of extreme values. One outlier can substantially increase SD.
4. **Adding a constant does not change SD:** If every farmer receives TZS 5,000 subsidy, their spending shifts up by 5 — but the spread stays the same. $\sigma(x + c) = \sigma(x)$.

Properties of Standard Deviation

Understanding these properties helps you work efficiently and spot errors:

1. **Always non-negative:** $\sigma \geq 0$. SD equals zero only when all values are identical.
2. **Same units as the data:** If x is in TZS, then σ is also in TZS — unlike variance which is in TZS².
3. **Sensitive to outliers:** Squaring amplifies the effect of extreme values. One outlier can substantially increase SD.
4. **Adding a constant does not change SD:** If every farmer receives TZS 5,000 subsidy, their spending shifts up by 5 — but the spread stays the same. $\sigma(x + c) = \sigma(x)$.
5. **Multiplying by a constant scales SD:** If all values double, SD also doubles. $\sigma(cx) = |c| \cdot \sigma(x)$.

Properties of Standard Deviation

Understanding these properties helps you work efficiently and spot errors:

1. **Always non-negative:** $\sigma \geq 0$. SD equals zero only when all values are identical.
2. **Same units as the data:** If x is in TZS, then σ is also in TZS — unlike variance which is in TZS².
3. **Sensitive to outliers:** Squaring amplifies the effect of extreme values. One outlier can substantially increase SD.
4. **Adding a constant does not change SD:** If every farmer receives TZS 5,000 subsidy, their spending shifts up by 5 — but the spread stays the same. $\sigma(x + c) = \sigma(x)$.
5. **Multiplying by a constant scales SD:** If all values double, SD also doubles. $\sigma(cx) = |c| \cdot \sigma(x)$.
6. **Basis for all further statistics:** The normal distribution, Z-scores, confidence intervals, and hypothesis tests are all built on SD.

Mean Deviation (MD) vs Standard Deviation (SD)

Feature	Mean Deviation (MD)	Standard Deviation (SD)
Formula	$\sum \ x - \bar{x}\ /n$	$\sqrt{\sum (x - \bar{x})^2/n}$
Uses absolute value?	Yes	No (uses squaring)
Penalises large deviations?	Equally	More (squares amplify large gaps)
Easy to interpret?	Yes — same units	Yes — same units
Used in advanced statistics?	Rarely	Extensively (basis for normal distribution, t-tests, regression)

Mean Deviation (MD) vs Standard Deviation (SD)

Feature	Mean Deviation (MD)	Standard Deviation (SD)
Formula	$\sum \ x - \bar{x}\ /n$	$\sqrt{\sum (x - \bar{x})^2/n}$
Uses absolute value?	Yes	No (uses squaring)
Penalises large deviations?	Equally	More (squares amplify large gaps)
Easy to interpret?	Yes — same units	Yes — same units
Used in advanced statistics?	Rarely	Extensively (basis for normal distribution, t-tests, regression)

Tip

Use SD for almost all purposes. It is the universal standard in statistics and is required for Topic III's empirical rule, probability distributions, and all inferential work in later courses.

Practice 4.4 — Standard Deviation

Your Turn

A seed company tests the **germination time (days)** of a new maize variety on 5 plots:

8, 12, 10, 15, 9

(a) Calculate the mean germination time.

(b) Use the shortcut formula $\sigma^2 = \frac{\sum x^2}{n} - \bar{x}^2$ to compute the variance.

(c) Calculate the standard deviation.

4.6 Coefficient of Variation

The Problem with Comparing Standard Deviations

A Comparison Problem

A ministry compares farm input spending in two regions:

	Region A (farm inputs)	Region B (crop yields)
Mean	TZS 46,600	380 kg/acre
SD	TZS 9,470	45 kg/acre

Which region has more variability?

We cannot compare TZS 9,470 to 45 kg/acre — different units, different scales!

The Problem with Comparing Standard Deviations

A Comparison Problem

A ministry compares farm input spending in two regions:

	Region A (farm inputs)	Region B (crop yields)
Mean	TZS 46,600	380 kg/acre
SD	TZS 9,470	45 kg/acre

Which region has more variability?

We cannot compare TZS 9,470 to 45 kg/acre — different units, different scales!

We need a **unitless, relative** measure of spread.

Deriving the Coefficient of Variation

The idea: Express the standard deviation as a *percentage* of the mean.

This removes units and puts variability on a comparable scale.

$$\text{Relative spread} = \frac{\text{Standard deviation}}{\text{Mean}}$$

Multiply by 100 to express as a percentage:

$$CV = \frac{\sigma}{\bar{x}} \times 100\%$$

What CV Tells You

- ▶ **Low CV** → data is **consistent** — values cluster tightly around the mean
- ▶ **High CV** → data is **variable** — values are spread widely relative to the mean
- ▶ **CV is unitless** → can compare datasets measured in

Coefficient of Variation — Examples

From Our Dataset:

$$CV = \frac{\sigma}{\bar{x}} \times 100\% = \frac{9.47}{46.6} \times 100\% = \mathbf{20.3\%}$$

CV in Action — Comparing Two Regions

Region Comparison

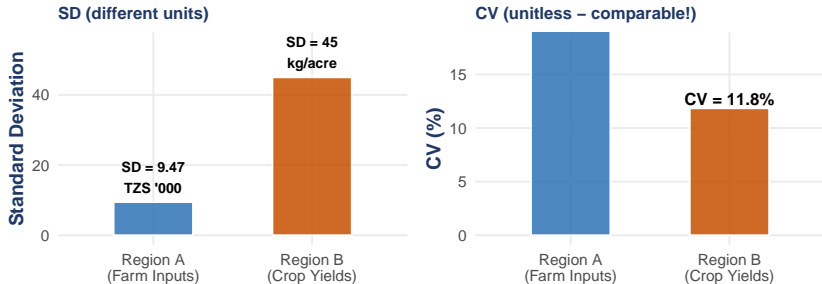


Figure 5

CV in Action — Comparing Two Regions

Region Comparison

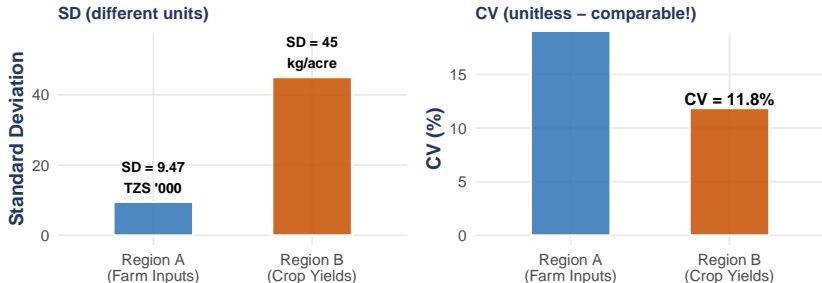


Figure 5

Region A: $CV = 9.47/46.6 \times 100 = 20.3\%$

Region B: $CV = 45/380 \times 100 = 11.8\%$

Region A has more relative variability than Region B, despite having a smaller SD in absolute terms.

Example: Factory Wage Comparison

Two factories A and B located in the same industrial area report the following weekly wages (in TZS):

Factory	Mean Wage (TZS '000)	Standard Deviation	Number of Workers
A	34.5	5.0	476
B	28.5	4.5	524

1. Which factory pays a larger average weekly wage?
2. Which factory has greater variability in individual wages?

Solution

(1) Comparing Average Weekly Wages

- ▶ **Factory A:** mean wage = TZS 34.5 thousand
- ▶ **Factory B:** mean wage = TZS 28.5 thousand

Since:

$$34.5 > 28.5$$

Factory A pays the higher average weekly wage.

(2) Comparing Variability Using the Coefficient of Variation

For Factory A:

$$CV_A = \frac{5}{34.5} \times 100\% = 14.49\%$$

For Factory B:

$$CV_B = \frac{4.5}{28.5} \times 100\% = 15.79\%$$

Interpretation:

- ▶ Since Factory B has the larger coefficient of variation, wages in Factory B are more variable.

Example: Commodity Prices in Two Cities

Prices of a particular commodity (in TZS '000) over five years in two cities are given below:

Year	City A	City B
1	20	22
2	19	23
3	16	10
4	20	18
5	12	15

Which city has more stable prices?

Additional Example 2: Comparing Price Stability

Commodity Prices in Two Cities

Prices of a particular commodity (in TZS '000) over five years in two cities are given below:

Year	City A	City B
1	20	22
2	19	23
3	16	10
4	20	18
5	12	15

Which city has more stable prices?

Solution

Step 1: Compute the Mean Prices

For City A:

$$\bar{x} = \frac{20 + 19 + 16 + 20 + 12}{5} = 17.4$$

For City B:

$$\bar{x} = \frac{22 + 23 + 10 + 18 + 15}{5} = 17.6$$

Step 2: Compute Deviations

City A

Price x	$x - \bar{x}$	$(x - \bar{x})^2$
20	2.6	6.76
19	1.6	2.56
16	-1.4	1.96
20	2.6	6.76
12	-5.4	29.16
Total		47.20

Standard deviation:

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{47.20}{5}} = \sqrt{9.44} = 3.07$$

City B

Price x	$x - \bar{x}$	$(x - \bar{x})^2$
22	4.4	19.36
23	5.4	29.16
10	-7.6	57.76
18	0.4	0.16
15	-2.6	6.76
Total		113.20

Standard deviation:

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{113.20}{5}} = \sqrt{22.64} = 4.76$$

Step 3: Compute the Coefficients of Variation

For City A:

$$CV = \frac{\sigma}{\bar{x}} \times 100\% = \frac{3.07}{17.4} \times 100\% = 17.64\%$$

For City B:

$$CV = \frac{\sigma}{\bar{x}} \times 100\% = \frac{4.76}{17.6} \times 100\% = 27.05\%$$

Interpretation

- ▶ Since City A has the smaller coefficient of variation, prices in City A are more stable.

When to Use CV

- ▶ **Comparing two datasets with different units** (TZS vs kg vs hours vs %)

When to Use CV

- ▶ **Comparing two datasets with different units** (TZS vs kg vs hours vs %)
- ▶ **Comparing two datasets with very different means** (e.g., adult weight vs infant weight)

When to Use CV

- ▶ **Comparing two datasets with different units** (TZS vs kg vs hours vs %)
- ▶ **Comparing two datasets with very different means** (e.g., adult weight vs infant weight)
- ▶ **Assessing consistency** — a factory wants low CV in product dimensions

When to Use CV

- ▶ **Comparing two datasets with different units** (TZS vs kg vs hours vs %)
- ▶ **Comparing two datasets with very different means** (e.g., adult weight vs infant weight)
- ▶ **Assessing consistency** — a factory wants low CV in product dimensions
- ▶ **Investment analysis** — CV measures risk *relative to return* (lower CV = better risk-adjusted return)

When to Use CV

- ▶ **Comparing two datasets with different units** (TZS vs kg vs hours vs %)
- ▶ **Comparing two datasets with very different means** (e.g., adult weight vs infant weight)
- ▶ **Assessing consistency** — a factory wants low CV in product dimensions
- ▶ **Investment analysis** — CV measures risk *relative to return* (lower CV = better risk-adjusted return)

When to Use CV

- ▶ **Comparing two datasets with different units** (T\$ vs kg vs hours vs %)
- ▶ **Comparing two datasets with very different means** (e.g., adult weight vs infant weight)
- ▶ **Assessing consistency** — a factory wants low CV in product dimensions
- ▶ **Investment analysis** — CV measures risk *relative to return* (lower CV = better risk-adjusted return)

Warning

CV should not be used when \bar{x} is zero or very close to zero — division by near-zero gives a misleadingly large or undefined CV.

Also avoid CV for data that can be negative (e.g., profit/loss — a negative mean gives a negative CV with no meaningful interpretation).

Practice 4.5 — Coefficient of Variation

Your Turn

A government nutritionist compares **protein intake (grams/day)** in two districts:

	District X	District Y
Mean	48 g	72 g
SD	6.2 g	8.5 g

- (a) Calculate the CV for each district.
- (b) District Y has a higher SD than District X. Does this mean District Y is more variable?
Use the CV to give a complete answer.
- (c) A nutritionist wants to target the district with **more consistent** protein intake for a simpler, standardised feeding programme. Which district should she target? Justify using CV.

What We Covered in Topic IV

1. **Why dispersion matters** — same mean can hide completely different distributions

What We Covered in Topic IV

1. **Why dispersion matters** — same mean can hide completely different distributions
2. **Range** — simplest measure; difference between max and min; sensitive to outliers

What We Covered in Topic IV

1. **Why dispersion matters** — same mean can hide completely different distributions
2. **Range** — simplest measure; difference between max and min; sensitive to outliers
3. **Mean Deviation** — average absolute distance from the mean; intuitive but rarely used in advanced work

What We Covered in Topic IV

1. **Why dispersion matters** — same mean can hide completely different distributions
2. **Range** — simplest measure; difference between max and min; sensitive to outliers
3. **Mean Deviation** — average absolute distance from the mean; intuitive but rarely used in advanced work
4. **Variance** — average squared deviation; basis for all further statistics; units are squared

What We Covered in Topic IV

1. **Why dispersion matters** — same mean can hide completely different distributions
2. **Range** — simplest measure; difference between max and min; sensitive to outliers
3. **Mean Deviation** — average absolute distance from the mean; intuitive but rarely used in advanced work
4. **Variance** — average squared deviation; basis for all further statistics; units are squared
5. **Standard Deviation** — square root of variance; same units as data; the universal measure of spread

What We Covered in Topic IV

1. **Why dispersion matters** — same mean can hide completely different distributions
2. **Range** — simplest measure; difference between max and min; sensitive to outliers
3. **Mean Deviation** — average absolute distance from the mean; intuitive but rarely used in advanced work
4. **Variance** — average squared deviation; basis for all further statistics; units are squared
5. **Standard Deviation** — square root of variance; same units as data; the universal measure of spread
6. **Coefficient of Variation** — SD expressed as % of mean; unitless; allows cross-dataset comparison

Final Practice — Comprehensive Question

Bringing It All Together

A rural health centre recorded the **number of patients seen per day** over 30 working days. The frequency distribution is:

Patients per day	1–10	11–20	21–30	31–40	41–50
f	4	8	11	5	2

- (a) Calculate the **mean** (use this as your \bar{x} throughout).
- (b) Calculate the **mean deviation**.
- (c) Calculate the **variance** and **standard deviation**.
- (d) Calculate the **coefficient of variation**.
- (e) A second clinic in a neighbouring village has mean = 18 patients/day and SD = 5.2.

Which clinic has more **relative** variability in daily patient load?
Use CV.

Next Topic

Topic V: Permutations & Combinations

Counting · Factorials

Permutations · Combinations

Applications to Business

Next Topic

Topic V: Permutations & Combinations

Counting · Factorials

Permutations · Combinations

Applications to Business



The connection forward:

Permutations and combinations count *how many ways* events can occur.

That counting will form the foundation of **probability** — which in turn builds on the distributions described by the mean and standard deviation you have mastered in Topics III and IV.

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., Salzman, 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 and Mathematics Statistics

Jordan University College

“The mean tells you where the centre is.

The standard deviation tells you how much to trust it.”