ECMT1010 Introduction to Economic Statistics

SCHOOL OF ECONOMICS · UNIVERSITY OF SYDNEY · FIRST-YEAR UNDERGRADUATE

Full guide + ask Sia about ECMT1010

SIDE 1/2 DEFINITIONS → DECISIONS → FORMULAS → CHECKS | EXAM REVISION SHEET | S2 2026 · whole-unit method map

1 · Exam Map

COURSE.JSON · BRIEF

Exact percentage weights: subject to confirmation

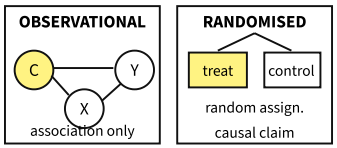
Permitted materials: not confirmed for this offering in the course materials we hold — check the exam instructions on the LMS.

ASSESSMENT	GROUNDED FORMAT
Week-7 test	35 MCQ · Weeks 1–6
Assignment	Individual data analysis
Final	Whole unit · Section B 50 marks; Section A implied
Weekly	Workshops / muddy cards

FINAL FOCUS *Classify → name parameter → select method → check conditions → calculate → decide → conclude in context. Formula sheet plus N(0,1)/t tables are described as supplied in the paper; verify current instructions.*

2 · Design Decides Claims

CH 1 · W1



Design figure: a confounder can drive X and Y; random assignment cuts the confounder→X link.

ASK	YES LICENCES	NO MEANS
Random selection?	Generalise	Sampling-bias risk
Random assignment?	Causal claim	Association only

★ *Confounder*: associated with both explanatory X and response Y; offers a rival explanation. Large n shrinks variance, never selection bias. Blinding: subjects removes placebo effect; researchers reduces observer bias; analysts reduces confirmation bias.



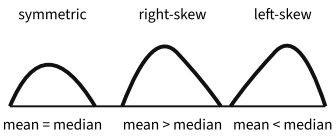
answer the two questions separately

Decision figure: selection governs external validity; assignment governs causal validity.

THREAT	DIAGNOSTIC CUE	CONSEQUENCE / REPAIR
Selection bias	Who was reachable or volunteered?	Sample unrepresentative; redesign selection
Non-response	Responders differ from silent invitees	Generalisation threatened

3 · Describe Before Inferring

CH 2 · W2



Shape figure: the mean follows the long tail; the median resists extreme values.

DATA	CENTRE	SPREAD / DISPLAY
Symmetric quantitative	Mean $\bar{x}$	SD s · histogram
Skewed / outliers	Median	IQR · boxplot
Categorical	Count / proportion	Bar chart
Two quantitative	Correlation r	Scatterplot first

CALCULATION BLOCK

$\bar{x} = \Sigma x_i / n$

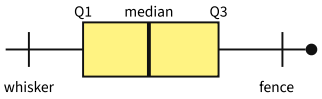
$s = \sqrt{[\Sigma (x_i - x)^2/(n-1)]}$

$IQR = Q3 - Q1$

$fences = Q1 - 1.5IQR,\ Q3 + 1.5IQR$

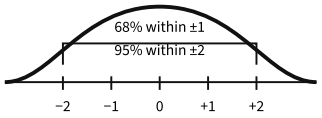
$z_i = (x_i - x)/s$

Five-number summary: min, Q1, median, Q3, max. For bell-shaped data: about 68% within ±1s, 95% within ±2s, 99.7% within ±3s. Check shape before using the 95% rule.



Boxplot figure: box = middle 50%; values beyond 1.5·IQR fences are flagged.

Correlation: $-1 \le r \le 1$; sign = direction, $|r|$ = linear strength. r is unit-free, symmetric, not resistant. $r = 0$ means no *linear* association; a curved relationship may remain. Correlation never establishes cause.



Standard-units figure: z counts SDs from the mean; the 68–95–99.7 rule needs a bell shape.

CHECK R	WHAT TO SAY / DO
Direction	Positive cloud rises; negative cloud falls
Strength	$\lvert r \rvert$ near 1 = tight linear band; near 0 = weak linear pattern
Outlier	Plot first: one point can dominate r
Curve	A strong nonlinear pattern may have $r \approx 0$
Units/order	Unit-free and $r(x,y) = r(y,x)$

4 · CI + Bootstrap

CH 3 · W3–4

★ *Sampling distribution*: values of a statistic over all possible samples. Its SD is the **standard error**. Bigger n → smaller SE → narrower interval.

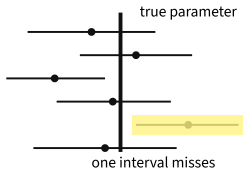
CONFIDENCE INTERVAL

estimate ± margin of error

ME = critical value × SE

95% shortcut = statistic ± 2·SE

general = statistic ± z*·SE

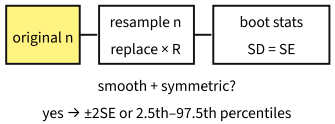


Coverage figure: the fixed parameter does not move; repeated intervals do.

Interpret: "We are 95% confident the interval … contains the true [parameter]." This describes long-run coverage, not a probability about the fixed parameter.

BOOTSTRAP ALGORITHM

STEP	DO	CHECK
1	Treat original sample as miniature population	Original sample representative
2	Resample same n **with replacement**	Values may repeat
3	Compute statistic; repeat many times	Plot distribution
4	SD of bootstrap statistics = SE	Smooth, roughly symmetric
5	±2SE or percentile endpoints	Skew/spikes → unreliable



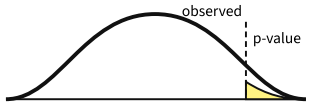
Bootstrap figure: resampling mimics repeated sampling; the spread estimates SE.

CHANGE	EFFECT	REASON
Larger original n	SE falls roughly $1/\sqrt{n}$; CI narrows	Statistic varies less
Higher confidence	Critical value rises; CI widens	More coverage requires more margin
More resamples	Simulation smoother; width essentially unchanged	Original information unchanged
Skewed / spiky bootstrap	Simple methods unreliable	Normal-shape approximation fails

5 · Hypothesis-Test Ritual

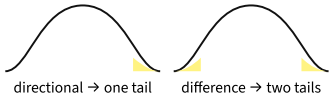
CH 4 · W5–6

1. Define population parameter.
2. Write H_0 with "=" and H_a with $<$, $>$ or $\neq$.
3. Check design/conditions; select statistic.
4. Compute p-value in the tail(s) named by H_a.
5. Compare p with α; decide and conclude in context.



Tail-area figure: small tail under H_0 means strong evidence against H_0.

RESEARCH WORDING	H_a	TAIL	SAFE CONCLUSION
Higher / increase	parameter > null	Right	Evidence it is higher
Lower / reduce	parameter < null	Left	Evidence it is lower
Different / changed	parameter ≠ null	Both	Evidence of a difference

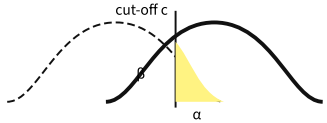


Tail-choice figure: H_a fixes the shaded region before the statistic is examined.

Randomization distribution: simulate a world where H_0 is true. Reallocate responses between experimentally assigned groups; for observational data use the guide's shift/shuffle idea. The observed statistic is compared with this null distribution. "As extreme" always follows the alternative: one direction or both.

p-value wording: probability of a statistic at least this extreme, *assuming H_0 is true*. It is not $P(H_0 \text{ true})$, not the effect size, and not the probability that the decision is wrong. Statistical significance also does not automatically mean practical importance.

$p < \alpha \rightarrow$ reject H_0. $p > \alpha \rightarrow$ do not reject; say "insufficient evidence", never that H_0 is proved. Directional claim → one tail. "Different/change" → two tails. Choose direction before seeing the data.



Error figure: α is rejection under a true H_0; β is non-rejection under a false H_0.

6 · Normal, CLT + Proportions

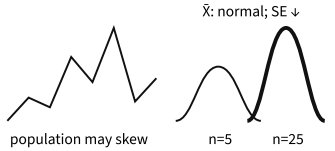
CH 5–6 · W6–9

NORMAL TRANSLATION

$z = (x - \mu)/\sigma \quad \rightleftarrows \quad x = \mu + z\sigma$

z*: 90% 1.645 · 95% 1.960 · 99% 2.576

$SD(\bar{X}) = \sigma/\sqrt{n}$



CLT figure: sample means centre at μ and narrow with SE = $\sigma/\sqrt{n}$ as n grows.

NORMAL CHECK	CONDITION	IF IT FAILS
Mean	$n \ge 30$, or population roughly normal	Simulation method
Proportion	$np \ge 10$ and $n(1-p) \ge 10$	Simulation method

Table discipline: shade first. Table gives left area: above = 1−left; between = larger left−smaller left. Distinguish an individual SD σ from a mean's SD $\sigma/\sqrt{n}$.

PROPORTION FORMULAS: THE P CHOICE

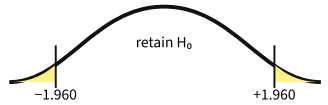
TASK	SE
One-p CI	$\sqrt{[\hat{p}(1-\hat{p})/n]}$
One-p test	$\sqrt{[p_0(1-p_0)/n]}$
Two-p CI	$\sqrt{[\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2]}$
Two-p test	$\sqrt{[\bar{p}\text{pool}(1-\bar{p}\text{pool})(1/n_1+1/n_2)]}$

TEST STATISTICS

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

pool: $\hat{p} = (x_1 + x_2)/(n_1 + n_2)$

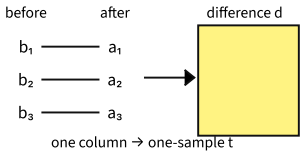
two p: $z = (\hat{p}_1 - \hat{p}_2)/SE\text{pool}$



Two-proportion test figure: at 5%, reject in either tail beyond ±1.960.

7 · Means + Paired Choice

CH 7 · W9



Paired-design figure: preserve matching; pairing removes between-subject variation and shrinks SE.

8 · Regression

CH 8 · W10

LEAST SQUARES

$\hat{y} = b_0 + b_1 x$ · $e = y - \hat{y}$

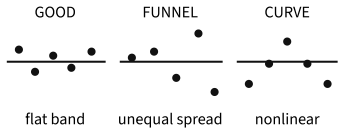
$b_1 = r(s_y/s_x)$ · $b_0 = \bar{y} - b_1 \bar{x}$

$t = b_1/SE(b_1)$, $df = n - 2$

$R^2 = r^2 = SSModel/SST$

$SST = SSModel + SSE$ · $s_e = \sqrt{[SSE/(n-2)]}$

OUTPUT	MARKER-READY MEANING
b_1	Predicted change in y per one-unit rise in x, with units
b_0	Predicted y at x=0; interpret only if sensible/in range
R^2	% of variation in y explained by x
Residual	Observed minus predicted; vertical gap



Residual figure: random flat band is good; funnel or curve means the linear model misses structure.

Predict only inside the observed x range. Extrapolation assumes the relationship continues without evidence. A significant slope supports a linear relationship, not causation.

8b · Covariance + Estimators

CH 11 · W13

RULES THAT BITE

$cov(X,Y) = E[(X - \mu_x)(Y - \mu_y)]$

$\rho = cov(X,Y)/(\sigma_x \sigma_y)$

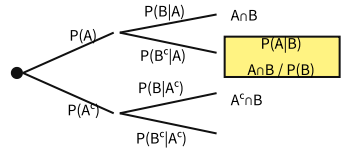
$Var(V+W) = Var(V) + Var(W) + 2cov(V,W)$

$Var(bZ) = b^2 Var(Z)$ · $Var(constant) = 0$

Independence makes covariance zero and removes the cross term; zero covariance does not prove independence. Before sampling, $\bar{X}$ is a random **estimator**; afterwards $\bar{x}$ is the realised **estimate**. For iid observations: $E(\bar{X}) = \mu$, $Var(\bar{X}) = \sigma^2/n$, $SD(\bar{X}) = \sigma/\sqrt{n}$.

9 · Probability + Random Variables

CH 9–11 · W11–13



Tree figure: multiply along paths; add B leaves for total probability; reverse with Bayes.

Check your exam instructions on the LMS for what you may bring; use this sheet to revise and self-test.

AskSia

FORWARD THIS TO YOUR TUTORIAL GROUP · SHARED BY _____

v2 · 2026-10-25

Semester 2 formal examination period, from 16 November 2026 (the Exams Office publishes the exact date).

ECMT1010 Worked Drills · Judgement Cards · Classic Traps

INTRODUCTION TO ECONOMIC STATISTICS · UNIVERSITY OF SYDNEY

Full guide + ask Sia about ECMT1010

SIDE 2/2 FIVE WORKED PATHS · IF-YOU-SEE-X CARDS · TRAPS · EXAM-MORNING PLAN

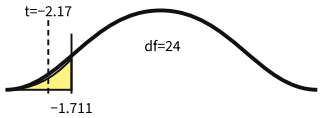
SELF-TEST: COVER THE ANSWERS

Show every line that earns a mark

10 · Drill 1 – One-Mean t

COURSE WORKED · 8 MARKS

Q. Bags labelled 250 g; random sample n=25, $\bar{x}$=247.4 g, s=6 g. At α=.05, test whether mean fill is below 250 g.



Mean-test figure: the observed t lies beyond the left-tail critical value.

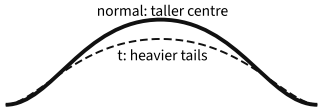
MARK PATH	WORKING
Parameter + H	μ=true mean fill; H_0:μ=250; H_a:μ<250
Method	σ unknown → one-sample t; df=25−1=24
SE	$6/\sqrt{25}$=1.2 g
Statistic	(247.4−250)/1.2=−2.17
Decision	−2.17<−1.711 → reject H_0
Context	Significant evidence at 5% that mean fill is below 250 g

COVER-AND-RECITE *Why t? Unknown σ. Why one tail? "Below". Where do df and tail appear? Beside the table value.*

10b · Mean-Method Card

CH 7

IF YOU SEE	DO
One quantitative sample	One-sample t
Two separate groups	Independent two-mean t
Same subject twice / matched	Differences d → one-sample t
Why paired?	Removes between-subject variability → smaller SE → higher power



smaller df → larger $|t^*|$ → wider CI

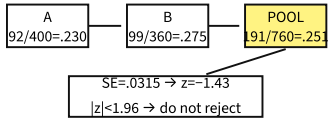
t-versus-normal figure: estimating σ adds uncertainty, so t uses heavier tails.

BEFORE CALCULATING	WRITE THE CHECK
Independence	Random sample/assignment and observations not linked, unless deliberately paired
Shape / size	Population roughly normal for small n; CLT support when n≥30
Unknown σ	Use sample s and t, not z
Degrees of freedom	n−1 for one mean or pairs; min(n_1,n_2) for two independent means
Direction	Read <, > or ≠ from the claim before using the table

11 · Drill 2 – Two Proportions

CH 6 WORKED · 8 MARKS

Q. Design A: 92/400 add to cart. Design B: 99/360. Test at 5% whether true proportions differ.



Pooled-test figure: H_0 supplies one common proportion for the test standard error.

MARK PATH	WORKING
Hypotheses	H_0:p_1=p_2; H_a:$p_1 \neq p_2$
Samples	$\hat{p}_1$=.230; $\hat{p}_2$=.275
Pool	$\hat{p}$=(92+99)/(400+360)=.251
SE	$\sqrt{[.251 \cdot .749(1/400+1/360)]}$=.0315
z	(.230−.275)/.0315=−1.43
Decision	1.43<1.96 → do not reject
Context	Insufficient evidence at 5% that rates differ

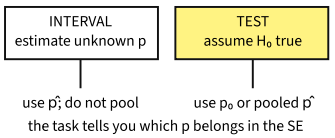
SELF-CHECK

```
Success/failure counts ≥10 in each
group? yes.
Test or interval? test → pool.
One or two tails? "differ" → two.
```

11b · Proportion Traps

CH 6 FAQ

TRAP	REPAIR
Use $\hat{p}$ in a one-p test SE	Use null p_0
Pool a two-p CI	Keep $\hat{p}_1$,$\hat{p}_2$ separate
Skip conditions	Check successes and failures in every group
Say "no difference"	Say "insufficient evidence of a difference"



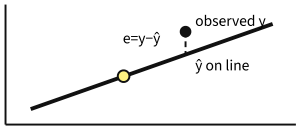
SE-choice figure: intervals estimate freely; tests calculate variability under the null.

PROMPT FRAGMENT	FAST SETUP
"Estimate the proportion"	$\hat{p} \pm z^* \sqrt{[\hat{p}(1-\hat{p})/n]}$
"Test p=p_0"	$z=(\hat{p}-p_0)/\sqrt{[p_0(1-p_0)/n]}$
"Estimate p_1-p_2"	Separate group SE terms
"Test p_1=p_2"	Pool counts first, then one null SE
Condition fails	Bootstrap CI or randomization test

12 · Drill 3 – Regression

CH 8 WORKED · 8 MARKS

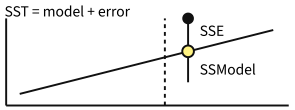
Q. n=22 cafés; weekly revenue y and advertising x both in $000. $\hat{y}$=9.6+1.40x, r=.72, SE(b_1)=.38.



line passes ($\bar{x}$,$\bar{y}$)

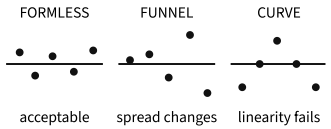
Least-squares figure: the fitted line minimises Σe^2 and passes through ($\bar{x}$,$\bar{y}$).

PROMPT	ANSWER PATH
Slope	Each extra $1,000 advertising is associated with $1,400 more predicted weekly revenue
x=5	$\hat{y}$=9.6+1.40(5)=16.6 → $16,600
Fit	R^2=.72²=.518 → about 52% of revenue variation explained by advertising
Slope test	t=1.40/.38=3.68; df=20
Decision	3.68>2.086 → reject H_0:β_1=0; strong evidence of a positive linear relationship



ANOVA figure: total variation splits into explained SSModel and residual SSE.

JUDGEMENT *A good numerical prediction can still be unjustified if x lies outside the observed range. Slope significance does not establish cause.*



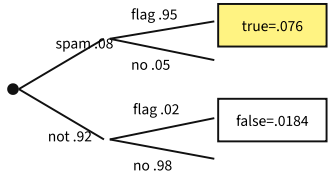
Residual-check figure: structure in residuals is evidence the fitted straight-line model is incomplete.

AUDIT ITEM	QUESTION TO ANSWER
Linearity	Does the residual plot bend?
Constant spread	Does the vertical band stay equally wide?
Influence	Does one unusual point control slope or r?
Scope	Is the requested x inside the observed range?
Slope test	H_0:β_1=0; t=b_1/SE(b_1); df=n−2
Fit	R^2 describes explained variation, not causal strength

13 · Drill 4 – Bayes

CH 9 WORKED · 5 MARKS

Q. P(spam)=.08; P(flagged|spam)=.95; P(flagged|not spam)=.02. Find P(spam|flagged).



Bayes tree: multiply down each flagged path, then divide the true path by all flagged paths.

STEP	WORKING
True-positive path	.95×.08=.0760
False-positive path	.02×.92=.0184
All flagged	.0760+.0184=.0944
Reverse conditional	P(spam|flag)=.0760/.0944=.805
Decision	A flagged email is genuinely spam about 80.5% of the time

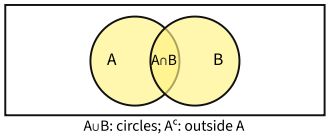
BAYES TEMPLATE

$$P(A|B) = P(B|A)P(A) \,/\, P(B)$$
$$P(B) = P(B|A)P(A)+P(B|A^c)P(A^c)$$

PROBABILITY JUDGEMENT CARDS

IF WORDING SAYS	MOVE
A **or** B	Add, subtract overlap
A **and** B	Multiply using conditional
A **given** B	Restrict denominator to B
Reverse the conditional	Bayes; include every evidence path
At least one	Often 1−P(none)

TRAP *A rare base rate can make many positives false even when sensitivity is high. Never swap P(A\|B) and P(B\|A).*



A∪B: circles; A^c: outside A

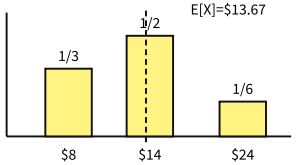
Event-set figure: subtract the overlap in the addition rule so it is not counted twice.

RULE	DIAGNOSTIC / INTERPRETATION
P(Ω)=1; P(A)≥0	Probabilities are bounded and exhaustive
$P(A^c)$=1−P(A)	Useful for "at least one" via none
P(A∪B)=P(A)+P(B)−P(A∩B)	"Or"; subtract double-counted overlap
P(A∩B)=P(A|B)P(B)	"And"; multiply along a tree path

14 · Drill 5 – Random Variable

CH 10 WORKED · 8 MARKS

Q. Fees: Basic $8, Plus $14, Pro $24. Counts are 2U, 3U, U. Find the distribution, E[X] and SD.



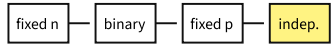
pmf figure: probability bars sum to 1; E[X] is the balance point, not the tallest bar.

X	P(X=X)	X·P
$8	2/6=1/3	2.667
$14	3/6=1/2	7.000
$24	1/6	4.000
Total / E[X]		**$13.67**

SPREAD

$$Var(X)=\Sigma(x-13.67)^2 p(x)$$
$$=10.72+.055+17.79=28.56$$
$$SD(X)=\sqrt{28.56}=\$5.34$$

Check probabilities sum to 1. Expected value is probability-weighted, not a plain average. Variance has squared units; SD returns to dollars.

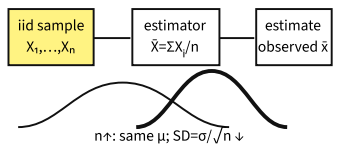


then X~Bin(n,p)

$$P(X=k)=C(n,k)p^k(1-p)^{n-k}$$
$$\text{mean}=np \cdot SD=\sqrt{[np(1-p)]}$$

Binomial gate: all four conditions must hold before using the formula.

CONTINUOUS RV	RULE
Density f(x)	Total area = 1; heights are not probabilities
Interval	P(a≤X≤b) = area from a to b
Exact point	P(X=x)=0
Uniform [a,b]	height=1/(b−a); mean=(a + b)/2; variance=(b−a)²/12

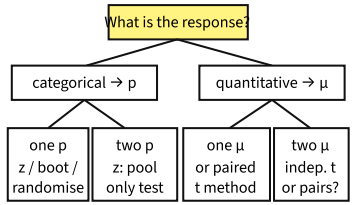


Estimator figure: $\bar{X}$ is random across samples; $\bar{x}$ is one realised draw from that distribution.

LINEAR-COMBINATION RULE	EXAM CONSEQUENCE
E[X+Y]=E[X]+E[Y]	Expectations add whether or not variables are independent
E[bX]=bE[X]	Scaling acts once on the mean

15 · Method Chooser

ALL CHAPTERS



Method tree: response type first; then number and dependence of samples; then CI versus test.

IF YOU SEE X	DO Y
Random selection	Generalise to population
Random assignment	Permit causal conclusion
Skew / outlier	Median + IQR; inspect before mean + SD
Unknown SE formula	Bootstrap CI if distribution smooth
H_0 world / tail area	Randomization test / p-value
Mean, σ unknown	t with explicit df
One/two proportions	z; $\hat{p}$ for CI, p_0/pool for test
Quantitative x and y	Regression; inspect residuals; avoid extrapolation
Reversed conditional	Bayes tree
Linear combination of RVs	EV/variance/covariance rules

16 · Classic Traps

FAQS · STRATEGY

WRONG REFLEX	CORRECT REPAIR
H_0 written with $\bar{x}$ or $\hat{p}$	Hypotheses use μ or p
p-value = chance H_0 true	Probability of data/extremer, assuming H_0
Not reject = prove no effect	Evidence insufficient
95% probability parameter lies here	95% is method coverage
More bootstrap resamples narrow CI	Larger original n narrows CI
σ used for $\bar{X}$	Use $\sigma/\sqrt{n}$
Pairing ignored	Compute per-pair d first
r=0 means unrelated	No linear association only
cov=0 means independent	Independence implies zero covariance, not converse
SD scaling by b	Variance scales by b^2

17 · Exam-Morning Plan

COURSE STRATEGY

1. Map the supplied formula sheet and N(0,1)/t tables before starting.
2. For every question, label data type, goal, sample structure, parameter and tail.
3. Write conditions and df beside the method; sketch/shade before using a table.