CMCE10002 Foundations of Business Analytics

THE UNIVERSITY OF MELBOURNE · UNDERGRADUATE LEVEL 1 · SEMESTER 2, 2026

Full guide + ask Sia about CMCE10002

SIDE 1/2 EXAM REVISION SHEET · DEFINITIONS → RULES → METHODS → CHECKS FINAL: 2 HOURS · 50% v2 · 2026-10-25

0 · Exam Map COURSE.JSON · BRIEF

10	20	20	FINAL 50

ASSESSMENT	WEIGHT
In-class quizzes, held during workshops	10%
Mid Semester Test	20%
Group Assignment	20%
2-hour end of semester examination	**50%**

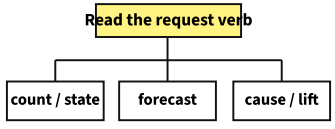
Permitted materials: not confirmed for this offering in the course materials we hold — check the exam instructions on the LMS.

Paper structure: hand written; **20 multiple-choice questions = 40%** and roughly **four short-answer questions with subparts = 60%**. Draws on lectures, workshops and assigned readings. The subject guide states that a pass in the final exam is required to complete the subject.

EXAM USE *For every code item, predict rows · columns · grain before the values. For every judgement item, write claim · limit · next check.*

1 · Name the Question First CH 1 · W1

TYPE	ASKS	EVIDENCE BAR
Descriptive	What happened?	clean, complete record; correct calculation
Predictive	What next?	performance on unseen rows
Causal	What produced it?	fair comparison with no-treatment outcome



same table ≠ same entitlement to claim

Question-type branch, redrawn from the guide: the verb fixes the evidence required.

SHORT-ANSWER STEM

```
"This is [type] because ___. It requires ___. The available data support ___, but not ___ because ___."
```

1b · Predictive Evidence CH 12 · W9–11

HONEST SEQUENCE

```
earlier rows: fit → later held-back rows: score → same later rows: score
trivial baseline
```

2 · Read R as Sentences CH 3 · W2–3

FORM	READ IT AS / RETURNS
`x ← value`	store right under name left
`c(a,b)`	combine same-type values → vector
`data$col`	take one column → vector
`data[rows, cols]`	subset; blank means keep all there
`==`	ask equality → TRUE/FALSE per value
`library()`	load package for this session
`read_csv()`	read file → tibble
`nrow(), ncol()`	one number each
`names()`	column-name vector

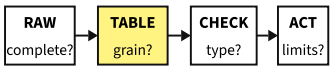
★ One = stores or names; two == ask. A filter needs the question form. Always ask what shape comes back: scalar, one value per row, or whole table.

3 · Grain Before Arithmetic CH 3

FOUR-QUESTION OPENING

```
1 How many rows? · 2 What are the columns? · 3 What does ONE row represent? · 4 What type is each column?
```

A row count counts records. If one customer has several orders, `nrow()` is not a customer count; count distinct customer identifiers. A row or column count of one after import can signal a misread separator. Confirm numeric types before adding or averaging; dates and identifiers often arrive as text.



Each layer discards information

gaps → grain → type → limits

Four layers and their failure modes, redrawn from the guide.

4 · Reproducible Workflow CH 2 · W2

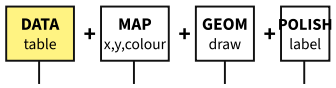
PLACE	WHAT SURVIVES
Console	nothing; for trying
Script	instructions
Quarto	instructions + explanation + output

TEST

```
clean session + this file + this data → this output; no manual step
```

If chunks work by hand but render fails, memory is hiding a dependency. Restart → run from top → read the first error. Usual causes: package loaded only in console or object created in an earlier experiment.

5 · Build a Plot in Four Parts CH 4 · W3



blank canvas = data only; no marks

adding layers ≠ proving the answer

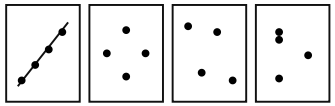
The ggplot grammar, redrawn from the guide: only the geometry draws marks.

QUESTION	GEOMETRY / PRE-STEP
count across categories	bar; height is tallied
one numeric distribution	histogram
two numeric variables	points
mean by week	**summarise first**, then plot

Mapping vs setting: put a column inside the aesthetic mapping to vary colour and create a legend; put a fixed constant outside to colour every mark the same. Labels and units are evidence, not decoration.

CHECK *Read the mapping back against the business question. A plot can be syntactically valid and answer the wrong quantity–especially count instead of mean.*

6 · Plot Before You Summarise CH 5 · W3



same mean + same SD; different shape

line · ring · cross · clumps

Guide principle: centre and spread cannot reveal clusters, bends or influential rows.

THREE-CLAUSE ANSWER

```
SUPPORTED: what the table does show.
HIDDEN: spread, shape, extreme rows or rival explanation.
NEXT: plot raw data or build the fairer comparison.
```

A mean locates a distribution; SD reports spread around its centre. Neither establishes shape or causality. Always carry n beside a grouped mean : an average over 2 rows prints like one over 400.

Exam response: do not stop at "correlation is not causation". Name the specific rival driver—price, category, brand strength or promotion in the guide's product example—and distinguish it from the separate problem of hidden within-band spread. Rival explanation needs a fairer comparison; hidden spread needs the underlying distribution.

Manager test: a useful answer finishes with the action that would settle the issue. "Supported, hidden, next step" is complete; a description alone is not.

7 · Six Verbs, Three Jobs CH 6 · W4

JOB	VERB	EFFECT
rows	`filter()`	keep rows meeting condition
	`arrange()`	reorder; descending is not default
columns	`mutate()`	add derived column
	`select()`	keep named/pattern/type columns
groups	`group_by()`	attach labels; visible shape unchanged
	`summarise()`	collapse to one row per group

Filter before ranking. Impossible or missing denominators otherwise rise to the top. Compute before narrowing. A dropped input cannot be used later. For comparable performance use a rate, not a size-driven dollar total.

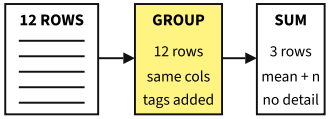
DERIVED MEASURE

```
return on invested capital = operating profit ÷ invested capital
guard: denominator > 0 before division
```

Read the verb by effect: filtering changes row count, selecting changes column count, arranging changes order only, mutating normally preserves rows and adds a column, grouping preserves the visible table, and summarising changes the grain. That effect is usually what a code-tracing question is really asking.

8 · Pipe + Grouped Summaries CH 7 · W4

Pipe = "then": pass the result on the left into the first slot on the right. It changes readability, not computation. Trace one line at a time.



Group tags; summary changes grain

ungroup before whole-table steps

Grouped-summary shape, redrawn from the guide's twelve-store example.

Wrong-level trap: if the question is average per customer but rows are orders, first collapse orders within customer, then customers within store. Skipping the first collapse answers an order-weighted question. Shares add a column to a summary; summarising again collapses to one row.

Pipe translation: read each connector as "then", not as a mathematical operator. Rewrite a confusing chain line by line and annotate the returned shape after every verb; nested calls and piped calls compute the same result.

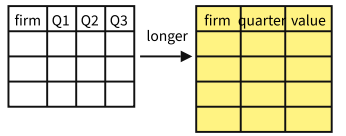
9 · Tidy Data + Pivots CH 8 · W5

TIDY CONTRACT

```
each column = one variable · each row = one observation · each cell = one value
```

IF...	DO...	WHY
variable trapped in headings	**lengthen**	period becomes values code can map/group/filter
variable names stored in cells	**widen**	each measure gets a column

WIDE: QTR COLS **LONG: FIRM-QTR**



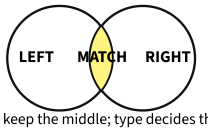
rows = old rows × gathered columns

Lengthening, redrawn from the guide: values stay; their address changes.

Pre-check: write the result's grain and predicted row count. Three firms across four quarters → **12 rows** after lengthening. Gather only columns carrying the same kind of value; sweeping in an identifier mixes types.

10 · Joins: Decide the Unmatched CH 9 · W5

JOIN	KEEPS	ORDER?
inner	matched keys only	no
left	every left row + matches	**yes**
anti	unmatched left rows only	**yes**
full	everything from either side	no



all joins keep the middle; type decides the sides

Three key regions, redrawn from the guide. Diagnose coverage before joining.

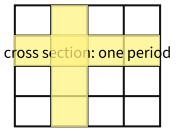
Match grain before key. Product-level data joined to product-week data repeat the product row once per week; totals then double-count. If output is larger than either input, key values are repeated and the join fanned out. Count rows per key on both sides first.

COVERAGE REPORT

```
unmatched count + unmatched fraction + which base was retained
```

Order test: inner and full joins retain the same key set whichever table is first. Left and anti joins are defined relative to the first table, so swapping inputs asks a different business question. Empty attached cells are evidence of no match, not a zero.

11 · Cross Section / Time / Panel CH 10 · W6



time series one unit across periods

panel every unit-period cell

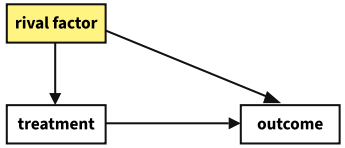
group away the untested dimension

Three readings of one grid, redrawn from the guide.

SHAPE	ONE ROW	AVERAGE MEANS
cross section	one unit	typical unit now
time series	one period	typical period for unit
panel	one unit-period	depends on collapse

Question compares stores → group by store, average time away. Question compares quarters → group by quarter, average stores away. Question asks which store bucks the trend → keep both dimensions . A period trapped in headings must be lengthened before it can be grouped.

12 · Causal Evidence CH 11 · W8



observed gap = effect + rival cause

Omitted-variable path, redrawn from the guide: association alone is not the treatment effect.

Potential outcomes: each unit has an outcome with treatment and without; the causal effect is their difference, but only one is observed. Model adjustment balances measured factors only. Random assignment makes arms alike on average, including on unmeasured differences.

ASSUMPTION	BREAKS WHEN...
comparability	unequal odds; arms differ beyond intervention; allocated condition not received
stability	spillovers; dropout; intervention varies

Effect forms: absolute percentage-point difference = practical count; relative change = difference ÷ control rate. Never call relative per cent "percentage points".

Predictive failure test: if an elaborate rule does not beat "next period equals the previous period" on the same held-back rows, that is a finding. Re-fitting until a nicer score appears turns the held-back rows into training information.

CMCE10002 Foundations of Business Analytics

THE UNIVERSITY OF MELBOURNE · SELF-TEST DRILLS · ANSWER PATHS · JUDGEMENT CARDS

Full guide + ask Sia about CMCE10002

SIDE 2/2 FIVE WORKED DRILLS · IF X → DO Y · CLASSIC TRAPS · EXAM-MORNING ROUTINE

ANSWERS CARRY THE PATH

v2 · 2026-10-25

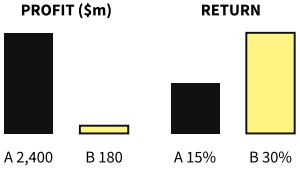
14 · Worked Drill 1: Rate, Not Size

CH 6 WORKED EXAMPLE

Given: Firm A operating profit = **$2,400m**, invested capital = **$16,000m**. Firm B profit = **$180m**, capital = **$600m**. Rank them and state what each ranking answers.

ANSWER PATH

```
A: 2,400 ÷ 16,000 = 0.15 = 15%
B: 180 ÷ 600 = 0.30 = 30%
```



The ranking reverses: dollars answer scale; the rate answers capital efficiency.

Conclusion: A earns roughly thirteen times the dollar profit, while B earns twice the return. Dollar ranking → A first; rate ranking → B first. Guard by filtering to a positive denominator before division; zero or negative capital creates misleading extreme ranks.

SELF-TEST *Which operation order? Filter unusable capital → mutate return → arrange descending → select the reporting columns.*

Full-credit explanation: profit scales with business size, so it answers who earned the most money. Dividing by invested capital removes that scale and answers who used each dollar of capital most efficiently. Name both questions; "B is better" is not a complete interpretation.

PAPER CODE LOGIC

```
keep capital > 0 → create return =
profit/capital → sort return
descending → retain firm, profit,
capital, return
```

15 · Worked Drill 2: Grouping Shape

CH 7 WORKED EXAMPLE

Given: twelve stores; columns are store, region and quarterly sales. State shape after grouping by region and after summarising.

STAGE	ROWS	GRAIN / REQUIRED FIELDS
start	12	one store; 3 columns
group region	**12**	still one store; same columns + group label
summarise	**3**	one region; mean sales + store count

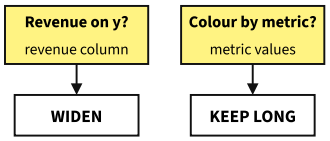
Grouping computes nothing. Summarising discards store detail. If the output still carries grouping, release it before a later whole-table share or sort.

16 · Worked Drill 3: Pivot by Purpose

CH 8 WORKED EXAMPLE

Given: one row per store-month-metric; metric cells contain *revenue, footfall, conversion*.

QUESTION	REQUIRED SHAPE	DECISION
plot revenue over time by store	revenue must be a column for y	**widen**: one row per store-month
compare all three metrics using colour	metric must be a value mapped to colour	**already long**: no pivot



targets must be column values

The pivot direction follows the variable the next operation must address.

Independent row-count check: if 3 firms × 4 quarters are lengthened, expect **12 rows**. Values have not changed; their address has. Name the two new columns for what they represent, not as generic "name" and "value".

Type guard: gather only columns holding the same kind of value. Pulling an identifier into the measured values forces names and numbers into one column and breaks the table's type discipline.

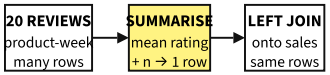
17 · Worked Drill 4: Join at One Grain

CH 9 WORKED EXAMPLE

Given: reviews = one row per review with product, date, rating; sales = one row per product-week with units sold. A product can have **twenty reviews in a week**.

ANSWER PATH

```
1 Direct join repeats weekly sales
once per review → false total.
2 Summarise reviews to product-week:
mean rating + review n.
3 Left-join from sales: every sales
row survives.
4 Check output rows = sales input
rows.
```



No reviews ≠ rating zero

blank rating; report match coverage

Prepare the many-side to the shared product-week grain before joining.

18 · Worked Drill 5: Effect in Two Forms

CH 11 WORKED EXAMPLE

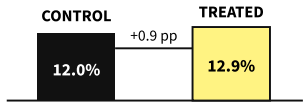
Given: randomised design; control next-day return = **12.0%**, treated = **12.9%**.

ABSOLUTE

```
12.9% - 12.0% = 0.9 percentage point
At 100,000 installs: 0.009 × 100,000
= 900 additional returning users
```

RELATIVE

```
(12.9% - 12.0%) ÷ 12.0% = 0.075 =
7.5% of control rate
```



same result: absolute 0.9 pp · relative 7.5%

Two scales, one experiment: practical count versus change relative to the base.

Judgement: a small effect in a large product can matter; state that judgement rather than letting a fraction of a point speak for itself. Do not inflate the result by calling 7.5% "7.5 percentage points".

FORM	USE IT FOR	SAY
absolute	how many users are affected	0.9 percentage point; 900 per 100,000
relative	compare effects on different bases	7.5% of the control rate

19 · Judgement Card: Forecast Claim

CH 12 WORKED EXAMPLE

Given: staffing rule predicts next month from the last **12 months** and "matches history almost exactly".

12 months in time order



FIT: first 9 **SCORE: last 3**

baseline = previous month, same 3

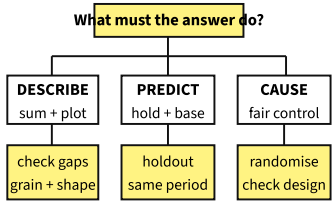
Held-back later rows make the score evidence about the future.

Answer: fit first 9 months; score last 3; evaluate a trivial "same as previous month" baseline on the same 3. Report both scores and the period. Exact fit to the training rows can be overfitting, not forecasting skill.

REPORT MUST CONTAIN	WHY
rule score on last 3	evidence on unseen future rows
baseline score on same 3	turns the score into a decision
the three-month period	defines where the comparison applies

20 · Method Chooser: If X → Do Y

CH 1–12 SYNTHESIS



Question first; table second.

Question → evidence → check: the course's recurring decision sequence.

IF YOU SEE…	DO THIS
"How many / what happened?"	define period + grain; count/summarise; state coverage
"Likely next month?"	time-ordered holdout + trivial baseline
"Improve / increase / lead to?"	name causal question; demand fair comparison
mean by group	group → summarise mean **and n**
rate/rank	filter denominator → mutate rate → arrange
periods in headings	lengthen; predict row count
measures stacked in cells	widen only if each measure must be a column
two tables	state grains + key + unmatched policy + expected rows
many units × many periods	keep compared dimension; group the other away

CHART CHOOSER

QUANTITY ASKED	ACTION
category counts	bar; count is produced by geometry
one numeric distribution	histogram; inspect shape before mean/SD
two numeric columns	points; inspect bends, clusters, extremes
average by week/group	summarise mean + n first, then plot

Question-to-grain rule: if the plotted or summarised object is customer, store-week or region, the input must reach that grain before the final operation. A chart that renders at the wrong grain can look entirely plausible.

Causal design check: random allocation supplies comparability only if assignment odds are equal and the arms differ only through the allocated condition. Spillovers, dropout or several versions of the intervention break stability and must be named.

21 · Short-Answer Scaffolds

GUIDE FAQS

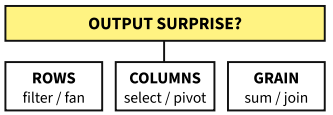
INTERPRET CODE

```
"This step [keeps/adds/collapses]
___. Output grain is one ___, with
___ rows/columns, because ___."
```

22 · Classic Traps: Diagnose Fast

ALL CHAPTERS

SYMPTOM	CAUSE	REPAIR
render fails; chunks worked	hidden session object/package	restart; top-down; first error
filter returns wrong rows	= used for equality	use ==
one import column	separator misread	inspect import and names
all points same colour	column set as constant	move column inside mapping
count drawn, mean asked	geometry used without summary	summarise first
top ranks impossible	bad denominator retained	filter before rate/rank
group line "did nothing"	grouping is invisible	inspect next verb
share becomes one row	summarised twice	mutate share onto summary
cannot group by quarter	quarter is a heading	lengthen first
joined rows multiply	key repeated / grains differ	count per key; pre-summarise
blank rating treated as zero	no match misread as value	report unmatched coverage
forecast score "perfect"	scored on training rows	later holdout + baseline



Find the first shape change.

Debug by shape before values: rows, columns, grain, then calculation.

23 · Claim-Limit-Repair Cards

CH 5 · 10–12

CLAIM	LIMIT	REPAIR
higher group mean	hidden spread / rival factors	plot rows; compare like with like
before-after improvement	seasonality, campaigns, news, competitors	concurrent fair control
adjusted causal estimate	only measured factors balanced	random assignment where feasible
low forecast error	uninterpretable alone	same-period trivial baseline
joined-table average	may describe matched subset only	unmatched count + fraction

PANEL AND PERCENTAGE TRAPS

PROMPT	CORRECT MOVE
Which store performs best?	keep stores; average quarters away
Is the business growing?	keep quarters; average stores away
Which store bucks the trend?	keep the full panel; collapse neither dimension

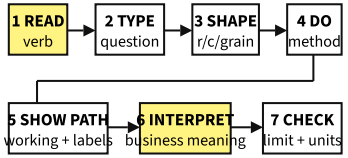
24 · Fast Recall Tables

GLOSSARY · FAQS

TERM	RECITE THIS
object	anything stored under a chosen name
vector	one-dimensional values of one type
tibble	equal-length vectors as table columns
argument	input passed inside function parentheses
grain	what one row represents
tidy	variable/column; observation/row; value/cell
join key	shared columns identifying the same row unit
fan out	row multiplication from repeated keys
panel	many units × many periods; one unit-period row
counterfactual	unobserved outcome without intervention
overfitting	learning training accidents that do not repeat
baseline	trivial alternative that gives a score meaning

25 · Exam-Morning Plan

COURSE STUDY STRATEGY



Cover → predict → reveal → correct.

Use the same routine for multiple-choice elimination and written answers.

1. **Confirm instructions** and read the command verb before the code or table.
2. **Classify:** descriptive, predictive or causal; then name the required evidence.
3. **Annotate shape:** rows, columns, grain and column types at each code step.
4. **Show the path:** operation order, calculation, units and labelled result.
5. **Close the answer:** business meaning, limit, and the next check or design.
6. **Final scan:** denominator; = vs ==; count vs distinct count; percentage points vs relative per cent; join fan-out; held-back period.

LAST-WEEK RULE *Work the weekly question sets; sit a mock cold and timed; sort mistakes into recall versus reasoning; deliberately revisit weeks 9–11 because earlier assessments do not test them.*