Statistics and Business Decisions

Lecture 1

QBA201 – Quantitative Business Analysis

Shi Qiu – Fall 2026

Your starting point

A quick, ungraded check

Which description fits your experience with statistics?

1. This is my first statistics course.
2. I know a few ideas, such as averages and percentages.
3. I have studied statistics before.
4. I have used statistics in a course or at work.

Discuss: What would help you learn most effectively this semester?

The course begins with a question

1. Describe the data we observe.
2. Use probability to describe uncertainty.
3. Learn about a population from a sample.
4. Explain the result in the language of the decision.

A calculation is useful when we know the question it answers.

Descriptive and inferential statistics

Descriptive

Summarize the data actually observed. Example: the average mark among last year's students.

Inferential

Use sample data to learn about a wider population, allowing for uncertainty. Example: estimate average spending among all customers from a customer sample.

First identify which group the statement describes.

One small set of illustrative marks

These six marks are a constructed teaching example:

$$55, \quad 65, \quad 70, \quad 70, \quad 80, \quad 90.$$

Mean

$$\frac{55 + 65 + 70 + 70 + 80 + 90}{6} = 71.67$$

Median and range

Median: $(70 + 70)/2 = 70$.

Range: $90 - 55 = 35$ marks.

The center and the spread describe different features of the same data.

Numerical summaries answer specific questions

Question	Useful summary
Where is the center?	Mean or median
How much do observations vary?	Range, variance or standard deviation
How common is a category?	Count or proportion

Choose the summary after stating what you want to learn.

A business decision: campus soft-drink sales

A university proposes an exclusive sales agreement to a soft-drink company.

- The university has about 50,000 students.
- The company would receive exclusive campus sales rights.
- The university would receive 35% of sales revenue and an annual fee of R200,000.

Discuss: What missing information would determine whether the agreement is attractive?

The quantities we know

Quantity	Value
Current sales	22,000 cans per week
Operating year	40 weeks
Selling price	R5.00 per can
Variable cost	R1.50 per can
University revenue share	35%
Annual fixed fee	R200,000

The missing quantity is total campus demand.

An assumed market share gives a demand estimate

Suppose current sales represent 25% of campus purchases.

$$\text{Weekly market} = \frac{22,000}{0.25} = 88,000 \text{ cans.}$$

$$\text{Annual sales} = 88,000(40) = 3,520,000 \text{ cans.}$$

$$\text{Annual revenue} = 3,520,000(5) = \text{R}17,600,000.$$

This is a conditional calculation. The 25% market share is an assumption.

The proposed agreement: profit calculation

Item	Rand per year
Revenue retained: $0.65(17,600,000)$	11,440,000
Less variable cost: $1.50(3,520,000)$	$-5,280,000$
Less annual fixed fee	$-200,000$
Profit under the stated assumptions	5,960,000

R200,000 is R0.2 million. Keep units and place values consistent.

Compare the proposal with current profit

Current profit:

$$40(22,000)(5-1.50) = \text{R}3,080,000.$$

Conditional increase under the proposal:

$$5,960,000 - 3,080,000 = \text{R}2,880,000.$$

Discuss: Which assumptions would you challenge before recommending the agreement?

A calculation can be precise while its input is uncertain

The arithmetic used a market share of 25%. We have not established that value.

- Students may change brands or buy off campus.
- The agreement may change prices or demand.
- A busy week's sales may not represent a typical week.

Separate a checked calculation from an assumption that still needs evidence.

A survey supplies evidence about demand

Ask 500 students to record their soft-drink purchases for seven days.

- One observation is one student's weekly can count.
- The sample mean summarizes those 500 students.
- The business question concerns all 50,000 students.

Discuss: Would surveying only students beside a drinks machine give a representative picture?

Population and sample

Population

All units relevant to the question.
In the campus case: all 50,000 students during the period of interest.

Sample

The units selected for observation.
In the campus case: the 500 students surveyed.

Define the population using the question, location and time period.

All oil tankers: a population or a sample?

Suppose a dataset contains transponder records for every oil tanker.

Discuss: What is the target group, and does the dataset cover it?

The target matters

It is a census of oil tankers only if every relevant tanker and period are covered.

For a question about all ships, oil tankers are only a subset.

Parameter and statistic

Term	Meaning
Parameter	A numerical characteristic of the population.
Statistic	A numerical summary calculated from a sample.

The population mean μ is a parameter.

The surveyed students' mean $\bar{x}$ is a statistic.

The same question at two scales

	Population	Sample
Group	50,000 students	500 surveyed students
Weekly mean	μ	$\bar{x}$
Role	Unknown target	Observed estimate

We calculate $\bar{x}$ from the sample to estimate μ for the population.

An inference needs a clear chain of reasoning

1. State the population quantity of interest.
2. Explain how the sample was selected.
3. Calculate the relevant sample statistic.
4. Quantify uncertainty and state the conclusion with its limitations.

Discuss: Which step would fail if the sample systematically excluded commuter students?

When can a sample support generalization?

- The sampled units must relate to the target population.
- The selection process should avoid systematic exclusions.
- Nonresponse and inaccurate measurements require attention.

A large biased sample can give a precise answer to the wrong population question.

Why an estimate needs uncertainty

Two properly selected samples can give different estimates.

An uncertainty statement describes how much the estimation procedure can vary across repeated samples.

- More data can reduce sampling variability.
- More data do not automatically remove selection or measurement bias.

What a 95% confidence level means

Imagine repeatedly drawing a sample and calculating an interval by the same valid method.

About 95% of those intervals would contain the fixed population parameter.

The 95% describes the long-run coverage of the interval procedure.

What a significance level means

A hypothesis test starts with a stated null claim about the population.

Its significance level α controls the probability of rejecting that claim *when the null is true*, under the test's assumptions.

A 5% significance level is not a 5% probability that every reported conclusion is wrong.

Reading a poll's uncertainty statement

Suppose a poll reports 52% support, with a margin of error of 3.4 percentage points at 95% confidence.

$$52\% - 3.4\% = 48.6\%, \qquad 52\% + 3.4\% = 55.4\%.$$

The reported interval

48.6% to 55.4%, for the stated target population and method.

Discuss: Why does this interval alone fail to establish majority support?

Using statistical software

Useful Excel functions

`AVERAGE`, `MEDIAN`, `STDEV.S` and `COUNTIF` summarize observed data.

Later, probability and inference functions will support more advanced questions.

Your responsibility

Identify the inputs, check units, choose the method and explain what the output answers.

Discuss: What would you check before trusting a spreadsheet's average?