HE2003 Econometrics I - Cheat Sheet (Quiz: Lectures 1–5 + Lecture 6 pp. 1–3) **QRS@NTU — qrsntu.org**

1. Causality & Probability Toolkit

3 correlation cases: X, Y correlated: (i) causality $X \to Y$; (ii) reverse causality $Y \to X$; (iii) common factor $C \to X$ and $C \to Y$ (confounder). Correlation alone never tells you which.

E/Var/Cov rules: $\mathbb{E}[aX+b] = a\mathbb{E}[X]+b$; $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$; $\mathrm{Cov}(aX+b, Y) = a\,\mathrm{Cov}(X,Y)$, $\mathrm{Cov}(X, Y+Z) = \mathrm{Cov}(X,Y)+\mathrm{Cov}(X,Z)$; $\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y + 2\,\mathrm{Cov}(X,Y)$.

Independence: $X \perp Y \Rightarrow \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \Rightarrow \mathrm{Cov}(X,Y) = 0$. Converse *false*: Cov = 0 (uncorrelated) does not imply independent (nonlinear dependence possible).

LIE: $\mathbb{E}[\mathbb{E}(Y \mid X)] = \mathbb{E}[Y]$. Fast route to $\mathbb{E}[Y]$ or $\mathbb{E}[g(X)Y]$: find $\mathbb{E}[Y \mid X]$ first, then average over X; cross-check against the direct expansion.

LIE worked example: $X \in \{0,1\}$ with $P(X{=}1) = 0.3$; $\mathbb{E}[Y \mid X{=}1] = 10$, $\mathbb{E}[Y \mid X{=}0] = 4$: $\mathbb{E}[Y] = 0.3(10) + 0.7(4) = 3 + 2.8 = 5.8$.

$N(0,1)$ **moments:** $X \sim N(0,1)$: $\mathbb{E}[X] = 0$, $\mathbb{E}[X^2] = 1$, $\mathbb{E}[X^3] = 0$ (all odd moments vanish by symmetry of ϕ).

Type I/II & significance: Type I = reject true H_0 (false positive); Type II = fail to reject false H_0 (false negative). Significance level := $P(\text{Type I} \mid H_0 \text{ true})$, usually 10/5/1% (0.1% legal/medical). Stricter cutoff ⇒ Type I↓, Type II↑; cannot shrink both at once.

RCT logic: Random assignment ⇒ $X \perp u$ (full independence) ⇒ $\mathbb{E}[u \mid X] = \mathbb{E}[u] = 0$. Causal effect ≈ $\mathbb{E}[Y \mid \text{treatment}] - \mathbb{E}[Y \mid \text{control}]$; RCTs often infeasible in economics ⇒ need observational-data tools (Ch. 3–).

Case examples: Beer sales vs. A/C sales: confounded by temperature (case iii). Windmill speed vs. wind speed: wind causes windmill (case i, not ii). Ice-cream sales vs. drowning: common factor temperature – regress each on temperature separately, never one on the other.

Conditional mean & population/sample: $\mathbb{E}(Y \mid X = x)$ = population mean of Y among units with $X = x$; e.g. gender wage gap = $\mathbb{E}(\text{wage} \mid \text{male}) - \mathbb{E}(\text{wage} \mid \text{female})$. *Population* = every unit; *sample* = the units actually drawn – keep the two distinct.

COVID ART example: Raise threshold 10 → 100 units: Type I↓ (fewer healthy people flagged), Type II↑ (more infected missed). Lower threshold 10 → 1: Type I↑, Type II↓ – illustrates the trade-off with a concrete knob.

2. The OLS Estimator

Model: $Y_i = \beta_0 + \beta_1 X_i + u_i$, $i = 1, \ldots, n$; Y_i dependent var, X_i regressor, u_i = every other factor affecting Y_i.

Three betas: (β_0, β_1) true unknown params; (b_0, b_1) generic candidates inside the minimization; $(\hat\beta_0, \hat\beta_1)$ the random OLS *estimators* (minimizers). Never conflate the three.

OLS objective: $(\hat\beta_0, \hat\beta_1) = \arg\min_{b_0, b_1} S(b_0, b_1)$, $S(b_0, b_1) = \sum_i [Y_i - (b_0 + b_1 X_i)]^2$ (sum of squared prediction errors).

FOCs: $\dfrac{\partial S}{\partial b_0} = -2\sum_i (Y_i - b_0 - b_1 X_i) = 0$; $\dfrac{\partial S}{\partial b_1} = -2\sum_i (Y_i - b_0 - b_1 X_i) X_i = 0$.

Normal equations: (4) $\bar Y - \hat\beta_0 - \hat\beta_1 \bar X = 0$; (5) $\frac{1}{n}\sum_i X_i Y_i - \hat\beta_0 \bar X - \hat\beta_1 \left(\frac{1}{n}\sum_i X_i^2\right) = 0$. Solve (4) for $\hat\beta_0$, substitute into (5), isolate $\hat\beta_1$.

$\hat\beta_1, \hat\beta_0$ **formulas:** $\hat\beta_1 = \dfrac{\frac{1}{n}\sum_i X_i Y_i - \bar X \bar Y}{\frac{1}{n}\sum_i X_i^2 - \bar X^2} = \dfrac{\sum_i (X_i - \bar X)(Y_i - \bar Y)}{\sum_i (X_i - \bar X)^2} = \dfrac{s_{XY}}{s_X^2}$; $\hat\beta_0 = \bar Y - \hat\beta_1 \bar X$. Needs $\sum_i (X_i - \bar X)^2 > 0$ (X must vary).

Summation identities: $\sum_i (X_i - \bar X)(Y_i - \bar Y) = \sum_i X_i Y_i - n\bar X \bar Y$; $\sum_i (X_i - \bar X)^2 = \sum_i X_i^2 - n\bar X^2$. Correct $\frac{1}{n}$ form (both sides): $\frac{1}{n}\sum_i (X_i - \bar X)(Y_i - \bar Y) = \frac{1}{n}\sum_i X_i Y_i - \bar X \bar Y$ – a lone $\frac{1}{n}$ on the RHS only is wrong by a factor of n.

Mean point: OLS line always passes through $(\bar X, \bar Y)$: $\hat\beta_0 + \hat\beta_1 \bar X = \bar Y$; recovers $\bar Y$ from $\bar X$ with no raw data.

Worked $\hat\beta_1, \hat\beta_0$ **example:** Given $\bar X = 10$, $\bar Y = 50$, $s_{XY} = 8$, $s_X^2 = 4$: $\hat\beta_1 = 8/4 = 2$; $\hat\beta_0 = 50 - 2(10) = 30 \Rightarrow \hat Y = 30 + 2X$.

Fitted value & residual: $\hat Y_i = \hat\beta_0 + \hat\beta_1 X_i$; $\hat u_i = Y_i - \hat Y_i = Y_i - (\hat\beta_0 + \hat\beta_1 X_i)$.

Prediction recipe: At $X = x_0$: $\hat Y = \hat\beta_0 + \hat\beta_1 x_0$. Change in prediction for ΔX: $\Delta\hat Y = \hat\beta_1\,\Delta X$, or the difference of two direct predictions – both must agree.

Convexity check: Hessian of $S(b_0, b_1)$: leading minor $2n > 0$, $\det H = 4n\sum_i (X_i - \bar X)^2 > 0 \Rightarrow$ positive definite ⇒ the FOC stationary point is the unique global minimum (not just a critical point).

California reference: $n = 420$: $\widehat{\text{TestScore}} = 698.9 - 2.28\,\text{STR}$. At STR= 20: $\hat Y = 653.3$; a 1-student cut in STR ⇒ +2.28 points.

Interpretation traps: Units of $\hat\beta_1$ = (units of Y)/(unit of X). Never quote $\hat\beta_0$ as a valid prediction at $X = 0$ outside the observed range of X (extrapolation); $\hat\beta_1$ is a *random* estimator, not the fixed β_1.

CEO salary residual example: $n = 209$: roe= 14.1 ⇒ $\widehat{\text{salary}} = 1224.05$; actual salary = 1095 ⇒ $\hat u = -129.06$ – an individual district/firm can deviate from the "on average" fitted line.

3. Least Squares Assumptions

LSA 1 (Least Squares Assumption 1): $\mathbb{E}[u_i \mid X_i] = 0$ (zero conditional mean).

LSA 2: (X_i, Y_i), $i = 1, \ldots, n$, are i.i.d.

LSA 3: Large outliers unlikely (informally: finite fourth moments).

$\mathbb{E}[Y \mid X]$ **line:** LSA1 ⇒ $\mathbb{E}[Y_i \mid X_i] = \beta_0 + \beta_1 X_i$: at any fixed X, the *average* Y sits exactly on the population line; individual u_i scatter points above/below.

LSA1 ⇒ moments, not ⇐: $\mathbb{E}[u \mid X] = 0 \Rightarrow \mathbb{E}[u] = 0$ and $\mathrm{Cov}(X, u) = 0$ (via LIE) – *not conversely*. Counterexample: $X \sim N(0,1)$, $u = X^2 - 1$: $\mathbb{E}[u] = 0$, $\mathrm{Cov}(X, u) = 0$, yet $\mathbb{E}[u \mid X] = X^2 - 1 \neq 0$ except at $X = \pm 1$.

Residual identities: $\sum_i \hat u_i = 0$ (FOC_{b_0}); $\sum_i \hat u_i X_i = 0$ (FOC_{b_1}); $\sum_i \hat u_i \hat Y_i = \hat\beta_0 \sum_i \hat u_i + \hat\beta_1 \sum_i \hat u_i X_i = 0$. Always true *by construction*, regardless of LSA1–3.

RCT ⇒ ATE: Binary $X_i \in \{0,1\}$ randomly assigned: $\beta_1 = \mathbb{E}[Y_i \mid X_i{=}1] - \mathbb{E}[Y_i \mid X_i{=}0]$ = average treatment effect.

Binary X **(general):** Any binary X: $\beta_0 = \mathbb{E}[Y \mid X{=}0]$, $\beta_1 = \mathbb{E}[Y \mid X{=}1] - \mathbb{E}[Y \mid X{=}0]$ – a *difference in means*, not a slope.

Selection/survivorship recipe: (1) name the unobserved factor sitting in u (ability, health, motivation); (2) explain why it drives *both* selection into the sample/into X *and* the outcome Y; (3) state the direction of bias. Hospital survey: sicker people select into hospital ⇒ $\mathrm{Cov}(\text{hospital}, u) \neq 0$. Wald bombers: only planes that *returned* are observed ⇒ reinforce the *least*-damaged areas, not the most.

RCT vs. observational: RCT: random X-assignment ⇒ LSA1 typically holds *by design*. Observational data (education, HSR lines): X not randomly handed out ⇒ LSA1 needs case-by-case justification.

Return-to-education example: wage = $\beta_0 + \beta_1 \text{educ} + u$, $u \ni$ innate ability: LSA1 needs $\mathbb{E}[\text{abil} \mid \text{educ}]$ constant across education levels – fails if higher-ability people self-select into more schooling.

LSA2/LSA3 violation examples: LSA2 fails: time series (firm revenue over 30 yrs, autocorrelated), non-random sampling (interviewing at city parks at 10am misses employed people). LSA3 fails: a handful of extreme outliers dominate the squared-error objective.

4. Sampling & Fit

SRS ⇒ i.i.d.: Simple random sampling: every population member equally likely drawn ⇒ independence across i; all drawn from the same population ⇒ identical distribution. Time series is the standard counterexample (autocorrelated).

TSS/ESS/SSR: $Y_i = \hat Y_i + \hat u_i$. $\text{TSS} = \sum_i (Y_i - \bar Y)^2$ (total); $\text{ESS} = \sum_i (\hat Y_i - \bar Y)^2$ (**explained**); $\text{SSR} = \sum_i \hat u_i^2$ (**residual**). TSS = ESS + SSR (cross term $2\sum_i \hat u_i \hat Y_i - 2\bar Y \sum_i \hat u_i = 0$). Never write "SSE".

R^2**:** $R^2 = \text{ESS/TSS} = 1 - \text{SSR/TSS} \in [0,1]$ since SSR ≥ 0; $R^2 = [\mathrm{Corr}(X,Y)]^2$. $\hat\beta_1 = 0 \Rightarrow \hat Y_i = \hat\beta_0 = \bar Y \Rightarrow \text{ESS} = 0 \Rightarrow R^2 = 0$.

$\mathrm{Corr}(X,Y)$ **algebra:** $\mathrm{Corr}(X,Y) = s_{XY}/(s_X s_Y)$; since $\hat\beta_1 = s_{XY}/s_X^2$, $\hat\beta_1 = \mathrm{Corr}(X,Y) \cdot (s_Y/s_X)$ – links the slope directly to the correlation coefficient and the two sample SDs.

Worked TSS/ESS/SSR: $n = 4$: $X = (1,2,3,4)$, $Y = (3,3,5,9)$, OLS fit $\hat Y = 2X$ (so $\hat\beta_0 = 0, \hat\beta_1 = 2$): $\bar Y = 5$, $\hat u = (1,-1,-1,1)$. TSS = 4+4+0+16 = 24; ESS = 9+1+1+9 = 20; SSR = 1+1+1+1 = 4; ESS + SSR = 24 = TSS; $R^2 = 20/24 = 0.833$.

Mean of $\hat Y$**:** $\frac{1}{n}\sum_i \hat Y_i = \bar Y$ (from $\sum_i \hat u_i = 0$, or directly from $\hat\beta_0 = \bar Y - \hat\beta_1 \bar X$).

$\bar Y$ **mean & variance:** Y_i i.i.d. (μ, σ^2): $\mathbb{E}[\bar Y] = \mu$ (needs only identical dist.); $\mathrm{Var}[\bar Y] = \sigma^2/n$ (needs independence: cross-covariances vanish). $\bar Y \overset{\text{approx}}{\sim} N(\mu, \sigma^2/n)$ for $n > 100$ (CLT).

$\mathrm{Var}(\bar Y)$ **derivation:** $\mathrm{Var}(\bar Y) = \mathrm{Var}\left(\frac{1}{n}\sum_i Y_i\right) = \frac{1}{n^2}\sum_i \mathrm{Var}(Y_i) + \frac{1}{n^2}\sum_{i \neq j} \mathrm{Cov}(Y_i, Y_j) = \frac{1}{n^2}(n\sigma^2) + 0 = \sigma^2/n$ (indep. kills the cross terms).

Sample variance s^2**:** $s^2 = \dfrac{1}{n-1}\sum_i (Y_i - \bar Y)^2$ estimates σ^2 ($n-1$ divisor corrects for using $\bar Y$ instead of μ); same idea underlies $\hat\sigma^2_{\hat\beta_1}$ in Sec. 6.

CAPM: $R - R_f = \alpha + \beta(R_m - R_f) + u$. Strict CAPM: $\alpha = 0$; $\hat\alpha > 0$ significant = "beats the market". $\beta < 1$: less systematic risk & lower expected excess return than market; $\beta > 1$: more risk & higher expected excess return.

Diff. of two indep. means: $\bar Y_1, \bar Y_2$ from independent samples, sizes N_1, N_2: $\mathbb{E}[\bar Y_1 - \bar Y_2] = \mu_1 - \mu_2$; $\mathrm{Cov}(\bar Y_1, \bar Y_2) = 0$; $\mathrm{Var}(\bar Y_1 - \bar Y_2) = \sigma_1^2/N_1 + \sigma_2^2/N_2$.

Hospital survey numbers: Hospital group: n = 7,774, mean health 3.21; non-hospital: n = 90,049, mean 3.93. Naive gap 0.72 is *not* the causal effect of hospitalisation – self-selection, not treatment, drives most of it.

Weight/height reference: $n = 100$: $\widehat{\text{Weight}} = -79.24 + 4.16\,\text{Height}$ (lb, in). At Height= 68: $\hat Y = 203.6$; a 2-inch growth spurt ⇒ $\Delta\hat Y = 8.32$ lb.

Low R^2 **is not a red flag:** TestScore on STR: $R^2 = 0.051$ – small, but says nothing about whether $\hat\beta_1$ is a valid *causal* estimate; that is an LSA1 question, not an R^2 question.

5. Unbiasedness, Efficiency & Dummies

Linear estimator of μ**:** $\hat\mu = \sum_i a_i Y_i$, Y_i i.i.d. (μ, σ^2): $\mathbb{E}[\hat\mu] = \mu \sum_i a_i$, so $\hat\mu$ unbiased $\iff \sum_i a_i = 1$.

Variance of $\hat\mu$**:** $\mathrm{Var}(\hat\mu) = \sigma^2 \sum_i a_i^2$ (independence kills cross terms). With $\sum_i a_i = 1$: $\sum_i a_i^2 = \sum_i (a_i - 1/n)^2 + 1/n \geq 1/n$, minimised at $a_i = 1/n\ \forall i$. $\bar Y$ **is BLUE:** Among all linear unbiased $\hat\mu = \sum_i a_i Y_i$: $\mathrm{Var}(\bar Y) = \sigma^2/n \leq \mathrm{Var}(\hat\mu)$ – Best Linear Unbiased Estimator of μ.

$\hat\beta_1$ **as a weighted sum:** $\hat\beta_1 = \sum_i w_i Y_i$, $w_i = \dfrac{X_i - \bar X}{\sum_j (X_j - \bar X)^2}$ ($\sum_i w_i = 0$, $\sum_i w_i X_i = 1$) – the OLS weights, used as the starting point of the Gauss–Markov proof.

OLS unbiasedness proof sketch: Substitute $Y_i = \beta_0 + \beta_1 X_i + u_i$ into $\hat\beta_1 = \sum_i w_i Y_i$: $\hat\beta_1 = \beta_1 \underbrace{\sum_i w_i X_i}_{=1} + \beta_0 \underbrace{\sum_i w_i}_{=0} + \sum_i w_i u_i = \beta_1 + \sum_i w_i u_i$.

Then $\mathbb{E}[\hat\beta_1 \mid X] = \beta_1 + \sum_i w_i \mathbb{E}[u_i \mid X_i] = \beta_1$ by LSA1.

$\mathrm{Var}(\hat\beta_1)$ **population formula:** Under homoskedasticity: $\mathrm{Var}(\hat\beta_1 \mid X) = \sigma_u^2 \sum_i w_i^2 = \dfrac{\sigma_u^2}{\sum_i (X_i - \bar X)^2}$ – matches Sec. 6: larger spread of X or smaller σ_u^2 ⇒ smaller $\mathrm{Var}(\hat\beta_1)$.

Gauss–Markov proof sketch: Any alternative linear unbiased $\tilde\beta_1 = \sum_i (w_i + c_i) Y_i$ needs $\sum_i c_i = 0, \sum_i c_i X_i = 0$ (for unbiasedness); then $\mathrm{Var}(\tilde\beta_1) = \mathrm{Var}(\hat\beta_1) + \sigma_u^2 \sum_i c_i^2 \geq \mathrm{Var}(\hat\beta_1)$, equality iff $c_i = 0\ \forall i$.

Gauss–Markov: Under LSA1–3 *plus* homoskedasticity $\mathrm{Var}(u_i \mid X_i) = \sigma_u^2$ (constant): $\hat\beta_0, \hat\beta_1$ are BLUE – $\mathrm{Var}(\hat\beta_0) \leq \mathrm{Var}(\tilde\beta_0)$, $\mathrm{Var}(\hat\beta_1) \leq \mathrm{Var}(\tilde\beta_1)$ for *any* linear unbiased $\tilde\beta_0, \tilde\beta_1$. Without homoskedasticity, OLS stays unbiased but need not be most efficient.

Binary regressor: $Y_i = \beta_0 + \beta_1 D_i + u_i$, $D_i \in \{0,1\}$: $\hat\beta_0 = \bar Y_{D=0}$, $\hat\beta_0 + \hat\beta_1 = \bar Y_{D=1}$, so $\hat\beta_1 = \bar Y_{D=1} - \bar Y_{D=0}$ (sample group-mean gap).

Common slope, diff. intercepts: $Y_i = \beta_0 + \delta_0 D_i + \beta_1 X_i + u_i$: $D_i{=}0$ line $\beta_0 + \beta_1 X_i$; $D_i{=}1$ line $(\beta_0 + \delta_0) + \beta_1 X_i$ – parallel lines, vertical gap δ_0.

Multi-dummy & base group: k categories ⇒ include $k-1$ dummies + intercept. Omitted category = **base group**, absorbed into β_0; each dummy's coefficient reads "vs. base group, holding other regressors fixed."

Dummy variable trap: All k category dummies *and* an intercept together ⇒ dummies sum to 1 for every i, an exact linear combo of the constant regressor ⇒ perfect multicollinearity, no unique OLS solution. Fix: drop one category.

Simulation intuition: $n = 10$ draws from $N(2,1)$, 20 sims: $Y_1 \in [-0.64, 4.27]$ vs. $\bar Y \in [1.16, 2.58]$ – $\bar Y$ closer to $\mu = 2$ in 16/20 sims, il-

lustrating $\text{Var}(Y_1) = \sigma^2 > \text{Var}(\bar{Y}) = \sigma^2/n$.
Law School model: $\text{salary}_i = \beta_0 + \beta_1 \text{top10}_i + \beta_2 \text{r11_25}_i + \beta_3 \text{r26_50}_i + \beta_4 \text{GPA}_i + u_i$; base group = ranked below 50th. Top-10 line: $(\beta_0 + \beta_1) + \beta_4 \text{GPA}$; each dummy coefficient = gap vs. base group, same GPA.
Hypothesis-test setup (L5): $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$; reject if $|\bar{Y} - \mu_0| > c$. Type I = reject true H_0; Type II = fail to reject false H_0; significance level = $P(\text{Type I} \mid H_0 \text{ true})$. Same logic sets up $H_0 : \beta_1 = 0$ (STR has no effect); made operational via $\text{SE}(\hat{\beta}_1)$ in Sec. 6.
Common-slope example: California regions: $D_i = 1$ if Region A, 0 if Region B, added alongside STR; δ_0 = region gap in test scores *not* explained by STR (a parallel-line shift).

6. Sampling Distribution of $\hat{\beta}_1$ (L6 pp.1–3)

$\hat{\beta}_1$ is a random variable: $\hat{\beta}_1$ is computed from a random sample, so it varies from sample to sample (a sampling distribution, not a fixed number); this is the object Sec. 6 studies, distinct from the within-sample spread of u_i.
Large-sample normality: Under LSA1–3, n large: $\hat{\beta}_1 \approx N(\beta_1, \sigma^2_{\hat{\beta}_1})$. Standardised version: $(\hat{\beta}_1 - \beta_1)/\sigma_{\hat{\beta}_1} \approx N(0,1)$.
$\sigma^2_{\hat{\beta}_1}$ & $\text{SE}(\hat{\beta}_1)$: A variance estimator $\hat{\sigma}^2_{\hat{\beta}_1}$ (built from $\hat{u}_i$ and $(X_i - \bar{X})^2$) exists – formula need not be recited. $\text{SE}(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}$ estimates the spread/dispersion of $\hat{\beta}_1$'s sampling distribution about β_1.
What makes $\text{SE}(\hat{\beta}_1)$ small: (qualitative) larger n; larger spread (variance) of X_i – more leverage to trace the slope; smaller error variance σ^2_u – points hug the line more tightly.
Standardised estimator: In large samples $\text{SE}(\hat{\beta}_1) \approx \sigma_{\hat{\beta}_1}$, so $\dfrac{\hat{\beta}_1 - \beta_1}{\text{SE}(\hat{\beta}_1)} \approx N(0,1)$ – same standardisation recipe as any $N(\mu, \sigma^2)$ variable.
Standardisation recipe: $Z = \dfrac{\text{value} - \text{mean}}{\text{sd}}$ turns any $N(\mu, \sigma^2)$ variable into $N(0,1)$. E.g. $\hat{\beta}_1 = 2.28$, $\beta_1 = 2$, $\sigma_{\hat{\beta}_1} = 0.5 \Rightarrow z = (2.28 - 2)/0.5 = 0.56$.
$P(\hat{\beta}_1 > c)$ **recipe:** Standardise: $P(\hat{\beta}_1 > c) = P\left(Z > \dfrac{c - \beta_1}{\sigma_{\hat{\beta}_1}}\right) = 1 - \Phi\left(\dfrac{c - \beta_1}{\sigma_{\hat{\beta}_1}}\right)$; read Φ off the table (use $\text{SE}(\hat{\beta}_1)$ in place of $\sigma_{\hat{\beta}_1}$ if only the estimate is known).
Symmetric-interval probability: $P(|\hat{\beta}_1 - \beta_1| < k\,\sigma_{\hat{\beta}_1}) = P(-k < Z < k) = 2\Phi(k) - 1$; e.g. $k = 1.96 \Rightarrow 0.95$, $k = 1 \Rightarrow 0.6827$. Read $\Phi(k)$ off the table.
Worked example: $\beta_1 = 2$, $\sigma_{\hat{\beta}_1} = 0.5$: $P(\hat{\beta}_1 > 2.98) = P\left(Z > \frac{2.98-2}{0.5}\right) = P(Z > 1.96) = 1 - \Phi(1.96) = 1 - .9750 = .0250$.
Big-picture recap: (1) model $Y_i = \beta_0 + \beta_1 X_i + u_i$ & target β_1 (Sec. 2); (2) estimator $\hat{\beta}_1$ (Sec. 2); (3) unbiasedness & efficiency under LSA1–3 (Sec. 5); (4) large-sample distribution $\hat{\beta}_1 \approx N(\beta_1, \sigma^2_{\hat{\beta}_1})$ and $\text{SE}(\hat{\beta}_1)$ (Sec. 6); (5) standardise & read $\Phi(z)$ for probability questions.

Reference Numbers (empirical examples)

Dataset (n)	Fitted line	SE	Used for
California TestScore/STR (420)	$698.9 - 2.28\text{STR}$	.52	fit, prediction
CEO salary/roe (209)	$963.2 + 18.5\,\text{roe}$	6.83	LSA1, residual, SE
Weight/Height (100)	$-79.2 + 4.16\text{Ht}$	–	prediction
Hospital (7,774/90,049)	survey means 3.21 vs. 3.93	–	group means, LSA1

Notation glossary: β: true param; b: candidate in minimization; $\hat{\beta}$: OLS estimate. u: true error; $\hat{u}$: residual. Y: observed; $\hat{Y}$: fitted. μ, σ^2: population mean/var; $\bar{Y}, s^2$: sample mean/var. w_i: OLS weight on Y_i; $\sigma^2_{\hat{\beta}_1}, \hat{\sigma}^2_{\hat{\beta}_1}$: (estimated) variance of $\hat{\beta}_1$'s sampling distribution; D_i: dummy/indicator regressor.

Standard Normal $\Phi(z)$ Table

z	.500	1.000	1.280	1.500	1.645
$\Phi(z)$	.6915	.8413	.8997	.9332	.9500
z	1.960	2.000	2.500	2.580	3.000
$\Phi(z)$	.9750	.9772	.9938	.9951	.9987

Using this table: $\Phi(z) = P(Z \leq z)$ for $Z \sim N(0,1)$; by symmetry $\Phi(-z) = 1 - \Phi(z)$. Plug $z = (\text{value} - \text{mean})/\text{sd}$ from the standardisation recipe, then read $\Phi(z)$ off here – not a test critical value, just a probability lookup.

Triggers & Traps

"causal effect": needs LSA1 $\mathbb{E}[u \mid X] = 0$; ask what is in u and whether it correlates with X.
"binary X": $\beta_1 = \mathbb{E}[Y \mid X{=}1] - \mathbb{E}[Y \mid X{=}0]$, a difference in means, not a slope.
"interpret a dummy coefficient": read it as "vs. the omitted/base group, holding other regressors fixed" (Sec. 5); never as a per-unit slope like a continuous X's coefficient.
"show unbiased": compute $\mathbb{E}[\sum_i a_i Y_i] = \mu \sum_i a_i$ and require $\sum_i a_i = 1$.
"$R^2 \leq 1$": chase the identity chain: $\text{TSS} = \sum_i [\hat{u}_i + (\hat{Y}_i - \bar{Y})]^2 \Rightarrow \text{TSS} = \text{ESS} + \text{SSR}$ (cross term = 0 via $\sum_i \hat{u}_i \hat{Y}_i = 0, \sum_i \hat{u}_i = 0$) $\Rightarrow \text{SSR} \geq 0 \Rightarrow R^2 \leq 1$.
"probability that the estimate exceeds/is below c": standardise with $\sigma_{\hat{\beta}_1}$ (or $\text{SE}(\hat{\beta}_1)$ if that's given): $z = (c - \beta_1)/\sigma_{\hat{\beta}_1}$, then use $\Phi(z)$ (upper tail $= 1 - \Phi(z)$).
"why is $\hat{\beta}_1$ random?": it depends on *which sample* is drawn – a different sample gives a different $\hat{\beta}_1$; that variability across hypothetical repeated samples is its sampling distribution (Sec. 6).
"E and Var of a difference of two independent means": $\mathbb{E}[\bar{Y}_1 - \bar{Y}_2] = \mu_1 - \mu_2$; independence $\Rightarrow \text{Cov}(\bar{Y}_1, \bar{Y}_2) = 0 \Rightarrow \text{Var}(\bar{Y}_1 - \bar{Y}_2) = \sigma^2_1/N_1 + \sigma^2_2/N_2$ (Sec. 4).
"H_0/H_1 for a mean" (L5): set up $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$; report Type I/Type II and the significance level $P(\text{Type I} \mid H_0)$ – this framework is in scope, the t-statistic machinery of L6 pp.4+ is not.
"efficient" / "smallest variance": compare only among *linear, unbiased* estimators (Gauss–Markov); never compare variance across a biased vs. unbiased pair to declare a winner.
"find $\text{Var}(\hat{\beta}_1)$" / "what does $\text{SE}(\hat{\beta}_1)$ estimate": population formula $\text{Var}(\hat{\beta}_1 \mid X) = \sigma^2_u / \sum_i (X_i - \bar{X})^2$ (homoskedastic case); $\text{SE}(\hat{\beta}_1)$ is its sample estimate – same small/large drivers as Sec. 6 (larger n or spread of X, smaller σ^2_u).
"$\bar{Y}$ is approximately normal" / CLT: Y_i i.i.d. $(\mu, \sigma^2) \Rightarrow \bar{Y} \overset{\text{approx}}{\sim} N(\mu, \sigma^2/n)$ for $n > 100$; the same CLT logic gives $\hat{\beta}_1 \approx N(\beta_1, \sigma^2_{\hat{\beta}_1})$ in Sec. 6.
"sample selects itself" / "survivorship": an LSA1 failure: name the omitted factor in u driving both selection and Y.
Trap: u_i vs. $\hat{u}_i$: only *residuals* $\hat{u}_i$ provably sum to zero (from the FOCs); the true error u_i need not.
Trap: $\mathbb{E}[u] = 0$ is not enough: LSA1 needs $\mathbb{E}[u \mid X] = 0$ at *every* X, not just the unconditional mean; $u = X^2 - 1$ is the standard counterexample.
Trap: k dummies + intercept: dummy variable trap – always use $k - 1$ dummies for k categories.
"5-step big picture" / "big picture framework": model + parameter $\rightarrow$ estimator $\hat{\beta}_1 \rightarrow$ its (large-sample) distribution $\rightarrow \text{SE}(\hat{\beta}_1) \rightarrow$ standardise & read $\Phi(z)$.
Trap: quoting $\hat{\beta}_0$ at $X = 0$: only a valid prediction within the observed range of X; extrapolating to $X = 0$ outside the sample range is unreliable.
"population line" vs. "OLS line": population regression line $\beta_0 + \beta_1 X$ is unknown/true; OLS line $\hat{\beta}_0 + \hat{\beta}_1 X$ is estimated from the sample – always keep the two distinct in an answer.
"sampling distribution" vs. "distribution of u": sampling distribution of $\hat{\beta}_1$ = variability across hypothetical repeated samples; distribution of u_i = spread *within* one sample. Different objects, both used in Sec. 4–6.
"holding other regressors fixed": multi-dummy or CAPM-style controls: read a coefficient as the gap/effect *at fixed values of the other regressors* (Sec. 5 Law School model, GPA held fixed across rank dummies).
Trap: dropping the sign: always keep the sign of $\hat{\beta}_1$ (and of any standardised z) in your final answer – the sign carries the economic/probability direction.