

0 • Exam Map

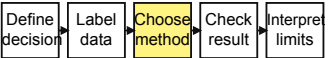
COURSE DATA · BRIEF

Quizzes 30%	Project 40%	Final 30%
COMPONENT	WEIGHT	FORMAT
Online Quizzes	30%	Individual online quizzes
Group-Based Analytics Project	40%	Group report + pre-recorded online presentation
Final Examination	30%	Restricted open book; one double-sided A4 memory aid

"This course has a **RESTRICTED OPEN BOOK** examination." "The memory aid is a single double sided A4 sheet of handwritten or typed notes for use during the examination." *University of Newcastle 2026 Handbook, BUSN1010*

1 • Answer Spine

ALL CHAPTERS



Exam sequence: decision → labels → method → independent check → bounded business claim.

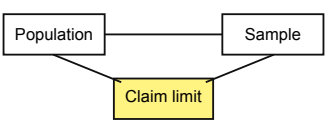
WRITE BEFORE CALCULATING	WRITE AFTER
Observational unit; variable and role; unit/category; population; comparison	Result in original units; uncertainty; practical consequence; scope limit
Sampling/model conditions; exact denominator	What changed data or design would change the action?

SAMPLE MEAN · DESCRIPTIVE ESTIMATE
 $\bar{x} = (1/n) \sum_{i=1}^n x_i$

More rows do not repair exclusion of the relevant population. Separate a sample description, a population claim and a causal claim.

EXAM MOVE Carry one labelled observation from raw row to final sentence; a changed unit or denominator reveals the first broken link.

SAMPLING GATE



Variable, population and sample must remain separate analytical roles.

Variable: a characteristic recorded for each unit that can take different values or categories. **Sample:** the observed population subset used to estimate or compare its features.

2 • Describe Without Hiding

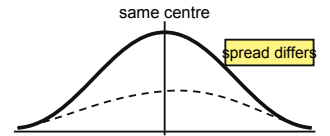
C2

Frequency distribution: counts or proportions across values, categories or intervals. It preserves category order or class meaning and exposes clustering, gaps, tails and unusual values.

Median: middle ordered observation; half at or below and half at or above. It resists distortion by a small number of extremes. **Standard deviation:** typical distance from the arithmetic mean, in the variable's original units.

SAMPLE STANDARD DEVIATION
 $s = \sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2 / (n - 1)]}$
Spread around $\bar{x}$; sample degrees-of-freedom divisor.

QUESTION	USE/CHECK	TRAP
Composition or categories?	Frequency table/display; preserve counts or proportions	Aggregation hides a group
Quantitative shape?	Distribution + centre + spread	Centre reported alone
Skew/extremes?	Median plus full shape	Mean treated as "typical"
Variability?	SD with mean and units	Different scales compared raw



Equal centres can conceal operationally important variation; inspect shape and spread together.

CHART/DATA AUDIT

BEFORE TRUSTING DISPLAY	ASK
Scale	Does the axis or binning change the apparent difference?
Rows	What is one observation? Are groups comparable?
Summary	Are centre, dispersion and shape all visible?
Scope	Observed records only, or a supported population claim?

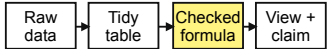
When summaries conflict: return to ordered values or the frequency display. Explain whether skew, gaps or rare extremes pull the mean away from the median; then state how the spread changes the operational comparison. Do not call an isolated unusual value an error without checking its source.

Report centre and dispersion together, in the original transaction, time, cost or demand units.

3 • Spreadsheet + Power BI

C3

Tidy data: each variable is a column, each observation a row, each value one cell. **Cell reference:** a formula address whose relative/absolute behaviour controls copying. **Filter context:** the active subset of data and relationships when a measure or visual is evaluated.



filter · type · aggregation
copied reference · traceability

Prepare a clean analysis table, control calculations, then turn the view into reproducible evidence.

FAILURE	DIAGNOSTIC CHECK
Dashboard changes after filter	Record active subset and relationships; recompute a small case
Copied formula drifts	Inspect relative/absolute row and column behaviour
Unexpected total	Check data type, grouping and aggregation level
Correct-looking visual	Trace every filter, formula and transformation to the question

STANDARD SCORE · SOFTWARE CROSS-CHECK
 $z = (x - \mu) / \sigma$
Distance from μ in population standard-deviation units.

Software output is not self-validating. Preserve raw data, document transformations, check calculations, and state what the visual permits a manager to conclude.

REPRODUCIBILITY TICKET

Question + unit	Raw source preserved
Variable types	Transformations logged
Formula copied correctly	Filter context named
Small case recalculated	Claim bounded

Declare → connect → check → conclude. Declare the data and decision scope; connect each formula or view to the business question; test the assumption most vulnerable to the active filter; conclude whether the action is intact, narrowed or requires a different choice.

For a dashboard claim, preserve the labels and denominator through every recode and subset. Compare the unfiltered result with the filtered view, then state whether the difference is intended business segmentation or an analysis defect.

EXAM MOVE If a value changes, isolate whether the cause is the business subset, formula, data type or aggregation before interpreting it.

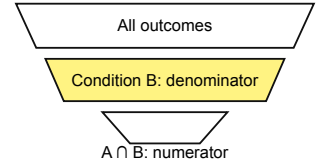
4 • Probability Rules

C4

Event: specified set of outcomes. Name events in words before symbols. A complement excludes the event; a union answers either-event; an intersection answers both-event. Remove overlap when events are not disjoint.

CONDITIONAL PROBABILITY
 $P(A | B) = P(A \cap B) / P(B)$
Condition B fixes the reference set and denominator.

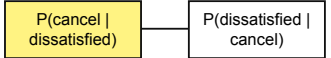
Independence: learning that one event occurred does not change the probability of the other. Treat independence as a model claim to be checked, not a synonym for disjoint events.



Condition first, then count the joint event inside that restricted group.

WORDS IN QUESTION	SET OPERATION	CHECK
not A	Complement	Everything outside A
A or B	Union	Remove shared overlap
A and B	Intersection	Both conditions hold
A given B	Conditional	Denominator is B
Independent?	Compare conditional and unconditional	Does knowing B alter A?

DO NOT REVERSE THE QUESTION



different denominators
→ different questions

Reversing the condition is invalid unless base rates and the joint event justify equality.

Transfer check: change one feature of the process, then recheck the event definitions, denominator and independence claim. Preserve conditions that did not change; revise only the calculation or interpretation that depends on the changed fact.

EVIDENCE	CLAIM
P(A B) differs from P(A)	Knowing B changes A; independence is not supported
Condition reversed	A new question; recalculate using the new reference group

5 • Random Variables

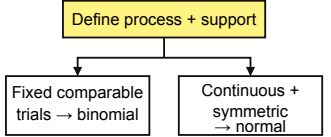
C5

Random variable: numerical mapping from uncertain outcomes to values. Define the value, support and process before choosing a distribution. **Expected value:** probability-weighted long-run average under the stated model.

EXPECTED VALUE
 $E(X) = \sum x P(X = x)$
Keep the payoff/cost/demand unit attached to each x.

NORMAL STANDARDISATION
 $z = (x - \mu) / \sigma$
Positive z is above μ ; magnitude is distance in σ units.

MODEL	USE WHEN	ASSUMPTION ALARM
Binomial count	Fixed number of comparable trials; X counts successes	Changing success probability or dependent outcomes
Normal	Continuous, symmetric bell-shaped process specified by μ and σ	Small discrete count or wrong process mechanism
General discrete X	Possible numeric outcomes and probabilities are stated	Probabilities/support incomplete



A familiar numerical fit cannot repair a distribution that mismatches the random mechanism.

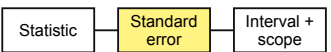
Interpret probability in business units. Expected value describes a long-run average, not a guaranteed single outcome.

6 • Estimation

C6

Sampling distribution: distribution of a statistic across repeated samples from the same sampling process. **Standard error:** its standard deviation.

CONFIDENCE INTERVAL FOR MEAN
 $\bar{x} \pm z\alpha/2 \cdot \sigma/\sqrt{n} = \text{estimate} \pm \text{critical value} \times \text{SE}$



Estimator → standard error → interval, interpreted only for the supported population.

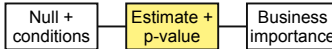
7 • Hypothesis Tests

C7

Null hypothesis: parameter claim used as the reference model. State null and alternative in parameter language; set the decision rule before calculation; then check conditions and compute.

ONE-SAMPLE TEST STATISTIC
 $z = (x^- - \mu_0) / (\sigma/\sqrt{n})$
Observed difference from μ_0 , scaled by the null standard error.

P-value: assuming the null model, probability of a statistic at least as incompatible with that model as observed. It is **not P(H₀ true)** and not the chance that a business action succeeds.



report as three distinct claims

Statistical evidence does not replace effect size, cost or operational importance.

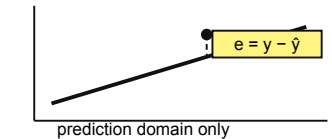
STAGE	WHAT EARNS THE LINE
Frame	Parameter, H ₀ /H ₁ , direction, population
Conditions	Sampling/model assumptions and denominator
Evidence	Estimate, uncertainty, test statistic, p-value
Decision	Apply the preset rule; use cautious language
Action	Compare effect with business consequences

8 • Regression + Forecasts

C8

SIMPLE LINEAR PREDICTION
 $\hat{y} = b_0 + b_1x$ · residual $e = y - \hat{y}$

Correlation is unit-free direction/strength of linear association. **Slope b₁** is fitted response/change per one predictor unit, within the model. Inspect residual structure and extremes; state the prediction domain.



Plot, interpret coefficients in units, inspect residuals, then limit prediction and causal language.

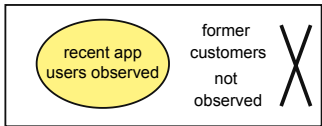
REVIEW	BOUNDARY
Association	Does not assign cause
Intercept	Interpret only if x = 0 is meaningful
Slope	Attached to response/predictor units and fitted design
Forecast	Separate trend, recurring seasonality and horizon
Extrapolation	Needs added stability/design evidence



W1 · Sampling Claim

C1 WORKED

Case. A retailer wants to explain declining repeat purchases but has a convenience sample of recent app users and no observations from former customers.



sample ≠ target population

Selection coverage is the first broken link; extra rows from the same frame do not repair it.

STEP	WORKED RESOLUTION
1 Define	Decision: explain declining repeat purchases. Unit: customer. Name variable roles and units.
2 Check	Population relevant to repeat purchases includes former customers; the convenience sample excludes them.
3 Conclude	Describe observed recent app users only; do not generalise explanation to all customers or claim cause.
4 Change needed	A sampling design covering the relevant population, including former customers.

IF A QUANTITATIVE VARIABLE IS OBSERVED

$\bar{x} = (1/n)\sum x_i$ describes the sample; representativeness governs population scope.

ANSWER LINE “The observed sample can describe __, but it cannot support __ because __ is excluded.”

SAMPLING ERROR VS DESIGN ERROR

RANDOM UNCERTAINTY	SELECTION/MEASUREMENT LIMIT
May shrink with an appropriate larger sample	Can remain even with many rows
Quantified by SE/CI under assumptions	Requires changed coverage, measurement or design

Always state which error your method addresses.

ONE-LINE DEFINITIONS

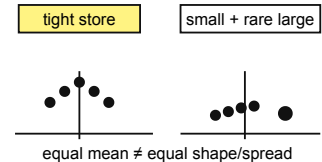
SE	SD of a statistic's sampling distribution
CI	Repeated-sampling range designed to capture a parameter at a stated long-run rate
Slope	Fitted response change per predictor unit
Residual	Observed response minus fitted response
Practical significance	Operational/economic importance of an estimated effect

When unsure, identify the first unsupported move and repair only the inference that depends on it.

W2 · Same Mean, Different Risk

C2 WORKED

Case. Two stores report the same average transaction value. One has a tightly grouped customer mix; the other combines many small purchases with rare large orders.



Compare the distribution, median and standard deviation before treating the stores as alike.

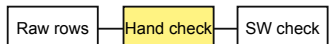
STEP	WORKED RESOLUTION
1 Display	Build a frequency display that preserves transaction scale and unusual values.
2 Centre	The shared mean alone is insufficient; report median where extremes distort “typical”.
3 Spread	Compute and interpret SD in transaction-value units; compare clustering and tail.
4 Decide	Operational variability differs even if average transaction value matches.

SPREAD CHECK

$s = \sqrt{[\sum (x_i - \bar{x})^2 / (n - 1)]}$

TRAP “Same average” is not “same customer mix”: A centre without shape and spread can hide the managerial difference.

SPREADSHEET VERIFICATION



Recalculate a small case; inspect formula references, filters, types and aggregation if results disagree.

Preserve raw data	Document recoding
Name filter context	Trace copied references
Check units	Bound final claim

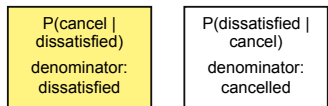
Debug in order: verify tidy rows and data types; test the copied reference; record the filter; inspect aggregation; reproduce one small total. Change one feature at a time.

IF THE NUMBER CHANGES WHEN...	INVESTIGATE
A filter is applied	Subset and relationships in filter context
A formula is filled down/across	Relative versus absolute reference behaviour
A visual groups records	Observation level and aggregation

W3 · Conditional Direction

C4 WORKED

Case. A service team confuses the chance that a dissatisfied customer cancels with the chance that a cancelled account belonged to a dissatisfied customer.



same intersection
different reference group

The numerator may involve the same joint event, but the denominator changes the meaning.

WRITE THE CONDITION FIRST

$P(A \mid B) = P(A \cap B) / P(B) \neq P(B \mid A)$ in general

STEP	WORKED RESOLUTION
1 Name	A = cancellation; B = dissatisfaction. Keep the outcome words attached.
2 Restrict	For “cancel given dissatisfaction”, restrict the reference set to B.
3 Count	Use the joint event $A \cap B$ as numerator and B as denominator.
4 Interpret	Read result as a proportion within dissatisfied customers; do not reverse.
5 Independence	Ask whether knowing dissatisfaction changes cancellation probability.

ANSWER LINE “Among customers in B, the probability of A is __.” The opening phrase protects the denominator.

PROBABILITY ERROR CLINIC

ERROR	REPAIR
Adds either-event counts with overlap	Identify and remove intersection
Multiplies automatically	Check whether conditioning/independence is justified
Treats disjoint as independent	Ask whether learning one event changes the other
Reverses A and B	Rewrite the conditioning group in words

Changed-assumption drill: if the events or reference group change, redraw the restricted set before recalculating. If only one condition changes, preserve unrelated facts and state whether the business action remains, narrows, or requires a different choice.

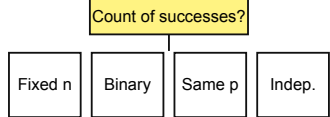
Among [condition B], the probability of [event A] is ...

Use the sentence as a final denominator check.

W4 · Choose the Random Model

C5 WORKED

Case. A manager treats a count of customer responses as normally distributed although the process has a small number of trials, changing success probabilities and dependent outcomes.



any failed condition:
do not force binomial or normal

Match the outcome definition and process assumptions before using a familiar distribution.

STEP	WORKED RESOLUTION
1 Define X	Number of customer responses; state its possible values and unit.
2 Diagnose	Small fixed count alone does not justify normal; changing p and dependence break comparable-trial logic.
3 Reject shortcut	Do not repair the wrong mechanism by a numerical fit.
4 Re-design	Specify a process/model whose assumptions match the trials before probability calculation.

IF OUTCOMES AND PROBABILITIES ARE VALID

$E(X) = \sum x P(X=x)$ · probability-weighted long-run average

IF A NORMAL MODEL IS JUSTIFIED

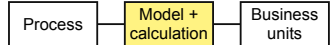
$z = (x - \mu) / \sigma$ · distance from μ in σ units

TRAP Expected value is a long-run centre, not the response guaranteed on the next trial.

DISTRIBUTION DECISION

OBSERVED VARIABLE	FIRST CHECK
Discrete count	Support, fixed trials, success definition, p, dependence
Continuous value	Shape, symmetry, μ/σ meaning, relevant domain
Payoff/cost table	All possible x and $P(X=x)$; probabilities match process

Interpretation ticket: name X and its units; justify the process; show the calculation; report probability, standardised distance or long-run average—not a guaranteed outcome.

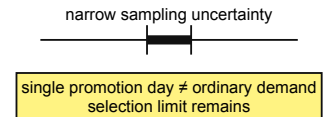


Define process → choose model → interpret the result in its original business unit.

W5 · Narrow Interval, Weak Scope

C6 WORKED

Case. A business reports a narrow confidence interval from a large transaction file, but the transactions come from a single promotional day and are not representative of ordinary demand.



precision is not representativeness

A tight interval can be precise for an irrelevant sampling process.

INTERVAL CALCULATION

$\bar{x} \pm z\alpha/2(\sigma/\sqrt{n}) = \text{estimate} \pm \text{critical value} \times \text{SE}$

STEP	WORKED RESOLUTION
1 Target	Parameter concerns ordinary demand, not only the promotional day.
2 Sampling model	Repeated samples must arise from the same relevant process; here coverage is misaligned.
3 Precision	Large file may reduce SE under its model and produce a narrow interval.
4 Boundary	CI does not absorb selection bias, measurement error or an irrelevant population.
5 Repair	Collect a representative sampling process for ordinary demand, then estimate.

ANSWER LINE “The interval is narrow under __, but the population claim fails because __.”

TEST INTERPRETATION DRILL

Case. A campaign team declares success because a test is statistically significant although the estimated improvement is too small to cover implementation cost.



both needed for a recommendation

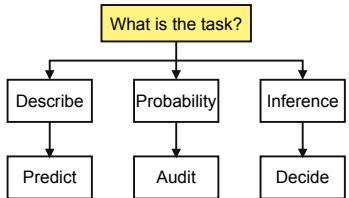
A test decision and a managerial decision answer different questions.

REPORT SEPARATELY	CORRECT BOUNDARY
Estimate + uncertainty	What effect was observed and how precisely?
P-value under H_0	Not the probability H_0 is true
Operational consequence	Does effect cover cost or matter in context?

9 · Method Chooser

FAST RECALL

IF ASKED TO...	CHOOSE / WRITE	FIRST TRAP
Describe observed quantitative values	Frequency shape + centre + spread; $\bar{x}$, median, s as appropriate	Centre alone
Standardise an observation	$z = (x - \mu) / \sigma$	Wrong μ/σ or model
Find long-run average of discrete X	$E(X) = \sum x P(X=x)$	Incomplete support/probabilities
Condition on an event	$P(A B) = P(A \cap B) / P(B)$	Reversing denominator
Estimate a population mean	$\bar{x} \pm \text{critical value} \times \text{SE}$	Selection/design error
Test a mean claim	H_0/H_1 , conditions, z, p-value, effect, consequence	P-value misread
Predict a response	$\hat{y} = b_0 + b_1x$; residual $y - \hat{y}$	Extrapolation/causal claim
Audit dashboard output	Tidy rows, references, filter context, aggregation, manual check	Trusting display

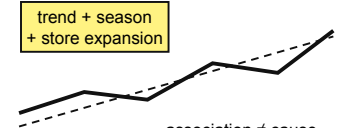


After choosing, name assumptions, calculate, cross-check and interpret within population/design scope.

10 · Regression + Forecast Clinic

C8

Scenario. Sales and advertising rise together over time, while the fitted line also absorbs expansion into new stores and a strong seasonal peak.



Time trend, recurring seasonality and store expansion can move with both variables.

WRITE	DO NOT WRITE
“Within the fitted data, one predictor unit is associated with b, response units.”	“Advertising caused b, more sales.”
Residual = observed – predicted; inspect pattern and extremes.	“A fitted line proves stability outside the observed range.”