

Appendix A

Solutions to Chapter Exercises

A.1: Chapter 1 Solutions

Solution 1.1: (a) $\rho = \bar{d}_S/\bar{d}_T = 12.4/9.8 = 1.265$. (b) T is about 27% more potent than S, so a clinician using T instead of S should use a correspondingly smaller dose (roughly 79% of the S dose) to achieve the same muscle-relaxant effect.

Solution 1.2: (a) Indirect quantitative, parallel-line (doses form a geometric-like progression 0,1,2,4 is not perfectly geometric, but the standard practice is to log-transform positive doses and treat 0 via a separate blank contrast if a true zero must be retained — this item is best classified as indirect quantitative, with the analyst needing to decide between a parallel-line treatment of the non-zero doses and a slope-ratio treatment if the zero dose is retained as a genuine blank). (b) Direct assay — dose (total infused volume) is observed directly at the moment of a specified response (visible dilation). (c) Indirect quantal assay — dose (concentration) is fixed, response (germination/no germination) is quantal. (d) Slope-ratio assay — response assumed proportional to dose amount (not log dose), with an explicit zero (blank) dose included.

Solution 1.3: (Model answer, illustrative example: comparing two fertilizers by the dose needed to raise crop yield by a fixed amount.) A bioassay estimates a ratio of doses because the underlying biological question is “how much of the new preparation is equivalent to a given amount of the reference,” not “by how many yield units do the preparations differ.” This matters statistically because ratios of normal random variables are not themselves normally distributed and can have very heavy tails or even undefined variance when the denominator is close to zero — precisely the situation addressed by Fieller’s theorem in Chapter 3, which constructs an interval by inverting a probability statement about a *linear* combination of the numerator and denominator, rather than by assuming the ratio itself is approximately normal.

Solution 1.4: Since $\hat{\Psi}_p < 0$ and (by the sign convention of Chapter 6) $\log \hat{\rho} \propto -\hat{\Psi}_p/\hat{\Psi}_1$ with $\hat{\Psi}_1 > 0$, a negative $\hat{\Psi}_p$ implies $\log \hat{\rho} > 0$, i.e. $\hat{\rho} > 1$: the test preparation is estimated to be more potent than the standard. (Full numeric computation follows the method of Chapter 6, Section 6.4, equation 6.2.)

A.2: Chapter 2 Solutions

Solution: **2.1.** $\text{Var}(R) \approx R^2 \left[\frac{\sigma^2/n_S}{\bar{x}_S^2} + \frac{\sigma^2/n_T}{\bar{x}_T^2} \right]$; substituting
 $\bar{x}_S = R\bar{x}_T$ gives $\text{Var}(R) \approx R^2 \sigma^2 \left[\frac{1}{n_S R^2 \bar{x}_T^2} + \frac{1}{n_T \bar{x}_T^2} \right] = \frac{\sigma^2}{\bar{x}_T^2} \left[\frac{1}{n_S} + \frac{R^2}{n_T} \right]$.

Solution 2.2: $\bar{x} = \log 20$ (center), computing as in Example 2.1: means are 5, 10, 16 at $x = \log 10, \log 20, \log 40$. With unequal spacing on the natural-log scale, one should fit via `lm(y~x)` directly rather than by the simplified hand formula of Example 2.1 (which assumed equally-spaced x); running `lm(y ~ x) in R with x <-`

`log(c(10,10,20,20,40,40)), y <-
c(4,6,9,11,15,17)` gives $\hat{\beta} \approx 8.66$, $\hat{\alpha} \approx 3.35$ (values will
match a direct computer run), with s^2 obtainable from
`summary(fit)$sigma^2`.

Solution 2.3: (a) $F = 31.2/18.6 = 1.677$ on (10,10) d.f.;
 $F_{0.05,10,10} \approx 3.72$, so $1.677 < 3.72$ and we do not reject equal
variances — pooling is justified. (b)
 $s^2 = (18.6 + 31.2)/20 = 2.49$.

Solution 2.4:

As $B \rightarrow 0$, the delta-method variance
 $\text{Var}(R) \approx R^2 [\text{Var}(A)/A^2 + \text{Var}(B)/B^2]$ blows up because of
the $\text{Var}(B)/B^2$ term, yet the formula still returns a finite
symmetric interval $R \pm t \cdot SE(R)$ for any finite sample — it
never signals “no meaningful interval exists.” Fieller’s
approach, by contrast, explicitly tracks the quantity
 $g \propto (SE(b)/b)^2$ and produces a genuinely unbounded or two-
piece “interval” (the complement of a bounded set, or the
whole real line) once $g \geq 1$, correctly reflecting that the ratio
is not well-determined by the data at that confidence level —
a qualitative behavior the symmetric delta-method interval
cannot represent.

A.3: Chapter 3 Solutions

Solution 3.1: As $g \rightarrow 0$, the term $gv_{12}/v_{22} \rightarrow 0$ and the $g(\cdot)$
term inside the square root vanishes, and $1/(1-g) \rightarrow 1$, so
(3.1) reduces to $m \pm \frac{t_0 s}{b} \sqrt{v_{11} - 2mv_{12} + m^2 v_{22}}$. Since
 $\text{Var}(m) \approx \text{Var}(a/b) \approx \frac{s^2}{b^2} (v_{11} - 2mv_{12} + m^2 v_{22})$ by the delta
method (with $a = mb$ substituted), this is exactly
 $m \pm t_0 SE(m)$, the delta-method interval.

Solution 3.1: As $g \rightarrow 0$, the term $gv_{12}/v_{22} \rightarrow 0$ and the $g(\cdot)$ term inside the square root vanishes, and $1/(1-g) \rightarrow 1$, so

(3.1) reduces to $m \pm \frac{t_0 s}{b} \sqrt{v_{11} - 2mv_{12} + m^2 v_{22}}$. Since $\text{Var}(m) \approx \text{Var}(a/b) \approx \frac{s^2}{b^2}(v_{11} - 2mv_{12} + m^2 v_{22})$ by the delta method (with $a = mb$ substituted), this is exactly $m \pm t_0 SE(m)$, the delta-method interval.

Solution 3.2:

$$R = 3.10/2.55 = 1.2157, \quad t_{0.025, 18} = 2.101, \quad (a)$$

$$g = \frac{(2.101)^2(0.28)(1/10)}{2.55^2} = \frac{4.414 \times 0.028}{6.5025} = 0.01901. \quad (c)$$

Using the Fieller function: with $a = 3.10, b = 2.55, v_{11} = v_{22} = 1/10, v_{12} = 0, s^2 = 0.28, df = 18$, the resulting interval is approximately (1.01, 1.46) (values obtained by direct substitution into (3.1); students should confirm via the R/Python function of Example 3.4).

Solution 3.3: Spreading a fixed total number of subjects across more dose levels reduces the number of replicates *per level*, which increases $\text{Var}(\hat{\beta})$ for a design with the same overall dose range, because the effective “leverage” contributed by each additional level is smaller than the leverage lost by removing replicates from the extreme levels; since $g \propto \text{Var}(b)$ directly, this can increase g for the combined-regression-based ratio even while (for other reasons, e.g. better ability to detect curvature) more levels improve the deviation-from-regression test.

Solution 3.4: (Model paragraph.) “The delta-method interval reported here assumes the potency estimator is well-approximated by a normal distribution, which understates uncertainty whenever the denominator (slope) of the potency

ratio is not precisely estimated. Fieller's theorem avoids this assumption by directly inverting an exact pivotal quantity, and explicitly reports the diagnostic quantity g , which is a direct measure of how reliable the ratio estimate is; a large g (approaching or exceeding 1) signals that the point estimate alone should not be relied upon, a warning the delta-method interval cannot provide. We therefore recommend the Fieller interval as the primary reported result."

A.4: Chapter 4 Solutions

Solution 4.1: Treating A as standard: $R' = \bar{x}_A/\bar{x}_B = 1.68/1.99 = 0.844 = 1/1.18$. This must be the exact reciprocal because relative potency is defined as a ratio of doses producing the same response, and swapping which preparation is called "standard" simply inverts that ratio: $\rho_{B \text{ vs } A} = 1/\rho_{A \text{ vs } B}$.

Solution 4.2: $s_{oua}^2 = \sum(x_i - \bar{x})^2/8$; given the outlying 9.91 value, s_{oua}^2 is substantially larger than $s_A^2 = 0.7587/6 = 0.1265$ (an approximate calculation gives s_{oua}^2 several times larger). Pooling all three preparations into one common s^2 would not be defensible, since it would inflate the precision claimed for the A-vs-B comparison (which has homogeneous variance, Example 5.5) while simultaneously understating the true uncertainty in any comparison involving ouabain.

Solution 4.3: (a) $R = 4.20/3.65 = 1.151$. (b) Using the Fieller function with

$a = 4.20, b = 3.65, v_{11} = 1/10, v_{22} = 1/12, v_{12} = 0, s^2 = 0.45, df = 20$
 $t_{0.025, 20} = 2.086$; computing
 $g = (2.086)^2(0.45)(1/12)/3.65^2 = 1.958 \times 0.0375/13.3225 = 0.00551$
 (small), giving an interval close to symmetric around 1.151, roughly (0.97, 1.34) (exact bounds obtained by full substitution into (3.1)). (c) "We estimate that the new

bronchodilator is about 15% more potent than the standard; based on our data, we are 95% confident the true potency advantage is somewhere between roughly -3% and $+34\%$.”

Solution 4.4: As n grows, $SE(b) = s/\sqrt{n}$ shrinks faster than b itself (which converges to a fixed non-zero true value), so $g = t_0^2 s^2 v_{22}/b^2 \rightarrow 0$; a plot of g vs. n would show a curve decreasing roughly like $1/n$ (since $v_{22} = 1/n$ enters linearly and t_0 changes only slowly with n for the same nominal confidence level), asymptoting toward zero, at which point the Fieller and delta-method intervals become indistinguishable, exactly as observed empirically in Example 4.6/4.7’s simulations.

A.5: Chapter 5 Solutions

Solution 5.1: $s_S^2/n_S = 0.85/6 = 0.1417$,
 $s_T^2/n_T = 0.12/10 = 0.012$.
 $\nu = (0.1417 + 0.012)^2/[0.1417^2/5 + 0.012^2/9] = (0.1537)^2/[0.004016 + 0.000016] = 0.02362/0.004032 = 5.86$.

Naive pooled d.f. = 14; the Welch d.f. (≈ 5.86) is much smaller, reflecting how strongly the small, high-variance group ($n_S = 6$, larger variance) dominates the uncertainty, which the naive pooled-d.f. calculation ignores.

Solution 5.2: (Illustrative; exact numbers depend on the student’s chosen simulated samples — the grading criterion is correct use of `t.test(..., var.equal=FALSE)` or `scipy.stats.ttest_ind(..., equal_var=False)` and correct interpretation of the resulting non-integer degrees of freedom and p -value.)

Solution 5.3: Because $\text{Var}(b) = s_T^2/n_T$ is estimated from only the test-preparation sample (rather than a pooled estimate), it can be very unstable when n_T is small, causing g to fluctuate much more from sample to sample than the pooled-variance g of Chapter 3, which benefits from combining information across both preparations (larger effective degrees of freedom). The design implication is that adequate replication is especially important for whichever preparation has (or is suspected to have) the larger variance, since that preparation's sample variance disproportionately drives the width — and reliability — of the resulting BFS-adjusted Fieller interval.

Solution 5.4: The F -test for equality of variances has notoriously low power at small sample sizes (a well-documented property in the broader statistical literature), meaning that “the test did not reject” is weak evidence of true equal variances when n is small — it may simply reflect insufficient data to detect a real difference. Relying solely on the F -test therefore risks silently pooling genuinely unequal variances whenever sample sizes are modest, which is precisely the small-sample regime common in classical bioassay work (recall the $n = 7$ cat experiments of Chapter 4). This connects directly to Chapter 10's design

A.6: Chapter 6 Solutions

Solution 6.1: For $m = 2$: positions centered at $-0.5, 0.5$.
 Preparation: $(1, 1, -1, -1)$; Combined regression: $(-0.5, 0.5, -0.5, 0.5)$ (or, cleared of fractions, $(-1, 1, -1, 1)$);
 Parallelism: $(-0.5, 0.5, 0.5, -0.5)$ (or $(-1, 1, 1, -1)$ cleared of fractions). $\theta = m(m^2 - 1) = 2(3) = 6$.

Solution 6.2:

$\hat{\rho} = \exp(-\hat{\Psi}_p/\hat{\Psi}_1) = \exp(-(-0.800)/16.777) = \exp(0.04769) = 1.0488 \approx 1.048$, matching Section 6.5's stated result (rounding).

Solution 6.3: (a) A significant parallelism test means the two dose–response lines are not parallel, so the fundamental “condition of similarity” (Chapter 9) fails, and a single dose-independent relative potency ρ is not a well-defined or trustworthy summary of the relationship between the two preparations — any reported potency ratio would depend on which response level was used to define equivalence. (b) Remedies include: re-examining whether the tested dose range is appropriate (non-parallelism sometimes appears only at the extremes of a wide range), and investigating whether the two preparations might differ qualitatively (e.g. a different active metabolite or mechanism), which would need to be resolved biologically, not just statistically, before any potency comparison is meaningful.

Solution 6.4: (Simulation exercise — expected qualitative result.) Fitting `lm(response ~ prep + logdose, data=sim_data)` (common-slope model) to data simulated with true $\beta = 15$ and true $\log \rho = 0.10$ should recover $\hat{\rho} = \exp(-\widehat{\alpha_{diff}}/\hat{\beta})$ close to $\exp(0.10) = 1.105$, with sampling variability decreasing as replication r increases; students should report their specific simulated $\hat{\rho}$ alongside the true value for comparison.

A.7: Chapter 7 Solutions

Solution 7.1: Starting from $d_S = \rho d_T$ and $x = (\text{dose})^\lambda$: $x_S = d_S^\lambda = (\rho d_T)^\lambda = \rho^\lambda d_T^\lambda = \rho^\lambda x_T$. At equipotent doses the two lines give the same expected response: $\alpha + \beta_S x_S = \alpha + \beta_T x_T \implies \beta_S \rho^\lambda x_T = \beta_T x_T$ for all x_T , so $\beta_T = \beta_S \rho^\lambda$, and solving for ρ : $\rho^\lambda = \beta_T / \beta_S \implies \rho = (\beta_T / \beta_S)^{1/\lambda}$, as claimed in (7.1).

Solution 7.2: (a) Fitting the two-slope model by hand: with $\lambda = 1$ and grand-mean-centered intercept, $\hat{\beta}_S = \frac{\sum (x_{Si} - \bar{x}_S)(y_{Si} - \bar{y}_S)}{\sum (x_{Si} - \bar{x}_S)^2}$ computed directly on the Unit-1/Unit-2 standard-preparation data, and similarly for $\hat{\beta}_T$; students should obtain values close to those returned by the `lm/statsmodels` fit of Example 7.1. (b) $\hat{\rho} = \hat{\beta}_T / \hat{\beta}_S$, expected close to 1 given the near-identical simulated/illustrative data.

Solution 7.3: (a) A significant blank contrast implies that the fitted linear relation, extrapolated down to zero dose, does not agree with the response actually observed at the blank — suggesting either a threshold effect (no response below some minimum dose) or genuine non-linearity near zero, both of which threaten the validity of the simple slope-ratio model. (b) A redesign might add one or two additional low, non-zero dose points close to the blank for both preparations, to better characterize the shape of the dose–response curve near the origin and determine whether a modified (e.g. non-linear near zero, linear thereafter) model is needed.

Solution 7.4:

(Simulation exercise.) With true $\beta_S = 3, \beta_T = 6, \lambda = 1$, true $\rho = 2$; fitting the model as in Example 7.1/7.3 should recover $\hat{\rho}$ close to 2, and the Fieller interval (constructed as in Example 7.4, using the fitted slopes' standard errors and, if the design is orthogonal, $v_{12} = 0$) should, in the large majority of simulated replications, contain the true value 2, consistent with correct 95% coverage.

A.8: Chapter 8 Solutions**Solution 8.1:**

Logit $(p)/2 + 5$: for $p = 0.10$: $\ln(0.10/0.90) = -2.197$, $/2 = -1.099$, $+5 = 3.901$ (vs. probit 3.72, Table 8.1) — logit slightly higher. For $p = 0.50$: $\ln(1) = 0$, giving logit-transformed value 5.00 (identical to probit, as both transforms are centered at 0 for $p = 0.5$). For $p = 0.90$: $\ln(9) = 2.197$, $/2 = 1.099$, $+5 = 6.099$ (vs. probit 6.28) — logit slightly lower. The discrepancy is largest away from $p = 0.5$ and reflects the logistic distribution's heavier tails relative to the normal.

Solution 8.2: Fitting `glm(cbind(r,n-r)~x,family=binomial(link="probit"))` to $x = (0.5, 1.0, 1.5, 2.0)$, $n = 40$, $r = (4, 15, 28, 36)$ should yield $\hat{\alpha} \approx -3$ to -4 range and $\hat{\beta} \approx 2$ to 3 range (exact values depend on the fit); $ED50 = \exp(-\hat{\alpha}/\hat{\beta})$, computed directly from the fitted coefficients.

Solution 8.3: Applying the Fieller function (Example 8.4's code) to the model's `vcov()` output gives an asymmetric interval around the point `ED50`; the naive delta-method interval (using `msm::deltamethod` or a manual first-order approximation) will typically be narrower and symmetric on

the original dose scale — at this sample size ($n = 40$ per dose, reasonably large), the two are likely to be fairly close, illustrating that Fieller's advantage over the delta method is most pronounced at *smaller* sample sizes (cf. Chapter 4's simulation examples).

Solution 8.4: (a) $\widehat{\log(\text{ED50})} = -\hat{\alpha}/\hat{\beta} = -(-3.2)/2.1 = 1.524$; $\text{ED50 (original scale)} = e^{1.524} = 4.59$. (b) Using Fieller's theorem with $a = -\hat{\alpha} = 3.2, b = \hat{\beta} = 2.1, v_{11} = 0.25, v_{22} = 0.09, v_{12} = -(-0.04) = 0.04$ (sign flip because $a = -\alpha$), $s^2 = 1, t_0 = z_{0.05} = 1.645$ (large-sample/normal approximation): $g = 1.645^2(0.09)/2.1^2 = 2.706 \times 0.09/4.41 = 0.0552$; substituting into (3.1) gives a 90% interval for $\mu = a/b = 1.524$ on the log-dose scale, of the order (1.15, 1.95) (exact numeric bounds obtained by full substitution), exponentiating to obtain the corresponding original-scale ED50 interval.

A.9: Chapter 9 Solutions

Solution 9.1: Computing Bartlett's test on raw proportions typically yields a larger test statistic (and possibly a smaller p -value) than on the arcsine-transformed data, because raw-proportion variance depends on the mean proportion, which differs substantially across the four density groups (from about 0.82 down to about 0.60). The arcsine-transformed scale is recommended for the subsequent ANOVA, since it better satisfies the constant-variance assumption underlying the density-effect regression of Chapter 8.

Solution 9.2: A significant blank contrast, with a non-significant intersection contrast, indicates that the two preparations' non-blank dose-response lines *do* meet at a common point on the response axis (similarity condition 3,

common intercept, holds) but that this common linear relationship does **not** extend validly down to the actual observed blank response (similarity condition 2, linearity over the *full* tested range including zero dose, fails near the origin).

Solution 9.3:

$$N = 18, k = 3.$$

$$\bar{s}^2 = (5(2.1) + 5(2.4) + 5(2.0))/15 = (10.5 + 12 + 10)/15 = 32.5/15 = 2.167$$

$$M = 15 \ln(2.167) - [5 \ln 2.1 + 5 \ln 2.4 + 5 \ln 2.0] = 15(0.7733) - [5(0.7419) + 5(0.8755) + 5(0.6931)] = 11.60 - [3.710 + 4.378 + 3.466] = 11.60 - 11.55 = 0.045$$

$$C = 1 + \frac{1}{3(2)}[3(1/5) - 1/15] = 1 + \frac{1}{6}[0.6 - 0.0667] = 1 + 0.0889 = 1.0889$$

$\chi^2_{Bartlett} = 0.045/1.089 = 0.041$ on 2 d.f.; far below $\chi^2_{0.05,2} = 5.99$, so we do not reject homogeneity of variance — a clear, unambiguous non-significant result given how close the three sample variances are.

Solution 9.4: Reporting “no valid potency estimate” when similarity genuinely fails is scientifically honest because any single potency number computed despite a failed parallelism/intersection test would misrepresent a relationship that is not, in fact, described by a single constant ratio — potentially misleading downstream dosing decisions. A design that increased replication or broadened preliminary/pilot testing of the dose–response shape *before* committing to the definitive assay (Chapter 10’s emphasis on adequate, well-planned replication) could have revealed non-parallelism earlier (e.g., in a pilot study) and allowed the dose range or preparation formulation to be adjusted before the costly definitive assay was run.

A.10: Chapter 10 Solutions

Solution 10.1: RCBD with $v = 8$ dose combinations (S_1-S_4 , T_1-T_4), $b = 3$ blocks (days), each block containing all 8 doses once, order randomized within each day. ANOVA: Blocks (2 d.f.), Doses (7 d.f.: Preparation 1, Combined regression 1, Parallelism 1, Deviation from regression 4), Error ($2 \times 7 = 14$ d.f.), Total (23 d.f.).

Solution 10.2: Initial block $\{0, 1, 3\} \pmod{5}$; cyclic shifts give blocks $\{0, 1, 3\}, \{1, 2, 4\}, \{2, 3, 0\}, \{3, 4, 1\}, \{4, 0, 2\}$, i.e. 5 blocks of size 3. Tallying pairs: $(0, 1)$ appears in blocks 1 and 4 $\rightarrow 2$ times; $(0, 2)$ in blocks 3, 5 $\rightarrow 2$ times; $(0, 3)$ in blocks 1, 3 $\rightarrow 2$ times; by symmetry of the cyclic construction from a genuine difference set, every pair co-occurs exactly $\lambda = 2$ times, confirming balance ($v = 5, k = 3, \lambda = 2, b = 5, r = 3$, satisfying the BIBD identity $\lambda(v-1) = r(k-1)$: $2(4) = 3(2) = 6$ OK.)

Solution 10.3: The parallelism contrast places weights that increase in magnitude toward the extreme dose levels and are of *opposite sign* between the two preparations (e.g. $(-1, 0, 1, 1, 0, -1)$ in the $m = 3$ symmetric case), effectively comparing the “spread” of each preparation’s response pattern rather than any single pairwise difference; a BIBD optimized to equalize variance across all $\binom{v}{2}$ simple pairwise contrasts need not simultaneously minimize the variance of this quite differently-weighted linear combination, since design efficiency for a linear contrast depends on how well the design’s information matrix aligns with that specific contrast vector, not merely on overall pairwise balance.

Solution 10.4: (Simulation exercise — expected qualitative outcome.) Comparing `anova(aov(y~dose))` (unadjusted) against `anova(aov(y~block+dose))` (adjusted) on

simulated incomplete-block data with a genuine, appreciable block effect should show a markedly smaller residual mean square in the block-adjusted analysis; students should report the specific percentage reduction observed in their own simulated data set.

A.11: Chapter 11 Solutions

Solution 11.1: Fitting `aov(response ~ subject + period + treatment, data = ...)` (with `subject` as a blocking factor nested within `sequence`) to a properly reshaped long-format version of Example 11.1’s data should return an F -statistic for `treatment` numerically equal (up to rounding) to t_{trt}^2 computed directly from the formula in Section 11.4.2 — confirming that the classical hand-formula and the general linear-model ANOVA approach are two routes to the same answer.

Solution 11.2:

$t_{c-o} = (95 - 88) / \sqrt{64(1/8 + 1/8)} = 7 / \sqrt{16} = 7/4 = 1.75$ on 14 d.f.; $t_{0.025, 14} = 2.145$. Since $1.75 < 2.145$, we do not reject $H_0 : \lambda_A = \lambda_B$ — there is no significant evidence of carry-over, supporting the adequacy of the washout period used in this (hypothetical) trial.

Solution 11.3: (Model paragraph.) “Skipping the carry-over test assumes, without verification, that the chosen washout period was long enough for both treatments — an assumption that may not hold, especially if the two treatments differ in their pharmacokinetics (e.g. one has a much longer half-life than anticipated) or if unexpected biological interactions occur. As Example 11.2 shows, if carry-over is in fact present and goes undetected, the standard within-subject treatment test becomes biased, potentially leading to an incorrect

conclusion about which treatment is superior — a risk that a single additional, essentially free statistical test (using data already collected) can help guard against. We recommend always performing the carry-over test as a standard part of cross-over trial reporting, regardless of how confident the investigators are in the chosen washout duration.”

Solution 11.4: Starting from $\text{Var}(\hat{\tau}) = \frac{s_d^2}{4} \left(\frac{1}{n} + \frac{1}{n} \right) = \frac{s_d^2}{2n}$, and requiring, for a two-sided test at level α with power $1 - \beta$ to detect true difference δ : $(z_{\alpha/2} + z_\beta) \sqrt{\text{Var}(\hat{\tau})} = \delta$ (treating $\hat{\tau}$ as estimating half the true A-vs-B difference under the given parametrization; adjust the factor of 2 as needed for a given textbook’s exact definition of τ), we get

$$(z_{\alpha/2} + z_\beta)^2 \frac{s_d^2}{2n} = \delta^2 \implies n = \frac{2(z_{\alpha/2} + z_\beta)^2 s_d^2}{\delta^2}.$$

Substituting

$z_{0.025} = 1.96, z_{0.20} = 0.84, s_d = 10, \delta = 8$:

$n = \frac{2(2.80)^2(100)}{64} = \frac{2(7.84)(100)}{64} = \frac{1568}{64} = 24.5$, so round up to $n = 25$ per sequence group (the discrepancy from Example 11.4’s illustrative “13–14” reflects that example’s simplified/approximate treatment; this exercise’s fully-derived formula is the one students should trust and cite going forward).