Table of Contents

# 2.1 Boxplot, Skewness, Kurtosis & Descriptive Statistics

## Descriptive statistics

Descriptive statistics summarise key features of a dataset using measures of central tendency and dispersion (see Unit 1). They help us understand the distribution’s shape, centre and spread.

## Boxplot

A **box‑and‑whisker plot** graphically summarises the distribution of a quantitative variable【444968039617338†L432-L446】. Its components are:

* The **median** (Q2) shown as a horizontal line inside the box.
* The **lower (Q1) and upper (Q3) quartiles**, forming the edges of the box.
* **Whiskers** extending to the smallest and largest non‑outlier values.
* **Outliers** plotted individually beyond the whiskers【444968039617338†L432-L446】.

Boxplots quickly convey skewness, spread and potential outliers【444968039617338†L448-L474】. They are ideal for comparing distributions across groups.

## Skewness

Skewness measures the asymmetry of a distribution. For a symmetric distribution, skewness equals zero【594094328782496†L15-L23】. Negative skewness indicates a long left tail; positive skewness indicates a long right tail.

## Kurtosis

Kurtosis quantifies the heaviness of the tails relative to the normal distribution. A normal distribution has kurtosis of 3【594094328782496†L67-L81】. Distributions with **high kurtosis** (heavy tails) have more extreme values than normal; those with **low kurtosis** (light tails) have fewer extremes.

## Example

Imagine plotting the exam scores of students from two classes using boxplots. One class might exhibit a longer upper whisker, indicating a few very high scores (positive skew), while the other might show a symmetric box, indicating balanced scores.

## Summary

Boxplots, skewness and kurtosis provide a succinct description of a dataset’s shape and spread. They help detect outliers and highlight asymmetry and tail behaviour【444968039617338†L432-L446】【594094328782496†L15-L23】.

## Reflection questions

1. What information can you glean from a boxplot that you might miss in a histogram?
2. How does positive skewness affect the mean relative to the median?
3. Explain the difference between kurtosis and skewness.

## References