MATH1. Part II

Probability and Statistics

Chapter 7

Comparison of Two Independent Samples

7.1 Hypothesis Testing: The Randomization Test

Comparison of Two Independent Samples

- Suppose that we have samples from two populations,
  - If the two samples look quite similar to each other, we might infer that the two populations are identical;
  - if the samples look quite different, we would infer that the populations differ.

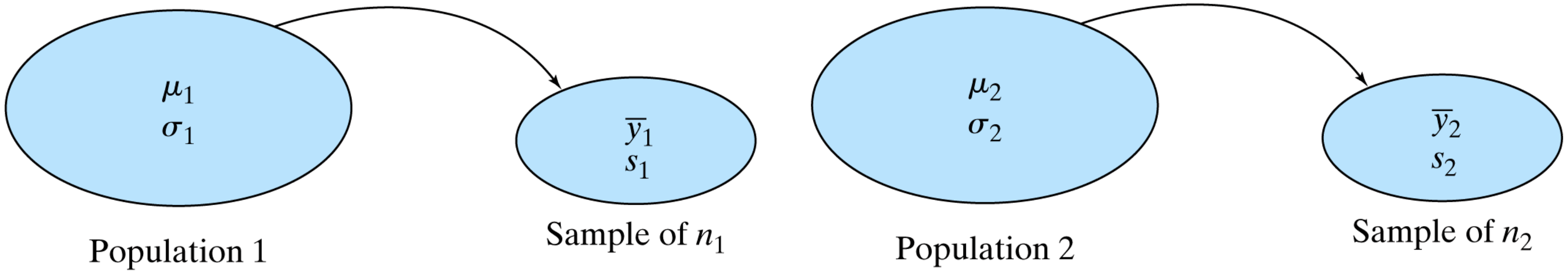


Figure 6.6.1 Notation for comparison of two samples

- *How different do two samples have to be,* in order for us to infer that the populations that generated them are actually different?

7.1 Hypothesis Testing: The Randomization Test

Comparison of Two Independent Samples

- The **randomization test** gives us a way to measure the variability in the difference of two sample means.

Example 7.1.1 Flexibility

- A researcher studied the flexibility of each of 7 women, 4 of whom were in an aerobics class and 3 of whom were dancers.
- $\bar{y}_1 = 51.25$, $\bar{y}_2 = 56$; $\bar{y}_1 - \bar{y}_2 = -4.75$
- Is being a dancer has no effect on flexibility?

Randomization test

- **First, assume no difference between groups.**
- **Claim:** the labels “aerobics” and “dance” are arbitrary and have nothing to do with flexibility (as measured by trunk flexion).

Table 7.1.1 Trunk flexion measurements

Aerobics	Dance
38	48
45	59
58	61
64	
Mean 51.25	56.00

7.1 Hypothesis Testing: The Randomization Test

Comparison of Two Independent Samples

Example 7.1.1 Flexibility

- $\bar{y}1 = 51.25$, $\bar{y}2 = 56$; $\bar{y}1 - \bar{y}2 = -4.75$
- Is being a dancer has no effect on flexibility?

Randomization test

- 35 possible ways (35 randomizations) to divide the 7 observations into two groups, of sizes 4 and 3.
  - $7!/(4!3!) = 35$
- 20 of 35 (20/35 = 57%) are at least as large in magnitude as 51.25-56 = -4.75
- Thus, the observed data are consistent with the claim that the "aerobics" and "dance" have nothing to do with flexibility.

Table 7.1.2 All 35 possible arrangements of the seven sample trunk flexion observations into two groups of sizes 4 and 3

Sample 1 ("aerobics")	Sample 2 ("dance")	Mean of sample 1	Mean of sample 2	Difference in means
38 45 58 64	48 59 61	51.25	56.00	**−4.75**
38 45 58 48	64 59 61	47.25	61.33	**−14.08**
38 45 58 59	64 48 61	50.00	57.67	**−7.67**
38 45 58 61	64 48 59	50.50	57.00	**−6.50**
38 45 64 48	58 59 61	48.75	59.33	**−10.58**
38 45 64 59	58 48 61	51.50	55.67	−4.17
38 45 64 61	58 48 59	52.00	55.00	−3.00
38 45 48 59	58 64 61	47.50	61.00	**−13.50**
38 45 48 61	58 64 59	48.00	60.33	**−12.33**
38 45 59 61	58 64 48	50.75	56.67	**−5.92**
38 58 64 48	45 59 61	52.00	55.00	−3.00
38 58 64 59	45 48 61	54.75	51.33	3.42
38 58 64 61	45 48 59	55.25	50.67	4.58
38 58 48 59	45 64 61	50.75	56.67	**−5.92**
38 58 48 61	45 64 59	51.25	56.00	**−4.75**
38 58 59 61	45 64 48	54.00	52.33	1.67
38 64 48 59	45 58 61	52.25	54.67	−2.42

7.2 Hypothesis Testing: The t Test

Hypothesis testing

- The general idea is to formulate a hypothesis statement:

  μ1 and μ2 have no difference/ differ

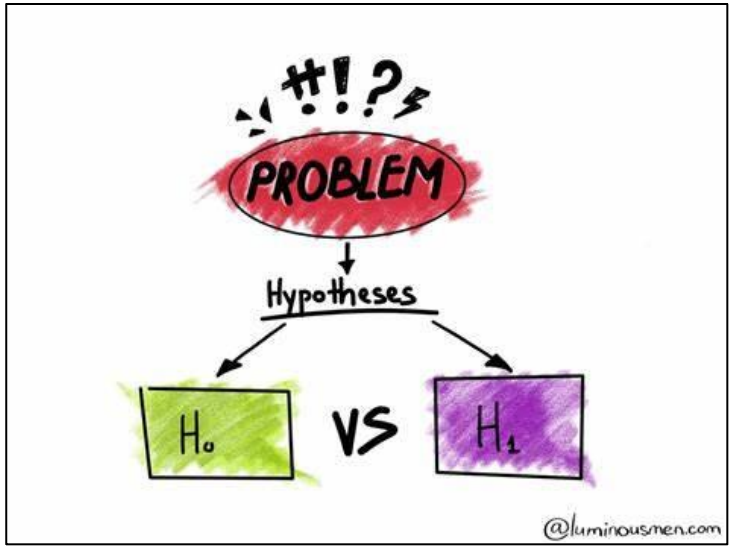
  and then to see whether the data provide sufficient evidence in support of that hypothesis.

- **Null hypothesis H_0**: the hypothesis that μ1 and μ2 are equal (no difference)

  $$H_0 : \mu 1 = \mu 2$$

- **Alternative hypothesis H_A**: the hypothesis that μ1 and μ2 are **NOT** equal (differ)

  $$H_A : \mu 1 \neq \mu 2$$



7.2 Hypothesis Testing: The t Test

Hypothesis testing

- **Null hypothesis H_0**: the hypothesis that μ1 and μ2 are equal (no difference)

$$H_0 : \mu 1 = \mu 2$$

- **Alternative hypothesis H_A**: the hypothesis that μ1 and μ2 are **NOT** equal (differ)

$$H_A : \mu 1 \neq \mu 2$$

Example 7.2.1 toluene and the brain

- Abuse of substances containing toluene (e.g., glue) can produce various neurological symptoms.
- The concentrations of the brain chemical norepinephrine (NE) in the medulla region of the brain, for six toluene-exposed rats and five control rats, are given in Table 7.2.1
- What are the corresponding hypotheses?

Table 7.2.1 NE concentration (ng/gm)

	Toluene (Group 1)	Control (Group 2)
	543	535
	523	385
	431	502
	635	412
	564	387
	549	
n	6	5
$\bar{y}$	540.8	444.2
s	66.1	69.6
SE	27	31

7.2 Hypothesis Testing: The t Test

Hypothesis testing

Example 7.2.1 toluene and the brain

- What are the corresponding hypotheses?
  - H_0 : μ1 = μ2, Toluene has **no** effect on NE concentration in rat medulla.
  - H_A : μ1 ≠ μ2, Toluene has **some** effect on NE concentration in rat medulla, which is caused by toluene.

Table 7.2.1 NE concentration (ng/gm)

	Toluene (Group 1)	Control (Group 2)
	543	535
	523	385
	431	502
	635	412
	564	387
	549	
n	6	5
$\bar{y}$	540.8	444.2
s	66.1	69.6
SE	27	31

- A **statistical test of hypothesis** is a procedure for assessing the strength of evidence present in the data in support of H_A.
- The data are considered to demonstrate evidence for H_A if any discrepancies from H_0 (the opposite of H_A) could not be readily attributed to chance (i.e., to sampling error).

7.2 Hypothesis Testing: The t Test

The t Statistic

- Null hypothesis: H_0: μ1 = μ2 ↔ **H_0: μ1 - μ2 = 0**
- Alternative hypothesis: H_A: μ1 ≠ μ2 ↔ **H_A: μ1 - μ2 ≠ 0**

- The **t test** is a standard method of choosing between these two hypotheses.
  - To carry out the t test, the first step is to compute the **test statistic**, which for a *t* test is defined as

$$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE_{(\bar{Y}_1 - \bar{Y}_2)}}$$

  - The subscript “s” on t_s serves as a reminder that this value is calculated from the data (“s” for “sample”).
  - Notice the structure of t_s : It is a measure of how far the difference between the sample means ($\bar{y}$1- $\bar{y}$2) is from 0 (the difference we would expect to see if H_0 were true - zero difference), expressed in relation to the SE of the difference (the amount of variation we expect to see in differences of means from random samples).

7.2 Hypothesis Testing: The t Test

The t Statistic

- Definition: The t test test statistic is defined as

$$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE_{(\bar{Y}_1 - \bar{Y}_2)}}$$

  – Notice the structure of t_s : It is a measure of how far the difference between the sample means ($\bar{y}$1- $\bar{y}$2) is from 0 (the difference we would expect to see if H_0 were true - zero difference), expressed in relation to the SE of the difference (the amount of variation we expect to see in differences of means from random samples).

Example 7.2.2 toluene and the brain (continued)

- What is the t statistic?
- Use the structure of t_s explain the meaning of t_s .

Table 7.2.1 NE concentration (ng/gm)

	Toluene (Group 1)	Control (Group 2)
	543	535
	523	385
	431	502
	635	412
	564	387
	549	
n	6	5
$\bar{y}$	540.8	444.2
s	66.1	69.6
SE	27	31

7.2 Hypothesis Testing: The t Test

The t Statistic

- Definition: The t test test statistic is defined as

$$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE_{(\bar{Y}_1 - \bar{Y}_2)}}$$

Example 7.2.2 toluene and the brain (continued)

- What is the t statistic?
- Use the structure of t_s explain the meaning of t_s .

  – $SE_{(\bar{Y}1 - \bar{Y}2)} = \sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}} = \sqrt{66.1^2/6 + 69.6^2/5} = 41.195$

  – $t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE_{(\bar{Y}_1 - \bar{Y}_2)}} = \frac{(540.8 - 444.2) - 0}{41.195} = 2.34$

  – The t statistic shows that the difference between ȳ1 and ȳ2 is about 2.3 SEs from zero

Table 7.2.1 NE concentration (ng/gm)

	Toluene (Group 1)	Control (Group 2)
	543	535
	523	385
	431	502
	635	412
	564	387
	549	
n	6	5
$\bar{y}$	540.8	444.2
s	66.1	69.6
SE	27	31

7.2 Hypothesis Testing: The t Test

The t Statistic

- Definition: The t test **test statistic** is defined as

$$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE_{(\bar{Y}_1 - \bar{Y}_2)}}$$

- How shall we judge whether our data are sufficient evidence for H_A?
  - If independent random samples from normally distributed populations, t_s falls in the Student's t distribution.

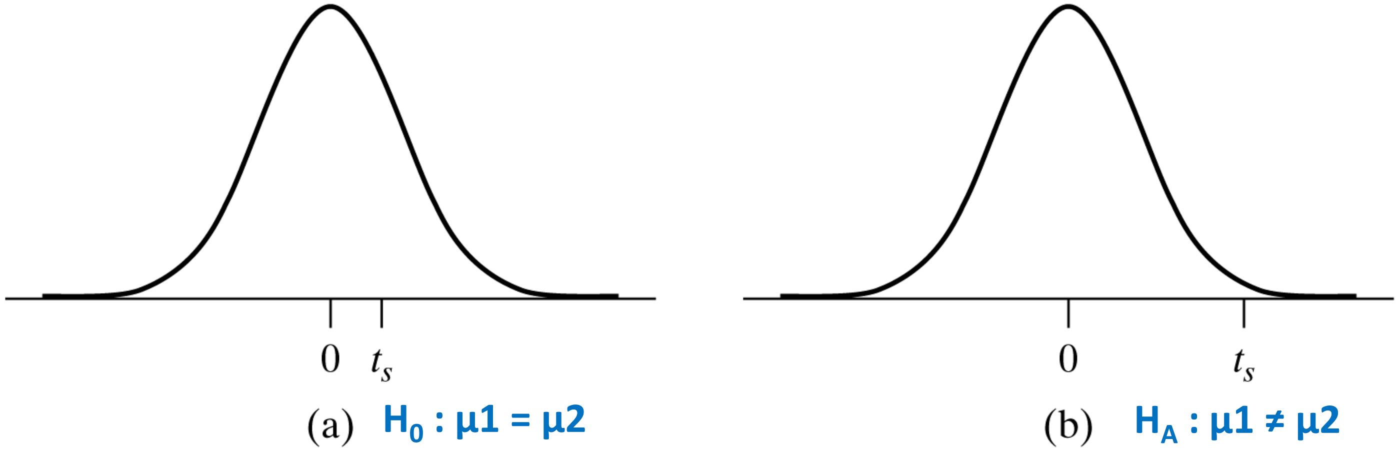


Figure 7.2.2 Essence of the *t* test. (a) Data compatible with H_0 (and thus a lack of significant evidence for H_A); (b) data incompatible with H_0 (and thus significant evidence for H_A).

7.2 Hypothesis Testing: The t Test

P-value

- The P-value represent whether an observed value t_s is "far" in the tail of the t distribution

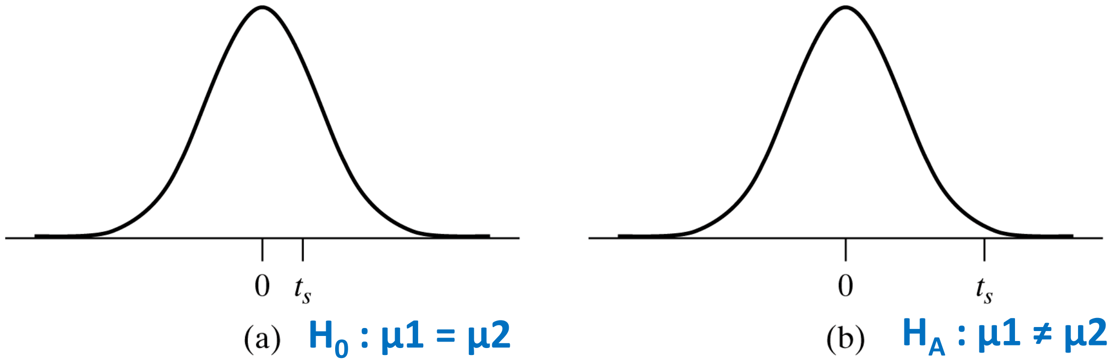


Figure 7.2.2 Essence of the t test. (a) Data compatible with H_0 (and thus a lack of significant evidence for H_A); (b) data incompatible with H_0 (and thus significant evidence for H_A).

- The **P-value** of the *t* test is the area under Student's t curve in the double tails beyond - t_s and + t_s .

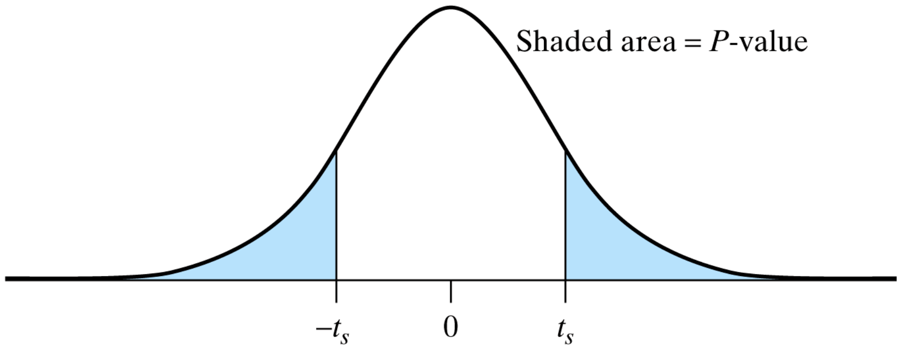
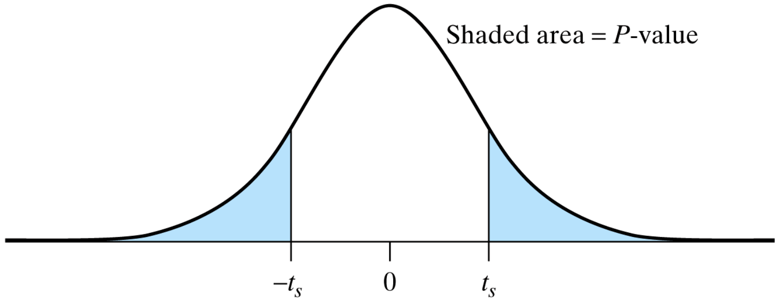


Figure 7.2.3 The two-tailed P-value for the t test

7.2 Hypothesis Testing: The t Test

P-value

- The **P-value** of the *t* test is the area under Student’s t curve in the double tails beyond - t_s and + t_s .



Example 7.2.2 toluene and the brain (continued)

- The $t_s = 2.34$
- What is the P-value?

Table 7.2.1 NE concentration (ng/gm)

	Toluene (Group 1)	Control (Group 2)
	543	535
	523	385
	431	502
	635	412
	564	387
	549	
n	6	5
$\bar{y}$	540.8	444.2
s	66.1	69.6
SE	27	31

7.2 Hypothesis Testing: The t Test

P-value

Example 7.2.2 toluene and the brain (continued)

- The t_s = 2.34
- What is the P-value?
  - Formula (6.7.1) df = $\dfrac{(SE_1^{\ 2}+SE_2^{\ 2})^2}{SE_1^{\ 4}/(n_1-1) + SE_2^{\ 4}/(n_2-1)}$ = 8.47
  - Thus, the P-value is the area under the t curve (with 8.47 degrees of freedom) beyond $\pm$2.34.
  - This area, which was found using a computer,* is shown in Figure 7.2.4 to be 0.0454.

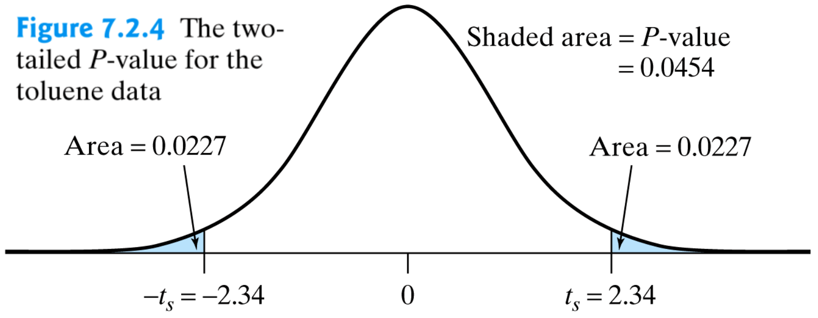
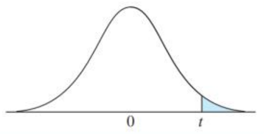


Figure 7.2.4 The two-tailed *P*-value for the toluene data

Table 4 Critical Values of Student's *t* distribution



df	0.20	0.10	0.05	0.04	0.03	0.025	0.02	0.01	0.005	0.0005
	UPPER TAIL PROBABILITY									
1	1.376	3.078	6.314	7.916	10.579	12.706	15.895	31.821	63.657	636.619
2	1.061	1.886	2.920	3.320	3.896	4.303	4.849	6.965	9.925	31.599
3	0.978	1.638	2.353	2.605	2.951	3.182	3.482	4.541	5.841	12.924
4	0.941	1.533	2.132	2.333	2.601	2.776	2.999	3.747	4.604	8.610
5	0.920	1.476	2.015	2.191	2.422	2.571	2.757	3.365	4.032	6.869
6	0.906	1.440	1.943	2.104	2.313	2.447	2.612	3.143	3.707	5.959
7	0.896	1.415	1.895	2.046	2.241	2.365	2.517	2.998	3.499	5.408
8	0.889	1.397	1.860	2.004	2.189	2.306	2.449	2.896	3.355	5.041
9	0.883	1.383	1.833	1.973	2.150	2.262	2.398	2.821	3.250	4.781

- P-value is a measure of compatibility between the data and H_0 and thus measures the evidence for H_A

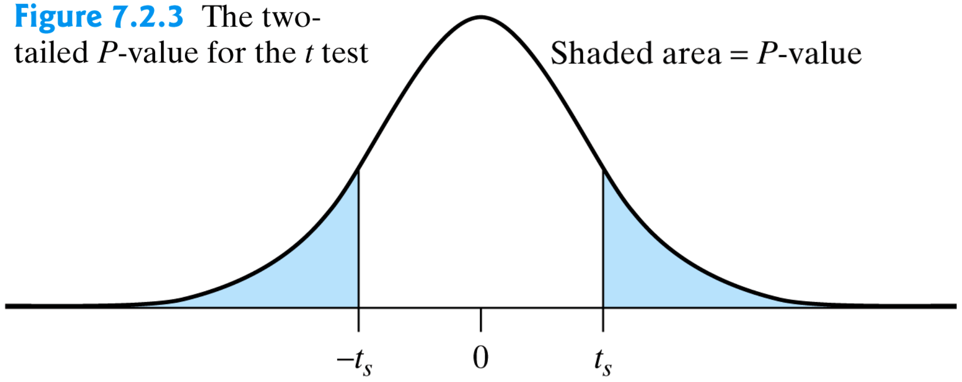
7.2 Hypothesis Testing: The t Test

P-value

DEFINITION The ***P*-value** for a hypothesis test is the probability, computed under the condition that the null hypothesis is true, of the test statistic being at least as extreme as the value of the test statistic that was actually obtained.

– A large P-value (close to 1) indicates a value of t_s near the center of the t distribution (lack of evidence for H_A);
– A small P-value (close to 0) indicates a value of t_s in the far tails of the t distribution (evidence for H_A).

Figure 7.2.3 The two-tailed *P*-value for the *t* test



e.g.
— P- value = 0.80 indicates a lack of evidence for H_A;
— P- value = 0.0001 indicates very strong evidence for H_A.
— How about P-value = 0.10?

7.2 Hypothesis Testing: The t Test

Drawing conclusions from a t test

- **Significance level α**: the threshold value, on the P-value scale, drawing a definite line between sufficient and insufficient evidence.
  - Significance level of the test is denoted by the Greek letter α (alpha).
  - We can think of a as **α** preset <u>threshold</u> of statistical significance.
  - <u>The value of α is chosen by whoever is making the decision</u>. Common choices are α = 0.10, 0.05, and 0.01.
  - If the **P-value ≤ α**, the data are judged to provide statistically significant evidence in favor of $\mathbf{H_A}$; we also may say that H_0 is rejected.
  - If the **P-value > α**, we say that the data provide insufficient evidence to claim that H_A is true, and thus $\mathbf{H_0}$ is not rejected.

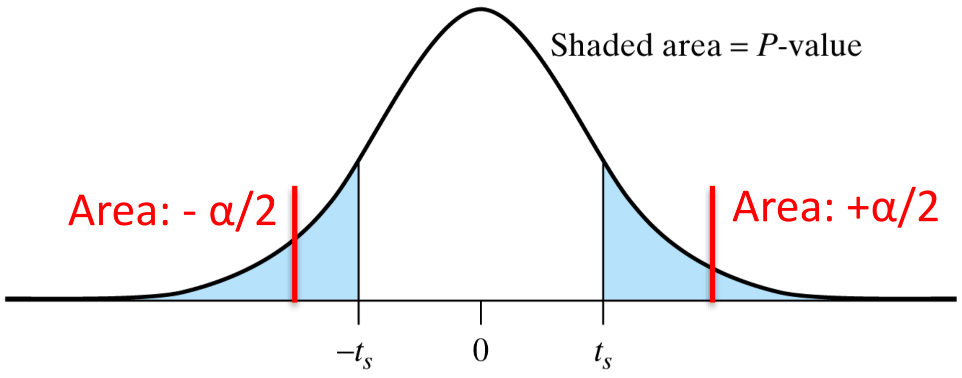


Figure 7.2.3 The two-tailed *P*-value for the *t* test

7.2 Hypothesis Testing: The t Test

Drawing conclusions from a t test

- **Significance level α**: the threshold value, on the P-value scale, drawing a definite line between sufficient and insufficient evidence.
  - If the **P-value ≤ α**, the data are judged to provide statistically significant evidence in favor of $\mathbf{H_A}$; we also may say that H_0 is rejected.
  - If the **P-value > α**, we say that the data provide insufficient evidence to claim that H_A is true, and thus $\mathbf{H_0}$ is not rejected.

Example 7.2.2 toluene and the brain (continued)

- The P-value = 0.0454
- Choose **α = 0.05,** is H_0 true or not?

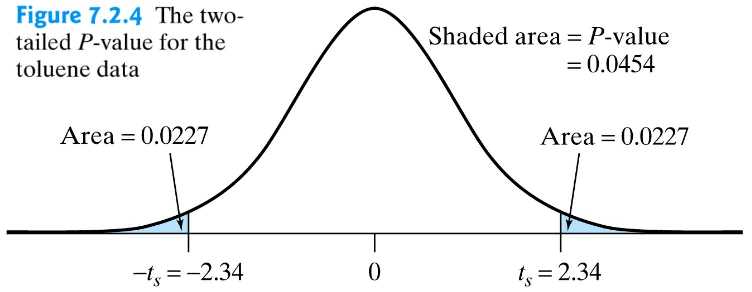


Figure 7.2.4 The two-tailed *P*-value for the toluene data

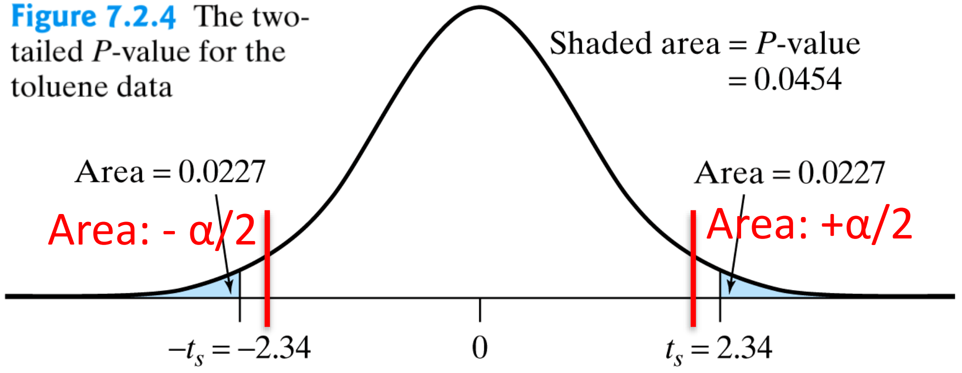
7.2 Hypothesis Testing: The t Test

Drawing conclusions from a t test

Example 7.2.2 toluene and the brain (continued)

- The P-value = 0.0454
- Choose **α = 0.05,** is H_0 true or not?
  - Because the P-value, 0.0454, is less than 0.05, **we reject H_0** .
  - We conclude that the data provide statistically significant evidence in favor of H_A at the 5% level.
  - <u>**Conclusion**</u>: The data provide sufficient evidence **at the 0.05 level of significance** (P-value = 0.0454) that toluene increases NE concentration.

Figure 7.2.4 The two-tailed *P*-value for the toluene data



7.2 Hypothesis Testing: The t Test

Drawing conclusions from a t test

- Significance Level **α** vs. P-value
  - For the t test, both **α** and the P-value are tail areas under Student's t curve.
  - But **α** is an arbitrary prespecified value; it can be (and should be) chosen before looking at the data.
  - By contrast, the **P-value** is determined from the data; indeed, giving the P-value is a way of describing the data.

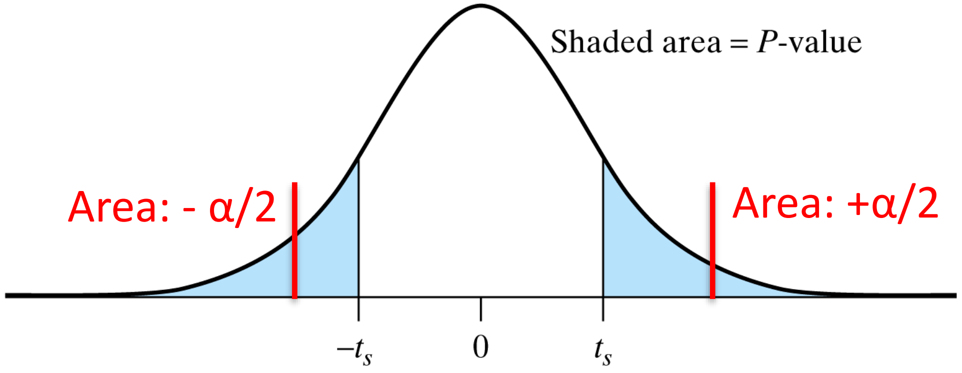


Figure 7.2.3 The two-tailed *P*-value for the *t* test

7.2 Hypothesis Testing: The t Test

Drawing conclusions from a t test

Example 7.2.5 Fast Plants

- From data in Table 7.2.3, can we conclude that the mean height of fast plants was smaller when ancy was used than when water (the control) was used? ($\alpha = 0.05$)

Hint:

- Formulate H_0 and H_A
- Calculate t_s
- Calculate P-value
- Compare P-value with α to get conclusion

Table 7.2.3 Fourteen-day height of control and of ancy plants

	Control	Ancy
n	8	7
$\bar{y}$	15.9	11.0
s	4.8	4.7


10 min Break

7.2 Hypothesis Testing: The t Test

Drawing conclusions from a t test

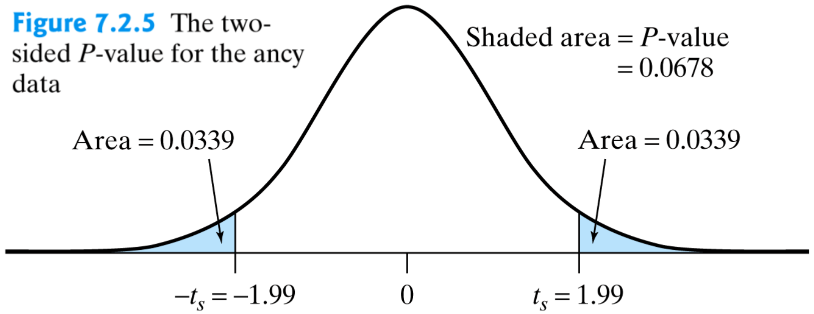
Example 7.2.5 Fast Plants

- From data in Table 7.2.3, can we conclude that the mean height of fast plants was smaller when ancy was used than when water (the control) was used? ($\alpha = 0.05$)
  - H_0: $\mu1 - \mu2 = 0$; H_A: $\mu1 - \mu2 \neq 0$
  - The value of the test statistic is $t_s = [(15.9 - 11.0) - 0]/2.46 = 1.99$
  - Formula (6.7.1) gives df=12.8 for the t distribution.
  - The P-value for the test is the probability of getting a t statistic that is at least as far away from zero as 1.99.
  - P-value = 0.0678

Table 7.2.3 Fourteen-day height of control and of ancy plants

	Control	Ancy
n	8	7
$\bar{y}$	15.9	11.0
s	4.8	4.7

Figure 7.2.5 The two-sided P-value for the ancy data



7.2 Hypothesis Testing: The t Test

Drawing conclusions from a t test

Example 7.2.5 Fast Plants

- Choose **α = 0.05,** is H_0 true or not?
  - The P-value = 0.0678 > 0.05
  - We conclude that the data do NOT provide statistically significant evidence in favor of H_A at the 5% level.
  - <u>Conclusion</u>: The data <u>do not provide sufficient evidence</u> (P-value = 0.0678) at the 0.05 level of significance to conclude that ancy and water differ in their effects on fast plant growth (under the conditions of the experiment that was conducted).
  - We do not say that there is evidence for H_0, but only that there is insufficient evidence against it (for H_A).

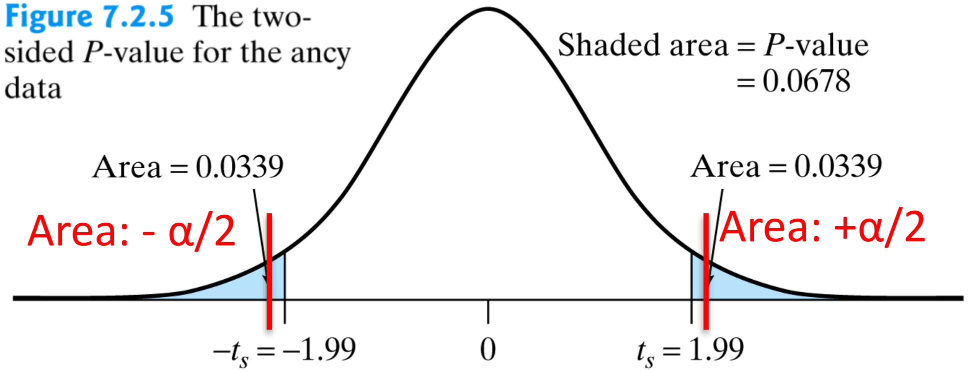


Figure 7.2.5 The two-sided P-value for the ancy data

7.3 Further Discussion of the t Test

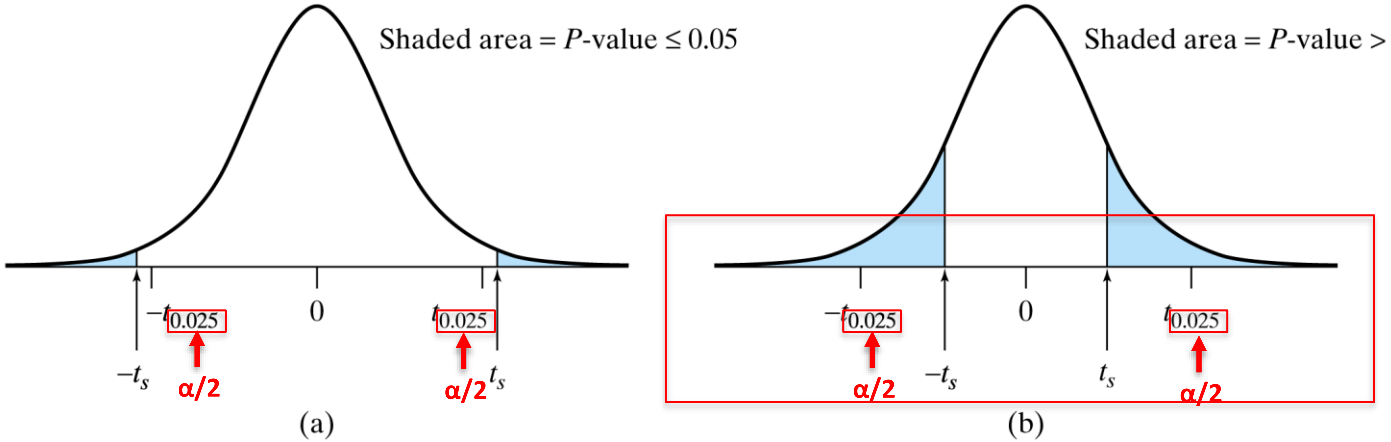
Relationship between test and confidence interval

- The t test at the 5% significance level
  - t test ($\alpha = 0.05$): lack significant evidence for H_A : $\mu1 \neq \mu2$, if and only if

$$|t_s| = \frac{|\bar{y}_1 - \bar{y}_2|}{SE_{(\bar{Y}_1 - \bar{Y}_2)}} < t_{0.025}$$

$$\rightarrow (\bar{y}1 - \bar{y}2) - t_{0.025} \times SE_{(\bar{Y}1 - \bar{Y}2)} < 0 < (\bar{y}1 - \bar{y}2) + t_{0.025} \times SE_{(\bar{Y}1 - \bar{Y}2)}$$

Figure 7.3.1 Possible outcomes of the t test at $\alpha = 0.05$. (a) If $|t_s| \geq t_{0.025}$ then P-value ≤ 0.05 and there is significant evidence for H_A (so H_0 is rejected). (b) If $|t_s| < t_{0.025}$, then P-value > 0.05 and there is a lack of significant evidence for H_A.



7.3 Further Discussion of the t Test

Relationship between test and confidence interval

- The t test at the 5% significance level
  - t test ($\alpha = 0.05$): lack significant evidence for $H_A : \mu 1 \neq \mu 2$, if and only if

$$|t_s| = \frac{|\bar{y}_1 - \bar{y}_2|}{SE_{(\bar{Y}_1 - \bar{Y}_2)}} < t_{0.025}$$

$$\rightarrow \underbrace{(\bar{y}1 - \bar{y}2) - t_{0.025} \times SE_{(\bar{Y}1 - \bar{Y}2)}} < 0 < \underbrace{(\bar{y}1 - \bar{y}2) + t_{0.025} \times SE_{(\bar{Y}1 - \bar{Y}2)}}$$

95% Confidence interval for $\mu 1 - \mu 2$

- Therefore, we lack significant evidence for $H_A : \mu 1 \neq \mu 2$ if and only if the confidence interval for $(\mu 1 - \mu 2)$ <u>includes zero</u>.
- Confidence interval approach and hypothesis testing approach are different ways of using the same basic information.

7.3 Further Discussion of the t Test

Interpretation of α

- If the observed value of t_s falls in the hatched (shaded) regions of the t_s axis, then there is significant evidence for H_A

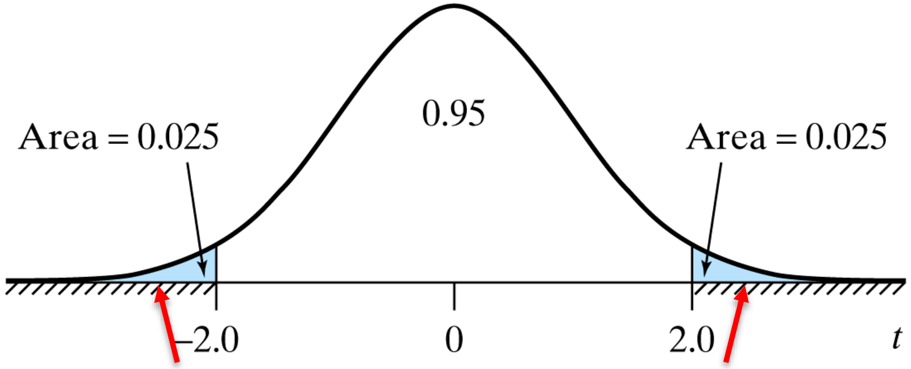


Figure 7.3.3 A t test at $\alpha = 0.05$. There is significant evidence for H_A if t_s falls in the hatched region

- α can be interpreted as a probability:
  - α = 0.05 = Pr { $\underbrace{\text{data provide significant evidence for } H_A}_{t_s \text{ : significant evidence for } H_A}$ }, if $\underbrace{H_0 \text{ is true}}_{\text{fact}}$ ⇨ **Type I error**

7.3 Further Discussion of the t Test

Type I and Type II errors

- **Type I error**: claiming that data provide evidence that significantly supports H_A when H_0 is true.
  - In choosing α, we are choosing our level of protection against Type I error.
  - $\alpha = 0.05 = \Pr\{\underbrace{\text{data provide significant evidence for } H_A}_{t_s\text{ : significant evidence for } H_A}\}$, $\underbrace{\text{if } H_0 \text{ is true}}_{\text{fact}}$ ⇒ **Type I error**
- **Type II error**: If H_A is true, but we do not observe sufficient evidence to support H_A

Table 7.3.2 Possible outcomes of testing H_0

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

7.3 Further Discussion of the t Test

Power

- The probability of making a Type II error is denoted by β:

$$\beta = \Pr\{\underbrace{\text{lack of significant evidence for } H_A}_{t_s\text{ : significant evidence for } H_A}\}, \text{ if } \underbrace{H_A}_{\text{fact}} \text{ is true}$$

Type II error

- **Power**: The chance of NOT making a Type II error, when H_A is true—that is, the chance of having significant evidence for H_A when H_A is true—is called the **power of a statistical test**:

$$\text{Power} = 1 - \beta = \Pr\{\text{significant evidence for } H_A\}, \text{ if } H_A \text{ is true}$$

7.3 Further Discussion of the t Test

Type I and Type II errors

Example 7.3.4 Immunotherapy

- H_0 : Immunotherapy has no effect on survival.
- H_A : Immunotherapy does affect survival.
- What is the Type I error?
- What is the Type II error?

Table 7.3.2 Possible outcomes of testing H_0

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

7.3 Further Discussion of the t Test

Type I and Type II errors

Example 7.3.4 Immunotherapy

- H_0 : Immunotherapy has no effect on survival.
- H_A : Immunotherapy does affect survival.
- What is the Type I error?
- What is the Type II error?

Table 7.3.2 Possible outcomes of testing H_0

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

— **Type I error** : if immunotherapy is not effective, but we conclude that our data provide significant evidence for H_A and thus conclude that immunotherapy is effective

— **Type II error** : if the immunotherapy is actually effective, but our data do not enable us to detect that fact and thus conclude that immunotherapy is not effective

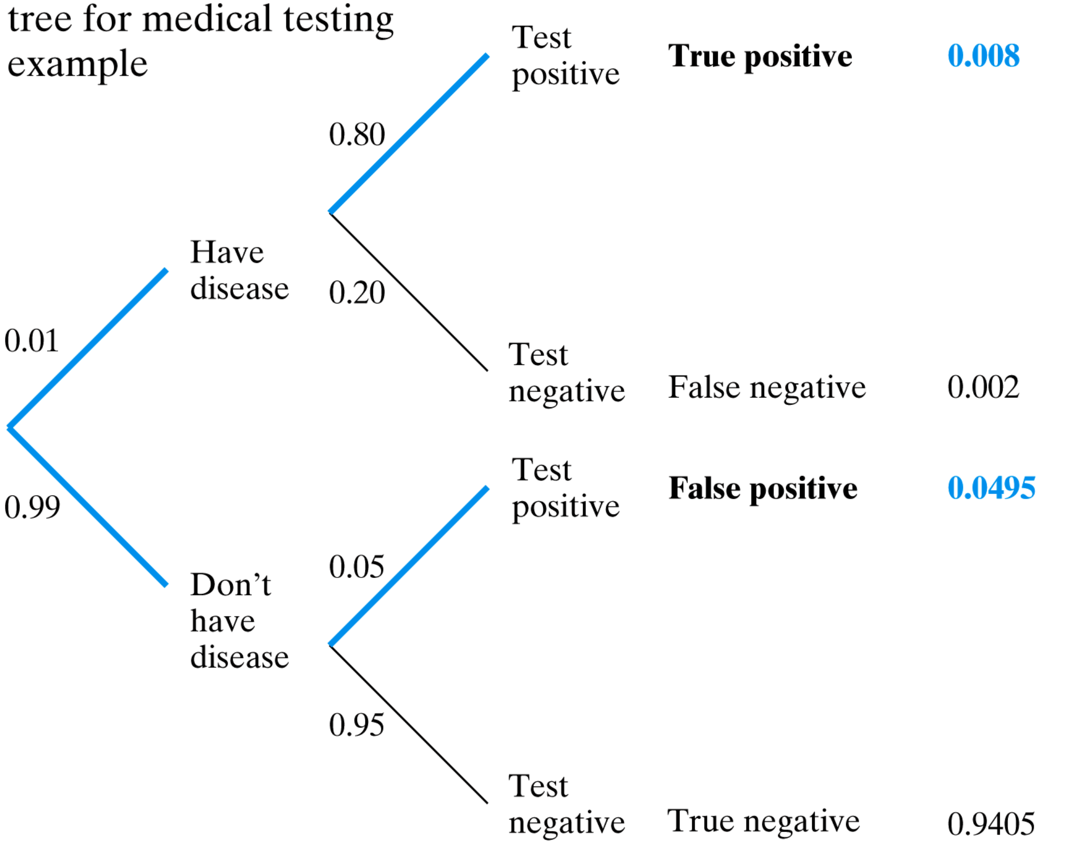
7.9 More on Principles of Testing Hypotheses

Interpretation of Error Probabilities

Example 7.9.1 Medical Testing

- Suppose a medical test is conducted to detect an illness.
- suppose that 1% of the population has the illness in question
- Suppose that the test has an 80% chance of detecting the disease if the person has it
- Suppose that the test has a 95% chance of correctly indicating that the disease is absent if the person really does not have the disease
- What is Type I error?
- What is Type II error?

Figure 7.9.1 Probability tree for medical testing example



7.9 More on Principles of Testing Hypotheses

Interpretation of Error Probabilities

Example 7.9.1 Medical Testing

- What is Type I error?
  - False positive
- What is Type II error?
  - False negative

Table 7.9.1 Hypothetical results of medical test of 100,000 persons

		fact → True situation		
		Healthy (H_0 true)	Ill (H_A true)	Total
TEST RESULT data	Negative (lack of significant evidence for H_A)	94,050	Type II 200	94,250
	Positive (significant evidence for H_A)	Type I 4,950	800	5,750
	Total	99,000	1,000	100,000

7.4 Association and Causation

Do changes in X cause changes in Y?

- response variable, Y: a variable that measures an outcome of interest
- explanatory variable X: a variable used to explain or predict an outcome.
- Experiment: can assess whether X affects the mean value of Y.
- Observational studies: can only conclude association between X and Y.

Association does not imply causation.

(a) Causation

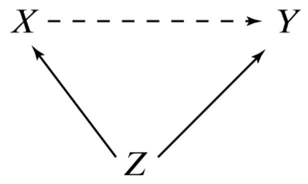
(b) Confounding

(c) Spurious association

(a)

(b) The effect of X on Y is confounded with the effect of Z on Y

Figure 7.4.1 Schematic representation of causation (a) and of confounding (b)



The association between X and Y is spurious; controlling for the lurking variable Z eliminates the X–Y link.

Figure 7.4.2 Schematic representation of spurious association

7.5 One-Tailed t Tests

Two-tailed t test (Review 7.2)

- Null hypothesis: H_0: μ1 = μ2 ↔ **H_0: μ1 - μ2 = 0**
- Alternative hypothesis: H_A: μ1 ≠ μ2 ↔ **H_A: μ1 - μ2 ≠ 0**
  - This alternative H_A is called a **nondirectional** alternative.

One-tailed t test

- Alternative hypothesis: **H_A: μ1 > μ2** or **H_A: μ1 < μ2**
  - This alternative H_A is called a **directional** alternative.
- Null hypothesis: **H_0: μ1 ≤ μ2** or **H_0: μ1 ≥ μ2**

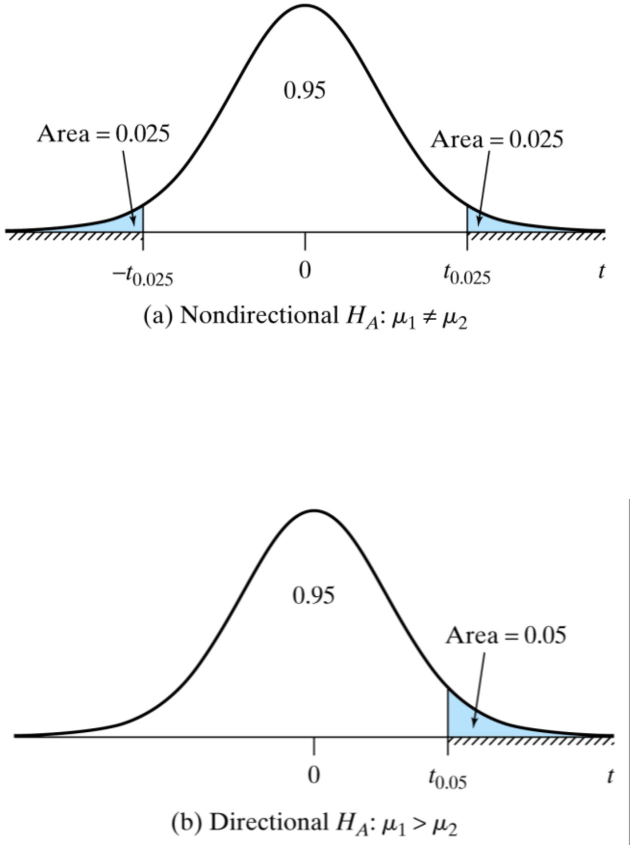


Figure 7.5.5 Two-tailed and one-tailed t test with $\alpha = 0.05$. The data provide significant evidence for H_A if t_s falls in the hatched region of the t-axis

7.5 One-Tailed t Tests

One-tailed t test

- Alternative hypothesis **H_A**: $\mu 1 > \mu 2$ or H_A: $\mu 1 < \mu 2$
  — This alternative H_A is called a <u>directional</u> alternative.
- Null hypothesis **H_0**: $\mu 1 \le \mu 2$ or $\mu 1 \ge \mu 2$

Rule for Directional Alternatives

It is legitimate to use a directional alternative H_A *only if* H_A is formulated before seeing the data and there is no scientific interest in results that deviate in a manner opposite to that specified by H_A.

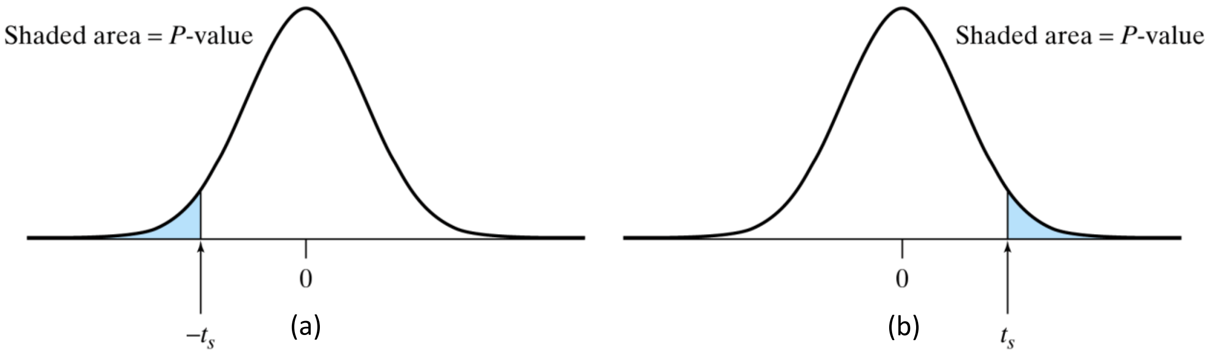


Figure 7.5.1 One-tailed P-value for a t test, (a) if the alternative is H_A: $\mu_1 < \mu_2$ and t_s is negative; (b) if the alternative is H_A: $\mu_1 > \mu_2$ and t_s is positive

7.5 One-Tailed t Tests

One-tailed t test

- Alternative hypothesis $\mathbf{H_A}$: μ1 > μ2 or H_A: μ1 < μ2
  — This alternative H_A is called a <u>directional</u> alternative.
- Null hypothesis $\mathbf{H_0}$: μ1 ≤ μ2 or μ1 ≥ μ2

Example 7.5.1 Niacin supplementation

- Consider a feeding experiment with lambs.
- Whether niacin will increase weight gain.
- How to formulate H_A and H_0 ?

7.5 One-Tailed t Tests

One-tailed t test

- Alternative hypothesis **H_A**: $\mu 1 > \mu 2$ or H_A: $\mu 1 < \mu 2$
  - — This alternative H_A is called a directional alternative.
- Null hypothesis **H_0**: $\mu 1 \leq \mu 2$ or $\mu 1 \geq \mu 2$

Example 7.5.1 Niacin supplementation

- Consider a feeding experiment with lambs.
- Whether niacin will increase weight gain.
- How to formulate H_A and H_0 ?
  - H_A : Niacin is effective in increasing weight gain ($\mu 1 > \mu 2$).
  - H_0 : Niacin is not effective in increasing weight gain ($\mu 1 \leq \mu 2$).

7.5 One-Tailed t Tests

One-tailed t test procedure

- **Step 1**. Check directionality—see if the data deviate from H_0 <u>in the direction specified by **H_A**</u>:
  - (a) If not, the P-value is greater than 0.50.
  - (b) If so, proceed to step 2.

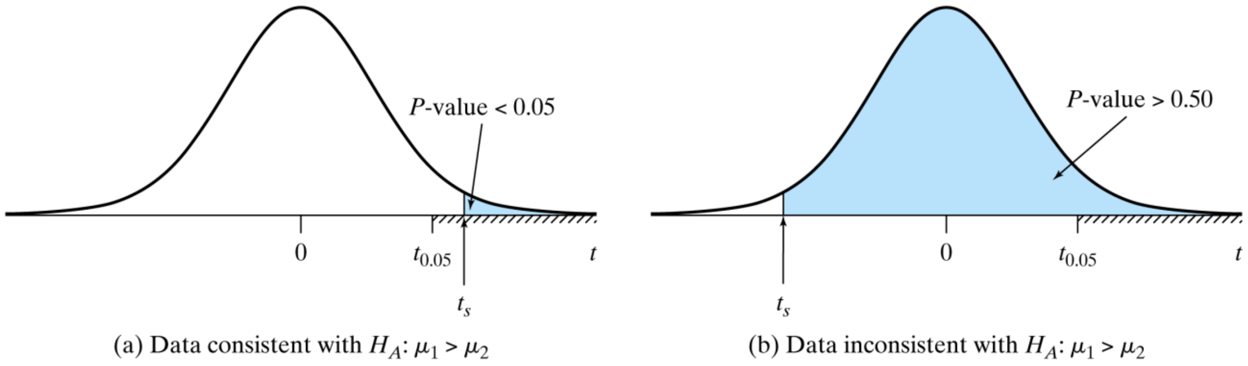


Figure 7.5.2 One-tailed P-value for a t test, (a) in which the data are consistent with H_A: $\mu_1 > \mu_2$; (b) in which the data are inconsistent with H_A: $\mu_1 > \mu_2$

- **Step 2**. The P-value of the data is the one-tailed area beyond t_s
- To conclude the test, one can make a decision at a prespecified significance level α:
  - H_0 is rejected if P-value ≤ α.

7.5 One-Tailed t Tests

One-tailed t test procedure

Example 7.5.2 Niacin supplementation (continued)

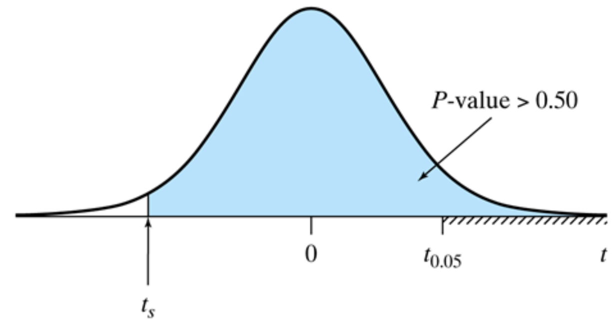
- H_0 : Niacin is not effective in increasing weight gain ($\mu1 \leq \mu2$).
- H_A : Niacin is effective in increasing weight gain ($\mu1 > \mu2$).
- Suppose the data give $\bar{y}1 = 10$ lb and $\bar{y}2 = 13$ lb.
- Given df = 18, $SE_{(\bar{Y}1 - \bar{Y}2)} = 2.2$ lb. $\alpha = 0.05$
- If there is sufficient evidence to support H_A?

7.5 One-Tailed t Tests

One-tailed t test procedure

Example 7.5.2 Niacin supplementation (continued)

- H_0 : Niacin is not effective in increasing weight gain ($\mu1 \le \mu2$).
- H_A : Niacin is effective in increasing weight gain ($\mu1 > \mu2$).
- Suppose the data give $\bar{y}1$ = 10 lb and $\bar{y}2$ = 13 lb.
- Given df = 18, $SE_{(\bar{Y}1 - \bar{Y}2)}$ = 2.2 lb. $\alpha = 0.05$
- If there is sufficient evidence to support H_A?
  - Check directionality: We have $\bar{y}1 < \bar{y}2$, but H_A asserts that $\mu1 > \mu2$.
  - This deviation from H_0 is <u>opposite</u> to the assertion of H_A
  - Consequently, P-value > 0.50, so we would not find significant evidence for H_A at any significance level.



(b) Data inconsistent with H_A: $\mu_1 > \mu_2$

7.5 One-Tailed t Tests

One-tailed t test procedure

Example 7.5.2 Niacin supplementation (continued)

- H_0 : Niacin is not effective in increasing weight gain ($\mu1 \le \mu2$).
- H_A : Niacin is effective in increasing weight gain ($\mu1 > \mu2$).
- Suppose the data give $\bar{y}1 = 14$ lb and $\bar{y}2 = 10$ lb.
- Given df = 18, $SE_{(\bar{Y}1 - \bar{Y}2)} = 2.2$ lb. $\alpha = 0.05$
- If there is sufficient evidence to support H_A?

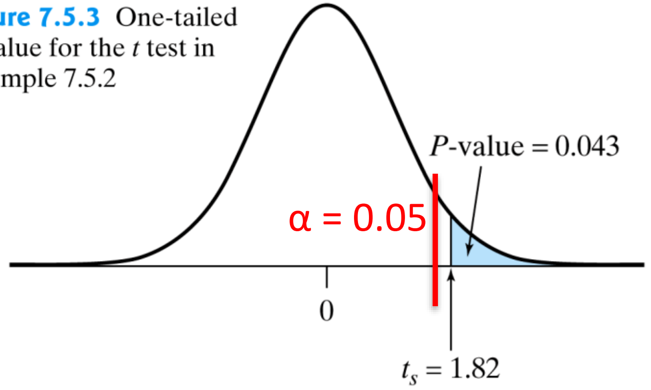
7.5 One-Tailed t Tests

One-tailed t test procedure

Example 7.5.2 Niacin supplementation (continued)

- H_0 : Niacin is not effective in increasing weight gain ($\mu1 \le \mu2$).
- H_A : Niacin is effective in increasing weight gain ($\mu1 > \mu2$).
- Suppose the data give $\bar{y}1 = 14$ lb and $\bar{y}2 = 10$ lb.
- Given df = 18, $SE_{(\bar{Y}1 - \bar{Y}2)} = 2.2$ lb. $\alpha = 0.05$
- If there is sufficient evidence to support H_A?
  - Check directionality: We have $\bar{y}1 > \bar{y}2$, H_A asserts that $\mu1 > \mu2$.
  - $t_s = [(14 - 10) – 0] /2.2 = 1.82$, This upper tail probability (found with a computer*) is 0.043.
  - Since P-value < α, we reject H_0 and conclude that there is some evidence that niacin is effective

Figure 7.5.3 One-tailed *P*-value for the *t* test in Example 7.5.2



7.5 One-Tailed t Tests

One-tailed t test procedure

Example 7.5.2 Niacin supplementation (continued)

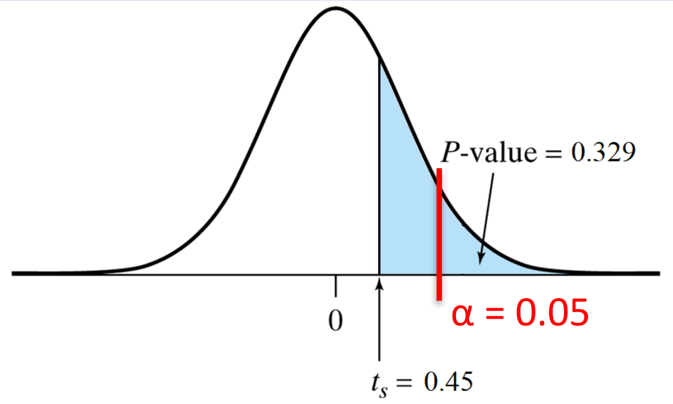
- H_0 : Niacin is not effective in increasing weight gain ($\mu1 \le \mu2$).
- H_A : Niacin is effective in increasing weight gain ($\mu1 > \mu2$).
- Suppose the data give $\bar{y}1 = 11$ lb and $\bar{y}2 = 10$ lb.
- Given df = 18, $SE_{(\bar{Y}1 - \bar{Y}2)} = 2.2$ lb. $\alpha = 0.05$
- If there is sufficient evidence to support H_A?

7.5 One-Tailed t Tests

One-tailed t test procedure

Example 7.5.2 Niacin supplementation (continued)

- H_0 : Niacin is not effective in increasing weight gain ($\mu1 \le \mu2$).
- H_A : Niacin is effective in increasing weight gain ($\mu1 > \mu2$).
- Suppose the data give $\bar{y}1$ = 11 lb and $\bar{y}2$ = 10 lb.
- Given df = 18, $SE_{(\bar{Y}1 - \bar{Y}2)}$ = 2.2 lb. $\alpha = 0.05$
- If there is sufficient evidence to support H_A?
  - Check directionality: We have $\bar{y}1 > \bar{y}2$, H_A asserts that $\mu1 > \mu2$.
  - $t_s = [(11 - 10) – 0] /2.2 = 0.45$, This upper tail probability (found with a computer*) is 0.329.
  - Since P-value > α, we do not find significant evidence for H_A .
  - We conclude that there is insufficient evidence to claim that niacin is effective.



7.6 More on Interpretation of Statistical Significance

Signifiant difference vs. important difference

- The term **significant** is often used in describing the results of <u>a statistical analysis</u>
  - e.g. “The difference was significant” ≈ “The null hypothesis of no difference was rejected.”
- The question of whether a difference is **important**, as opposed to (statistically) significant, cannot be decided on the basis of the P-values alone but must also include an examination of the magnitude of the estimated population difference as well as specific expertise in the research area or practical situation.

Example 7.6.2 Body Weight

- For these data, $t_s = 0.93$ and P-value ≈ 0.45.
- It is not statistically significant for any reasonable choice of α.
  - The lack of statistical significance does not imply that the sex difference in body weight is small or unimportant.
  - It means only that the data are inadequate to characterize the difference in the population means, especially with such small sample sizes

Table 7.6.2 Body weight (lb)

	Males	Females
n	2	2
$\bar{y}$	175	143
s	35	34

7.6 More on Interpretation of Statistical Significance

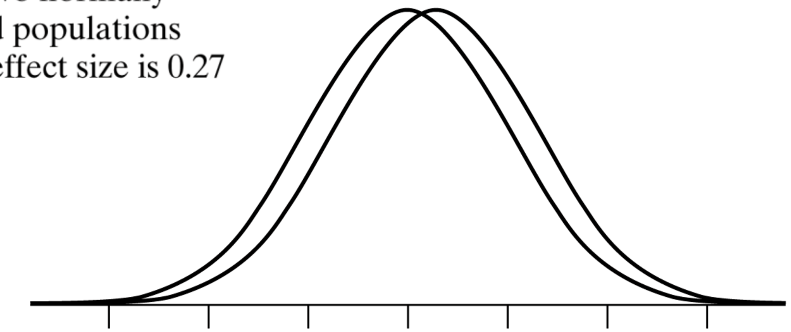
Effect size

- The effect size in a study is the difference between μ1 and μ2, expressed relative to the standard deviation of one of the populations.

$$\text{Effect size} = \frac{|\mu_1 - \mu_2|}{\sigma}$$

  – when working with sample data we can only calculate an estimated effect size by using sample values in place of the unknown population values.

Figure 7.6.1 Overlap between two normally distributed populations when the effect size is 0.27



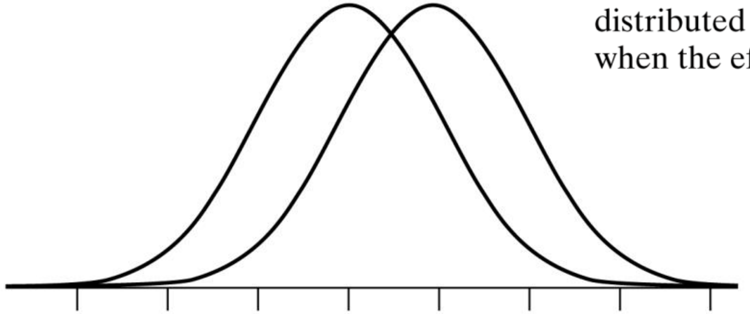


Figure 7.6.2 Overlap between two normally distributed populations when the effect size is 0.91

7.8 Student’s t: Conditions and Summary

Conditions

The test and confidence interval procedures we have described are appropriate if the following conditions* hold:

1. Conditions on the design of the study

a) It must be reasonable to regard the data as random samples from their respective populations. The populations must be large relative to their sample sizes. The observations within each sample must be independent.

b) The two samples must be independent of each other.

2. Condition on the form of the population distributions

- The sampling distributions of $\bar{Y}_1$ and $\bar{Y}_2$ must be (approximately) normal.
- This can be achieved via normality of the populations or by appealing to the Central Limit Theorem (recall Section 6.5) if the populations are nonnormal but the sample sizes are large, where “largeness” depends on the degree of nonnormality of the populations.

7.8 Student's t: Conditions and Summary

Summary of t Test Mechanics

t Test

$$
\begin{aligned}
H_0&: \mu_1 = \mu_2 \\
H_A&: \mu_1 \neq \mu_2 \text{ (nondirectional)} \\
H_A&: \mu_1 < \mu_2 \text{ (directional)} \\
H_A&: \mu_1 > \mu_2 \text{ (directional)}
\end{aligned}
$$

$$\text{Test statistic: } t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\mathrm{SE}_{(\bar{Y}_1 - \bar{Y}_2)}}$$

$$P\text{-value} = \text{tail area under Student's } t \text{ curve with}$$

$$\mathrm{df} = \frac{(\mathrm{SE}_1^2 + \mathrm{SE}_2^2)^2}{\mathrm{SE}_1^4/(n_1 - 1) + \mathrm{SE}_2^4/(n_2 - 1)}$$

Nondirectional H_A: P-value = two-tailed area beyond t_s and $-t_s$

Directional H_A: Step 1. Check directionality.

Step 2. P-value = single-tail area beyond t_s

Decision: Significant evidence for H_A if P-value $\leq \alpha$

Summary

Chapter 7 Comparison of Two Independent Samples

- 7.1 Hypothesis Testing: The Randomization Test
- 7.2 Hypothesis Testing: The t Test
- 7.3 Further Discussion of the t Test
- 7.9 More on Principles of Testing Hypotheses
- 7.4 Association and Causation
- 7.5 One-Tailed t Tests
- 7.6 More on Interpretation of Statistical Significance
- 7.8 Student’s t: Conditions and Summary

Homework

Chapter 7

- 7.1.1 ;
- 7.2.4 ; 7.2.5 ;
- 7.3.5 ; 7.3.7 ;
- 7.4.2 ;
- 7.5.2 ; 7.5.4 ;
- 7.6.3 ; 7.6.6 ;
- 7.8.1 ;
- 7.9.1