Chapter 8 A test is a model comparison

Detailed notes pp. 1–3 and 14–15
Lecture slides PDF pp. 2–40

Name the population parameter before writing hypotheses. The **null hypothesis** H_0 specifies the reference claim; the **alternative** H_a specifies the departure the scientific question asks about. Choose its direction before inspecting the result.

$$\text{test statistic} = \frac{\text{estimate} - \text{null value}}{\text{standard error under the test model}}$$

A **null distribution** describes this statistic across repeated samples when H_0 is true. A large absolute value is meaningful only relative to the correct distribution and tail.

A p-value is conditional on the null model

The p-value is the probability, *assuming the null model*, of a statistic at least as extreme as the observed one in the direction(s) of H_a. It is not $P(H_0 \mid \text{data})$ or the probability the result is "due to chance."

For an observed statistic z with an appropriate standard-normal null reference,

Alternative	p-value	Reject at level α when
$\theta > \theta_0$	$1 - \Phi(z)$	$z > z_\alpha$
$\theta < \theta_0$	$\Phi(z)$	$z < -z_\alpha$
$\theta \neq \theta_0$	$2[1 - \Phi(\lvert z \rvert)]$	$\lvert z \rvert > z_{\alpha/2}$

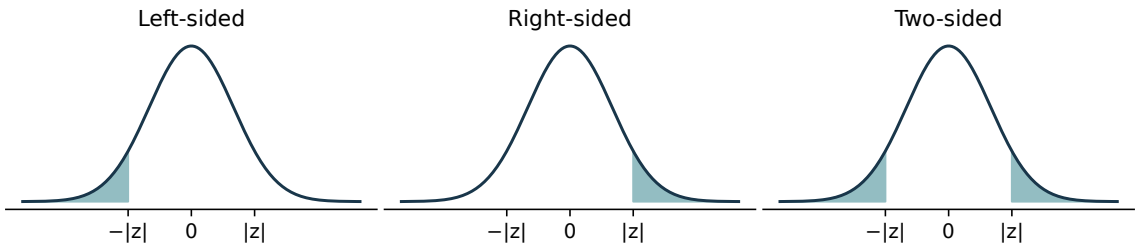
Detailed notes pp. 14–15: critical values and rejection regions

Equivalent decision rule: reject when the p-value is at most the prespecified significance level α. Boundary equality has probability zero under a continuous reference distribution.

Write a conclusion about the parameter

- **Reject:** evidence against H_0 in favor of the stated alternative, at the chosen significance level and under the assumptions.
- **Fail to reject:** insufficient evidence against H_0; this does not prove it true or establish equality.

Small p-values do not measure effect size, scientific importance or causation. Report the observed estimate and context as well as the decision.



The alternative determines which outcomes count as at least as extreme.

Keep in mind. A sampling design determines what can be inferred. A calculation cannot make a convenience sample representative or an observational comparison causal.

Chapter 8 Errors and power

Detailed notes pp. 16–19 and 25–27
Lecture slides PDF pp. 41–53; 74–82

Decision	H_0 true	Specified alternative true
Reject H_0	Type I error	Detection
Fail to reject	Correct non-rejection	Type II error

$$\alpha = P(\text{reject} \mid H_0), \quad \beta(\theta_1) = P(\text{fail to reject} \mid \theta = \theta_1), \quad \text{Power}(\theta_1) = 1 - \beta(\theta_1).$$

For an approximate test, the actual Type I probability is approximately the nominal α. Power depends on a specified true alternative, sample size, variability and rejection rule. There is no single β for every value in a composite alternative.

Compute power in two stages

1. Use the null distribution and α to establish the rejection region on the original estimate scale.
2. Keep that region fixed. Compute its probability under the specified alternative's sampling distribution.

For a two-sided mean test with constant SE s_0, set

$$L = \mu_0 - z_{\alpha/2} s_0, \qquad U = \mu_0 + z_{\alpha/2} s_0.$$

If $\bar{X}$ is approximately $N(\mu_1, s_1)$ under a specified alternative,

$$\beta(\mu_1) \approx \Phi\left(\frac{U - \mu_1}{s_1}\right) - \Phi\left(\frac{L - \mu_1}{s_1}\right).$$

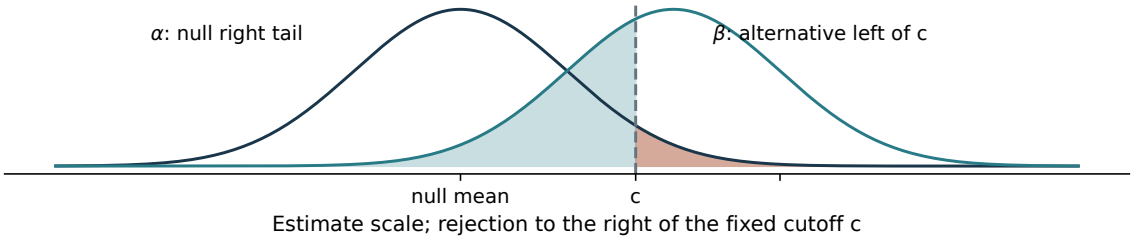
For known, unchanged population SD, $s_1 = s_0 = \sigma/\sqrt{n}$. If variance changes with the parameter, the alternative SE must change too.

A left-sided proportion test illustrates that change

The null cutoff is $c = p_0 - z_\alpha \sqrt{p_0(1-p_0)/n}$. At true proportion p_1, the direct normal power approximation is

$$P_{p_1}(\hat{p} < c) \approx \Phi\left(\frac{c - p_1}{\sqrt{p_1(1-p_1)/n}}\right).$$

Larger n, smaller variability, or a larger effect in the tested direction generally improves power. Increasing α expands rejection and increases power but permits more Type I errors.



Keep the rejection cutoff fixed; change the distribution to compute power.

Keep in mind. Power is not 1 − p-value. Revisit the carpenter calculation, p. 18, and fitness calculation, p. 26, if the two-distribution distinction is unclear.

Chapter 8 Tests for means and paired data

Detailed notes pp. 20–22; one mean on 1–3 and 16
Lecture slides PDF pp. 54–65; one mean on 2–30

One population mean

Detailed notes pp. 2 and 16

For $H_0 : \mu = \mu_0$,

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

With independent normal observations and known σ, this is exactly $N(0,1)$ under H_0. With an adequate iid finite-variance CLT, the normal reference is approximate. Chapter 8 may replace σ by s for a suitable large-sample approximation; that substitution is not an exact normal result.

Two independent population means

For $H_0 : \mu_1 - \mu_2 = \Delta_0$,

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}.$$

Use independent groups and adequate within-group sampling. Known SDs and normal populations give an exact normal reference. With unknown SDs and sufficiently large samples, replace σ_i by s_i for the course's approximate normal procedure. Unequal SDs are allowed in this large-sample formula; no pooled variance assumption is imposed.

The numerator order determines the sign of the estimate, alternative and conclusion. A negative observed difference favors a left-sided alternative, not automatically a two-sided rejection.

Paired observations are one sample of differences

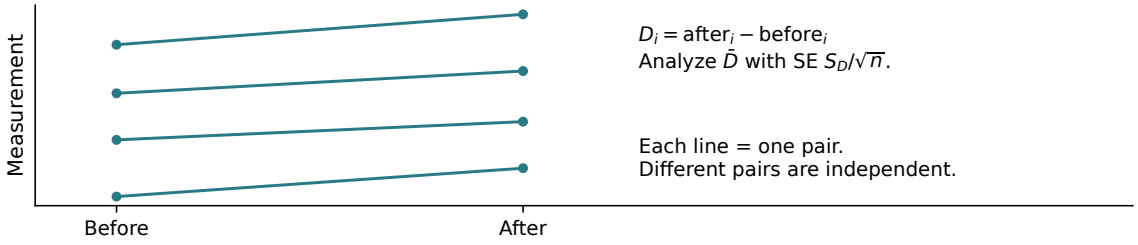
For the same individuals measured twice, or explicitly matched pairs, define

$$D_i = X_i - Y_i, \qquad \mu_D = \mathrm{E}(D_i).$$

Analyze $D_1, \ldots, D_n$, where n is the number of pairs. The estimated SE is $s_D/\sqrt{n}$. A suitable large-sample normal test uses $(\bar{D} - d_0)/(s_D/\sqrt{n})$ approximately; Chapter 9 gives exact t_{n-1} inference for iid normal differences.

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X, Y).$$

Pairing carries information about covariance. Do not replace s_D by $\sqrt{s_X^2 + s_Y^2}$ unless an appropriate covariance justification exists.



Pairing keeps the connection between two measurements on the same unit.

Keep in mind. The design decides paired versus independent. Equal sample sizes alone do not create pairs; the same individuals or meaningful matches do.

Chapter 8 Tests for proportions

Detailed notes pp. 23–29
Lecture slides PDF pp. 66–90

One proportion uses the null probability in its SE

For independent Bernoulli trials and $H_0 : p = p_0$, $\widehat{p} = x/n$ and

$$\boxed{Z = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}.}$$

Check expected null successes $np_0 > 5$ and failures $n(1-p_0) > 5$, plus independence and an appropriate sampling fraction. This is a large-sample normal approximation. The Chapter 7 interval instead estimates SE with $\widehat{p}$; do not silently exchange the denominators.

Two proportions pool only under equality in the null

For independent groups and $H_0 : p_1 = p_2$, the null imposes a shared probability. Following the slides, denote its pooled estimate by $\widehat{p}$ (and $\widehat{q} = 1 - \widehat{p}$):

$$\widehat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\widehat{p}_1 + n_2\widehat{p}_2}{n_1 + n_2}.$$

It is weighted by sample size, not the simple average unless $n_1 = n_2$.

$$\boxed{Z = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})(1/n_1 + 1/n_2)}}.}$$

Check the null expected success/failure counts in each group using the common pooled estimate. The observations and groups must be independent. The normal reference is approximate.

Do not transfer pooling to the confidence interval

A CI for an unknown difference uses the unpooled estimated SE

$$\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}.$$

The two procedures ask different questions: the test evaluates an imposed equality; the CI estimates an unrestricted difference. Therefore a pooled test and an unpooled normal CI need not yield identical decisions at the same nominal level.

Carry the interpretation back to the population

Name the two population proportions, keep a fixed subtraction order and report a percentage-point difference. Statistical evidence for a difference does not itself show why it occurred.

		Variance used (SE is its square root)
Equality-null test	$\widehat{p}_1, \widehat{p}_2 \longrightarrow \widehat{p}$	$\widehat{p}(1-\widehat{p})(1/n_1 + 1/n_2)$
Difference CI	$\widehat{p}_1, \widehat{p}_2$ stay separate	$\widehat{p}_1(1-\widehat{p}_1)/n_1 + \widehat{p}_2(1-\widehat{p}_2)/n_2$

Pool a common probability under equality; keep separate estimates for a difference CI.

Keep in mind. Do not use this equality-null pooled formula mechanically for a nonzero null difference. That is a different constrained problem.

Chapter 8 Choosing and checking a method

Detailed notes pp. 30–32
Lecture slides PDF pp. 2–90

Four questions before any calculation

1. **Target:** a mean, proportion, or difference? Name the population parameter.
2. **Design:** one sample, independent groups or paired observations?
3. **Reference:** why normal, and is that exact or approximate? Which SE is justified under the null?
4. **Direction:** left, right or two-sided, selected from the scientific question?

Confidence intervals and tests share a framework

For a two-sided normal test and matching interval that use the *same* SE and reference, rejecting $H_0 : \theta = \theta_0$ at level α is equivalent to excluding θ_0 from the $1 - \alpha$ CI. This equivalence requires matching procedures; the proportion formulas on p. 4 illustrate why it is not automatic.

Additional published connection: a Poisson mean

This connection appears on detailed p. 31; it is not a separate worked example in the Chapter 8 slides. For iid counts $X_i \sim \text{Poisson}(\lambda)$ with equal exposure,

$$\text{E}(X_i) = \text{Var}(X_i) = \lambda, \qquad \text{Var}(\bar{X}) = \lambda/n.$$

A sufficiently large-sample test of $H_0 : \lambda = \lambda_0 > 0$ can therefore use

$$Z = \frac{\bar{X} - \lambda_0}{\sqrt{\lambda_0/n}} \quad \text{with an approximate normal null reference.}$$

If calculating power at λ_1, retain the rejection rule and use the alternative SE $\sqrt{\lambda_1/n}$. Unequal exposure, dependence or variance exceeding the Poisson mean can invalidate this SE.

Use the existing practice in stages

The Friday questions begin on detailed p. 4; their released solutions follow later. The assigned two-proportion question is separate. Check the course page for released solutions rather than assuming an unreleased answer is available.

What Chapter 9 changes

Estimating variability from a small sample can change the reference distribution substantially. Under normal sampling, t inference accounts for that uncertainty. The estimate-minus-null structure and careful interpretation remain.



Matching two-sided test and CI: same SE, reference and alpha

The CI–test connection requires matching SE, reference and significance level.

Keep in mind. Never turn "failed to reject" into "proved no effect," or a small p-value into a measure of a large effect.