BIOS6902 Statistical Inference

CITY UNIVERSITY OF HONG KONG · SEMESTER A 2026/27

Full guide + ask Sia about BIOS6902

SIDE 1 / 2 · COPY THIS BY HAND

HANDWRITTEN-A4 PLANNER · targets → assumptions → rule → interpretation

PRINTED SHEET NOT PERMITTED

Exam map · official facts

PRIORITY 0

Assignments 30% | Midterm 30% | Final 40%

Final examination (2 hours) · 40% · closed book, with ONE A4 sheet of HANDWRITTEN notes permitted (official course materials).

PUT ON YOUR PAGE	WHY IT EARNS SPACE
Target + estimand	Prevents mean/median and association/effect swaps
Assumptions	Decide whether a method is defensible
SE + reference law	Turns estimate into interval/test
Failure + repair	Shows diagnostic judgement
Reporting sentence	Stops probability reversals

Separate pass conditions: at least 40% of possible continuous-assessment marks and at least 35% of possible final-exam marks; both apply.



Course spine: population target, sample estimator, then error accounting.

1 · Estimation skeleton

L1

Name the target: $\theta=T(F)$. Estimator $\hat{\theta}$ is a function of the sample. $\text{Bias}=E(\hat{\theta})-\theta$; $\text{Var}=E[(\hat{\theta}-E\hat{\theta})^2]$; MSE=bias²+variance. Consistency means MSE→0; bias can lower MSE.

MEAN UNDER INDEPENDENCE

$E(\bar{X})=\mu$ · $\text{Var}(\bar{X})=\sigma^2/n$ · $SE=s/\sqrt{n}$

dependence adds covariance terms ⇒ naïve SE too small

CI rule: coverage belongs to the procedure; after observation, endpoints and θ are fixed.

Model → target → report

FIRST 30 SECONDS

SCIENTIFIC QUANTITY	MODEL SHORTHAND	TARGET
Disease prevalence	Bernoulli	p=E(X)
Biomarker centre	Normal / nonparametric	μ or median
Event count	Poisson	λ
Survival time	Gamma / nonparametric	mean or median

Do not let convenience rename the target. A median question is not a mean question merely because $\bar{X}$ has an easy SE. Name the sampling unit and independence assumption; repeated patients or clinics add covariance and widen uncertainty.

Model bargain: parametric precision costs a defensible family; nonparametric robustness usually costs width.

2 · Empirical distribution

L2

STAIRCASE

$\hat{F}_n(x)=n^{-1}\Sigma I(X_i\le x)$

$E(\hat{F}_n)=F(x)$ · $\text{Var}=F(x)\{1-F(x)\}/n$

Plug-in: estimate T(F) by $T(\hat{F}_n)$. Linear functional $\int r(x)dF$ becomes $n^{-1}\Sigma r(X_i)$; medians, correlations and ratios stay computable when analytic SEs are awkward.

SIMULTANEOUS DKW BAND

$P\{\sup_x|\hat{F}_n-F|>\varepsilon\}\le 2e^{-2n\varepsilon^2}$

$\varepsilon_n=\sqrt{\log(2/\alpha)/(2n)}$; clip $\hat{F}\pm\varepsilon$ to $[0,1]$

Pointwise ≠ simultaneous. Pay for a whole-curve claim only when you intend to search across x.

SAMPLE
F ⟶ n vals ⟶ stat

RESAMPLE
EDF ⟶ n draws ⟶ stat

SUBSTITUTE
resample spread ≈ sampling spread if EDF is close

Bootstrap substitution: resampled spread stands in for repeated-sample spread.

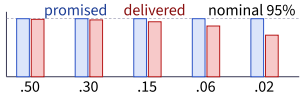
Bootstrap recipe

RESAMPLE

1. Draw n observations with replacement.
2. Recompute identical T*; repeat B times.
3. $SE_{boot}=\sqrt{[B^{-1}\Sigma(T^*_b-\bar{T}^*)^2]}$.

INTERVAL	RULE / RISK
Normal	$\hat{T}\pm1.96$ SE; symmetric
Percentile	Use 2.5%,97.5% resample quantiles
Pivotal	$(2\hat{T}-q_{.975},2\hat{T}-q_{.025})$

Check: bootstrapped mean SE should be near $s/\sqrt{n}$. Bootstrap estimates variability; it need not remove boundary bias.



Wald coverage deteriorates as a proportion approaches a boundary.

Method choice

L4

Ask: target, sample size/tails, plausible family, finite-sample guarantee? Small n/heavy tails favour resampling or a bound; large n/rough symmetry supports a normal interval. Impossible probability endpoints need Wilson or Agresti–Coull.

Four failures: boundary Wald; symmetric interval for skew; bootstrap trusted to erase bias; parametric bootstrap from the wrong family.

Coverage simulation: repeatedly sample, build the interval and count captures; it checks the procedure's promised level, not whether one realised interval is "correct".

3 · Moments & likelihood

L3

Moments: match sample and population moments, solve. **MLE:** maximise $L(\theta)=\Pi f(X_i;\theta)$, usually via $\ell(\theta)=\Sigma\log f$. MLE is invariant: estimate $g(\theta)$ with $g(\hat{\theta})$.

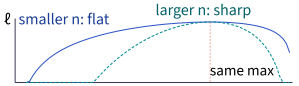
BERNOULLI / POISSON

$\hat{p}=\bar{X}$ · $SE(\hat{p})=\sqrt{\hat{p}(1-\hat{p})/n}$

$\hat{\lambda}=\bar{X}$ · $I_1(\lambda)=1/\lambda$ · $SE(\hat{\lambda})=\sqrt{(\hat{\lambda}/n)}$

Uniform(0,θ): $MLE=X_{(n)}$; $E=n\theta/(n+1)$, bias $-\theta/(n+1)$; corrected form $(n+1)X_{(n)}/n$.

Log scale: products become sums, differentiation is easier, and log preserves the maximiser.



Information is curvature: sharper likelihood means tighter estimation.

Information route

SCORE → SE

FOUR REUSABLE LINES

$s(X;\theta)=\partial\log f/\partial\theta$

$I_1(\theta)=\text{Var}\{s\}=-E(\partial^2\log f/\partial\theta^2)$

$I_n=nI_1$ · $SE(\hat{\theta})\approx1/\sqrt{I_n(\hat{\theta})}$

CI: $\hat{\theta}\pm z_{\alpha/2}SE$

Delta: $SE\{g(\hat{\theta})\}\approx|g'(\theta)|SE(\hat{\theta})$. If g is sharply curved or $g'\approx0$, simulate from the fitted model; efficiency then depends on correct specification.

Logit: $SE(\hat{\psi})=\sqrt{[1/\{n\hat{p}(1-\hat{p})\}]}$. **Exponential rate:** $I_1=1/\lambda^2\Rightarrow SE=\lambda/\sqrt{n}$.

Assumption ledger

CHOOSE

TOOL	NEEDS	FAILURE
Wald	Approx normality	Boundary/skew/small n
Nonparam bootstrap	Representative sample	Bias reproduced
Param bootstrap	Correct family	Confident misspecification
Hoeffding/DKW	Bounded variable / CDF	Extra width

Information is additive: independent observations give $I_n=nI_1$, so SE falls as $1/\sqrt{n}$. Halving an SE takes four times the data. Use observed information at $\hat{\theta}$ for an estimated SE; use the null value in a null-based test when that test's derivation requires it.

Transformation checklist: state g; differentiate; evaluate $|g'|$ at $\hat{\theta}$; multiply by $SE(\hat{\theta})$; transform the estimate; inspect whether a symmetric interval can leave the transformed parameter space. The delta method is local linearisation, not a licence to ignore curvature.

MLE diagnostics: check parameter boundaries, verify the score root is a maximum, and remember regular asymptotics can fail when the true value lies on a boundary or the model is not identifiable.

5 · Sufficiency

L5

T(X) is sufficient when the conditional law of full data given T=t is free of θ. **Factorisation:** $p(x;\theta)=h(x)g(T(x);\theta)$; θ touches data only through T.

FAMILY	SUFFICIENT SUMMARY
Bernoulli / Poisson	ΣX_i
Normal, μ and σ^2 unknown	$(\Sigma X_i,\Sigma X_i^2)$
Uniform(0,θ)	$X_{(n)}$, not $\bar{X}$

Rao–Blackwell: $E(\theta^*|T)$ cannot raise MSE. One-to-one functions preserve sufficiency; many-to-one functions may lose it.



Rao–Blackwell: conditioning on a sufficient summary reduces spread.

6 · Bayesian update & CRB

L6

BAYES FOR PARAMETERS

posterior ∝ likelihood × prior

$\text{Beta}(\alpha,\beta)+Y$ of n ⇒ $\text{Beta}(\alpha+Y,\beta+n-Y)$

posterior mean=$(Y+\alpha)/(n+\alpha+\beta)$

Read α,β as prior successes/failures. Report the prior and show sensitivity when n is small. Normal–Normal posterior precision is prior precision plus sample precision.

CRAMÉR–RAO

$\text{Var}(\hat{\theta})\ge1/\{nI_1(\theta)\}$ for unbiased estimators only

Bernoulli: $I_1(p)=1/[p(1-p)]$ ⇒ bound $p(1-p)/n$

Attaining the bound means efficient. A biased estimator may beat this variance floor in MSE.



EM: E-step responsibilities → M-step weighted means → repeat with updated parameters.

EM loop

LATENT LABELS

E: compute responsibilities under current parameters.

M: update parameters with responsibility-weighted averages. Iterate until stable; use several starts because mixture likelihoods can have several peaks.

TWO EQUAL-VARIANCE COMPONENTS

$\tau_i=\varphi(y_i-\mu_0)/\{\varphi(y_i-\mu_0)+\varphi(y_i-\mu_1)\}$

$\mu_0\leftarrow\Sigma\tau_iy_i/\Sigma\tau_i$ · $\mu_1\leftarrow\Sigma(1-\tau_i)y_i/\Sigma(1-\tau_i)$

Trap: EM can settle at a local peak; compare several starts and likelihoods.

Posterior ≠ point estimate. The posterior is a distribution; its mean is one chosen functional. A strong prior can reduce variance by adding bias, so compare posterior mean with the MLE and describe the prior in units of observations.

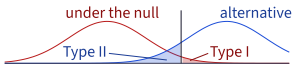
7 · Tests, errors, power

L7–8

State H_0,H_1 in parameter language; choose statistic and rejection region before seeing direction. Power $\beta(\theta)=P_\theta(\text{reject }H_0)$. Type I rejects true H_0; Type II misses H_1.

SIZE α	$\sup_{\theta\in\Theta_0}\beta(\theta)=\alpha$
LEVEL α	$\sup_{\theta\in\Theta_0}\beta(\theta)\le\alpha$
POWER	1−Type II at stated alternative

Neyman–Pearson: simple vs simple, reject for large $L(\theta_1)/L(\theta_0)$; most powerful at level α against that one alternative.



Moving the threshold trades Type I error (α) against Type II error (β).

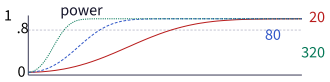
General tests

WALD / LR

Wald: $(\hat{\theta}-\theta_0)/SE$ under H_0; two-sided reject when $|T|\ge z_{\alpha/2}$. Proportion test uses p_0 in the null spread; interval uses $\hat{p}$.

LR: $\lambda=\sup_{\Theta_0}L/\sup_{\Theta}L$. Under regularity, $-2\log\lambda\to\chi^2_d$, d=constrained parameters, not observations.

P-value: probability under H_0 of a statistic at least as extreme as observed; neither $P(H_0|\text{data})$ nor importance. Large p means insufficient evidence, not proof.



Power rises with effect size; curves are n=20, 80, 320 (red, blue, green); at the null it equals α.

Fit & permutation

L8

PEARSON FIT

$\chi^2=\Sigma(O_i-np_{0i})^2/(np_{0i})$ · df=k-1

Permutation: under exchangeable labels, recompute any statistic over relabellings; p is the share at least as extreme. Exact if all are enumerated.

NORMAL TESTS

upper-tail $H_1:\theta>\theta_0$ rejects if $\bar{X}>\theta_0+z_\alpha/\sqrt{n}$

two-sided rejects if $|\bar{X}-\theta_0|>z_{\alpha/2}/\sqrt{n}$

Power: rises with $\sqrt{\{nI_1\}}$ times effect size. Near H_0 it equals α; with large n, tiny effects can be detectable, so report magnitude and interval with the verdict.

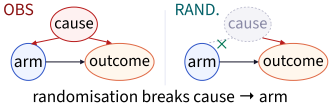
Goodness-of-fit conditions: compute expected counts before the statistic; combine categories if some are too small. If parameters are estimated from the same data, the reference degrees of freedom must account for them; for the fully specified course case it is k−1.

9 · Potential outcomes

L9

Observe $Y_i=A_iY_i(1)+(1-A_i)Y_i(0)$; individual effect is never observed. Target $\tau=E\{Y(1)-Y(0)\}$.

Association ≠ effect: $E(Y|A=1)-E(Y|A=0)=\tau$ only with design/assumptions. Randomisation makes A independent of potential outcomes, turning the difference in means into a causal estimator.



Randomisation severs the common cause → assignment arrow.

Experiment rule

BEFORE LAUNCH

Pre-specify primary outcome, α, smallest worthwhile effect, sample size and stopping rule; randomise; check balance; report estimate + interval + p-value; distinguish practical from statistical significance.

SIZING RULE

80% power, two-sided $\alpha=.05$: $n\approx16\sigma^2/\delta^2$ per arm

10 · Adjustment

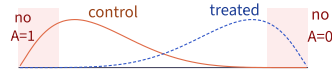
L10

Ignorability: $A\perp\{Y(0),Y(1)\}\,|\,X$. **Positivity:** $0<P(A=1|X)<1$ where inference is claimed. Ignorability is not testable from observed data.

PROPENSITY / IPW

$\pi_i=P(A_i=1|X_i)$

$\tau=E[A_iY_i/\pi_i-(1-A_i)Y_i/(1-\pi_i)]$



No overlap: left tail has no treated units; right tail has no controls. Restrict the target or redesign.

Model treatment chance → plot overlap → estimate → sandwich SE/resampling → inspect weights → sensitivity analysis. Near-zero scores create enormous weights; poor overlap changes the population you can study.

11 · Trustworthy tests

L11

Peeking and many metrics buy repeated chances to be wrong. K independent looks: $FWER=1-(1-\alpha)^K$. Bonferroni tests each planned look/metric at α/K. A correction chosen after inspection cannot restore the guarantee.

Pre-specify: primary outcome and stopping rule.

Design sentence: "Random allocation, pre-specified outcome and stopping rule; report estimate, interval and practical meaning."

Forward this to your tutorial group

V2 · 2026-10-11

Semester A examination period, 7–19 December 2026.

Your exam allows one HANDWRITTEN A4 — copy side 1 by hand; do not bring this printout.

BIOS6902 Statistical Inference · Drills & traps

CITY UNIVERSITY OF HONG KONG · SEMESTER A 2026/27

Full guide + ask Sia about BIOS6902

SIDE 2 / 2 · DO THESE BEFORE THE EXAM

WORKED DRILLS · calculate → diagnose → write the conclusion

REVISION PRINTOUT ONLY

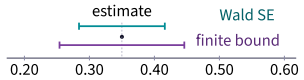
D1 · Prevalence from counts

L1 WORKED

Given: 70/200 ⇒ $\hat{p}$=.35. SE=$\sqrt{\{.35\times.65/200\}}$=.0337. 95% CI=.35±1.96(.0337)=(.284,.416).

Write: data are consistent with prevalence about 28%–42%; the procedure captures truth in 95% of repeated studies of this size. **Never:** "95% probability p is in this realised interval."

Test H_0:p=.30: z=(.35−.30)/$\sqrt{\{.30\times.70/200\}}$=1.54; p≈.12 ⇒ do not reject at 5%.



Same estimate, two intervals: validity without shape assumptions costs width.

D2 · Bias–variance proof

SHOW STEPS

Mean: E($\bar{X}$)=μ; Var=nσ²/n²=σ²/n; MSE→0. **Uniform max:** E($X_{(n)}$)=nθ/(n+1); bias=−θ/(n+1), yet MSE→0: biased and consistent. **Poisson mean:** unbiased; Var=MSE=λ/n.

Choice rule: compare estimators by MSE, not bias alone. Slight off-centring can win when it greatly reduces sampling variance; consistency is a limit statement, not finite-sample unbiasedness.



Equal MSE can arise from high variance or squared bias; the remedy differs.

Estimator table

RECALL

MODEL	ESTIMATE · SE
Bernoulli p	$\bar{X}$ · $\sqrt{\hat{p}(1-\hat{p})/n}$
Poisson λ	$\bar{X}$ · $\sqrt{(\bar{X}/n)}$
Normal μ	$\bar{X}$ · s/$\sqrt{n}$
Uniform upper θ	$X_{(n)}$; corrected ×(n+1)/n

Boundary drill: for a small observed proportion close to zero, decline the ordinary normal interval because its approximation is poor and can leave [0,1]; choose a distribution-free or named boundary-respecting interval and state the reason.

Dependence drill: if several outcomes come from the same patient or clinic, Var(ΣX_i) contains covariance terms. Positive within-cluster correlation makes s/$\sqrt{n}$ too optimistic; naming independence is part of the answer, not decoration.

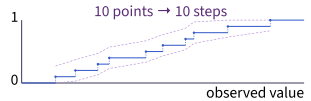
Interval denominator: for an estimate-centred Bernoulli interval use $\hat{p}(1-\hat{p})$; for a score-type null test derived under H_0 use $p_0(1-p_0)$. Mixing the two midway produces a statistic with no stated reference law.

D3 · EDF & DKW

L2

Fix x; B_i=I(X_i≤x)~Bernoulli(F(x)). F̂=$n^{-1}\Sigma B_i$; E(F̂)=F and Var(F̂)=F(1−F)/n. The supremum in DKW gives whole-curve control.

At fixed x: uncertainty is widest where F(x)=.5 because F(1−F) is maximal there. The DKW half-width is constant in x, shrinks at $n^{-1/2}$, and is clipped to [0,1]. Unlike pointwise intervals, it controls the whole staircase in one statement.



Ten observations make ten EDF steps; a simultaneous band has constant vertical width.

D4 · Three bootstrap CIs

CHOLESTEROL

Median gap T̃≈18.5; SE≈7.4. Normal: (3.7,33.3). Percentile: (5.0,33.3). Pivotal: (5.0,34.0). All exclude zero; lower ends expose asymmetry. If resampling is skewed, avoid symmetric normal CI; state why.

Ratio estimate −.07 is inside ±.20, but CI (−.24,.15) is not contained: equivalence is *not* shown. The whole interval must fit.

Failure → repair

L4

CI OUTSIDE [0,1]	Boundary; Wilson/Agresti–Coull
SKEWED RESAMPLES	Percentile/pivotal
MAX BIAS PERSISTS	Bootstrap is not automatic bias repair
PARAM CI FAR NARROWER	Check family specification

Right-skewed survival: exponential fit targets mean; nonparametric median is a different functional. Report the median and resampled interval when shape is not defensible.

Course data drill: two cholesterol groups give a plug-in difference of means about 20.9 with SE about 5.55, hence an approximate 95% interval about (9.8,32.0). State that the disease group is higher on this evidence; do not convert an observational group difference into an intervention claim.

Band reading: a DKW band controls the maximum vertical gap everywhere. To estimate P(a<X≤b), read F̂(b)−F̂(a); uncertainty for a searched-over contrast must respect simultaneity. A collection of separately valid pointwise intervals does not automatically protect the whole curve.

Bootstrap coding check: every resample has size n and is drawn with replacement from the empirical distribution; recompute the entire statistic, including any fitted nuisance parameters. Resampling residual pieces while holding part of the calculation fixed changes the procedure.

D5 · Bernoulli MLE

DERIVE

ℓ(p)=Ylogp+(n−Y)log(1−p). Score Y/p−(n−Y)/(1−p)=0 ⇒ p̂=Y/n. For 35/100: p̂=.35; SE=.0477; CI≈(.256,.444). Moments agrees.

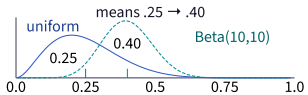
Poisson: −E(second derivative)=1/λ ⇒ SE=$\sqrt{(\hat{\lambda}/n)}$.

Exponential rate: I_1=1/λ² ⇒ SE=λ̂/$\sqrt{n}$. **Delta:** for τ=e^{μ}, τ̂=$e^{\hat{X}}$, SE≈$e^{\hat{X}}$/$\sqrt{n}$.

D6 · Two priors

Y=2, N=10

MLE=.20. Beta(1,1)→Beta(3,9), mean .25. Beta(10,10)→Beta(12,18), mean .40. Strong prior contributes 20 imaginary observations—twice the real sample—and pulls toward .50. Report sensitivity.



Two successes in ten: prior strength moves the posterior mean from .25 to .40.

D7 · CRB check

EFFICIENT?

Bernoulli I_1(p)=1/[p(1−p)]. Bound=1/[nI_1]=p(1−p)/n=Var($\bar{X}$); sample proportion is efficient. **Trap:** bound restricts unbiased estimators only.

D8 · Poisson test twice

N=20

$\bar{X}$=1.4, H_0:λ=1. Wald z=.4/$\sqrt{(1/20)}$=1.79; p≈.073. LR=2n{($\lambda_0-\bar{\lambda}$)−$\bar{\lambda}$log($\lambda_0/\bar{\lambda}$)}≈2.84; compare χ^2_1. Neither rejects at 5%.

Report: both tests point towards a higher rate but do not reach the 5% threshold; give the estimate and interval so the reader sees the remaining uncertainty. "Retain H_0" must never become "prove λ=1".

D8b · Normal prior weighting

PRECISION

For X_i~N(μ,σ²), σ known, and μ~N(m,τ²), posterior variance={1/τ²+n/σ²}$^{-1}$. Posterior mean is a weighted average of $\bar{x}$ and m; each source receives weight proportional to its precision. As n grows, σ²/n shrinks and the prior weight vanishes. More precision means more weight in the posterior mean.

EM drill: write responsibilities first, then weighted means. A mixture likelihood is a product of sums; the complete-data version would separate if labels were known. EM uses expected labels, not recovered labels. Run different initial means; agreement is a diagnostic, disagreement exposes multiple peaks.

Prior sensitivity sentence: "The posterior mean moved from ___ to ___ across the stated priors; the stronger prior contributed the equivalent of ___ observations." This is more informative than calling a prior merely weak or strong.

D9 · Factorise & improve

L5

Poisson joint mass=[Π1/x_i!]$e^{-n\lambda}\lambda^{\Sigma x_i}$; only T=Σx touches λ ⇒ sufficient. Given T=t, E(X_1|T)=t/n=$\bar{X}$; variance falls λ→λ/n.

Uniform: θ^{-n}I($X_{(n)}$≤θ), so max sufficient, mean not. **Normal:** expanding squares leaves Σx and Σx².

3 trials → totals

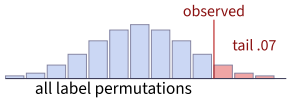
Σ=0	Σ=1	Σ=2	Σ=3
000	one 1	two 1s	111
1 seq	3 seq	3 seq	1 seq

For three Bernoulli trials, the total partitions eight sequences into sufficient cells.

D10 · Goodness of fit

N=200

p_0=(.40,.35,.15,.10); O=(92,64,28,16); E=(80,70,30,20). Contributions 1.800,.514,.133,.800; sum=3.248, df=3. Below upper 5% point ⇒ compatible, not proof of equality. Estimating parameters reduces the reference degrees of freedom.



Permutation p-value: share of relabellings at least as extreme as observed.

D11 · Exact permutation

BY HAND

(X_1,X_2,Y_1)=(1,9,3); observed |$\bar{X}-\bar{Y}$|=2. Of 3!=6 label reshuffles, four are at least as extreme; exact p=4/6≈.67. For large samples, sample many relabellings and report the count.

Exchangeability is key. Random assignment guarantees it under the sharp null; otherwise shuffling may answer an unsupported question.

D11b · Testing language

MARKS IN WORDS

Tea tasting: all four correct among four-of-eight choices occurs once in 70 random attempts, about .014. The logic is: null claim, statistic, probability under the null, line fixed in advance. Do not skip which outcomes count as at least as extreme.

Size vs level: size α means the worst null rejection probability equals α; level α means it is at most α. Discrete outcomes may not allow exactly .05, so a valid test can be conservative.

Observed z=1.70: two-sided p≈.089. It does not reject at 5% (critical 1.96) but does at 10% (critical 1.645). Reporting .089 preserves both facts and shows how close the evidence lies to either chosen line.

LR degrees of freedom: count restrictions: dim(Θ) −dim(Θ_0). Do not use sample size or number of categories unless the model's restrictions make that count correct.

D12 · Potential outcomes

L9

Each unit reveals Y(1) or Y(0), never both. Individual effect is unobserved; estimate an average across randomised groups. Do not call an observational difference a treatment effect without identifying assumptions.

U	A	Y1	Y0	Δ
1	T	seen	—	?
2	C	—	seen	?
3	T	seen	—	?
4	C	—	seen	?

Fundamental problem: one potential outcome is permanently hidden for each unit.

D13 · Binary experiment

500/ARM

$\hat{p}_1$=.15, $\hat{p}_0$=.12 ⇒ effect=.03. SE=$\sqrt{\{.15(.85)/500+.12(.88)/500\}}$=.0216. CI=(−.012,.072). Conclusion: compatible with small loss through substantial gain; study did not settle the question.

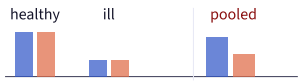
Sizing: σ=20, δ=4 ⇒ n≈16(400)/16=400/arm. Halve δ and n quadruples.

D14 · Confounding reversal

L10

Healthy: 320 treated,80 control, rate .80 both. Unhealthy: 80 treated,320 control, rate .30 both. Pooled treated=.70; control=.40; naïve gap=.30. Within-stratum gaps=0; standardised effect=0.

The naïve number compared different people, not interventions. Adjustment changes the comparison; it cannot verify all common causes were measured.



Blue/orange = treated/control. Healthy: n=320/80, both .80; ill: n=80/320, both .30; pooled .70/.40.

Adjustment traps

DIAGNOSE

EXTREME Π	Huge weights; inspect/trim and report
NO OVERLAP	Restrict support; target changes
HIDDEN CAUSE	Sensitivity analysis; qualify
POST-TREATMENT X	May remove part of effect

Five-step observational answer: state ignorability and positivity; model π=P(A=1|X); plot overlap by arm; form the weighted/standardised estimate with appropriate uncertainty; finish with weight diagnostics and sensitivity to an unmeasured cause.

Randomisation test: under the sharp null, keep outcomes fixed, rerandomise labels by the actual mechanism and count contrasts at least as extreme.

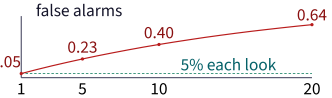
Covariates: adjust for pre-treatment common causes. A mediator removes part of the effect; a post-treatment variable is not a baseline confounder.

D15 · Peeking & metrics

L11

Ten looks: 1−.95^{10}=.401; Bonferroni .005/look. Real looks correlate, so independent arithmetic overstates inflation but shows direction.

Twelve metrics: 1−.95^{12}=.460; threshold .05/12=.0042. p=.031 fails; call it hypothesis-generating and confirm as a new primary outcome.



Independent repeated looks inflate the chance of at least one false alarm.

Fast test selector

PROMPT → TOOL

SIMPLE VS SIMPLE	Neyman–Pearson LR
GENERAL PARAMETER	Wald / composite LR
COUNTS VS PROPORTIONS	Pearson χ^2, df k−1
RANDOMISED ODD SHAPE	Permutation
MANY LOOKS/METRICS	Pre-specify; adjust α

Sentence traps

WRITE PRECISELY

"95% CHANCE Θ INSIDE"	95% is repeated-study coverage
"P=P(H_0 TRUE)"	Tail probability under H_0
"NOT SIGNIFICANT=NO EFFECT"	Insufficient evidence; inspect CI/power
"SIGNIFICANT=IMPORTANT"	Judge magnitude separately
"ASSOCIATION=EFFECT"	Need design/identification
"CRB LIMITS ALL"	Unbiased only

Formula belt

REBUILD

MSE=bias²+Var · SE($\bar{X}$)=s/√n

ε_{DKW}=√{log(2/α)/(2n)}

I_1=Var(score)=−E(second derivative) · SE≈1/√(nI_1)

SE{g(θ̂)}≈|g′(θ̂)|SE(θ̂)

posterior ∝ likelihood×prior; Beta adds counts

Wald=(θ̂−θ_0)/SE_0 · −2logLR→$\chi^2_{constraints}$

χ^2 fit=Σ(O−E)²/E · ATE=E{Y(1)−Y(0)}

n≈16σ²/δ²/arm · FWER=1−(1−α)^K

Final five checks

DISCIPLINE

1. Target and denominator named?
2. Assumption buying result stated?
3. p̂ for CI, p_0 for null test?
4. Units, df, reference law given?
5. Conclusion avoids probability reversal and causal overreach?

Answer: method → numbers → reference → scientific conclusion.

Forward this to your tutorial group

V2 · 2026-10-11

Semester A examination period, 7–19 December 2026.

Your exam allows one HANDWRITTEN A4 — copy side 1 by hand; do not bring this printout.