جامعة آشور
الكلية التقنية الهندسية
قسم هندسة تقنيات الاتصالات

الاحتمالية والإحصاء

Probability and Statistics

INTRODUCTION TO STATISTICS

Descriptive Statistics, Data Organization and Visualization

المرحلة الثالثة
Third Stage

مدرس المادة

مدرس مساعد كرم ضافر عبد الله

Assistant Lecturer Karam Dhafer Abdullah

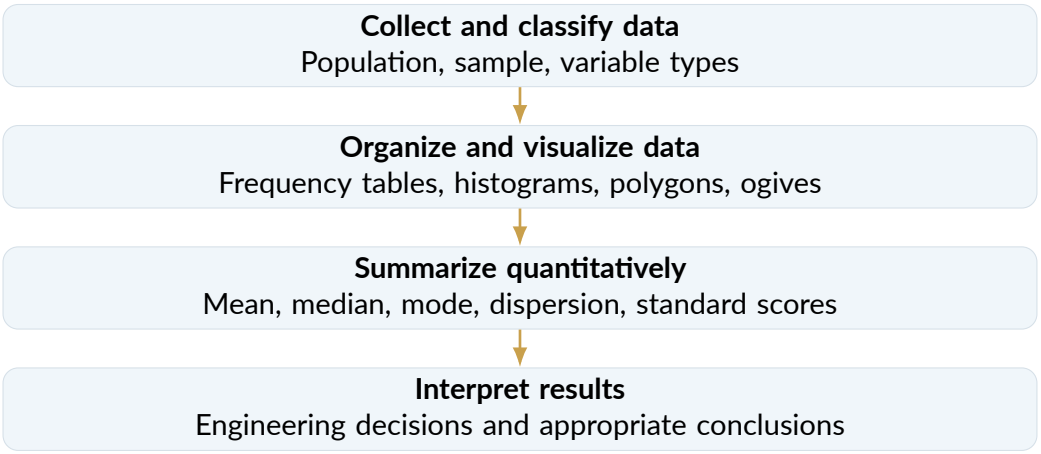
About These Lecture Notes

Scope, learning outcomes and study roadmap

Scope. This revised unit develops the **statistics** material in the supplied *Introduction to Statistics* lecture file. It covers data concepts, frequency distributions, graphical representations, measures of central tendency, measures of dispersion, standard scores and grouped-data applications. Although the course is titled *Probability and Statistics*, the supplied file does **not** develop probability axioms or probability distributions; those topics require subsequent lecture material.

Learning outcomes. On completing this unit, students should be able to:

1. Explain descriptive versus inferential statistics and distinguish a population from a sample.
2. Classify qualitative and quantitative variables and organize raw observations in frequency tables.
3. Draw and interpret a histogram, a frequency polygon and ascending/descending ogives.
4. Calculate mean, median, mode, range, mean absolute deviation, variance, standard deviation and coefficient of variation.
5. Standardize an observation using a z-score and interpret variations in engineering measurements.
6. Compare a result with the original data and identify when a grouped-data result is only an approximation.



Editorial approach. Repetitive source slides have been consolidated; terminology and equations have been standardized. Several source mistakes or misleading shortcuts (including a coefficient-of-variation percentage and the grouped-mode formula) have been corrected. Clearly identified *engineering illustrations* are additional teaching examples.

Contents

1 The Role of Statistics in Engineering 3

2 Branches of Statistics and Key Terms 3

3 Classification of Statistical Data 4

4 Frequency Distributions and Tabulation 5

5 Data Visualization: Histogram, Polygon and Ogive 6

6 Measures of Central Tendency 8

7 Measures of Dispersion 9

8 Standard Deviation, CV and Standard Scores 11

9 Comprehensive Grouped-Data Case Study 12

10 How to Interpret Statistical Results 14

11 Exercises and Self-Assessment 15

12 Formula Reference Sheet 17

1 The Role of Statistics in Engineering

1.1 What is statistics?

Statistics is the discipline concerned with the systematic **collection, organization, presentation, analysis and interpretation of data** to support descriptions, comparisons and decisions under uncertainty.

In ordinary English, *statistics* can also mean numerical summaries such as percentages, averages or counts. As an academic discipline, it refers to the methods used to extract meaningful information from data.

1.2 Why do engineers study statistics?

Engineering systems rarely produce perfectly identical observations. Measurement noise, environmental effects, tolerances, manufacturing variability and operating conditions cause data to vary. Statistical analysis helps distinguish a genuine pattern from ordinary variation.

- **Experimental design and testing:** summarize repeated laboratory measurements.
- **Quality control:** assess manufacturing tolerances and failure rates.
- **Communication systems:** describe measured delay, received power or bit-error counts.
- **Research:** evaluate whether observations support a technical conclusion.

Engineering illustration (added): network latency. Suppose a network connection is tested five times, producing delays of 8, 9, 9, 10 and 14 ms. A descriptive analysis reports a mean of 10 ms and a maximum of 14 ms. A further inferential study would require a representative sampling strategy before extending conclusions to the entire network operating period.

1.3 Four activities in a statistical study

1. **Collection:** select appropriate instruments and record observations consistently.
2. **Organization and presentation:** use tables and graphs to reveal structure.
3. **Analysis:** compute numerical summaries or fit suitable statistical models.
4. **Interpretation:** communicate the meaning and limitations of the results.

A visible trend does not automatically establish its cause. Statistical conclusions depend on measurement quality, study design and whether the collected observations represent the group of interest.

2 Branches of Statistics and Key Terms

2.1 Descriptive and inferential statistics

	Descriptive statistics	Inferential statistics
Purpose	Summarize or display observed data	Use sample evidence to draw conclusions about a larger population
Typical methods	Tables, charts, mean, median, variability	Estimation, hypothesis tests, predictions with uncertainty
Example	Mean delay of 30 recorded packets	Estimate network-wide mean delay using sampled packets

2.2 Population, sample, parameter and statistic

- **Population:** the entire set of individuals, devices or measurements under investigation.
- **Sample:** a subset selected from the population for observation or analysis.
- **Parameter:** a numerical characteristic of a population (for example, the population mean μ).
- **Statistic:** a number computed from a sample (for example, the sample mean $\bar{x}$).

Example. A department has 600 laboratory test records. If every record is analyzed, the 600 records form the population of interest. If 60 records are selected and analyzed, these constitute a sample. The mean of those 60 values is a *sample statistic*.

2.3 Variables and observations

A **variable** is a characteristic that may take different values. A dataset contains observed values of one or more variables. A **frequency** is the number of observations belonging to a specified value or class.

Important distinction. A sample is not necessarily representative simply because it contains many observations. Convenience sampling can produce systematic bias.

3 Classification of Statistical Data

3.1 Qualitative (categorical) variables

Qualitative observations are labels or categories. Examples include the protocol used by a device, antenna type or pass/fail status. Common forms are:

- **Binary/dichotomous:** two categories, such as *connected/disconnected*.
- **Multicategory:** more than two categories, such as *Wi-Fi/ethernet/cellular*.

A numerical code assigned to a category does not automatically make the variable quantitative; e.g. codes 1, 2 and 3 for equipment type are still categories.

3.2 Quantitative variables

Quantitative variables are numerical measurements or counts.

- **Discrete:** counts with separated possible values, such as 0, 1, 2, ... packet losses.
- **Continuous:** measurements that may vary over a continuum, such as signal amplitude or propagation delay.

Variable	Type	Example value
Transmission medium	Qualitative	Optical fiber
Link state	Qualitative, binary	Active
Lost packets	Quantitative, discrete	4
Received power	Quantitative, continuous	-73.2 dBm
Measured delay	Quantitative, continuous	12.4 ms

3.3 Raw and grouped data

Raw (ungrouped) data are recorded observations prior to grouping. **Grouped data** aggregate observations into classes with frequencies. Grouping simplifies large datasets, but individual values are not fully preserved.

Source-style example. Observed marks:

$$12, 15, 15, 17, 17, 17, 19, 19, 19, 19, 19, 21, 21, 21, 23, 23, 23, 23, 25, 25.$$

A simple frequency table summarizes these $n = 20$ observations:

Mark x_i	12	15	17	19	21	23	25
Frequency f_i	1	2	3	5	3	4	2

Check: $\sum_i f_i = 1 + 2 + 3 + 5 + 3 + 4 + 2 = 20$.

4 Frequency Distributions and Tabulation

4.1 Frequency, relative frequency and cumulative frequency

A **frequency distribution** organizes observations according to distinct values or class intervals. The principal quantities are:

$$n = \sum_{i=1}^{k} f_i \qquad \text{(number of observations)},$$

$$r_i = \frac{f_i}{n} \qquad \text{(relative frequency)},$$

$$p_i = 100\frac{f_i}{n}\% \qquad \text{(percentage frequency)},$$

$$F_j = \sum_{i=1}^{j} f_i \qquad \text{(cumulative frequency)}.$$

Relative frequencies must sum to 1 (or percentages to 100%, apart from rounding).

4.2 Class interval, limits, boundaries and midpoint

For an interval a–b, the **class limits** are a and b. For whole-number measurements, the corresponding continuous **boundaries** are often $a - 0.5$ and $b + 0.5$. The **class midpoint** is

$$\boxed{x_i = \frac{\text{lower boundary} + \text{upper boundary}}{2}}.$$

The **class width** h is the difference between successive lower boundaries (or between a class's upper and lower boundaries). All class intervals must be mutually exclusive and cover the intended measurement range.

> For the mark class 50–59, the boundaries are 49.5 and 59.5, the midpoint is 54.5 and the width is 10.

4.3 How to construct a grouped frequency table

1. Determine the minimum and maximum observations and the range $R = x_{\max} - x_{\min}$.
2. Select a reasonable number of classes, k (Sturges' suggestion is $k \approx 1 + 3.322 \log_{10} n$; this is a guideline, not a requirement).
3. Choose a convenient class width, approximately $h \geq R/k$, and define nonoverlapping classes.
4. Count the observations in each class and compute r_i and F_i.
5. Verify $\sum f_i = n$, $\sum r_i = 1$ (subject to rounding), and that the final cumulative frequency is n.

4.4 Worked example: 20 starting salaries

The supplied lecture lists starting salaries (in thousands of dollars) for 20 graduates:

34	29	27	39	41	28	32	37	35	36
23	31	33	34	29	27	35	29	30	32

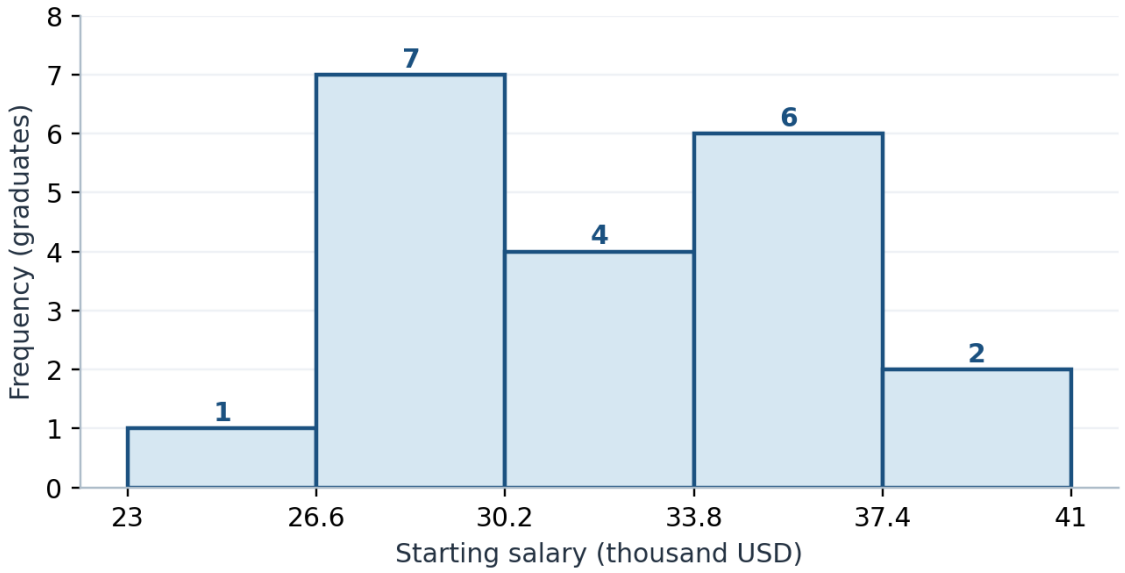
Minimum $= 23$, maximum $= 41$, and $R = 18$. Using five intervals yields width $h \approx 18/5 = 3.6$. Define the first four as left-closed/right-open, with the last including 41.

Salary class	f	$r = f/20$	Cumulative F
$[23.0, 26.6)$	1	0.05	1
$[26.6, 30.2)$	7	0.35	8
$[30.2, 33.8)$	4	0.20	12
$[33.8, 37.4)$	6	0.30	18
$[37.4, 41.0]$	2	0.10	20
Total	**20**	**1.00**	

5 Data Visualization: Histogram, Polygon and Ogive

5.1 Histogram

A histogram displays grouped numerical data using adjoining bars. With *equal class widths*, the bar height can represent the frequency. With *unequal class widths*, use frequency density (frequency divided by width) so bar *area* corresponds to frequency.



> Unlike an ordinary bar chart for categories, histogram bars are adjacent because the horizontal axis is a numerical scale. Read the axis boundaries carefully: a bar does not denote a single exact observation.

5.2 Frequency polygon

Plot (x_i, f_i) for each class midpoint x_i, connect adjacent points with straight lines and, when suitable, close the polygon toward zero frequency beyond the end classes. It is useful for comparing patterns among distributions.

5.3 Cumulative frequency curve (ogive)

A **less-than ogive** plots cumulative counts at the upper class boundaries. It begins at cumulative frequency 0 at the first lower boundary. A **more-than ogive** plots counts at or above class boundaries; it begins at n and falls to 0.

> For the salary data, the less-than cumulative frequencies are $0, 1, 8, 12, 18, 20$ at boundaries $23.0, 26.6, 30.2, 33.8, 37.4, 41.0$. The graph should be nondecreasing.

5.4 Reading graphs responsibly

- Inspect axis scales, units, sample sizes and class widths.
- Do not infer a probability model solely from a rough histogram shape.
- Grouping can hide local details; different bin choices may change the appearance.
- A graph does not explain why the observations have their observed pattern.

6 Measures of Central Tendency

6.1 Arithmetic mean

The **arithmetic mean** uses all observations and measures a numerical center. For n sample values $x_1, \ldots, x_n$,

$$\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}.$$

For a population of N observations, the mean is commonly denoted $\mu = (1/N)\sum_{i=1}^{N} x_i$. For a discrete frequency table or grouped midpoint approximation,

$$\boxed{\overline{x} \approx \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}}.$$

The expression is exact for a discrete-value frequency table; when x_i are class midpoints, it is generally an *approximation*.

Example A (source marks). For nine test scores

$$86, 97, 84, 73, 63, 88, 97, 100, 95,$$

the sum is 783 and the mean is $783/9 = \boxed{87}$ marks.

Example B (frequency-weighted mean). Given values $x = 5, 6, 7, 8, 9$ with frequencies $f = 4, 8, 14, 11, 3$, compute

$$\sum f = 40, \qquad \sum fx = 20 + 48 + 98 + 88 + 27 = 281,$$

so $\overline{x} = 281/40 = \boxed{7.025}$.

Sensitivity to outliers. The mean reacts to extreme measurements. The values $4.2, 4.3, 4.7, 4.8, 5.0, 5.1, 9.0$ have mean 5.3, but removing the last (extreme) observation gives a mean near 4.7. Excluding observations, however, requires a justified rule and must not be used to conceal valid data.

6.2 Median for ungrouped data

The **median** is the central value after sorting, with two cases:

$$\text{Med} = \begin{cases} x_{((n+1)/2)}, & n \text{ odd}, \\ \frac{1}{2}\left(x_{(n/2)} + x_{(n/2+1)}\right), & n \text{ even}, \end{cases}$$

where $x_{(j)}$ means the jth observation in ascending order.

Example. The sorted nine marks are

$$63, 73, 84, 86, \mathbf{88}, 95, 97, 97, 100.$$

The fifth value is in the center; therefore Med $= \boxed{88}$. For $[3, 5, 8, 12]$, the median is $(5 + 8)/2 = 6.5$.

6.3 Median for grouped data

Identify the **median class** as the first class whose cumulative frequency reaches or exceeds $n/2$. For an equal-width grouped table,

$$\boxed{\widetilde{x} \approx L + \frac{n/2 - F_{\text{prev}}}{f_{\text{med}}}\, h},$$

where L is its lower class boundary, F_{prev} is the cumulative frequency preceding the median class, f_{med} is its frequency and h is the class width. The result assumes approximately uniform distribution within that class.

6.4 Mode for ungrouped and grouped data

The **mode** is the most frequent observed value. A dataset can have no repeated values, one mode, or more than one mode. For a grouped table, the class with the largest frequency is the **modal class**. Under equal class widths, a commonly used interpolation estimate is

$$\boxed{\text{Mode} \approx L + \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})}\, h},$$

where f_m is the modal-class frequency and f_{m-1}, f_{m+1} are adjacent frequencies.

The *midpoint of the modal class* is a rough representative value, **not** the general formula for the grouped mode. The above interpolation is an estimate and requires care for open-ended classes or unequal widths.

6.5 Mean, median and mode compared

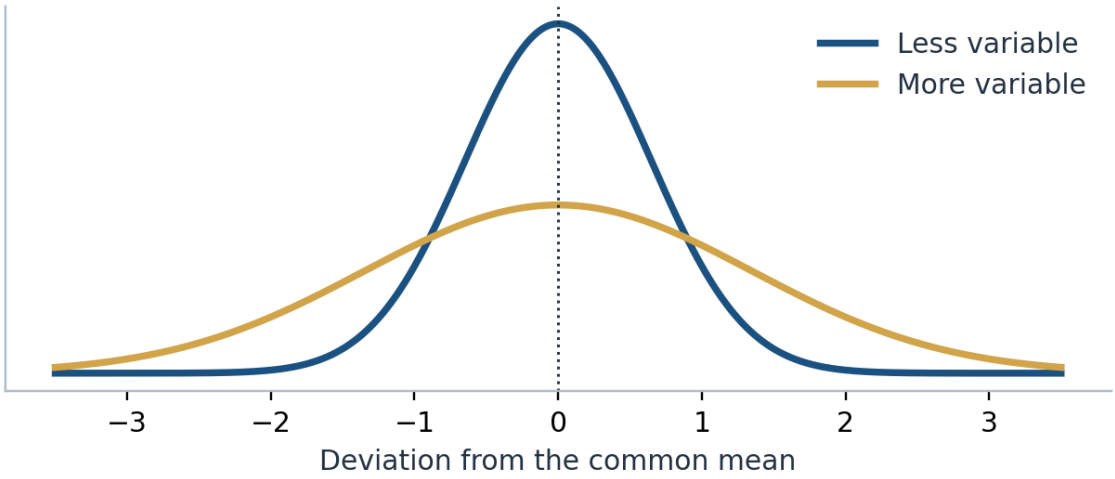
Measure	Major strength	Limitation
Mean	Uses every quantitative observation	Sensitive to outliers
Median	Resistant to extreme values	Ignores exact distance from most observations
Mode	Useful for categories and typical value	Can be ambiguous or unstable

For a unimodal, approximately symmetric distribution, mean, median and mode may be close. *However, exact equality of the three does not by itself prove symmetry.* The empirical relationship Mode $\approx 3\,$Median $- 2\,$Mean is only a rough rule for certain moderately skewed distributions, not an identity.

7 Measures of Dispersion

7.1 Why dispersion matters

A mean alone does not explain how tightly measurements cluster. Two sets can have the same mean but very different spread. Dispersion quantifies this variability.



7.2 Range

$$\boxed{R = x_{\max} - x_{\min}}.$$

For data with minimum 35 and maximum 48, the range is $48 - 35 = 13$. The range is easy to find but is especially sensitive to the two extreme values.

7.3 Mean absolute deviation (MAD)

The mean absolute deviation about the sample mean is

$$\boxed{\text{MAD} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{x}|}.$$

For grouped observations represented by midpoints, replace this by $\sum f_i |x_i - \overline{x}| / \sum f_i$ (an approximation).

Example. For $5, 9, 16, 17, 18$,

$$\overline{x} = \frac{65}{5} = 13, \quad |x_i - 13| = 8, 4, 3, 4, 5.$$

Hence MAD $= (8 + 4 + 3 + 4 + 5)/5 = \boxed{4.8}$.

7.4 Population variance and standard deviation

Variance uses squared deviations. For a **population**,

$$\boxed{\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}, \qquad \boxed{\sigma = \sqrt{\sigma^2}}.$$

7.5 Sample variance and standard deviation

For a **sample**, the customary unbiased estimator of the population variance is

$$\boxed{s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}, \qquad \boxed{s = \sqrt{s^2}}.$$

The denominator $n-1$ is used for sample variance; the population formula uses N. The unit of variance is the square of the original unit, whereas standard deviation has the original unit.

Example. Using $5, 9, 16, 17, 18$ and $\overline{x} = 13$, the squared deviations sum to

$$64 + 16 + 9 + 16 + 25 = 130.$$

Population variance: $\sigma^2 = 130/5 = 26$ and $\sigma \approx 5.10$.
Sample variance: $s^2 = 130/4 = 32.5$ and $s \approx 5.70$.

7.6 Variance for a frequency table

With frequency f_i and midpoint x_i, let $n = \sum_i f_i$ and $\overline{x} = \sum_i f_i x_i / n$. Then

$$\boxed{s^2 \approx \frac{\sum_i f_i (x_i - \overline{x})^2}{n-1}}.$$

A convenient computational form is

$$\boxed{s^2 \approx \frac{\sum_i f_i x_i^2 - \frac{(\sum_i f_i x_i)^2}{n}}{n-1}}.$$

Use $n-1$ when treating grouped data as a *sample*; use n for a complete population. Grouped variance based on midpoints is approximate because within-class differences are lost.

8 Standard Deviation, CV and Standard Scores

8.1 Worked grouped-variance example (50 test scores)

The original lecture includes a frequency distribution summarized below.

Class (marks)	Frequency f	Midpoint x	fx	$f(x-34.6)^2$
21–25	4	23	92	538.24
26–30	9	28	252	392.04
31–35	16	33	528	40.96
36–40	10	38	380	115.60
41–45	10	43	430	705.60
46–50	1	48	48	179.56
Totals	**50**		**1,730**	**1,972.00**

$$\begin{aligned} \overline{x} &= \frac{1,730}{50} = 34.6, \\ s^2 &\approx \frac{1,972}{49} = 40.2449, \\ s &\approx \sqrt{40.2449} = \boxed{6.344}. \end{aligned}$$

8.2 Coefficient of variation

The **coefficient of variation** compares relative spread among positive, ratio-scale measurements with meaningful nonzero means:

$$\boxed{\text{CV} = \frac{s}{|\overline{x}|} \times 100\%}.$$

Using the example above,

$$\text{CV} = \frac{6.344}{34.6} \times 100\% = \boxed{18.34\%}.$$

> **Correction to source calculation.** A value near 1.83 is not the correct percentage: the CV is about 18.34%. CV becomes difficult to interpret near a mean of zero, and it is generally unsuitable for interval scales without a meaningful zero.

8.3 Standard scores (z-scores)

A z-**score** represents distance from the mean in units of standard deviation:

$$\boxed{z = \frac{x - \mu}{\sigma}} \quad \text{or} \quad \boxed{z = \frac{x - \overline{x}}{s}}.$$

The first formula uses population parameters; the second standardizes with sample summaries.

> **Example.** For a distribution with mean 50 and standard deviation 10, a measurement $x = 65$ has
>
> $$z = (65 - 50)/10 = \boxed{1.5}.$$
>
> The value lies 1.5 standard deviations above the mean. A z-score by itself is *not* a probability: computing a tail probability requires an assumed or estimated distribution.

8.4 Engineering illustration (added): compare unlike units

The source gives five weights (kg) and five lengths (cm) for material blocks:

Weight	3.52	3.14	3.35	3.42	3.23
Length	20.36	20.15	20.52	20.20	19.68

Using sample standard deviations, the weights have $\overline{x}_w = 3.332$ kg, $s_w \approx 0.151$ kg, and $\text{CV}_w \approx 4.52\%$. Lengths have $\overline{x}_l = 20.182$ cm, $s_l \approx 0.316$ cm, and $\text{CV}_l \approx 1.57\%$. In *relative* terms, the weights vary more. Directly comparing 0.151 kg with 0.316 cm is not meaningful because their units differ.

9 Comprehensive Grouped-Data Case Study

9.1 Test-score distribution of 150 students

The source file provides the following grouped mathematical scores. This section consolidates its several example slides into one self-contained problem.

Class	Frequency f	Midpoint x	Cumulative F	fx	fx^2
10–19	1	14.5	1	14.5	210.25
20–29	6	24.5	7	147.0	3,601.50
30–39	9	34.5	16	310.5	10,712.25
40–49	31	44.5	47	1,379.5	61,387.75
50–59	42	54.5	89	2,289.0	124,750.50
60–69	32	64.5	121	2,064.0	133,128.00
70–79	17	74.5	138	1,266.5	94,354.25
80–89	10	84.5	148	845.0	71,402.50
90–99	2	94.5	150	189.0	17,860.50
Total	**150**			**8,505**	**517,407.50**

9.2 Step 1: grouped mean

$$\boxed{\overline{x} \approx \frac{\sum fx}{n} = \frac{8,505}{150} = 56.7 \text{ marks}}.$$

9.3 Step 2: grouped median

Since $n/2 = 75$, the median lies in the class 50–59 (F moves from 47 to 89). Here $L = 49.5$, $F_{\text{prev}} = 47$, $f_{\text{med}} = 42$ and $h = 10$:

$$\boxed{\widetilde{x} \approx 49.5 + \frac{75 - 47}{42}(10) = 56.17 \text{ marks}}.$$

9.4 Step 3: grouped mode

The highest class frequency is 42 in 50–59. Adjacent frequencies are 31 and 32:

$$\boxed{\text{Mode} \approx 49.5 + \frac{42 - 31}{(42 - 31) + (42 - 32)}(10) = 54.74}.$$

> The midpoint 54.5 is a coarse modal-class representative, not the interpolation-based estimate. The more precise estimate above is 54.74 under the adopted grouped-data model.

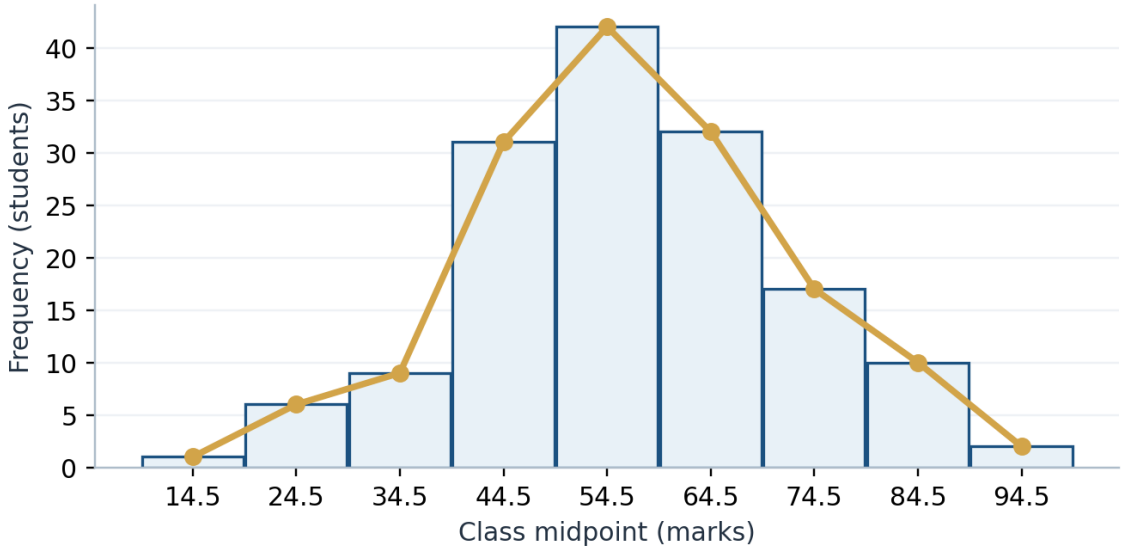
9.5 Step 4: sample variance and standard deviation

The computational formula gives

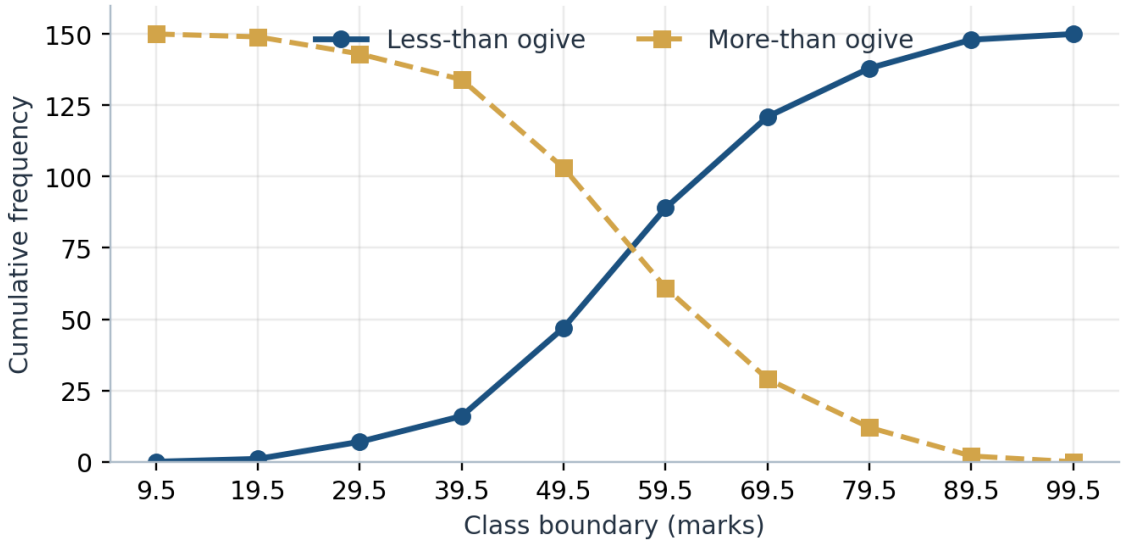
$$s^2 \approx \frac{517,407.5 - (8,505)^2/150}{149} = \boxed{236.067},$$
$$s \approx \sqrt{236.067} = \boxed{15.364 \text{ marks}},$$
$$\text{CV} \approx \frac{15.364}{56.7} \times 100\% = \boxed{27.10\%}.$$

These are approximate because the raw within-class observations are unavailable.

9.6 Step 5: plot and interpret the distribution



The modal class is 50–59; most observations are located in the central score classes. The polygon uses class midpoints. The plot is descriptive and cannot, by itself, justify a particular theoretical probability distribution.



The less-than ogive climbs from 0 to 150 as class boundaries increase, while the more-than ogive declines from 150 to 0. Both use *class boundaries*, not class midpoints. The ascending curve makes the approximate median visible at cumulative frequency 75.

10 How to Interpret Statistical Results

A complete interpretation answers four questions: **What was measured? How was it sampled? How large is the typical value? How much does it vary?**

10.1 A practical reporting template

A defensible engineering sentence includes the variable and unit, n, center, spread and any key limitations. For example:

> *Based on 50 grouped test scores, the estimated mean was 34.6 marks and the estimated sample standard deviation was 6.34 marks. Because class midpoints were used, the values are approximate.*

10.2 Common mistakes to avoid

1. **Using the wrong variance denominator:** population and sample standard deviations are not interchangeable.
2. **Forgetting the percentage conversion:** CV $= 0.1834$ is 18.34%, not 1.834%.
3. **Confusing class midpoint with grouped mode:** the highest-frequency class identifies the modal class, not an exact observed value.
4. **Missing endpoint conventions:** adjacent classes must not double-count boundary observations.
5. **Confusing samples and populations:** inference requires consideration of sampling and uncertainty.
6. **Using tables without units:** always label axes, measurement scales and results.

11 Exercises and Self-Assessment

11.1 Conceptual questions

1. Explain why a *packet count* is discrete while a measured *propagation delay* is continuous.
2. Distinguish a *population parameter* from a *sample statistic* using an engineering example.
3. Why does a histogram represent numerical class intervals but a bar chart usually represents categories?
4. Can two different datasets have the same mean and different standard deviations? Explain.
5. Does $z = 2$ automatically mean the observation has a 2% chance of occurring? Why not?

11.2 Calculation problems

1. Find the mean, median, mode, range and sample standard deviation of $[3, 4, 4, 6, 8]$.
2. The delays $[9, 11, 12, 14, 14]$ ms are observed. Compute the sample mean, sample variance and sample CV.
3. For the grouped scores below, estimate the mean, median and mode using equal-width grouped-data formulas:

Class	0–9	10–19	20–29	30–39	40–49	50–59
Frequency	2	3	7	12	8	3

4. For the grouped latency distribution $[0, 2), [2, 4), [4, 6), [6, 8)$ with frequencies $2, 5, 9, 4$, construct the frequency, relative-frequency and cumulative-frequency columns. Describe the correct endpoints for a less-than ogive.

5. The source's final homework table gives classes 155–159, 160–164, 165–169, 170–174, 175–179, 180–184, 185–189, 190–194 with frequencies $35, 42, 47, 54, 41, 35, 28, 18$. Construct its grouped-data summary and three types of plot (histogram, polygon and less-than ogive).

11.3 Answers to the first four calculation problems

1. For $[3, 4, 4, 6, 8]$: mean $= 5$, median $= 4$, mode $= 4$, range $= 5$ and sample variance $= (4+1+1+1+9)/4 = 4$, so $s = \boxed{2}$.

2. The delay sum is 60 and $\overline{x} = 12$ ms. Squared deviations: $9+1+0+4+4 = 18$, so $s^2 = 18/4 = 4.5$ ms^2; $s = \sqrt{4.5} = 2.121$ ms; CV $\approx 17.68\%$.

3. Total $n = 35$; midpoint weighted sum $\sum fx = 1,157.5$; mean ≈ 33.07. Median class is 30–39, so $\widetilde{x} = 29.5 + (17.5 - 12)10/12 \approx 34.08$. Modal class is 30–39, so mode $\approx 29.5 + (12-7)10/[(12-7)+(12-8)] = 35.06$. These are grouped estimates.

4. $n = 20$; $f = [2, 5, 9, 4]$; relative frequencies $[0.10, 0.25, 0.45, 0.20]$; cumulative frequencies $[2, 7, 16, 20]$. A less-than ogive uses coordinates $(0, 0), (2, 2), (4, 7), (6, 16), (8, 20)$.

Independent practice. For Exercise 5, confirm first that $\sum f = 300$ and use continuous class boundaries beginning at 154.5. Always label each graph's axes and units.

12 Formula Reference Sheet

Quantity	Formula or definition
Relative frequency	$r_i = f_i/n$; percentage $= 100 f_i/n\%$
Cumulative frequency	$F_j = \sum_{i=1}^{j} f_i$
Range	$R = x_{\max} - x_{\min}$
Sample mean	$\bar{x} = \sum_i x_i/n$
Frequency mean	$\bar{x} \approx \sum_i f_i x_i / \sum_i f_i$
Grouped median	$\tilde{x} \approx L + [(n/2 - F_{\text{prev}})/f_{\text{med}}]h$
Grouped mode	$L + [d_1/(d_1 + d_2)]h$, $d_1 = f_m - f_{m-1}$, $d_2 = f_m - f_{m+1}$
Mean absolute deviation	MAD $= \sum_i \lvert x_i - \bar{x}\rvert/n$
Population variance	$\sigma^2 = \sum_i (x_i - \mu)^2/N$
Sample variance	$s^2 = \sum_i (x_i - \bar{x})^2/(n-1)$
Grouped sample variance	$s^2 \approx [\sum f_i x_i^2 - (\sum f_i x_i)^2/n]/(n-1)$
Standard deviation	$s = \sqrt{s^2}$ or $\sigma = \sqrt{\sigma^2}$
Coefficient of variation	CV $= 100 s/\lvert\bar{x}\rvert\%$ (suitable nonzero ratio-scale mean)
Standard score	$z = (x - \mu)/\sigma$ or $z = (x - \bar{x})/s$

Symbol key. x_i denotes a value or class midpoint; f_i denotes its frequency; n denotes sample size; N denotes population size; L is the lower class boundary; h is the class width; F_{prev} is cumulative frequency preceding a selected class; μ and σ denote population mean and standard deviation; $\bar{x}$ and s denote sample mean and standard deviation.

Source and adaptation note. These classroom notes are an edited, corrected and reorganized adaptation of the uploaded 102-page *Introduction to Statistics* source. Numerical practice datasets and major concepts are retained, with corrections where mathematically necessary. Any *engineering illustration (added)* is newly written for the Communications Engineering Technologies audience. This is an introductory *statistics* unit, not a substitute for probability-theory lectures.