ETF1100 Business Statistics

MONASH UNIVERSITY · LEVEL 1 · SEMESTER 2 2026

Full guide + ask Sia about ETF1100

SIDE 1/2 METHODS · RULES · DECISIONS | **ONE A4 NOTE SHEET** | Final exam · 50% · two-hour supervised on-campus eExam

Tutorial 10% | Quiz 5% | Case 20% | Group 15% | Final exam 50%

PERMITTED MATERIALS: "One double sided A4 sheet of notes, on paper."

ETF1100 2026 S2 Week 12 Exam Revision page (links to s2-2026 resources) (MD_Week_12_Exam_Revision.md)

1 · Start With the Row W1

one row → variable → type → method

Guide redraw: define the observational unit before choosing a calculation.

FIELD ROLE	USE / TRAP
Identifier	duplicate checks; do not average
Nominal	labels, no order → count/proportion
Ordinal	ordered levels; spacing not guaranteed
Discrete	counts; numerical arithmetic
Continuous	measurements; units essential
Missing	never silently zero

★ Audit sequence: unit → variable role → missing/impossible values → duplicate identifiers → target population. Preserve the untouched source; put derived fields in new columns and reconcile row counts after filters.

EXAM BITE *A numeric code such as postcode can be nominal. Storage format does not determine statistical type.*

2 · Choose the Display W2

What is the question?

category bars | distribution histogram | time line | 2 numeric scatter

unequal groups → % + denominator

Guide redraw: purpose and variable type decide the chart, not Excel defaults.

QUESTION	REPORT
How many?	count: workload / volume
What share?	relative frequency = count ÷ relevant total
Unequal exposure?	normalised rate, with exposure unit
Audit	title, axes, unit, order, missing groups, selected range

Bars start from a common baseline. Histograms use adjacent numerical bins. Lines preserve genuine order, especially time. A truncated axis may exaggerate a length difference; 3-D geometry obscures magnitude. Normalisation repairs the denominator, not uncontrolled product mix, hours or case complexity.

2b · Spreadsheet Audit W1–2

STAGE	ACTION	EVIDENCE RETAINED
land	save raw tab; one header row	untouched source

3 · Categorical Probability W3

	Online	Store	Total
New	36	24	60
Return	18	42	60

inside = joint; edge = denominator

Guide redraw: shade the conditioning total before dividing.

CORE RULES

$P(A^c)=1-P(A)$

$P(A\cup B)=P(A)+P(B)-P(A\cap B)$

$P(A|B)=P(A\cap B)/P(B)$

independence: $P(A|B)=P(A)$, hence $P(A\cap B)=P(A)P(B)$

ASKED	DENOMINATOR	CHECK
joint A and B	grand total	cell ÷ all
marginal A	grand total	A total ÷ all
A given B	B total	B-conditioned shares sum 1

Association: compare outcome proportions across groups using the same conditioning logic. A large raw cell may merely reflect a large group. Report a difference between percentages as percentage points; relative percent change divides that gap by a named baseline.

EXAM BITE *Reversing the words after "given" reverses the denominator. Write the fraction before the decimal. Association in recorded data is not causation.*

3b · Denominator Checklist W3

BEFORE SUBMIT	QUESTION
Unit	ticket, visit, order or customer?
Eligibility	who could experience "success"?
Missing	excluded, shown, or separate?
Scope	sample description or population inference?

Repeated visits are not automatically independent customers. A pooled difference can also hide a lurking group pattern; investigate before choosing the preferred comparison.

3c · Probability Drill Grid W3

FROM ONE TWO-WAY TABLE	REQUIRED WORKING
joint	interior cell ÷ grand total; name both events
reversed conditionals	same cell, two different group totals

4 · Centre, Spread, Shape W4

min | Q1 | median | Q3 | max

IQR = Q3 − Q1; inspect tail/outlier

Guide redraw: the five-number summary shows centre, middle-half spread and tails.

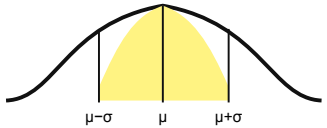
SHAPE / PURPOSE	CENTRE	SPREAD
symmetric, no strong outlier	mean	standard deviation
skewed / influential outlier	median	IQR
workload / total planning	mean may still matter	show valid extremes

STANDARDISE AN OBSERVATION

z = (x - mean) / standard deviation

z > 0 above mean; z < 0 below; unit-free

Units: mean, median, range, IQR and standard deviation use the original unit; variance uses squared units; z has no unit. Standardising does not make skewed data normal. Use a normal model only after checking shape and whether the model is suitable.



Guide redraw: interpret distance and sign before seeking a normal area.

OUTLIER RULE *Verify the record and business origin. Correct a confirmed error; retain and explain a valid exception; never delete merely because the answer changes.*

4b · Distribution Workflow W4

STEP	QUESTION	EXAM OUTPUT
1 shape	symmetric, skewed, clusters, gaps?	name visible form
2 centre	typical case or resource average?	median or mean, with unit
3 spread	middle half or mean-distance?	IQR or SD, with unit
4 unusual	error, valid exception, second process?	investigate; sensitivity
5 compare	same unit, period, inclusion rule?	difference plus context

Resistance test: predict which summaries change if the largest value doubles. Mean, range and SD react; median and IQR are usually more resistant. Resistance is not permission to ignore operational extremes.

5 · Sampling & Intervals W5

population → sample → estimate → interval

Guide redraw: the parameter is fixed; the estimate varies across random samples.

INTERVAL ARCHITECTURE

estimate ± critical value × SE

margin = critical value × SE

width = 2 × margin

MEAN · LARGE-SAMPLE PRACTICE

$SE(\bar{x})=s/\sqrt{n}$

95% CI: $\bar{x}\pm 1.96(s/\sqrt{n})$

PROPORTION · USUAL INTERVAL

$\hat{p}=x/n$

$SE(\hat{p})=\sqrt{[\hat{p}(1-\hat{p})/n]}$

CI: $\hat{p}$ ± critical × $SE(\hat{p})$

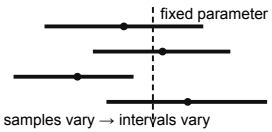
CHANGE	INTERVAL EFFECT
n increases	SE falls; narrower (4×n ≈ ½SE)
confidence increases	critical value rises; wider
variability increases	SE rises; wider
selection bias	not repaired by a narrow interval

Interpretation: the confidence procedure captures the true population parameter at its stated long-run rate under assumptions. For one realised interval, state plausible population parameter values, not the percentage of individual observations and not a probability attached to a fixed parameter.

Endpoint checksum: midpoint = estimate; half-width = margin; lower < upper; both carry the parameter's unit. Confirm whether the entry asks for an endpoint, margin, width or full interval.

EXAM BITE *Random sampling supports generalisation. Independence, selection, nonresponse and measurement quality remain separate from arithmetic precision.*

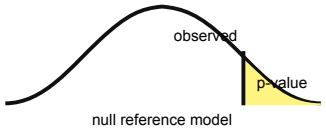
5b · CI Decision Checks W5



Guide redraw: confidence is a repeated-sampling property of the procedure.

INTERVAL LOCATION	DECISION READING
entirely above target	plausible means all exceed target
crosses target	values on both sides remain plausible

6 · Tests: Claim → Decision W6–7



Guide redraw: the p-value is the tail at least as extreme under H_0.

1. State population H_0 and H_1; direction comes from the claim before data.
2. Check sampling/independence and approximation conditions.
3. Compute SE and standardised statistic; use the matching tail(s).
4. If $p \le \alpha$, reject H_0; otherwise do not reject H_0.
5. Conclude about the represented population; report effect size and practical importance.

MEAN TEST · LARGE-SAMPLE PRACTICE

H_0: $\mu=\mu_0$

$z=(\bar{x}-\mu_0)/(s/\sqrt{n})$

ONE-PROPORTION Z TEST

H_0: $p=p_0$

$SE_0=\sqrt{[p_0(1-p_0)/n]}$

$z=(\hat{p}-p_0)/SE_0$

CLAIM WORDING	ALTERNATIVE / TAIL
exceeds / above	> · right
below / reduced	< · left
differs / changed	≠ · two-sided

Test SE uses p_0; usual proportion-interval SE uses $\hat{p}$.

For the normal approximation, check expected success and failure counts using the proportion appropriate to the method.

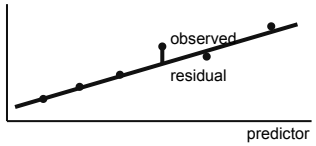
DECISION	H_0 TRUE	H_0 FALSE
reject	Type I error	correct
do not reject	correct	Type II error

LANGUAGE *A small p-value is not P(H0 true). "Do not reject" is not "prove equal". Statistical evidence is not automatically operational importance.*

6b · Test Output Decoder W6–7

OUTPUT	MEANING	COMMON MIX-UP
estimate	observed sample effect	population truth
SE	sampling fluctuation scale	sample SD
z	null distance in SE units	effect magnitude
p-value	null-tail probability	P(H_0 true)
α	pre-set Type I threshold	move after seeing p
CI	plausible parameter range	individual-value range

7 · Regression & Models W7–10



Guide redraw: inspect direction, form, strength and unusual points before r.

SIMPLE REGRESSION

$\hat{y}=b_0+b_1x$

residual e=y-ŷ

R^2 = proportion of sample y-variation explained

OUTPUT	INTERPRETATION / CHECK
b_1	predicted y change per +1 x; name both units
b_0	predicted y at x=0; may lack meaning
e > 0	observed lies above fitted value
residual plot	curve, funnel or influential case warns
prediction	avoid extrapolation outside observed x range

Correlation: unit-free linear association from −1 to 1; sensitive to outliers and blind to some curved relationships. Regression/correlation alone do not establish causation.

MULTIPLE REGRESSION

$\hat{y}=b_0+b_1x_1+\cdots+b_kx_k$

Each slope compares predicted outcomes for a one-unit predictor change while holding the other included predictors fixed. For a 0/1 indicator, coefficient = predicted difference from the omitted reference category at the same included controls. Ordinary R^2 cannot fall when predictors are added; adjusted R^2 penalises weak additions. Neither validates causation or new-data prediction.

OPTIMISATION

decision variables → constraints → objective → verify solution

Guide redraw: formulate in words and units before opening Solver.

Define variables, objective and every capacity/demand/policy constraint; add non-negativity and integrality where relevant. Then substitute the returned values into every constraint, compute slack and mark binding limits. **Feasible** satisfies constraints; **optimal** is best among feasible choices. Stress-test contribution, capacity and demand. A sensitive mix changes sharply after a plausible input change.

ETF1100 Business Statistics · Worked Drills

MONASH UNIVERSITY · CALCULATIONS TO THE CONCLUSION

Full guide + ask Sia about ETF1100

SIDE 2/2 WORKED EXAMPLES · TRAPS · FAST RECALL

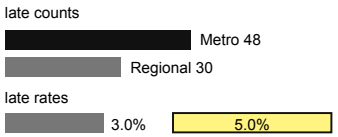
ONE A4 NOTE SHEET

Keep full precision → final requested format

8 · Counts vs Rates W2 WORKED

Given: metropolitan 48 late of 1,600; regional 30 of 600.

```
Metro rate = 48/1600 = 0.030 = 3.0%
Regional rate = 30/600 = 0.050 = 5.0%
Observed difference = 2.0 percentage points
```



Count answers workload; rate answers relative performance.

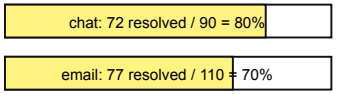
Conclusion: metropolitan operations have more late cases to resolve, but regional operations have the worse late-delivery rate. Show both when the decision needs workload and performance.

TRAP	REPAIR
compare 48 with 30 only	divide by exposure totals
say "2% higher"	say 2 percentage points
claim cause	name uncontrolled mix / complexity

9 · Conditional Rates W3 WORKED

Given: chat resolves 72/90; email 77/110.

```
P(resolve|chat)=72/90=80.0%
P(resolve|email)=77/110=70.0%
gap = 10.0 percentage points
```



Each channel total supplies its own conditional denominator.

Conclusion: channel and first-contact resolution are associated in the recorded sample. Do not claim changing channel causes improvement: ticket complexity and staffing may differ.

9b · File Classification Drill W1 WORKED

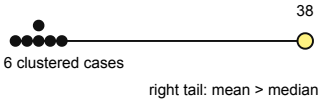
File: one row per order; OrderID, Channel, LoyaltyTier, Items, Spend, OrderDate, DeliveryDate.

FIELD	ROLE	DEFENSIBLE ACTION
OrderID	identifier	check duplicates; never average
Channel	nominal	count / conditional proportion
LoyaltyTier	ordinal	preserve order; spacing unknown

10 · Extreme Value W4 WORKED

Data: 8, 9, 10, 10, 11, 12, 38 minutes.

```
sum = 98
mean = 98/7 = 14.0 min
median = 4th ordered value = 10 min
range = 38-8 = 30 min · IQR = 12-8 = 4 min
```



Guide-derived drill: the valid extreme matters operationally even when median is typical.

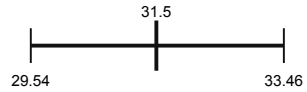
Conclusion: right-skewed; 38 pulls the mean up, so median better describes a typical case. Investigate 38 rather than deleting it. If 38 is verified as a data-entry error for 13, corrected sum = 73, mean = 10.43, median remains 10.

AUTOMATIC-ENTRY CHECK	EXPECTED
mean / median	original unit
variance	squared unit
standard deviation	non-negative, original unit
z-score	sign agrees with above/below mean

11 · Mean Confidence Interval W5 WORKED

Given: n=64 orders, x̄=31.5 h, s=8.0 h; use 1.96.

```
SE=8/√64=1.00 h
margin=1.96×1.00=1.96 h
95% CI=31.5±1.96=[29.54, 33.46] h
```



Midpoint = estimate; equal half-widths = 1.96 hours.

Interpret for the **population mean fulfilment time**, assuming random sampling and approximation conditions—not individual orders.

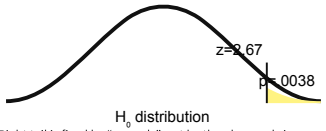
11b · Interval Repairs W5 FAQ

FAULTY CLAIM	REPAIR
"95% of orders lie from 29.54 to 33.46 h."	the interval concerns the population mean, not individual orders
"There is 95% probability μ is inside."	the procedure has 95% long-run coverage under assumptions
"Narrow means unbiased."	width covers sampling fluctuation, not selection or measurement bias
"Lower endpoint is 33.46."	order endpoints and recover midpoint 31.5

12 · Mean Test W6 WORKED

Given: n=49 calls, x̄=6.8 min, s=2.1 min; test at α=0.05 whether μ exceeds 6.

```
H₀: μ=6 · H₁: μ>6
SE=2.1/√49=0.30 min
z=(6.8-6.0)/0.30=2.67
right-tail p≈0.0038 < 0.05
```



Right tail is fixed by "exceeds", not by the observed sign.

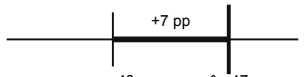
Decision: reject H_0. **Conclusion:** the sample supports a population mean wait above 6 minutes for the represented call process, under the sampling conditions.

WRONG	CORRECT
"H_0 has probability .0038"	result this extreme is rare under H_0
"accept H_1 because z is big"	compare p with pre-set α
switch to two tails	use claim-set direction

13 · Proportion Test W7 WORKED

Given: 94 conversions among 200 eligible visits; test p>0.40 at 5%.

```
H₀:p=.40 · H₁:p>.40
p̂=94/200=.470
SE₀=√(.40×.60/200)=.03464
z=(.47-.40)/.03464=2.02
p≈.0217 → reject H₀
```



Observed lift = 7 percentage points; the test SE still uses p_0.

Conclusion: evidence that the represented population conversion proportion exceeds 40%. Keep the visit-level denominator distinct from a customer-level claim if people can repeat.

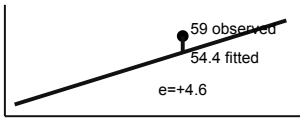
13b · Test Completion Grid W6–7 DRILLS

MARKING STEP	MEAN EXAMPLE	PROPORTION EXAMPLE
parameter	population mean wait μ	population conversion p
benchmark	6 minutes	0.40
direction	exceeds → right	exceeds → right
SE input	$s/\sqrt{n}$	p_0 under H_0
effect	$\bar{x}-\mu_0$ = 0.8 min	$\hat{p}-p_0$ = 0.07 = 7 pp
conclusion	population mean claim	population proportion claim

14 · Simple Regression W7–8 WORKED

Model: Timê = 18.4 + 0.72 Distance, minutes and kilometres. At x=50 km, observed y=59 min.

```
ŷ=18.4+0.72(50)=54.4 min
e=y-ŷ=59-54.4=+4.6 min
```



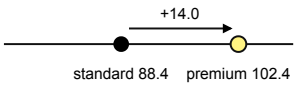
Positive residual: the trip took longer than the fitted line predicted.

Slope sentence: within the observed range, each additional kilometre is associated with 0.72 predicted minute more travel time, on average. The intercept is meaningful only if zero kilometres is relevant. With R^2=0.64, 64% of sample time variation is explained by this fitted distance model—not 64% prediction accuracy or causality.

RESIDUAL PATTERN	CONCERN
curve	linear form misses structure
funnel	spread changes with fitted value
isolated leverage point	verify; compare fits transparently

15 · Indicator Coefficient W9 WORKED

```
Sale$=72-1.8Price+9.5Rating+14Premium
Price=12, Rating=4, Premium=1:
72-21.6+38+14=102.4 thousand
Premium=0: 88.4 thousand
```



Same price and rating; Premium changes 0→1 against standard reference.

Premium sites are associated with $14 thousand higher predicted weekly sales at the same included price and rating. Do not say "causes". Omitted site characteristics remain uncontrolled.

15b · Regression Output Audit W8–9 FAQ

BEFORE INTERPRETING	WRITE
outcome	name and unit of y
predictor	name, unit, meaningful increment
coefficient	direction + predicted change
indicator	coded-1 group vs reference-0 group
controls	only predictors actually included
scope	observed range; sample / represented population
diagnostic	residual form, spread, influence

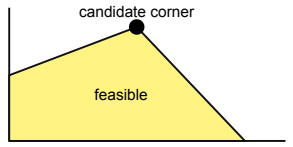
16 · Product-Mix Model W10 WORKED

Variables: Standard kits x; Premium kits y.

```
maximise Z=30x+45y
labour: 2x+3y≤120
materials: x+2y≤70
demand: y≤30
x,y≥0 [integer if whole kits required]
```

Verify candidate x=30, y=20:

CHECK	SUBSTITUTE	STATUS
labour	2(30)+3(20)=120	binding
materials	30+2(20)=70	binding
premium demand	20≤30	slack 10
objective	30(30)+45(20)	$1,800



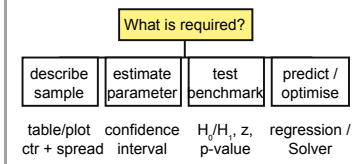
Guide redraw: feasibility comes before comparing objective values at corners.

The candidate is feasible and uses all labour and materials. It is optimal only after comparison with all feasible corners or a verified Solver run. Solver success validates the entered mathematics, not a missing business restriction.

WORDS	SYMBOL	MEANING
at most	≤	capacity / upper limit
at least	≥	minimum requirement
exactly	=	fixed balance
whole units	integer	do not round after solve

EXAM BITE *Recompute the objective and each constraint. Report quantities, objective value, bottleneck, spare capacity and the assumption to monitor.*

17 · Method Chooser W1–10



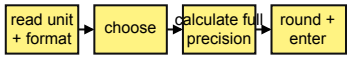
Finish: unit + scope + limit

New exam route: identify the task before selecting an output or formula.

18 · Fast Recall Matrix ALL WEEKS

SIGNAL	USE	DO NOT SAY / DO
category comparison	count + within-group %	ignore denominator
numerical distribution	shape + centre + spread	mean by default
parameter range	estimate ± critical×SE	% of individuals
claim vs benchmark	H_0/H_1, statistic, p	P(H_0 true)
binary population share	$\hat{p}$; interval or test SE	mix $\hat{p}$ and p_0
two numerical variables	scatter → r / regression	causation from fit
several predictors	controlled coefficient	control omitted factors
best choice under limits	objective + constraints	trust Solver blindly

PRECISION ROUTINE



Guide redraw: automatic marking rewards exact reading and final-format discipline.

1. Underline requested quantity: count, probability, percentage, percentage points, SE, statistic, p-value, endpoint, coefficient or objective.
2. Write denominator / benchmark / reference category and units before calculating.
3. Keep full precision through intermediate steps. Estimate sign and rough magnitude.
4. Convert decimal ↔ percentage only when requested; round once at the final line.
5. Check sign, unit, plausible range and whether the answer is sample description or population conclusion.

19 · Last-Minute Traps FAQ + DRILLS

IF YOU SEE…	THINK…
postcode / ID	label, not measurement
unequal group sizes	rate plus count
given / conditional	restrict denominator
skew / outlier	median + IQR; investigate
95% confidence	procedure coverage; parameter
p>α	insufficient evidence; not equality
significant	also report effect size
residual	y−ŷ, not ŷ−y
indicator	coded 1 vs reference 0
feasible	not necessarily optimal

FINAL SWEEP *Method fits variable type? Assumptions named? Denominator correct? Tail chosen before result? Units and reference group stated? No causal overclaim? Requested precision and entry convention followed?*