



0 • Exam Map

COURSE.JSON · BRIEF

Quizzes 20% Group 30% Final 50%

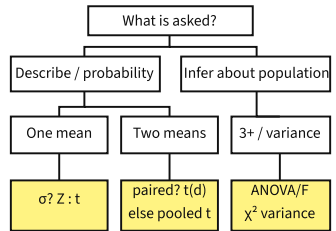
"This is a restricted open book exam. The following materials and provisions are permitted: All paper-based, i.e. written and printed, materials, including lecture notes, textbook, tutorial exercises, practice exams or your own notes ... You are allowed to use Excel, Word and PDF notes." (Info about the Final Exam, 26134, Spring 2026)

COMPONENT	WEIGHT	FORMAT
Online quizzes	20%	Best 4 of 6 count
Group assignment	30%	Data-analysis project
Final examination	50%	Online; 25 questions

Execution: roughly 4–5 minutes per equal-weight question; route first, compute, attach units, interpret, then move. With no negative marking, attempt every question.

1 • Master Route

CH 12 · FIG 12.2



Route before calculating: target → samples/groups → parameter → assumptions.

PROMPT	USE	TABLE
One mean, σ known	Z	Z
One mean, σ unknown	t, df n-1	t
One variance / SD	χ^2 , df n-1	χ^2
Two means, matched	paired t on d	t
Two means, independent; equal variances	pooled t	t
Two variances	F = larger/lower	F
Three+ means	one-way ANOVA	F

TABLE WORKFLOW Standardise first; then select the Z, t, χ^2 , F or binomial table. For t and χ^2 use $df=n-1$; for pooled t use n_1+n_2-2 ; for F preserve numerator/denominator df. A two-sided test puts $\alpha/2$ in each tail. Write statistic, df and tail beside the lookup before comparing. Under equal weighting, bank quick computes, flag slow items, keep a review buffer, and attempt all 25 because there is no negative marking.

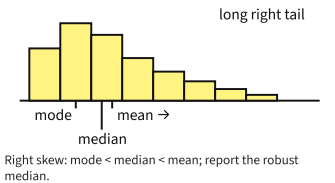
2 • Describe First

CH 1 • FINAL COMPUTE

VARIABLE	VALID SUMMARY / DISPLAY
Nominal	counts, proportions, mode; bar
Ordinal	median/ranks; ordered bar
Discrete numerical	mean/median, variance; bars
Continuous	mean/median, s; histogram/boxplot

CENTRE + SPREAD
 $\bar{x} = \sum x_i / n$ · range = max - min
 $s^2 = \sum (x_i - \bar{x})^2 / (n-1)$ · $s = \sqrt{s^2}$
CV = $(s/\bar{x}) \times 100\%$

Choose centre: symmetric/no severe outliers → mean + s; skew/outliers → median + IQR. CV is unit-free, so compare relative spread across scales.

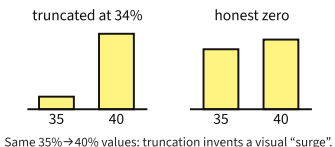


3 • Communicate Honestly

CH 2 • SHORT ANSWER

DATA / TASK	CHART	TRAP
Categorical comparison	bar / dot	pie angles hard to rank
One numerical distribution	histogram / boxplot	bars touch in histogram
Two numerical variables	scatterplot	r measures linear association only

$r = \text{cov}(x, y) / (s_x s_y)$, $-1 \leq r \leq 1$
Correlation ≠ causation. Inspect the scatterplot: $r \approx 0$ can hide a strong curved relation. For bars, the value axis starts at zero. Distinguish percentage points from relative percent: (new - old) / old $\times 100\%$.



FREQUENCY FIELD	DEFINITION + CHECK
frequency f	class count; $\sum f = n$
relative frequency	f/n ; column sums to 1
percentage frequency	100f/n; column sums to 100%
continuous frequency	equal-width bins; touching histogram bars

Report centre with spread: two datasets may share $\bar{x}$ but differ sharply in s. A responsible statistic states sample or population, centre used, period, units, source and limitations.

4 • Probability Rules

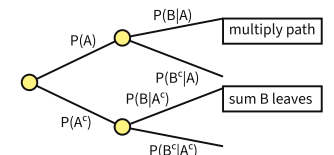
CH 3 • TABLE ITEM

CORE RULES
 $P(A^c) = 1 - P(A)$
 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 $P(A|B) = P(A \cap B) / P(B)$
 $P(A \cap B) = P(A|B)P(B)$

Read a contingency table: joint = cell/grand total; marginal = row or column total/grand total; conditional = cell / conditioning total.

RELATIONSHIP	DIAGNOSTIC EQUALITY
Mutually exclusive	$P(A \cap B) = 0$
Independent	$P(A \cap B) = P(A)P(B)$, or $P(A B) = P(A)$
Collectively exhaustive	union probability = 1

BAYES + TOTAL PROBABILITY
 $P(A_1|B) = P(B|A_1)P(A_1) / P(B)$
 $P(B) = \sum P(B|A_i)P(A_i)$



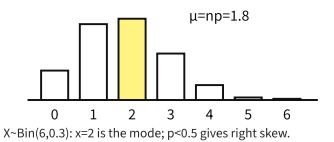
Multiply along a path for a joint; sum matching leaves for a marginal.

5 • Binomial

CH 4 • EXACT/CUMULATIVE

Check four conditions: fixed n; two outcomes; constant p; independent trials. Then $X \sim \text{Bin}(n, p)$.
 $P(X=x) = C(n, x) p^x (1-p)^{n-x}$
 $C(n, x) = n! / [x!(n-x)!]$
 $E(X) = np$ · $\text{Var}(X) = np(1-p)$ · $SD = \sqrt{np(1-p)}$

WORDS	PROBABILITY
exactly k	$P(X=k)$
at most k	$P(X \leq k)$
at least k	$1 - P(X \leq k-1)$
more than k	$1 - P(X \leq k)$



GENERAL DISCRETE RV
 $0 \leq P(X=x) \leq 1$ · $\sum P(X=x) = 1$
 $F(x) = P(X \leq x)$ · $E(X) = \sum x P(X=x)$
 $\text{Var}(X) = \sum (x - \mu)^2 P(X=x) = E(X^2) - \mu^2$ · $SD = \sqrt{\text{Var}(X)}$

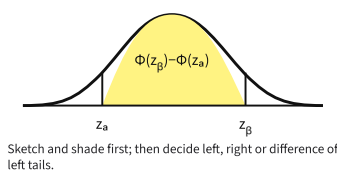
MODEL TRAP A binomial is only the special case meeting all four conditions; otherwise use the listed PMF rather than forcing $\text{Bin}(n, p)$.

6 • Normal + Z

CH 5 • TABLE LOOKUP

$X \sim N(\mu, \sigma^2)$: continuous, symmetric, mean=median=mode; total density area=1 and $P(X=\text{exact value})=0$. Empirical check: $\approx 68\%$, 95%, 99.7% within 1, 2, 3 SD.

$Z = (X - \mu) / \sigma \sim N(0, 1)$
left: $P(X \leq x) = \Phi(z)$
right: $P(X > x) = 1 - \Phi(z)$
interval: $P(a \leq X \leq b) = \Phi(z_b) - \Phi(z_a)$
symmetry: $\Phi(-z) = 1 - \Phi(z)$

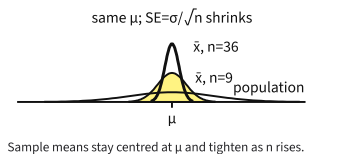


7 • Sampling Distribution

CH 6 • MEAN PROBABILITY

Parameter μ, σ, p is fixed/unknown; **statistic** $\bar{x}, s, \hat{p}$ varies sample to sample. For the sample mean, $E(\bar{x}) = \mu$ and:

$SE(\bar{x}) = \sigma / \sqrt{n}$
 $Z = (\bar{x} - \mu) / (SE(\bar{x}))$ when σ known
CLT: for large n (rule of thumb $n \geq 30$), $\bar{x}$ is approximately normal regardless of population shape; if the population is normal, $\bar{x}$ is normal for any n. A fourfold n halves SE.



7b • Ethical Decision

CH 11 • VIGNETTE

Four principles: respect for persons (consent/autonomy); beneficence (minimise harm, maximise benefit); justice (fair distribution); respect for law and public interest (lawful, transparent, accountable). **Consequentialism** judges outcomes; **deontology** judges duties and means.

safe projects	safe people	safe data	safe settings	safe outputs
---------------	-------------	-----------	---------------	--------------

all five checked for the scenario
Safety belongs to the whole access arrangement, not the dataset alone.
Apply every element to the facts. De-identification reduces but does not remove re-identification risk.

8 • Confidence Intervals

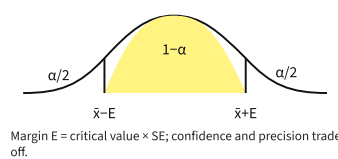
CH 7 • CORE FINAL

GENERAL
estimate $\pm$ critical value $\times$ SE

SITUATION	CI FOR M
σ known	$\bar{x} \pm z_{\alpha/2} \cdot \sigma / \sqrt{n}$
σ unknown	$\bar{x} \pm t_{\alpha/2, n-1} \cdot s / \sqrt{n}$

Common z*: 90% 1.645 · 95% 1.96 · 99% 2.576. Higher confidence → larger critical value → wider CI. Larger n → smaller SE → narrower CI. Student t has heavier tails; as $df \rightarrow \infty$, $t \rightarrow Z$.

Interpret the procedure: 95% of intervals built this way capture μ . Do not assign 95% probability to the fixed μ after this interval is observed.



9 • Hypothesis Test Engine

CH 8 • EVERY TEST

- State H_0 with equality and H_1 from the claim.
- Choose statistic/table; verify assumptions.
- Compute SE and standardised statistic.
- Find critical value(s) or p-value.
- Reject H_0 if $p \leq \alpha$ / statistic enters rejection region; otherwise **fail to reject**.
- Conclude in the problem's units and context.

$z = (\bar{x} - \mu_0) / (\sigma / \sqrt{n})$
 $t = (\bar{x} - \mu_0) / (s / \sqrt{n})$, $df = n - 1$

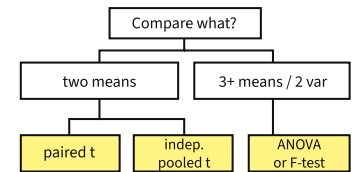
H ₁	TAIL	CRITICAL ALLOCATION
$\mu \neq \mu_0$	two	$\alpha/2$ each side
$\mu > \mu_0$	right	all α right
$\mu < \mu_0$	left	all α left

Errors: Type I = reject true H_0 , probability α . Type II = fail to reject false H_0 , probability β . Power = $1 - \beta$. Lower α raises β unless n rises.

CI BUILD ORDER Identify the parameter; decide whether σ is known; set $\alpha = 1 - \text{confidence}$ and use $\alpha/2$ per side; compute SE; read z^* or t^* with $df = n - 1$; find $E = \text{critical} \times SE$; report $\bar{x} \pm E$ with bounds, units and the population mean. Higher confidence or greater spread widens the interval; larger n narrows it only at the $\sqrt{n}$ rate. A point estimate alone gives no sampling uncertainty. For σ^2 or σ , route to χ^2 ; its interval is asymmetric because χ^2 is right-skewed.

10 • Compare Groups

CH 9 • CHOOSE CORRECTLY



Same units before/after → paired; separate groups → independent.

INDEPENDENT, EQUAL VARIANCES
 $s_p^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2] / (n_1 + n_2 - 2)$
 $SE = \sqrt{[s_p^2 (1/n_1 + 1/n_2)]}$
 $t = (\bar{x}_1 - \bar{x}_2) / SE$ · $df = n_1 + n_2 - 2$

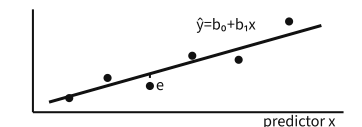
PAIRED
 $d_i = \text{before}_i - \text{after}_i$
 $t = d / (s_d / \sqrt{n})$ · $df = n - 1$

ANOVA + VARIANCES
 $SST = SSB + SSW$ · $F = MSB / MSW$
variance test: $F = s_1^2 / s_2^2$ (Larger on top)
 $df = (n_1 - 1, n_2 - 1)$
ANOVA H_0 : all group means equal. Large upper-tail F → at least one differs; it does not say which. Avoid many pairwise t-tests: they inflate overall Type I error.

11 • Regression

CH 10 • FIT + READ

$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ · $\hat{y} = b_0 + b_1 x$
 $b_1 = r(s_y / s_x) = \sum (x - \bar{x})(y - \bar{y}) / \sum (x - \bar{x})^2$
 $b_0 = \bar{y} - b_1 \bar{x}$ · $e_i = y_i - \hat{y}_i$
 $R^2 = SSR / SST = 1 - SSE / SST$; simple regression: $R^2 = r^2$



OLS minimises squared vertical residuals and passes through $(\bar{x}, \bar{y})$.

Interpret b_1 as estimated change in y per one-unit x, with units. Treat b_0 cautiously if $x=0$ is outside the data. R^2 is explained variation, not causal proof; confounding or reverse causation may drive association.

SLOPE TEST
 $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$
 $t = b_1 / se(b_1)$, $df = n - 2$

Reject when $|t|$ exceeds the two-tailed critical value or $p < \alpha$. Then interpret b_1 in x/y units. Significance supports association, not causation; predict only inside the observed x range.

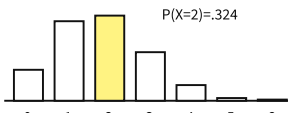


W1 · Binomial Calls

CH 4 WORKED · 4 MARKS

Given: 6 independent calls, sale probability $p=0.3$; X = sales. Find $P(X=2)$, mean and variance.

- Fixed $n=6$, two outcomes, constant $p=3$, independent → **X-Bin(6, 3)**.
- $C(6,2)=15$; $p^2=.09$; $(1-p)^4=.74=$.2401.
- $P(X=2)=15 \times .09 \times .2401 =$ **0.324** (3 dp).
- $\mu=np=6 \times .3=$ **1.8**; $\sigma^2=np(1-p)=6 \times .3 \times .7=$ **1.26**; $\sigma=\sqrt{1.26}=$ **1.12**.



The exact event is one highlighted bar, not a cumulative area.

TRAP p belongs to successes; $(1-p)^n$ to failures. "At least k " = $1-P(X \leq k-1)$.

Recall · Descriptive

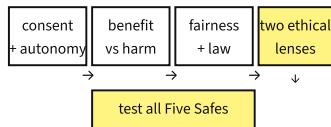
CH 1-2

PROMPT	FAST RESPONSE
right skew	mean > median; report median
sample spread	divide squared deviations by $n-1$
compare scales	$CV=s/\bar{x} \times 100\%$
bar chart	zero baseline
34% → 38.5%	+4.5 percentage points; +13.2% relative

Ethics Vignette Order

CH 11 WORKED

- Persons:** original purpose, consent, reasonable expectations.
- Benefit + justice:** weigh benefit against profiling/re-identification harm; ask who bears each.
- Law/public interest:** lawful, transparent, accountable; compare outcomes with duties.
- Five Safes:** projects, people, data, settings, outputs—apply each to the facts.



Marks come from framework-to-fact mapping, not a bare yes/no verdict.

Consequentialist: are net outcomes beneficial?
Deontological: are consent, confidentiality and duties respected regardless of benefit?

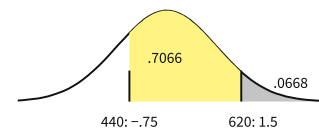
FIVE SAFES Is the project appropriate? Are people authorised? Can data identify anyone? Is the setting secure? Are outputs safe? Safety is not a property of the dataset alone.

W2 · Normal Sales

CH 5 WORKED · 4 MARKS

Given: $X \sim N(500, 80^2)$. Find (a) $P(X > 620)$, (b) $P(440 < X < 620)$. Use $\Phi(1.5) = .9332$, $\Phi(.75) = .7734$.

- 620: $z=(620-500)/80=$ **1.5**.
- $P(X > 620) = 1 - \Phi(1.5) = 1 - .9332 =$ **.0668**.
- 440: $z=(440-500)/80=$ **-.75**; $\Phi(-.75) = 1 - .7734 =$ **.2266**.
- $P(440 < X < 620) = .9332 - .2266 =$ **.7066**.



Middle band uses a difference; upper tail uses $1-\Phi(z)$.

TRAP The Z table gives left area. Sketch the requested region before subtracting.

Recall · Probability

CH 3

QUESTION WORD	OPERATION
not	complement: $1-P(A)$
A or B	add, subtract overlap
A and B	multiply conditional path
A given B	joint ÷ $P(B)$
independent?	joint = product of marginals

	A	A ^c
B	AnB	A ^c nB
B ^c	AnB ^c	A ^c nB ^c

$$\div P(B) = P(A|B)$$

Conditioning changes the denominator to the stated row or column.

REVERSE + PARTITION
 $P(A_i | B) = P(B | A_i) P(A_i) / P(B)$
 $P(B) = \sum P(B | A_i) P(A_i)$

Do not merge concepts: mutually exclusive non-zero events cannot be independent. If A occurs, an exclusive B becomes impossible; independence would leave $P(B)$ unchanged.

NORMAL / SAMPLING CLUE	DENOMINATOR OR AREA
one observation X	standardise with σ
sample mean $\bar{x}$	standardise with $\sigma/\sqrt{n}$
left of cut	$\Phi(z)$
right of cut	$1-\Phi(z)$
between two cuts	$\Phi(z_2) - \Phi(z_1)$

NORMAL SANITY CHECK For continuous X , $P(X=a)=0$, so endpoints do not change an interval probability. A z-score counts SDs from μ . The Z table returns left area $\Phi(z)$; sketch first, then use a complement or difference. Check against 68-95-99.7%.

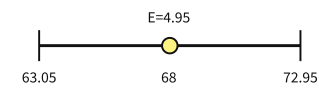
W3 · 95% CI for μ

CH 7 WORKED · 4 MARKS

Given: $n=25$, $\bar{x}=68$ seconds, $s=12$ seconds; $t_{.025,24}=2.064$.

- σ unknown → **t**, $df=25-1=$ **24**.
- $SE=s/\sqrt{n}=12/\sqrt{25}=$ **2.4 s**.
- Margin $E=t^*SE=2.064 \times 2.4=$ **4.95 s**.
- CI= $68 \pm 4.95 =$ **(63.05, 72.95) seconds**.

Interpret: the 95% interval-building procedure captures the true mean in 95% of repeated samples.



Point estimate in the centre; margin extends equally on both sides.

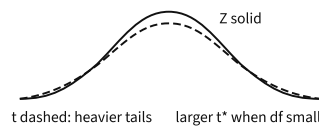
TRAP Unknown σ means **t**, and a two-sided interval uses $\alpha/2$ with $df=n-1$.

CI / Test Link

CH 7-8

INTERVAL / TEST	EQUIVALENT TWO-SIDED DECISION
95% CI excludes μ_0	reject H_0 at $\alpha=.05$
95% CI contains μ_0	fail to reject H_0 at $\alpha=.05$

Both use estimate-null measured in SEs. CI shows plausible parameter values and precision; the test gives a decision at chosen α . Neither statistical significance nor a narrow interval proves practical importance or causation.



t dashed: heavier tails larger t^* when df small
Estimating σ with s adds uncertainty, so t critical values are larger.

CONFIDENCE	A	Z*
90%	.10	1.645
95%	.05	1.96
99%	.01	2.576

INFERENCE TARGET	REFERENCE DISTRIBUTION
mean; σ known	Z
mean; σ unknown	t, $df=n-1$
one variance / SD	χ^2 , $df=n-1$; right-skewed, non-negative

Width check: confidence $\uparrow$ widens; $n \uparrow$ narrows. A more precise interval is narrower, but precision does not repair a biased sample or prove the model assumptions.

W4 · Pooled Two-Sample t

CH 12 WORKED · 4 MARKS

Given: independent groups: $n_1=12$, $\bar{x}_1=52$, $s_1=6$; $n_2=15$, $\bar{x}_2=48$, $s_2=5$; equal variances; $\alpha=.05$ two-tailed; $t_{.025,25}=2.060$.

- Route: two independent means, σ unknown, equal variances → pooled t ; $df=12+15-2=$ **25**. $H_0:\mu_1=\mu_2$; $H_1:\mu_1 \neq \mu_2$.
- $s^2p=[(11 \times 36 + 14 \times 25)/25]=746/25=$ **29.84**.
- $SE=s_p/\sqrt{1/12+1/15}=\sqrt{4.476}=$ **2.116**.
- $t=(52-48)/2.116=$ **1.89**. Since $|1.89| < 2.060$, **fail to reject H_0** : insufficient evidence of a mean difference at 5%.



The observed t remains inside the non-rejection region.

TRAP Paired observations are not two independent samples: calculate d , and run one-sample t on d .

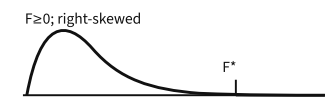
Group-Test Recall

CH 9

CLUE	STATISTIC	DF
same units twice	$t=d/(s_d/\sqrt{n})$	$n-1$
2 separate means	pooled t	n_1+n_2-2
2 variances	s_1^2/s_2^2	n_1-1, n_2-1
3+ means	MSB/MSW	$k-1, N-k$

ANOVA / F Read

CH 9 · FIG 9.3



ANOVA and the two-variance F-test use the upper tail.

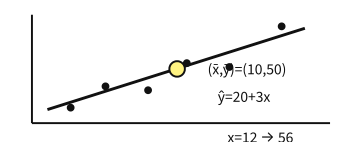
ANOVA: H_0 says all means are equal; a large $F=MSB/MSW$ means between-group variation dominates within-group noise. Rejection says at least one mean differs, not which. **Variance test:** put the larger sample variance on top so $F \geq 1$ and keep numerator/denominator df in the same order.

METHOD	CHECK BEFORE COMPUTING
paired t	natural one-to-one matching; analyse d .
pooled t	separate groups; equal variances assumed
ANOVA	three or more means; one upper-tail F
F variance test	ratio of two variances; preserve df order

W5 · Regression from Summaries

CH 10 WORKED · 4 MARKS

- Given:** advertising x (\$000s) vs sales y : $\bar{x}=10$, $\bar{y}=50$, $s_x=4$, $s_y=15$, $r=.8$. Find line, predict at $x=12$, state R^2 .
- $b_1=r(s_y/s_x)=.8(15/4)=.8 \times 3.75=$ **3.0**. Each extra \$1,000 advertising is associated with 3 more sales units.
 - $b_0=\bar{y}-b_1\bar{x}=50-3(10)=$ **20**. Thus **$\hat{y}=20+3x$** .
 - $x=12$: $\hat{y}=20+3(12)=$ **56 units**.
 - $R^2=r^2=.8^2=$ **.64**: 64% of sales variation is explained; 36% remains residual.



The line passes through $(\bar{x}, \bar{y})$; use this as the intercept check.

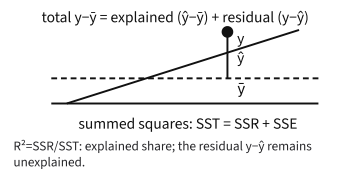
TRAP The slope ratio is s_y/s_x , not s_x/s_y . Association and high R^2 do not prove causation.

Interpretation Stems

CH 8-10

OUTPUT	WRITE
test rejects	"At α , evidence that $[H_1 \text{ in context}]$."
test does not reject	"At α , insufficient evidence that..."
slope b_1	"Each 1-unit x rise is associated with estimated b_1 -unit y change."
R^2	" $R^2 \times 100\%$ of variation in y is explained by the model."
CI	state bounds, units, parameter and long-run coverage

Never "accept H_0 ". Never call p the probability H_0 is true. Never extrapolate the fitted line beyond observed x without warning.



REGRESSION CHECK	MEANING
$\sum e_i^2$ minimised	least-squares criterion
line through $(\bar{x}, \bar{y})$	intercept arithmetic check
$R^2=r^2$	simple regression only

CONFOUNDER CHECK A confounder is related to x and influences y ; omission can bias the simple slope. Multiple regression reads b_1 holding measured predictors constant, but still does not prove causation. A causal claim asks what y would be if x changed and all else stayed the same.

13 · Trap Matrix

CH 1-10 FAQs

IF YOU SEE...	DO THIS
sample variance	divide by $n-1$
skew/outlier	prefer median + IQR
sample mean probability	use $q/\sqrt{n}$, not σ
σ unknown	use t , $df=n-1$
two-sided claim	split $\alpha/2$ per tail
$p \leq \alpha$	reject H_0
same subjects twice	paired t on differences
3+ means	ANOVA, not many t-tests
F variance ratio	larger variance on top
$r \approx 0$	check scatter for non-linearity
significant slope	do not infer causation
de-identified data	risk reduced, not eliminated

	H_0 true	H_0 false
reject	Type I, α	correct, power
fail reject	correct	Type II, β

Your decision is observable; the truth of H_0 is not.

15 · Final 20-Second Check

CH 12

- Correct method and table? df and tail right?
- SE divides by $\sqrt{n}$? Units carried?
- Probability lies 0-1; variance non-negative?
- Decision matches p /critical value?
- Conclusion names population, parameter and direction?
- Every question attempted?

OFFLINE EXCEL RECALL	FUNCTION NAMED IN GUIDE
centre / sample spread	AVERAGE, MEDIAN, STDEV.S, VAR.S
binomial	BINOM.DIST
normal probability / cut	NORM.DIST, NORM.S.DIST, NORM.S.INV
t confidence margin	CONFIDENCE.T

EXAM MOVE Route → formula → substitution → value → decision → contextual sentence.

LAST-PASS SYMPTOM	REPAIR
probability outside 0-1	recheck area, complement and arithmetic
variance / SD confused	variance has squared units; SD has original units
exact value for continuous X	probability is 0; use an interval
"accept H_0 "	write "fail to reject H_0 "
ANOVA rejected	say at least one mean differs, not which
R^2 treated as causation	state explained variation only
paired rows treated separately	form d , first; test the differences
F-table lookup	keep numerator and denominator df in ratio order