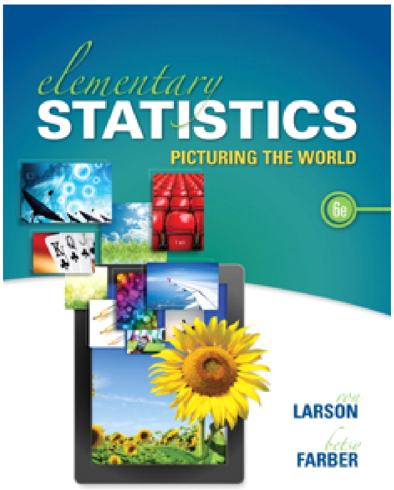
Elementary Statistics: Picturing The World

Sixth Edition



Chapter 8

Hypothesis Testing with Two Samples

8.1 pg. 424 9,10,11,12,14,15,17,19,20,24

8.2 pg. 432 8,12,13,14,17,18,21

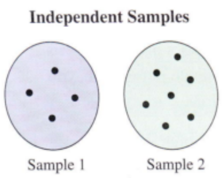
8.3 pg. 442 10,12 Pick one problem from 14-20

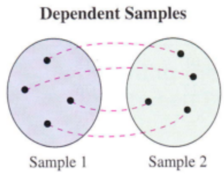
8.4 pg. 451 7,9 OR 10,19,20

Lesson 8-1 Testing the difference between Means (Large independent samples)

Two samples are independent if the samples selected from one population is not related to the sample selected from the second population.

Two samples are dependent if each member of one sample corresponds to a member of the other sample they are also called Paired Samples or matched samples

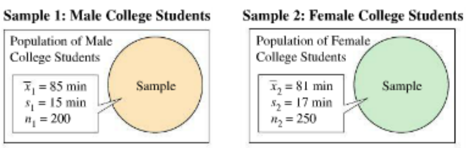


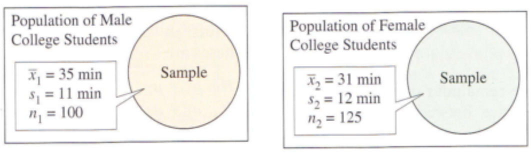


Example 1 Independent and dependent
Classify each pair of samples as independent or dependent justify your answer
1. Sample 1: Resting heart rates of 35 people before drinking coffee
   Sample 2: Resting heart rates of the same people after drinking two cups

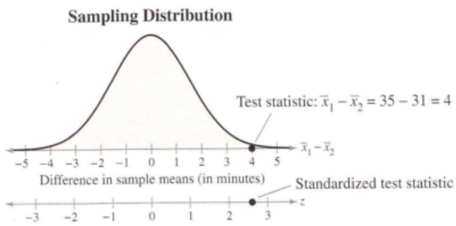
2. Sample 1: Test scores for 35 stats students
   Sample 2: Test scores for 42 biology students who do not take stats

An overview of Two-Sample hypothesis Testing
Test the CLAIM comparing the means of two different populations using independent samples.
Marketing plan for an Internet service provider and want to determine whether there is a difference in the amount of time male and female college students spend online each day WHY??
YOU could take a census but that is not practical sooooooo we can assume that there is no difference in the mean populations $\mu_1 - \mu_2 = 0$
then by taking a random sample from each population, and using the resulting two-sample $\overline{x}_1 = \overline{x}_2$
you can perform a two-sample hypothesis test





The graph below shows the sampling distribution of $\bar{x}_1 - \bar{x}_2$ for many similar samples taken from each population, under the assumption that $\mu_1 - \mu_2 = 0$. From the graph, you can see that it is quite unlikely to obtain sample means that differ by 4 minutes if the actual difference is 0. The difference of the sample means would be more than 2.5 standard errors from the hypothesized difference of 0! So, you can conclude that there is a significant difference in the amount of time male college students and female college students spend online each day.



It is important to remember that when you perform a two-sampled hypothesis test using independent samples, you are testing a claim concerning the difference between the parameters in two populations, NOT the values of the parameters themselves

DEFINITION

For a two-sample hypothesis test with independent samples,

1. the **null hypothesis** H_0 is a statistical hypothesis that usually states there is no difference between the parameters of two populations. The null hypothesis always contains the symbol $\leq$, $=$, or $\geq$.
2. the **alternative hypothesis** H_a is a statistical hypothesis that is true when H_0 is false. The alternative hypothesis contains the symbol $>$, $\neq$, or $<$.

$$\begin{cases} H_0\colon \mu_1 = \mu_2 \\ H_a\colon \mu_1 \neq \mu_2 \end{cases}, \quad \begin{cases} H_0\colon \mu_1 \leq \mu_2 \\ H_a\colon \mu_1 > \mu_2 \end{cases}, \quad \text{and} \quad \begin{cases} H_0\colon \mu_1 \geq \mu_2 \\ H_a\colon \mu_1 < \mu_2 \end{cases}.$$

Regardless of which hypothesis you use, you always assume there is no difference between the population means. $\mu_1 - \mu_2 = 0 \, or \, \mu_1 = \mu_2$

8-1 Continued

Two-Sample z-test for the difference between means

WE will now perform a z-test for the difference between two population means

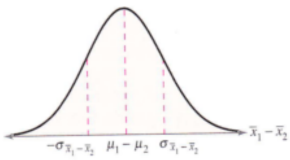
These conditions MUST BE MET

1. Both of the population Standard deviations must be known σ_1 σ_2
2. The sample must be randomly selected
3. The sample must be independent
4. Each sample size must be at least 30 or, if not, each population must have a normal distribution with a known standard deviation.

IF THOSE conditions are met then the sampling distribution for $\bar{x}_1 - \bar{x}_2$ (the difference of the sample means) is a normal distribution with mean and standard error as follows

In Words	In Symbols
The mean of the difference of the sample means is the assumed difference between the two population means. When no difference is assumed, the mean is 0.	Mean $= \mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$
The variance of the sampling distribution is the sum of the variances of the individual sampling distributions for $\bar{x}_1$ and $\bar{x}_2$. The standard error is the square root of the sum of the variances.	Standard error $= \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Sampling Distribution for $\bar{x}_1 - \bar{x}_2$



$$z = \frac{(\text{Observed difference}) - (\text{Hypothesized difference})}{\text{Standard error}}.$$

TWO-SAMPLE z-TEST FOR THE DIFFERENCE BETWEEN MEANS

A **two-sample z-test** can be used to test the difference between two population means μ_1 and μ_2 when these conditions are met.

1. Both σ_1 and σ_2 are known.
2. The samples are random.
3. The samples are independent.
4. The populations are normally distributed *or* both $n_1 \geq 30$ and $n_2 \geq 30$.

The **test statistic** is $\bar{x}_1 - \bar{x}_2$. The **standardized test statistic** is

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} \quad \text{where} \quad \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

GUIDELINES

Using a Two-Sample z-Test for the Difference Between Means (Independent Samples, σ_1 and σ_2 Known)

IN WORDS	IN SYMBOLS
1. Verify that σ_1 and σ_2 are known, the samples are random and independent, and either the populations are normally distributed *or* both $n_1 \geq 30$ and $n_2 \geq 30$.	
2. State the claim mathematically and verbally. Identify the null and alternative hypotheses.	State H_0 and H_a.
3. Specify the level of significance.	Identify α.
4. Determine the critical value(s).	Use Table 4 in Appendix B.
5. Determine the rejection region(s).	
6. Find the standardized test statistic and sketch the sampling distribution.	$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$
7. Make a decision to reject or fail to reject the null hypothesis.	If z is in the rejection region, then reject H_0. Otherwise, fail to reject H_0.
8. Interpret the decision in the context of the original claim.	

Example 2 Two-Sample z-test for the difference between means

A consumer education organization claims that there is a difference in the mean credit card debt of males and females in the US. The results of a random survey of 200 individuals from each group are shown. The two samples are independent. Do the results support the organizations claim? Use $\alpha = 0.05$

Sample Statistics for Credit Card Debt

Females	Males
$\bar{x}_1 = \$2290$	$\bar{x}_2 = \$2370$
$s_1 = \$750$	$s_2 = \$800$
$n_1 = 200$	$n_2 = 200$

```
2-SampZTest
 μ1≠μ2
 z=-1.031721408
 p=.3022026809
 x̄1=2290
 x̄2=2370
↓n1=200
```

Example 3 USING THE CALCULATOR for two-sample z-test

The American Automobile Association claims that the average daily cost for meals and lodging for vacationing in Texas is less than the same average costs for vacationing in Virginia. The table shows the results of a random survey of vacationers in each state. the two samples are independent. At $\alpha = 0.01$, is there enough evidence to support the claim?

$$H_0 : \mu_1 \geq \mu_2$$

$$H_a : \mu_1 < \mu_2 \; Claim$$

Sample Statistics for Daily Cost of Meals and Lodging for a Family of Four

Texas	Virginia
$\bar{x}_1 = \$248$	$\bar{x}_2 = \$252$
$s_1 = \$15$	$s_2 = \$22$
$n_1 = 50$	$n_2 = 35$

TI-83/84

2-SampZTest
Inpt:Data Stats
σ1:15
σ2:22
x1:248
n1:50
x2:252
↓n2:35

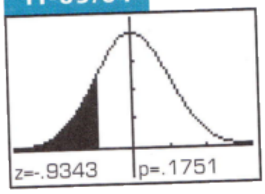
TI-83/84

2-SampZTest
↑σ2:22
x1:248
n1:50
x2:252
n2:35
μ1:≠μ2 <μ2 >μ2
Calculate Draw

TI-83/84

2-SampZTest
μ1<μ2
z= -.9343200811
p=.1750693839
x1:248
x2:252
↓n1:50

TI-83/84



EXAMPLE 3

Using Technology to Perform a Two-Sample *z*-Test

A travel agency claims that the average daily cost of meals and lodging for vacationing in Texas is less than the average daily cost in Virginia. The table at the left shows the results of a random survey of vacationers in each state. The two samples are independent. Assume that $\sigma_1 = \$19$ for Texas and $\sigma_2 = \$24$ for Virginia, and that both populations are normally distributed. At $\alpha = 0.01$, is there enough evidence to support the claim? [H_0: $\mu_1 \geq \mu_2$ and H_a: $\mu_1 < \mu_2$ (claim)]

Solution Note that σ_1 and σ_2 are known, the samples are random and independent, and the populations are normally distributed. So, you can use the z-test. The top two displays show how to set up the hypothesis test using a TI-84 Plus. The remaining displays show the results of selecting *Calculate* or *Draw*.

Sample Statistics for Daily Cost of Meals and Lodging for Two Adults

Texas	Virginia
$\bar{x}_1 = \$234$	$\bar{x}_2 = \$240$
$n_1 = 25$	$n_2 = 20$

NORMAL FLOAT AUTO REAL DEGREE MP
PRESS [<] OR [>] TO SELECT AN OPTION
2-SampZTest
↑σ1:19
σ2:24
x̄1:234
n1:25
x̄2:240
n2:20
μ1:≠μ2 <μ2 >μ2
Color: BLUE
Calculate Draw

NORMAL FLOAT AUTO REAL DEGREE MP
2-SampZTest
μ1<μ2
z=-0.912448597
p=0.1807662795
x̄1=234
x̄2=240
n1=25
n2=20

8.1 pg. 424 9,10,11,12,14,15,17,19,20,24

Lesson 8-2 Testing the difference between means When the population SD is not known

t-test to test the difference between two populations. The following conditions are necessary to use a t-test for SMALL independent samples

1. The population Standard deviation is NOT KNOWN *σ is NOT Known*

2. The samples must be randomly selected

3. The samples must be independent. Recall that two samples are independent if the samples selected from one population is not related to the sample selected from another

4. Each population must have a normal distribution or the same is at least 30

TWO-SAMPLE *t*-TEST FOR THE DIFFERENCE BETWEEN MEANS

A **two-sample *t*-test** is used to test the difference between two population means μ_1 and μ_2 when (1) σ_1 and σ_2 are unknown, (2) the samples are random, (3) the samples are independent, and (4) the populations are normally distributed *or* both $n_1 \geq 30$ and $n_2 \geq 30$. The **test statistic** is $\bar{x}_1 - \bar{x}_2$, and the **standardized test statistic** is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}.$$

Variances are equal: If the population variances are equal, then information from the two samples is combined to calculate a **pooled estimate of the standard deviation $\hat{\sigma}$.**

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

The standard error for the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is

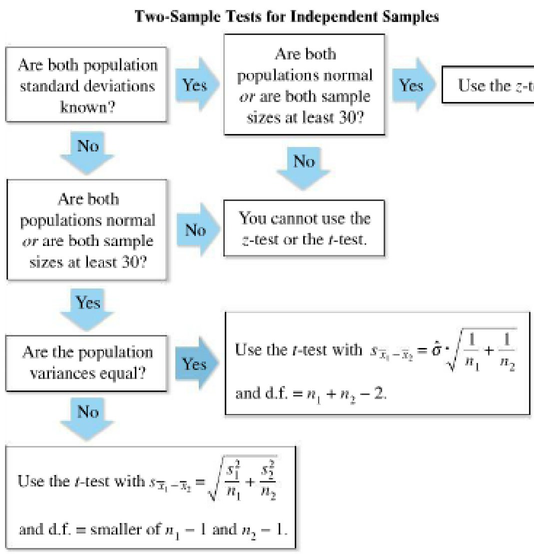
$$s_{\bar{x}_1 - \bar{x}_2} = \hat{\sigma} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad \text{Variances equal}$$

and d.f. $= n_1 + n_2 - 2$.

Variances are not equal: If the population variances are not equal, then the standard error is

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad \text{Variances not equal}$$

and d.f. = smaller of $n_1 - 1$ and $n_2 - 1$.



GUIDELINES

Using a Two-Sample *t*-Test for the Difference Between Means (Independent Samples, σ_1 and σ_2 Unknown)

IN WORDS	IN SYMBOLS
1. Verify that σ_1 and σ_2 are unknown, the samples are random and independent, and either the populations are normally distributed *or* both $n_1 \geq 30$ and $n_2 \geq 30$.	
2. State the claim mathematically and verbally. Identify the null and alternative hypotheses.	State H_0 and H_a.
3. Specify the level of significance.	Identify α.
4. Determine the degrees of freedom.	d.f. $= n_1 + n_2 - 2$ or d.f. = smaller of $n_1 - 1$ and $n_2 - 1$
5. Determine the critical value(s).	Use Table 5 in Appendix B.
6. Determine the rejection region(s).	
7. Find the standardized test statistic and sketch the sampling distribution.	$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}$
8. Make a decision to reject or fail to reject the null hypothesis.	If t is in the rejection region, then reject H_0. Otherwise, fail to reject H_0.
9. Interpret the decision in the context of the original claim.	

EXAMPLE 1

See Minitab steps on page 464.

A Two-Sample *t*-Test for the Difference Between Means

The results of a state mathematics test for random samples of students taught by two different teachers at the same school are shown at the left. Can you conclude that there is a difference in the mean mathematics test scores for the students of the two teachers? Use $\alpha = 0.10$. Assume the populations are normally distributed and the population variances are not equal.

Sample Statistics for State Mathematics Test Scores

Teacher 1	Teacher 2
$\bar{x}_1 = 473$	$\bar{x}_2 = 459$
$s_1 = 39.7$	$s_2 = 24.5$
$n_1 = 8$	$n_2 = 18$

8-2 Continued

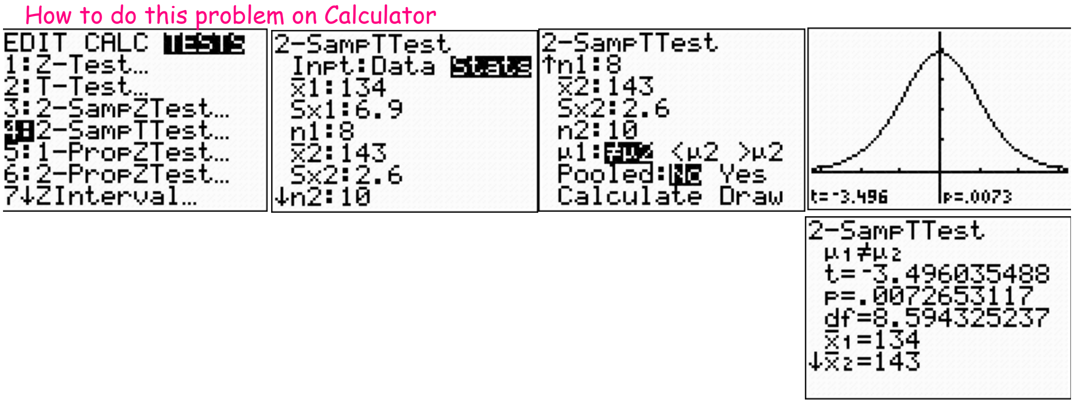
Example 1 A two-sampled t-test for the difference between Means

The braking distancs of 8 volkswagen GTI's and 10 Ford Focuses were tested when traveling at 60mph on dry pavement. The results are shown at the left. Can you conclude that there is a difference in the mean braking distances of the two types of cars? Use $\alpha = 0.01$ .Assume the populations are normally distributed and the population variances are not equal

Sample Statistics for Braking Distance on Dry Pavement

GTI	Focus
$\bar{x}_1 = 134$ ft	$\bar{x}_2 = 143$ ft
$s_1 = 6.9$ ft	$s_2 = 2.6$ ft
$n_1 = 8$	$n_2 = 10$

How to do this problem on Calculator

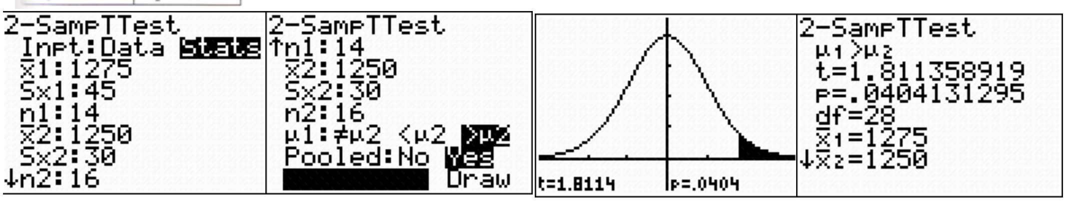


Example 2 A two-sample t-test for the difference between means

A manufacturer claims that the calling range (in feet) of its 2.4 GHz cordless telephone is greater than that of its leading competitor. You perform a study using 14 randomly selected phones from the manufacturer and 16 random selected phones from its competitor. The results are shown. At $\alpha = 0.05$ can you support the manufacturer's claim? Assume normal and variances are equal

Sample Statistics for Calling Range

Manufacturer	Competition
$\bar{x}_1 = 1275$ ft	$\bar{x}_2 = 1250$ ft
$s_1 = 45$ ft	$s_2 = 30$ ft
$n_1 = 14$	$n_2 = 16$



pg. 432 9,12,13,14,17,18,21

Lesson 8-3

Testing the difference between Means (DEPENDENT SAMPLES)

8-1 and 8-2 was testing the difference between two means of independent samples so the test statistic was $\bar{x}_1 - \bar{x}_2$

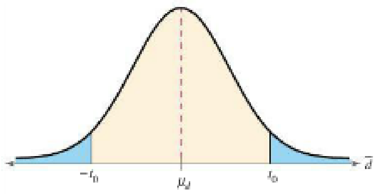
today we are performing a two-sample hypothesis test with dependent samples. So we first must find the difference d for each data pair $d = x_1 - x_2$

The test statistic is the mean $\bar{d} = \frac{\sum d}{n}$ of these differences

The following conditions are required to conduct the test

1. The samples must be randomly selected
2. The samples must be dependent (PAIRED)
3. Both populations must be normally distributed n at least 30

When these conditions are met, the **sampling distribution for $\bar{d}$, the mean of the differences of the paired data entries in the dependent samples,** is approximated by a t-distribution with $n - 1$ degrees of freedom, where n is the number of data pairs.



The symbols listed in the table are used for the t-test for μ_d. Although formulas are given for the mean and standard deviation of differences, you should use technology to calculate these statistics.

Symbol	Description
n	The number of pairs of data
d	The difference between entries in a data pair
μ_d	The hypothesized mean of the differences of paired data in the population
$\bar{d}$	The mean of the differences between the paired data entries in the dependent samples $\bar{d} = \frac{\Sigma d}{n}$
s_d	The standard deviation of the differences between the paired data entries in the dependent samples $s_d = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n - 1}}$

t-TEST FOR THE DIFFERENCE BETWEEN ME

A t-test can be used to test the difference of two population means whe conditions are met.

1. The samples are random.
2. The samples are dependent (paired).
3. The populations are normally distributed *or* $n \geq 30$.

The **test statistic** is

$$\bar{d} = \frac{\Sigma d}{n}$$

and the **standardized test statistic** is

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}.$$

The degrees of freedom are

d.f. $= n - 1$.

GUIDELINES

Using the t-Test for the Difference Between Means (Dependent Samples)

IN WORDS	IN SYMBOLS
1. Verify that the samples are random and dependent, and either the populations are normally distributed *or* $n \geq 30$.	
2. State the claim mathematically and verbally. Identify the null and alternative hypotheses.	State H_0 and H_a.
3. Specify the level of significance.	Identify α.
4. Identify the degrees of freedom.	d.f. $= n - 1$
5. Determine the critical value(s).	Use Table 5 in Appendix B.
6. Determine the rejection region(s).	
7. Calculate $\bar{d}$ and s_d.	$\bar{d} = \frac{\Sigma d}{n}$ $s_d = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n - 1}}$
8. Find the standardized test statistic and sketch the sampling distribution.	$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$
9. Make a decision to reject or fail to reject the null hypothesis.	If t is in the rejection region, then reject H_0. Otherwise, fail to reject H_0.
10. Interpret the decision in the context of the original claim.	

8.3 Continued

EXAMPLE 1

See Minitab steps on page 464.

The *t*-Test for the Difference Between Means

A shoe manufacturer claims that athletes can increase their vertical jump heights using the manufacturer's training shoes. The vertical jump heights of eight randomly selected athletes are measured. After the athletes have used the shoes for 8 months, their vertical jump heights are measured again. The vertical jump heights (in inches) for each athlete are shown in the table. At $\alpha = 0.10$, is there enough evidence to support the manufacturer's claim? Assume the vertical jump heights are normally distributed.

Athlete	1	2	3	4	5	6	7	8
Vertical jump height (before using shoes)	24	22	25	28	35	32	30	27
Vertical jump height (after using shoes)	26	25	25	29	33	34	35	30

Before	After	d	d^2
24	26	−2	4
22	25	−3	9
25	25	0	0
28	29	−1	1
35	33	2	4
32	34	−2	4
30	35	−5	25
27	30	−3	9
		$\Sigma = -14$	$\Sigma = 56$

EXAMPLE 2

The *t*-Test for the Difference Between Means

The campaign staff for a state legislator wants to determine whether the legislator's performance rating (0–100) has changed from last year to this year. The table below shows the legislator's performance ratings from the same 16 randomly selected voters for last year and this year. At $\alpha = 0.01$, is there enough evidence to conclude that the legislator's performance rating has changed? Assume the performance ratings are normally distributed.

Voter	1	2	3	4	5	6	7	8
Rating (last year)	60	54	78	84	91	25	50	65
Rating (this year)	56	48	70	60	85	40	40	55

Voter	9	10	11	12	13	14	15	16
Rating (last year)	68	81	75	45	62	79	58	63
Rating (this year)	80	75	78	50	50	85	53	60

Example 1 The t-test for the difference between means

A Golf club manufacturer claims that golfers can lower their score by using their newly designed golf clubs. Eight golfers are randomly slected and each is asked to give his or her most recent score. After using the new clubs for one month, the golfers are shown at the table. Assuming the golf scores are normally distributed, is there enough evidence to support the claim at

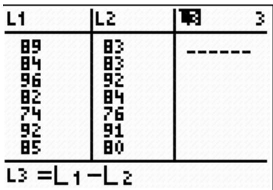
Golfer	1	2	3	4	5	6	7	8
Score (old design)	89	84	96	82	74	92	85	91
Score (new design)	83	83	92	84	76	91	80	91

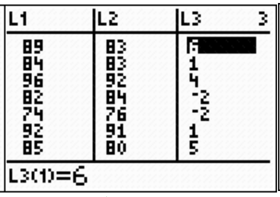
$H_0: \mu_d \le 0$ reject

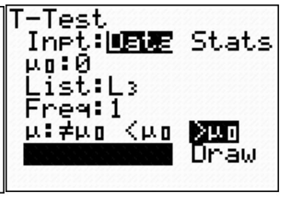
$H_a: \mu_d > 0$ claim

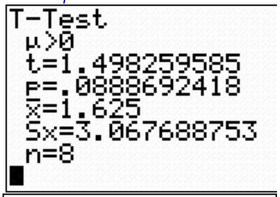
$d = (\text{old} - \text{New})$

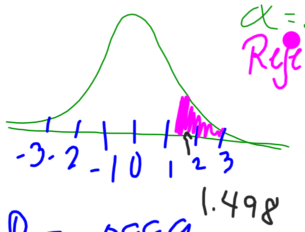
$L3 = L1 - L2$





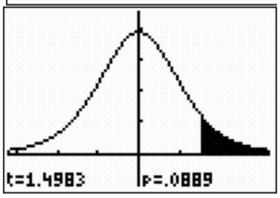






$\alpha = .10$ d.f. $= 8-1=7$

Reject $t > 1.415$



Reject Null at 1% Significance there is enough evidence to Support the claim that the golfers would lower their score by using the new clubs

$P = .0889$

$\alpha = .1$ $\quad p < \alpha$

Reject the Null

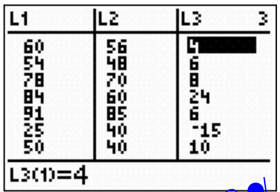
EXAMPLE 2

The t-Test for the Difference Between Means

A state legislator wants to determine whether her performance rating (0–100) has changed from last year to this year. The following table shows the legislator's performance rating from the same 16 randomly selected voters for last year and this year. At $\alpha = 0.01$, is there enough evidence to conclude that the legislator's performance rating has changed? Assume the performance ratings are normally distributed.

Voter	1	2	3	4	5	6	7	8	
Rating (last year)	60	54	78	84	91	25	50	65	L1
Rating (this year)	56	48	70	60	85	40	40	55	L2

Voter	9	10	11	12	13	14	15	16	
Rating (last year)	68	81	75	45	62	79	58	63	L1
Rating (this year)	80	75	78	50	50	85	53	60	L2



$L3 = L1 - L2$

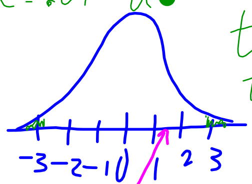
$H_0: \mu_d = 0$ failed to reject

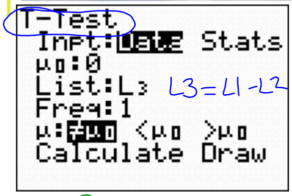
$H_a: \mu_d \ne 0$ Claim

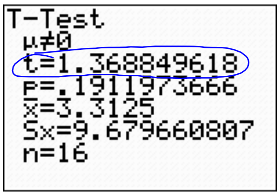
$\alpha = .01$ $\quad$ d.f. $= 16-1 = 15$

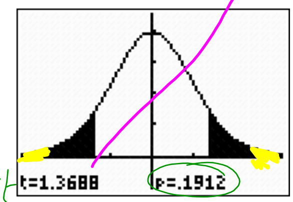
$t = -2.947$

$t = 2.947$









fail to reject 1% Significance level there is not enough evidence to Support the claim that there is No difference In her rating.

$P = .1912$ $\quad \alpha = .01$

$P > \alpha$

fail to reject

Assign: pg. 442 10,12 PICK one

14-20 ONLY DUE 1

Lesson 8-4 Testing the difference between proportions

we will learn how to use a z-test to test the difference between two population proportions

If the claim is about two population parameters p_1 *and* p_2 the some possible pairs of null and alterative hypothesis are

$$\begin{cases} H_0: p_1 = p_2 \\ H_a: p_1 \neq p_2 \end{cases}, \quad \begin{cases} H_0: p_1 \leq p_2 \\ H_a: p_1 > p_2 \end{cases}, \quad \text{and} \quad \begin{cases} H_0: p_1 \geq p_2 \\ H_a: p_1 < p_2 \end{cases}.$$

Regardless of which hypotheses you use always assume there is no difference between the population proportions $p_1 = p_2$

1. The samples must be randomly selected.
2. The samples must be independent.
3. The samples must be large enough to use a normal sampling distribution. That is, $n_1 p_1 \geq 5$, $n_1 q_1 \geq 5$, $n_2 p_2 \geq 5$, and $n_2 q_2 \geq 5$.

If these conditions are met, then the **sampling distribution for $\hat{p}_1 - \hat{p}_2$, the difference between the sample proportions,** is a normal distribution with mean

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$$

and standard error

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}.$$

Notice that you need to know the population proportions to calculate the standard error. Because a hypothesis test for $p_1 - p_2$ is based on the assumption that $p_1 = p_2$, you can calculate a weighted estimate of p_1 and p_2 using

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2}, \text{ where } x_1 = n_1 \hat{p}_1 \text{ and } x_2 = n_2 \hat{p}_2.$$

With the weighted estimate $\overline{p}$, the standard error of the sampling distribution for $\hat{p}_1 - \hat{p}_2$ is

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \text{ where } \overline{q} = 1 - \overline{p}.$$

Also observe that you need to know the population proportions in verifying that the samples are large enough to be approximated by the normal distribution. But when determining whether the z-test can be used for the difference between proportions for a binomial experiment, you should use $\overline{p}$ in place of p_1 and p_2 and use $\overline{q}$ in place of q_1 and q_2.

TWO-SAMPLE z-TEST FOR THE DIFFERENCE BETWEEN PROPORTIONS

A two-sample z-test is used to test the difference between two population proportions p_1 and p_2 when these conditions are met.

1. The samples are random.
2. The samples are independent.
3. The quantities $n_1\overline{p}$, $n_1\overline{q}$, $n_2\overline{p}$, and $n_2\overline{q}$ are at least 5.

The **test statistic** is $\hat{p}_1 - \hat{p}_2$. The **standardized test statistic** is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\overline{p} = \frac{x_1 + x_2}{n_1 + n_2}$ and $\overline{q} = 1 - \overline{p}$.

If the null hypothesis states $p_1 = p_2$, $p_1 \leq p_2$, or $p_1 \geq p_2$, then $p_1 = p_2$ is assumed and the expression $p_1 - p_2$ is equal to 0 in the preceding test.

GUIDELINES

Using a Two-Sample z-Test for the Difference Between Proportions

IN WORDS	IN SYMBOLS
1. Verify that the samples are random and independent.	
2. Find the weighted estimate of p_1 and p_2. Verify that $n_1\overline{p}$, $n_1\overline{q}$, $n_2\overline{p}$, and $n_2\overline{q}$ are at least 5.	$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2}$, $\overline{q} = 1 - \overline{p}$
3. State the claim mathematically and verbally. Identify the null and alternative hypotheses.	State H_0 and H_a.
4. Specify the level of significance.	Identify α.
5. Determine the critical value(s).	Use Table 4 in Appendix B
6. Determine the rejection region(s).	
7. Find the standardized test statistic and sketch the sampling distribution.	$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
8. Make a decision to reject or fail to reject the null hypothesis.	If z is in the rejection region, then reject H_0. Otherwise, fail to reject H_0.
9. Interpret the decision in the context of the original claim.	

To use P-values in a test for the difference between proportions, use the same guidelines as above, skipping Steps 3 and 4. After finding the standardized test statistic, use the Standard Normal Table to calculate the P-value. Then make a decision to reject or fail to reject the null hypothesis. If P is less than or equal to α, reject H_0. Otherwise fail to reject H_0.

Lesson 8-4 Continued

EXAMPLE 1

See TI-84 Plus steps on page 465.

A Two-Sample *z*-Test for the Difference Between Proportions

A study of 150 randomly selected occupants in passenger cars and 200 randomly selected occupants in pickup trucks shows that 86% of occupants in passenger cars and 74% of occupants in pickup trucks wear seat belts. At $\alpha = 0.10$, can you reject the claim that the proportion of occupants who wear seat belts is the same for passenger cars and pickup trucks?

Study Tip

To find x_1 and x_2, use $x_1 = n_1\hat{p}_1$ and $x_2 = n_2\hat{p}_2$.

Sample Statistics for Vehicles

Passenger cars	Pickup trucks
$n_1 = 150$	$n_2 = 200$
$\hat{p}_1 = 0.86$	$\hat{p}_2 = 0.74$
$x_1 = 129$	$x_2 = 148$

```
NORMAL FLOAT AUTO REAL DEGREE MP
2-PropZTest
x1:129
n1:150
x2:148
n2:200
p1:≠p2 <p2 >p2
Color: BLUE
Calculate Draw
```

```
NORMAL FLOAT AUTO REAL DEGREE MP
2-PropZTest
p1≠p2
z=2.734478928
p=0.0062480166
p̂1=0.86
p̂2=0.74
p̂=0.7914285714
n1=150
n2=200
```

EXAMPLE 2

A Two-Sample *z*-Test for the Difference Between Proportions

A medical research team conducted a study to test the effect of a cholesterol-reducing medication. At the end of the study, the researchers found that of the 4700 randomly selected subjects who took the medication, 301 died of heart disease. Of the 4300 randomly selected subjects who took a placebo, 357 died of heart disease. At $\alpha = 0.01$, can you support the claim that the death rate due to heart disease is lower for those who took the medication than for those who took the placebo?

Sample Statistics for Cholesterol-Reducing Medication

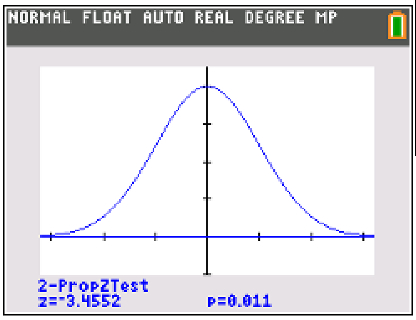
Received medication	Received placebo
$n_1 = 4700$	$n_2 = 4300$
$x_1 = 301$	$x_2 = 357$
$\hat{p}_1 \approx 0.0640$	$\hat{p}_2 \approx 0.0830$

Study Tip

To find $\hat{p}_1$ and $\hat{p}_2$ use $\hat{p}_1 = \frac{x_1}{n_1}$ and $\hat{p}_2 = \frac{x_2}{n_2}$.

```
NORMAL FLOAT AUTO REAL DEGREE MP
2-PropZTest
x1:301
n1:4700
x2:357
n2:4300
p1:≠p2 <p2 >p2
Color: BLUE
Calculate Draw
```

```
NORMAL FLOAT AUTO REAL DEGREE MP
2-PropZTest
p1<p2
z=-3.455163927
p=2.750267359E-4
p̂1=0.0640425532
p̂2=0.0830232558
p̂=0.0731111111
n1=4700
n2=4300
```



EXAMPLE 1

A Two-Sample z-Test for the Difference Between Proportions

See TI-83/84 steps on page 489.

In a study of 200 randomly selected adult female and 250 randomly selected adult male Internet users, 30% of the females and 38% of the males said that they plan to shop online at least once during the next month. At $\alpha = 0.10$, test the claim that there is a difference between the proportion of female and the proportion of male Internet users who plan to shop online.

Sample Statistics for Internet Users

Females	Males
$n_1 = 200$	$n_2 = 250$
$\hat{p}_1 = 0.30$	$\hat{p}_2 = 0.38$
$n_1\hat{p}_1 = 60$	$n_2\hat{p}_2 = 95$

Study Tip

To find x_1 and x_2, use $x_1 = n_1\hat{p}_1$ and $x_2 = n_2\hat{p}_2$.

$H_0: P_1 = P_2$ Reject

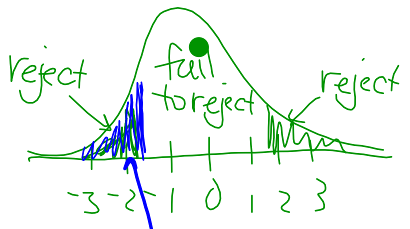
$H_a: P_1 \neq P_2$ Claim

two-tailed $\alpha = .1$

$\text{Invnorm}(\frac{.10}{2}) = -1.645$

$X_1 = 200(.30) = 60$ $200(.70) = 140$

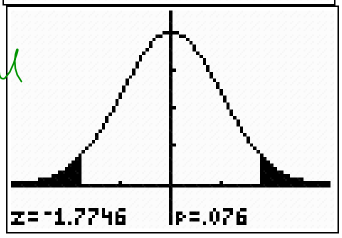
$X_2 = 250(.38) = 95$ $250(.62) =$



$z = -1.77$ Reject

```
2-PropZTest
 x1:60
 n1:200
 x2:95
 n2:250
 p1:≠p2 <p2 >p2
 Calculate Draw
```

```
2-PropZTest
 p1≠p2
 z=-1.774615984
 p=.0759612188
 p̂1=.3
 p̂2=.38
↓p̂=.3444444444
```



at 1% Significance level there is enough evidence to reject Null Support the claim. There is a difference between proportions of male & female Internet Shoppers.

EXAMPLE 2

A Two-Sample z-Test for the Difference Between Proportions

A medical research team conducted a study to test the effect of a cholesterol-reducing medication. At the end of the study, the researchers found that of the 4700 randomly selected subjects who took the medication, 301 died of heart disease. Of the 4300 randomly selected subjects who took a placebo, 357 died of heart disease. At $\alpha = 0.01$, can you conclude that the death rate due to heart disease is lower for those who took the medication than for those who took the placebo? *(Adapted from New England Journal of Medicine)*

Sample Statistics for Cholesterol-Reducing Medication

Received medication	Received placebo
$n_1 = 4700$	$n_2 = 4300$
$x_1 = 301$	$x_2 = 357$
$\hat{p}_1 = 0.064$	$\hat{p}_2 = 0.083$

Study Tip

To find $\hat{p}_1$ and $\hat{p}_2$ use

$\hat{p}_1 = \frac{x_1}{n_1}$ and

$\hat{p}_2 = \frac{x_2}{n_2}$.

$\hat{P}_1 = \frac{301}{4700}$

$\hat{P}_2 = \frac{357}{4300}$

$H_0: P_1 \geq P_2$

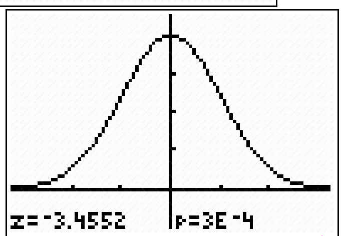
$H_a: P_1 < P_2$ (Claim)

$\alpha = .01$ left-tailed

$\text{Invnorm}(.01) = -2.326$

```
2-PropZTest
 x1:301
 n1:4700
 x2:357
 n2:4300
 p1:≠p2 <p2 >p2
 Calculate Draw
```

```
2-PropZTest
 p1<p2
 z=-3.455163927
 p=2.7502674E-4
 p̂1=.0640425532
 p̂2=.0830232558
↓p̂=.0731111111
```



Reject the Null at 1% Significance there is enough evidence to Support the Claim that the death rate is lower for the people that took the Medo vs the people who took the placebo

Assign: pg. 451 7,9 OR 10,19,20

Attachments

8-1 book.pdf

8-2 book.pdf

8-3 Book.pdf

8-4 book.pdf

t distribution table.pdf

HomeworkCh8_Larson&Farber.pdf