Sections 7.3, 7.4, and 7.9

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 205: Elementary Statistics for the Biological and Life Sciences

Hypothesis tests and confidence intervals

t_s is the test statistic. A 95% CI for $\mu_1 - \mu_2$ is given by

$$\bar{y}_1 - \bar{y}_2 \pm t_{0.025} SE_{\bar{Y}_1 - \bar{Y}_2}.$$

This interval contains zero (the hypothesized value of $\mu_1 - \mu_2$) when

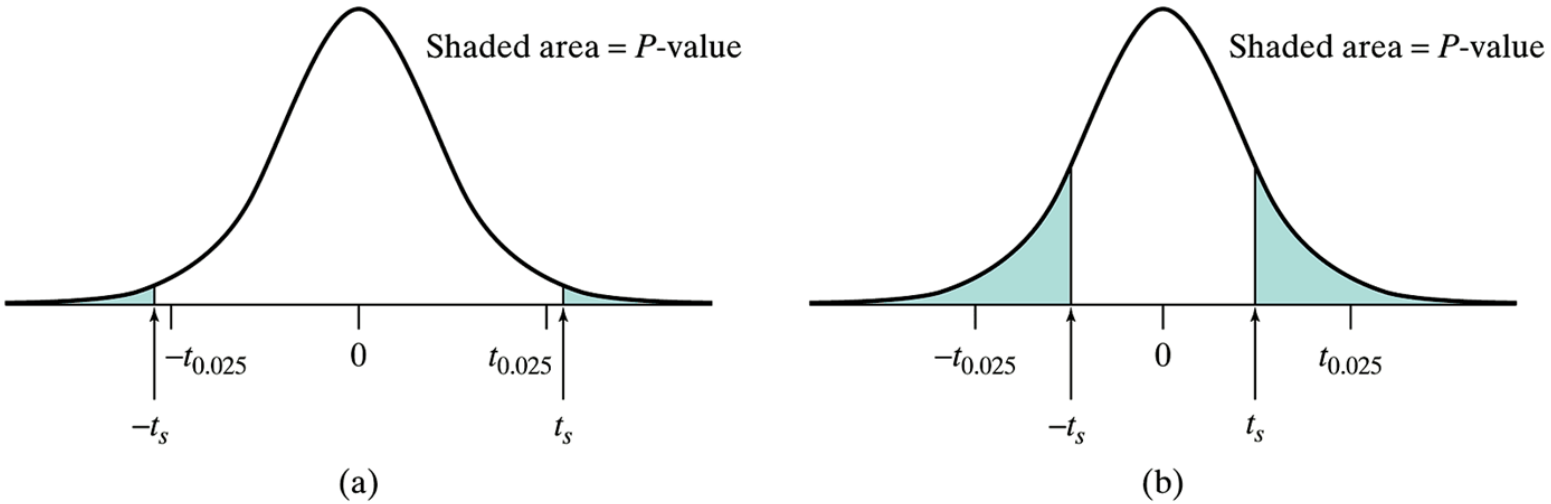
$$|\bar{y}_1 - \bar{y}_2| < t_{0.025} SE_{\bar{Y}_1 - \bar{Y}_2},$$

that is, when

$$\left| \frac{\bar{y}_1 - \bar{y}_2}{SE_{\bar{Y}_1 - \bar{Y}_2}} \right| < t_{0.025},$$

i.e. $|t_s| < t_{0.025}$. This last statement implies P-value > 0.05.

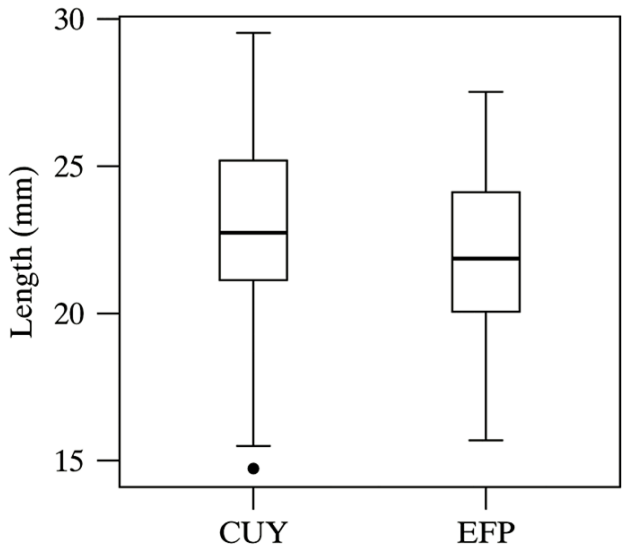
A 95% CI for $\mu_1 - \mu_2$ does not contain zero $\Leftrightarrow$ we reject $H_0 : \mu_1 = \mu_2$ in favor of $H_0 : \mu_1 \neq \mu_2$ at the 5% level.



(a) 0 not in 95% CI for $\mu_1 - \mu_2 \Leftrightarrow |t_s| > t_{0.025} \Leftrightarrow$ P-value < 0.05,
(b) 0 in 95% CI for $\mu_1 - \mu_2 \Leftrightarrow |t_s| < t_{0.025} \Leftrightarrow$ P-value > 0.05

Example 7.3.1

Biologists took samples of crawfish *Orconectes sanborii* from two rivers in central Ohio, the Cuyahoga River and East Fork of Pine Creek and measured their lengths (mm).



The data appear to be approximately normal in from each river. The resulting 95% CI for $\mu_1 - \mu_2$ (Cuyahoga vs. East Fork) is $(-2.68, 0.81)$ mm. Since this interval includes zero, we accept $H_0 : \mu_1 = \mu_2$ at the 5% level. There is no statistical evidence that the mean crawfish lengths are different across rivers.

Interpretation of α

- We reject $H_0 : \mu_1 = \mu_2$ when P-value $< \alpha$.
- When the null $H_0 : \mu_1 = \mu_2$ *is true*, we *wrongly reject* $H_0 : \mu_1 = \mu_2$ with probability α.
- α is called the **Type I error rate**

$$\alpha = \Pr\{\text{Reject } H_0 | H_0 \text{ is true}\}.$$

- Wrongly rejecting the null is a Type I error.

Type II error rate β

- When the alternative $H_A : \mu_1 \neq \mu_2$ *is true*, we *wrongly accept* $H_0 : \mu_1 = \mu_2$ with probability β.
- β depends on μ_1, μ_2, σ_1, σ_2, n_1, and n_2. We never actually know β but we can guess it.
- β is called the **Type II error**

$$\beta = \Pr\{\text{Accept } H_0 | H_A \text{ is true}\}.$$

- Wrongly accepting the null is a Type II error.
- The **power** of the test is

$$1 - \beta = \Pr\{\text{Reject } H_0 | H_A \text{ is true}\}.$$

Possible outcomes of a hypothesis test

Table 7.3.2 Possible outcomes of testing H_0

		True situation	
		H_0 true	H_A true
OUR DECISION	Lack of significant evidence for H_A	Correct	Type II error
	Significant evidence for H_A	Type I error	Correct

Four possibilities.

Example 7.3.3 Marijuana and the pituitary

- Cannabinoids can be transmitted from the mother to fetus (through the placenta) and to the infant through milk. One group of mice are given cannabinoids, the other group are controls. Say μ_1 is mean pituitary function among cannabinoid mice and μ_2 is mean pituitary function among controls.
- We test $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$.
- If we make a Type I error, then we are (wrongly) saying marijuana affects the pituitary of offspring and there could be uneccessary widespread panic.
- If we make a Type II error, then we are (wrongly) saying that marijuana use *does not* affect offspring pituitary function. Then marijuana-using mothers may choose to continue marijuana use, and ultimately negatively affect their kid(s).

Experiments vs. observational studies

- The **response variable** Y measures the outcome of interest, and
- the **explanatory variable** X is used to explain or predict the outcome. So far this has been "group," e.g. treatment or control.
- In an **experiment** we can tease out whether changing X *affects* the distribution of Y (usually focus on mean). This implies a causal relationship; changing X *causes* Y to change.
- With **observational studies** we cannot discuss causality, but rather only association. That is we can find out whether X and Y are *related*, but not whether X causes Y to change.
- Whether we can discuss how X causes Y, or only how X is related to Y has to do with how the data were collected.

Example 7.4.1 hematocrit in males and females

Hematocrit is a measure of the concentration of red blood cells in blood. $n_1 = 489$ 17-year-old males measured and $n_2 = 469$ females.

Table 7.4.1 Hematocrit (percent)

	Males	Females
Mean	45.8	40.6
SD	2.8	2.9

Does being male *cause* mean hematocrit to go up? Is gender and hematocrit *related*?

Observational study of two naturally occurring populations. We merely *observe* an existing relationship.

Example 7.4.2 Pargyline and sucrose consumption

Experiment carried out to see how the psychoactive drug Pargyline affects feeding behavior in the black blowfly. Response Y is amount of sucrose drunk in 30 minutes. $n_1 = 905$ given Pargyline and $n_2 = 900$ given saline (controls).

Table 7.4.2 Sucrose consumption (mg)

	Control	Pargyline
Mean	14.9	46.5
SD	5.4	11.7

Does Pargyline *cause* sugar consumption to increase?

Controlled experiment with treatments administered to two essentially identical populations; manipulations give two "man-made" populations: Pop'n: 1 blowflies given Pargyline, Pop'n 2: given saline.

Experiment vs. observational study: cholesterol

- Your book has a nice example illustrating the difference.
- In a clinical trial, experiments *randomly* assign the same population to a cholesterol-lowering drug or a control. At the end of the study $n_1 = 100$ treatment and $n_2 = 100$ controls have their blood cholesterol measured and a two-sample t-test is conducted to determine if there's a difference.
- If there is a difference, we can infer that the drug *causes* cholesterol to go down; that's the only difference in the populations!

Experiment vs. observational study: cholesterol

- In an observational study, a random sample of people from Camden, SC are measured for cholesterol; several other variables are also recorded, including age, gender, weight, height, blood pressure, marriage status, etc.
- It's found that those under 30 have lower cholesterol than those over 50 years old using a two-sample t-test, from samples of size $n_1 = 453$ and $n_2 = 229$.

Experiment vs. observational study: cholesterol

- Can we conclude that the cholesterol increase is due to age?
- Not necessarily; age *might be* directly related to cholesterol, but it might be that those over 50 ate more bacon and eggs their whole life than those under 30, due to dietary changes in the American diet over time.
- Here, diet is said to be *confounded* with age. Diet is really the causal factor, not age.
- In other words: (a) diet is related to age, and (b) diet is related to cholesterol, so (c) cholesterol is related to age.
- How would we determine whether age is related to cholesterol? Hint: we'd have to conduct a very expensive experiment over a long time...

Observational studies...oops!

- Young and Karr (2011) examined 52 claims based on observational studies that were later studied rigorously via experiments.
- These include hormone replacement therapy for post-menopausal women, vitamins E and C, selenium, low fat diets, folic acid, B6, B12, calcium, beta-carotine.
- In *every instance* the experiment found either no evidence for the claim, or else the association *was in the opposite direction*.
- Example: Bairati et al. (2005) in "A Randomized Trial of Antioxidant Vitamins to Prevent Second Primary Cancers in Head and Neck Cancer Patients" found that Vitamin E and β-carotine *made cancer worse*.

From Bairati et al. (2005)

Although low dietary intakes of antioxidant vitamins and minerals have been associated with higher risks of cancer, results of trials testing antioxidant supplementation for cancer chemoprevention have been equivocal. We assessed whether supplementation with antioxidant vitamins could reduce the incidence of second primary cancers among patients with head and neck cancer. Methods: We conducted a multicenter, double-blind, placebo-controlled, randomized chemoprevention trial among 540 patients with stage I or II head and neck cancer treated by radiation therapy between October 1, 1994, and June 6, 2000. Supplementation with α-tocopherol (400 IU/day) and β-carotene (30 mg/day) or placebo began on the first day of radiation therapy and continued for 3 years after the end of radiation therapy.

From Bairati et al. (2005)

In the course of the trial, β-carotene supplementation was discontinued after 156 patients had enrolled because of ethical concerns. The remaining patients received α-tocopherol or placebo only. Survival was evaluated by Kaplan-Meier analysis. Cox proportional hazards models were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs). All statistical tests were two-sided. Results: After a median follow-up of 52 months, second primary cancers and recurrences of the first tumor were diagnosed in 113 and 119 participants, respectively. The effect of supplementation on the incidence of second primary cancers varied over time. Compared with patients receiving placebo, patients receiving α-tocopherol supplements had a higher rate of second primary cancers during the supplementation period (HR $= 2.88$, 95% CI $= 1.56$ to 5.31)...Conclusions: α-Tocopherol supplementation produced unexpected adverse effects on the occurrence of second primary cancers and on cancer-free survival.

Overview of hypothesis test

- We have a null hypothesis H_0 and the alternative H_A.
- The P-value gives evidence against H_0.
- We reject H_0 if P-value $< \alpha$, where α is the significance level of the test, usually $\alpha = 0.05$.
- α is the probability of a Type I error
  P-value $= \Pr\{\text{reject } H_0 | H_0 \text{ true}\}$.
- R carries out the test $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$ using `t.test(sample1,sample2)`.
- The power of a test is $1 - \beta = \Pr\{\text{reject } H_0 | H_A \text{ true}\}$. This depends on the unknown μ_1 and μ_2.

How to pick H_0 and H_A?

- Since the P-value only gives evidence toward H_A, H_A is *what we want to show*. Also called the "research hypothesis."
- H_0 is the "status quo" – what we want to disprove.
- In an experiement, H_0 will always be that there is no mean difference between treatment and control.

More on P-values (p. 279)

- The P-value of the data is the probability (assuming H_0 is true) of getting a result as extreme as, or more extreme than, the result that was actually observed.
- The P-value is the probability that, if H_0 were true, a result would be obtained that would deviate from as much as (or more than) the actual data do.
- The P-value of the data is the probability (assuming H_0 is true) of getting a result as deviant as, or more deviant than, the result actually observed where deviance is measured as discrepancy from H_0 in the direction of H_A.
- The P-value **is not** the probability that the null hypothesis is true.

Review of important ideas so far

- Test $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$.
- Test statistic is $t_s = (\bar{Y}_1 - \bar{Y}_2)/SE_{\bar{Y}_1 - \bar{Y}_2}$.
- P-value $\Pr\{|T_{df}| > |t_s|\}$ is probability of seeing bigger difference in sample means than actually saw, given H_0 is true.
- Small P-value gives evidence towards $H_A : \mu_1 \neq \mu_2$.
- Reject $H_0 : \mu_1 = \mu_2$ if P-value $< \alpha$, usually $\alpha = 0.05$.
- α is called the significance level of the test, it is the probability of a Type I error.
- Type I error is rejecting $H_0 : \mu_1 = \mu_2$ when H_0 is really true.
- Type II error is accepting $H_0 : \mu_1 = \mu_2$ when $H_A : \mu_1 \neq \mu_2$ is really true.
- The power of the test is $\Pr\{\text{reject } H_0 | H_A \text{ true}\}$.

Review of important ideas so far

- A 95% confidence interval for $\mu_1 - \mu_2$ includes zero **if and only if** we accept $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$ at the 5% significance level.
- The t-test needs normal data *for small sample sizes*, say $n_1 < 30$ or $n_2 < 30$. Check this with a probability plot.
- In large samples, we don't worry about the normality assumption.
- In small samples, the permutation test in Section 7.1 always gives the correct P-value, even when data are not normal.

Review of important ideas so far

- Rejecting $H_0 : \mu_1 = \mu_2$ means that the outcome is *associated* with group membership (e.g. treatment or control) in observational studies.
- In a carefully controlled experiment Rejecting $H_0 : \mu_1 = \mu_2$ may confer a causal relationship.
- "Association is not causation" necessarily.
- A confounding variable is one that changes with group membership, but really causes the response to change.
- In general a P-value is the probability of getting any test statistic more "deviant" than what we saw in the direction of H_A, assuming H_0 is true.