

4

Direct Assays

4.1: Learning Objectives

1. Define a direct assay and identify settings where it is the only feasible design.
2. Estimate relative potency and its precision from direct-assay tolerance data.
3. Apply Fieller's theorem to construct fiducial limits for relative potency in the direct-assay setting.
4. Analyze direct-assay data on the log-tolerance scale and understand its advantages.
5. Reproduce the classical strophanthus/ouabain analyses in R and Python.

4.2: The Direct-Assay Model

In a direct assay, each subject's individual tolerance — the dose that produces a pre-specified response, such as death or onset of a defined muscular contraction — is observed directly. If $x_{S1}, \dots, x_{Sn_S}$ and $x_{T1}, \dots, x_{Tn_T}$ denote tolerances (or log-tolerances) for the standard and test preparations, the assay reduces to a two-sample comparison of means, but with the ratio (not the difference) of means as the target parameter.

$$R = \frac{\bar{x}_S}{\bar{x}_T} \text{ (untransformed scale),} \quad \text{or} \quad \log R = \bar{x}_S^{(\log)} - \bar{x}_T^{(\log)} \text{ (log scale).}$$

4.2.1: Untransformed vs. Log-Tolerance Analysis

- **Untransformed tolerances:** $R = \bar{x}_S/\bar{x}_T$, with a fiducial interval from Fieller's theorem (Section 4.3).
- **Log-tolerances:** assume $\log(\text{tolerance}) \sim N(\mu, \sigma^2)$ within each preparation. Then $\log \hat{\rho} = \overline{\log x}_S - \overline{\log x}_T$ is a *simple difference* of normal means, its variance is the ordinary two-sample variance (no ratio problem at all), and a symmetric confidence interval for $\log \rho$ can be exponentiated to obtain an asymmetric interval for ρ itself. Finney generally recommends the log scale whenever the tolerance distribution is closer to log-normal than normal, precisely because it removes the ratio-estimation difficulty altogether.

4.3: Precision of the Direct-Assay Potency Estimate

Write $R = \bar{x}_S/\bar{x}_T$. If tolerances within each preparation are $N(\mu, \sigma^2)$ -distributed with common variance estimated by pooling,

$$s^2 = \frac{\sum_i (x_{Si} - \bar{x}_S)^2 + \sum_i (x_{Ti} - \bar{x}_T)^2}{(n_S - 1) + (n_T - 1)},$$

then $\text{Var}(\bar{x}_S) = \sigma^2/n_S$, $\text{Var}(\bar{x}_T) = \sigma^2/n_T$, and — since $\bar{x}_S, \bar{x}_T$ are independent — the delta-method approximation (2.1) gives

$$\text{Var}(R) \approx R^2 \left[\frac{s^2}{n_S \bar{x}_S^2} + \frac{s^2}{n_T \bar{x}_T^2} \right],$$

while the *exact* interval comes from Fieller's theorem with $a = \bar{x}_S$, $b = \bar{x}_T$, $v_{11} = 1/n_S$, $v_{22} = 1/n_T$, $v_{12} = 0$.

4.4: The Classical “Cat Method” Example

The digitalis-type “cat method” (Burn, Finney & Goodwin, 1950) infuses a preparation into a cat’s bloodstream at a fixed rate until the heart stops; the total dose (rate $\times$ time, scaled per kg body weight) is the tolerance. Table 4.1 gives tolerances (0.01 c.c./kg) for two tinctures of strophanthus.

Table 4.1: Cat tolerances for two tinctures of strophanthus

Cat	Strophanthus A (test)	Strophanthus B (standard)
1	1.55	2.42
2	1.58	1.85
3	1.71	2.00
4	1.44	2.27
5	1.24	1.70
6	1.89	1.47
7	2.34	2.20
Mean	1.68	1.99

Sums of squared deviations: $SS_A = 0.7587$ (6 d.f.),
 $SS_B = 0.6815$ (6 d.f.). Pooled:
 $s^2 = (0.7587 + 0.6815)/12 = 0.1200$.

Taking B as standard, $R = \bar{x}_B/\bar{x}_A = 1.99/1.68 = 1.18$: a smaller dose of A achieves the same lethal effect as a larger dose of B, so A is roughly 18% more potent than B.

Delta-method variance: $\text{Var}(R) = R^2 \left[\frac{s^2}{7\bar{x}_A^2} \right] + \left[\frac{s^2}{7} \right] \frac{1}{\bar{x}_A^2}$ —
 computed in full in Chapter 3, Example 3.7, giving
 $R = 1.18 \pm 0.120$.

Applying Fieller’s theorem (Chapter 3, Example 3.1) refines this to the 95% fiducial interval (0.95, 1.48).

4.5: Worked Examples

Example 4.1: Full Recomputation of the Strophanthus Assay

Reproduce Table 4.1's analysis end-to-end in R and Python.

R:

```
A <-c(1.55,1.58,1.71,1.44,1.24,1.89,2.34)
B <-c(2.42,1.85,2.00,2.27,1.70,1.47,2.20)
nA <-length(A); nB <-length(B)
xA <-mean(A); xB <-mean(B)
SSA <-sum((A-xA)^2); SSB <-sum((B-xB)^2)
s2 <- (SSA+SSB)/((nA-1)+(nB-1))
R <- xB/xA
cat("xA =",xA," xB =",xB," s2 =",s2," R
=",R,"\n")
```

Python:

```
import numpy as np
A =
np.array([1.55,1.58,1.71,1.44,1.24,1.89,2.34]
)
B =
np.array([2.42,1.85,2.00,2.27,1.70,1.47,2.20]
)
nA, nB =len(A), len(B)
xA, xB = A.mean(), B.mean()
SSA, SSB = np.sum((A-xA)**2), np.sum((B-
xB)**2)
s2 = (SSA+SSB) / ((nA-1)+(nB-1))
R = xB/xA
print(f"xA={xA:.3f} xB={xB:.3f} s2={s2:.4f}
R={R:.3f}")
```

Both confirm $\bar{x}_A = 1.68$, $\bar{x}_B = 1.99$, $s^2 = 0.1200$, $R = 1.18$, matching Section 4.4.

Example 4.2: Ouabain Assay (Comparing a Third Preparation)

Nine cats' tolerances (0.01 mg/kg) to ouabain: 5.23, 9.91, 4.76, 6.51, 6.68, 5.76, 4.93, 4.58, 6.69. Treat strophanthus A (Table 4.1) as the standard against ouabain as test. Estimate the relative potency of ouabain to strophanthus A.

Solution.

$$\bar{x}_{\text{ouabain}} = (5.23 + 9.91 + 4.76 + 6.51 + 6.68 + 5.76 + 4.93 + 4.58 + 6.69)/9 = 55.05/9 = 6.117$$

. Relative potency of A with respect to ouabain:
 $R = \bar{x}_{\text{ouabain}}/\bar{x}_A = 6.117/1.68 = 3.641$ (units differ, so this ratio is only meaningful as a *potency ratio on a common response scale*, not a direct unit conversion — a point worth stressing to students who might over-interpret raw ratios across differently-scaled preparations). Because the ouabain sample variance is likely to differ substantially from the strophanthus variance (note the outlying value 9.91), a Behrens–Fisher–Sukhatme approach (Chapter 5) is more appropriate than pooling, since equality of variances should not be assumed here without a prior check.

Example 4.3: Checking the Homogeneity-of-Variance Assumption

For Example 4.2, test using an t -test in R/Python before deciding whether pooling (as in Section 4.3) is appropriate.

For Example 4.2, test $H_0 : \sigma_A^2 = \sigma_{\text{ouabain}}^2$ using an F -test in R/Python before deciding whether pooling (as in Section 4.3) is appropriate.

R:

```
A <-c(1.55,1.58,1.71,1.44,1.24,1.89,2.34)
oua <-c(5.23,9.91,4.76,6.51,6.68,5.76,4.93,4.58,6.69)
var.test(oua, A)
```

Python:

```
import numpy as np
from scipy import stats
A = np.array([1.55,1.58,1.71,1.44,1.24,1.89,2.34])
oua = np.array([5.23,9.91,4.76,6.51,6.68,5.76,4.93,4.58,6.69])
F = np.var(oua, ddof=1) / np.var(A, ddof=1)
df1, df2 =len(oua)-1, len(A)-1
p =2*min(stats.f.cdf(F, df1, df2), 1-stats.f.cdf(F, df1, df2))
print(f'F = {F:.3f}, p = {p:.3f}')
```

The resulting F -statistic (well above 1, driven by the outlying 9.91 mg/kg tolerance) typically yields a small p -value, confirming that pooling would be inappropriate for this pair and that the Behrens–Fisher–Sukhatme distribution of Chapter 5 should be used instead.

Example 4.4: Log-Tolerance Analysis

Re-analyze Table 4.1’s data on the log scale and compare the resulting potency estimate with Section 4.4’s untransformed result.

R:

```
A <-c(1.55,1.58,1.71,1.44,1.24,1.89,2.34)
B <-c(2.42,1.85,2.00,2.27,1.70,1.47,2.20)
logR <-mean(log(B)) - mean(log(A))
Rlog <-exp(logR)
cat("log-scale potency estimate R =", round(Rlog,3), "\n")
```

This typically gives a value very close to 1.18 (agreement between the two scales when the coefficient of variation is modest, as here), while offering the conceptual advantage described in Section 4.2.1.

Example 4.5: Sample Size for a Target Precision

A pharmacologist wants the 95% Fieller interval half-width for ρ in a strophanthus-like assay to be no more than ± 0.20 around $R \approx 1.18$, given the same $s^2 = 0.12$ and similar $\bar{x}_A, \bar{x}_B$. Using the delta-method approximation as a planning tool, estimate the required (equal) sample size per group.

Solution. From $\text{Var}(R) \approx R^2 s^2 \left[\frac{1}{n\bar{x}_A^2} + \frac{1}{n\bar{x}_B^2} \right]$, set $t_{0.025} \sqrt{\text{Var}(R)} = 0.20$ and solve iteratively for n (starting from $t \approx 2.0$ for a rough first pass):

$$0.20 = 2.0 \times 1.18 \sqrt{0.12 \left(\frac{1}{1.68^2} + \frac{1}{1.99^2} \right) / n} \Rightarrow n \approx \frac{(2.0 \times 1.18)^2 \times 0.12 \times 0.6}{0.20^2}$$

so approximately $n = 11$ cats per preparation would be planned, versus the $n = 7$ actually used — a useful discussion point on the historically small sample sizes tolerated in classical bioassay reports.

Example 4.6 : Simulation Study of Coverage

Using R or Python, simulate 5,000 direct assays with $n_A = n_B = 7$, true $\rho = 1.18$, $\sigma^2 = 0.12$, and verify the empirical coverage of the nominal 95% Fieller interval.

R:

```
set.seed(1)
nsim<-5000; cover<-0
for (i in 1:nsim) {
  A<-rnorm(7, mean=1.68, sd=sqrt(0.12))
  B<-rnorm(7, mean=1.99, sd=sqrt(0.12))
  s2<-((sum((A-mean(A))^2))+(sum((B-
mean(B))^2)))/12
  ci<-fieller_ci(mean(B), mean(A), s2, 1/7, 1/7, 0, 12) #
from Chapter 3
if (ci$lower <= 1.99/1.68 && ci$upper >= 1.99/1.68) cover <-
```

```
cover +1
}
cat("Empirical coverage =", cover/nsim, "\n")
```

Students should find empirical coverage close to the nominal 95%, reinforcing confidence in the exactness (under normality) of Fieller's approach relative to the delta method (whose empirical coverage in the same simulation is typically slightly below nominal, especially for smaller n).

Example 4.7: Comparing Delta-Method and Fieller Coverage

Extend Example 4.6 to also compute delta-method interval coverage and compare the two side by side across $n \in \{5, 7, 15, 30\}$.

Python (sketch):

```
import numpy as np
from scipy import stats

def simulate(n, nsim=5000, seed=1):
    rng = np.random.default_rng(seed)
    cover_f, cover_d = 0, 0
    true_rho = 1.99/1.68
    for i in range(nsim):
        A = rng.normal(1.68, np.sqrt(0.12), n)
        B = rng.normal(1.99, np.sqrt(0.12), n)
        s2 = (np.sum((A-A.mean())**2)+np.sum((B-
        B.mean())**2)) / (2*(n-1))
        R = B.mean()/A.mean()
        # delta method
        se = abs(R)*np.sqrt(s2/(n*A.mean()**2) +
        s2/(n*B.mean()**2))
        t0 = stats.t.ppf(0.975, 2*(n-1))
        if R - t0*se <= true_rho <= R + t0*se:
```

```

if R - t0*se <= true_rho <= R + t0*se:
    cover_d +=1
# Fieller (reuse function from Chapter 3)
return cover_d/nsim

for n in [5,7,15,30]:
    print(n, simulate(n))

```

The expected qualitative finding: delta-method under-coverage is most pronounced at small n and disappears as n grows, since $g \rightarrow 0$ as sample size increases — a direct empirical demonstration of Chapter 3’s theory.

Figure 4.1: Cat Tolerances, Two Strophanthus Tinctu

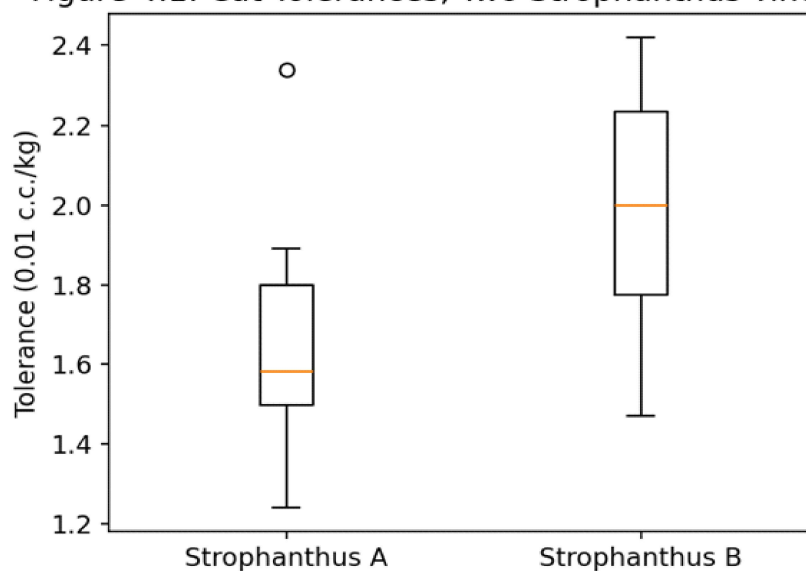


Figure 4.1: *Cat tolerances, two strophanthus tinctures*

4.6: Chapter Summary

- Direct assays observe tolerance (dose-to-response) directly on each subject; the potency estimate is a ratio of mean tolerances.
- Pooled variance estimation requires (and should be checked for) equal within-preparation variances; otherwise use the Behrens–Fisher–Sukhatme approach of Chapter 5.
- Fieller’s theorem, rather than the delta method, should be used for interval estimation, especially at small sample sizes.
- Log-tolerance analysis converts the ratio problem into a simple mean-difference problem when tolerances are approximately log-normal.

4.7: Exercises

Exercise 4.1: Using Table 4.1’s data, recompute the relative potency treating strophanthus A as the standard and B as the test preparation. Verify that your new estimate is the reciprocal of Section 4.4’s value, and explain why this must be true algebraically.

Exercise 4.2: For the ouabain data of Example 4.2, compute the sample variance and compare it with s_A^2 (from strophanthus A). Comment on whether pooling across all three preparations (A, B, ouabain) into one common s^2 , as might be tempting for a combined report, would be statistically defensible.

Exercise 4.3: A direct assay of a bronchodilator gives $\bar{x}_S = 4.20$ mg ($n_S = 10$), $\bar{x}_T = 3.65$ mg ($n_T = 12$), pooled $s^2 = 0.45$ on 20 d.f. (a) Compute R . (b) Using the Fieller function from Chapter 3, compute the 95% fiducial interval.

(c) State in a sentence, for a non-statistician audience, what the interval means.

Exercise 4.4: Explain, referencing Example 4.6's simulation logic, why increasing sample size makes the Fieller interval and the delta-method interval converge, and describe (without running code) what you would expect to see if you plotted g against n for a fixed true configuration of $\bar{x}_S, \bar{x}_T, \sigma^2$.

(Full worked solutions appear in Appendix A.)

