Lecture Four

Section 4.1 – Inferences about Two Population Proportions

Objectives

Test a claim two population proportions or construct a confidence interval estimate of the difference between two population properties.

Distinguish between Independent and Dependent Sampling

A sampling method is ***independent*** when the individuals selected for one sample do not dictate which individuals are to be in a second sample. A sampling method is ***dependent*** when the individuals selected to be in one sample are used to determine the individuals to be in the second sample.

Dependent samples are often referred to as matched-pairs samples. It is possible for an individual to be matched against him- or herself.

Example

For each of the following, determine whether the sampling method is independent or dependent.

A researcher wants to know whether the price of a one night stay at a Holiday Inn Express is less than the price of a one night stay at a Red Roof Inn. She randomly selects 8 towns where the location of the hotels is close to each other and determines the price of a one night stay.

> ***The sampling method is dependent since the 8 Holiday Inn Express hotels can be matched with one of the 8 Red Roof Inn hotels by town***

A researcher wants to know whether the "state" quarters (introduced in 1999) have a mean weight that is different from "traditional" quarters. He randomly selects 18 "state" quarters and 16 "traditional" quarters and compares their weights.

> ***The sampling method is independent since the "state" quarters which were sampled had no bearing on which "traditional" quarters were sampled.***

Notation for Two Proportions

For population 1, we let:

p_1 = population proportion

n_1 = size of the sample

x_1 = number of successes in the sample (the sample proportion)

The corresponding notations apply to which come from population 2.

Pooled Sample Proportion

The pooled sample proportion is denoted by p and is given by:

$$\bar{p} = \frac{x_1 + n_2}{n_1 + n_2} \qquad \bar{q} = 1 - \bar{p}$$

- We denote the complement of p by q, **so q = 1 – p**

Requirements

We have proportions from two independent simple random samples.
For each of the two samples, the number of successes is at least 5 and the number of failures is at least 5.

Test Statistic for Two Proportions

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$ where $p_1 - p_2 = 0$ (assumed in the null hypothesis)

$$\hat{p}_1 = \frac{x_1}{n_1} \quad and \quad \hat{p}_2 = \frac{x_2}{n_2}$$ (sample proportions)

$$\bar{p}_1 = \frac{x_1 + x_2}{n_1 + n_2}$$ (pooled sample proportion)

$$\bar{q} = 1 - \bar{p}$$

P-value: Use Standard Normal Distribution Table. (Use the computed value of the test statistic z and find its P-value)

Critical values: Use Standard Normal Distribution Table. (Based on the significance level a, find critical values.)

Sampling Distribution of the Difference between Two Proportions

Use the positive and negative values of z (for two tails) and solve for $p_1 - p_2$. The results are the limits of the confidence interval given earlier.

The difference can be approximated by a normal distribution with mean $p_1 - p_2$ and variance

$$\sigma^2_{(\hat{p}_1 - \hat{p}_2)} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

The variance of the *differences* between two independent random variables is the *sum* of their individual variances.

The preceding variance leads to $\sigma_{\left(\hat{p}_1-\hat{p}_2\right)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

Provided that $n_1 \hat{p}_1 \hat{q}_1 \geq 10; \quad n_2 \hat{p}_2 \hat{q}_2 \geq 10$ and each sample size no more than 5% pf the population size.

When constructing the confidence interval estimate of the difference between two proportions, we don't assume that the two proportions are equal, and we estimate the standard deviation as

$$\sigma = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

In the test statistic $z = \dfrac{\left(\hat{p}_1 - \hat{p}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$

The best point estimate of p is called the ***pooled estimate*** of p, denoted $\hat{p}$, where

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Test statistic for Comparing Two Population Proportions

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sigma_{\hat{p}_1 - \hat{p}_2}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

McNemar's Test

	Success	**Failure**
Success	f_{11}	f_{12}
Failure	f_{21}	f_{22}

$$z_0 = \frac{\left|f_{12} - f_{21}\right| - 1}{\sqrt{f_{12} + f_{21}}}$$

Hypothesis Test Regarding the Difference between Two Population Proportions

To test hypotheses regarding two population proportion:

- ✓ The samples are independently obtained using simple random sampling
- ✓ $n_1 \hat{p}_1 \hat{q}_1 \geq 10; \quad n_2 \hat{p}_2 \hat{q}_2 \geq 10$
- ✓ $n_1 \leq 0.05 N_1 \quad and \quad n_2 \leq 0.05 N_2$ (the sample size is more than 5% of the population size); this requirement ensures the independence necessary for a binomial experiment.

***Step* 1**: Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways:

Two–Tailed	***Left–Tailed***	***Right–Tailed***
$H_0 : p_1 = p_2$	$H_0 : p_1 = p_2$	$H_0 : p_1 = p_2$
$H_1 : p_1 \neq p_2$	$H_1 : p_1 < p_2$	$H_1 : p_1 > p_2$

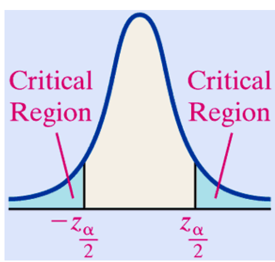
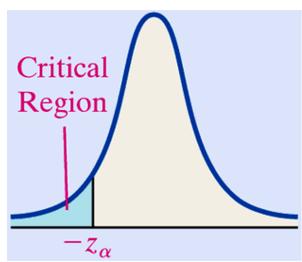
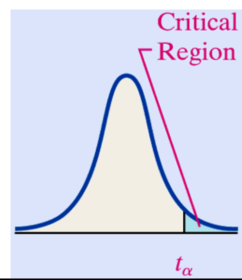
Note: p_1 is the population proportion for population 1, and p_2 is the population proportion for population 2.

***Step* 2:** Select a level of significance, α, based on the seriousness of making a Type I error.

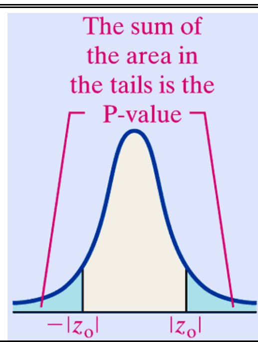
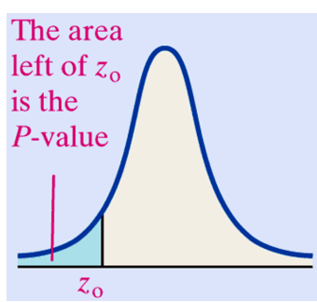
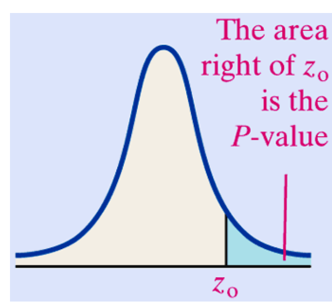
***Step* 3**: Compute the test statistic

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad where \quad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Classical Approach

Two–Tailed	***Left–Tailed***	***Right–Tailed***
$z_0 < -z_{\alpha/2}$ *or* $z_0 > z_{\alpha/2}$ *Reject the null hypothesis*	$z_0 < -z_\alpha$ *Reject the null hypothesis*	$z_0 > z_\alpha$ *Reject the null hypothesis*
		

***Step* 3**: Estimate the *P*–value

Two–Tailed	***Left–Tailed***	***Right–Tailed***
		

If ***P-value < α****, reject the null hypothesis*

Example

A recent General Social Survey asked the following two questions of a random sample of 1483 adult Americans under the hypothetical scenario that the government suspected that a terrorist act was about to happen:

Do you believe the authorities should have the right to tap people's telephone conversations?

Do you believe the authorities should have the right to detain people for as long as they want without putting them on trial?

The results of the survey are shown below

		Detain	
		Agree	**Disagree**
Tap	**Agree**	572	237
Phone	**Disagree**	224	450

Do the proportions who agree with each scenario differ significantly? Use the α = 0.05 level of significance.

Solution

The sample proportion of individuals who believe that the authorities should be able to tap phones is

$$\hat{p}_1 = \frac{572+237}{1483} = 0.5455$$

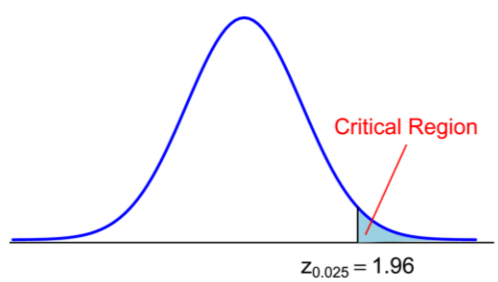
The sample proportion of individuals who believe that the authorities should have the right to detain people is

$$\hat{p}_2 = \frac{572+224}{1483} = 0.5367$$

We want to determine whether the difference in sample proportions is due to sampling error or to the fact that the population proportions differ.

The samples are dependent and were obtained randomly. The total number of individuals who agree with one scenario, but disagree with the other is 237 + 224 = 461, which is greater than 10. We can proceed with McNemar's Test.

***Step* 1**: The hypotheses are as follows $\begin{cases} H_0: & \hat{p}_1 = \hat{p}_2 \\ H_1: & \hat{p}_1 \neq \hat{p}_2 \end{cases}$



***Step* 2:** The level of significance is α = 0.05

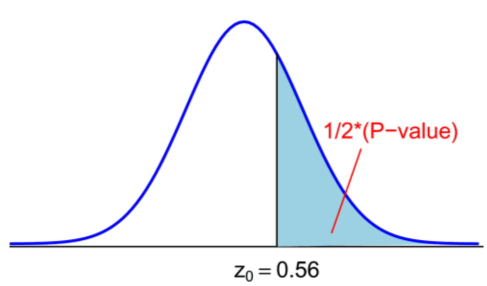
***Step* 3:** The test statistic is:

$$z_0 = \frac{|f_{12} - f_{21}| - 1}{\sqrt{f_{12} + f_{21}}} = \frac{|237-224|-1}{\sqrt{237+224}}$$

$$z_{0.025} = 1.96 > 0.56 = z_0$$

$\Rightarrow$ We fail to reject the null hypothesis

$$P-value = 2 \cdot P(z > 0.56) \approx 0.5754$$



Since the P-value is greater than the level of significance $\alpha = 0.05$, we fail to reject the null hypothesis.

Conclusion

There is insufficient evidence at the $\alpha = 0.05$ level to conclude that there is a difference in the proportion of adult Americans who believe it is okay to phone tap versus detaining people for as long as they want without putting them on trial in the event that the government believed a terrorist plot was about to happen.

Confidence Interval Estimate of $p_1 - p_2$

The confidence interval estimate of the difference $p_1 - p_2$ is:

$$\left(\hat{p}_1 - \hat{p}_2\right) - E < \left(p_1 - p_2\right) < \left(\hat{p}_1 - \hat{p}_2\right) + E$$

Where the margin of error E is given by $E = z_{\alpha/2}\sqrt{\dfrac{\hat{P}_1\hat{q}_1}{n_1} + \dfrac{\hat{P}_2\hat{q}_2}{n_2}}$

Rounding: Round the confidence interval limits to three significant digits,

Example

The table below lists results from a simple random sample of front-seat occupants involved in car crashes. Use a 0.05 significance level to test the claim that the fatality rate of occupants is lower for those in cars equipped with airbags.

	Airbag Available	No Airbag Available
Occupant Fatalities	41	52
Total number of occupants	11,541	9,853

Solution

Requirements are satisfied: two simple random samples, two samples are independent; Each has at least 5 successes and 5 failures (11,500, 41; 9801, 52).
Use the P-value method.

Step 1: Express the claim as $p_1 < p_2$.

Step 2: If $p_1 < p_2$ is false, then $p_1 \geq p_2$.

Step 3: $p_1 < p_2$ does not contain equality so it is the alternative hypothesis. The null hypothesis is the statement of equality.

$H_0: \quad p_1 = p_2$

$H_a: \quad p_1 < p_2$ (original claim)

Step 4: Significance level is 0.05

Step 5: Use normal distribution as an approximation to the binomial distribution. Estimate the common values of p_1 and p_2

Step 6: Find the value of the test statistic.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

$$= \frac{\left(\frac{41}{11{,}541} - \frac{52}{9{,}853}\right) - 0}{\sqrt{\frac{(0.004347)(0.995653)}{11{,}541} + \frac{(0.004347)(0.995653)}{9{,}853}}}$$

$$= -1.91$$

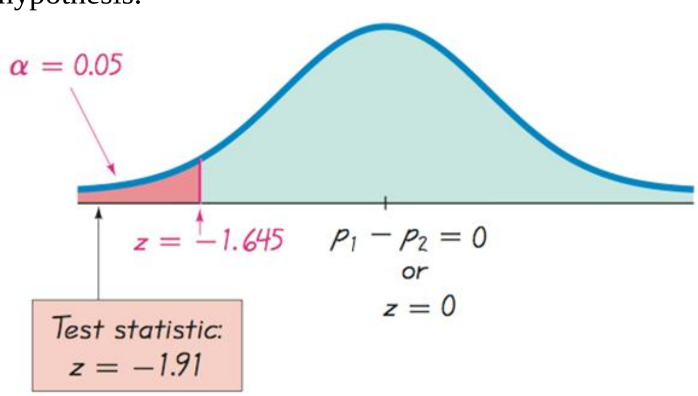
Left-tailed test. Area to left of $z = -1.91$ is 0.0281 (Standard Normal Distribution Table), so the *P*-value is 0.0281.

Step 7: Because the *P*-value of 0.0281 is less than the significance level of $\alpha = 0.05$, we reject the null hypothesis of $p_1 = p_2$.

Because we reject the null hypothesis, we conclude that there is sufficient evidence to support the claim that the proportion of accident fatalities for occupants in cars with airbags is less than the proportion of fatalities for occupants in cars without airbags. Based on these results, it appears that airbags are effective in saving lives.

Example: **Using the Traditional Method**

With a significance level of $\alpha = 0.05$ in a left- tailed test based on the normal distribution, we refer to Standard Normal Distribution Table and find that an area of $\alpha = 0.05$ in the left tail corresponds to the critical value of $z = -1.645$. The test statistic of does fall in the critical region bounded by the critical value of $z = -1.645$.
We again reject the null hypothesis.



Caution

When testing a claim about two population proportions, the *P*-value method and the traditional method are equivalent, but they are *not* equivalent to the confidence interval method. If you want to test a claim about two population proportions, use the *P*-value method or traditional method; if you want to estimate the difference between two population proportions, use a confidence interval.

Example

Use the sample data given in the preceding Example to construct a 90% confidence interval estimate of the difference between the two population proportions. (As shown in Table 8-2 on page 406, the confidence level of 90% is comparable to the significance level of $\alpha = 0.05$ used in the preceding left-tailed hypothesis test.) What does the result suggest about the effectiveness of airbags in an accident?

Solution

Requirements are satisfied as we saw in the preceding example.

90% confidence interval: $z_{\alpha/2} = 1.645$ Calculate the margin of error, E

$$E = z_{\alpha/2}\sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1} + \frac{\hat{P}_2\hat{q}_2}{n_2}}$$

$$= 1.645\sqrt{\frac{\frac{41}{11{,}541}\cdot\frac{11{,}500}{11{,}541}}{11{,}541} + \frac{\frac{52}{9{,}853}\cdot\frac{9801}{9{,}853}}{9{,}853}}$$

$$= 0.001507$$

TI-83 / 84 Calculator – For Hypothesis and confidence intervals

Press **STAT**

Select **TESTS**

Choose the option of **2-PropZTest** (for hypothesis test)

Or **2-PropZInt** (for confidence test)

Result:

Calculator will display a P-value instead of critical values, so the P-value method of hypothesis is used.

Exercises *Section* 4.1 – Inferences about Two Population Proportions

1. A Student surveyed her friends and found that among 20 males, 4 smoke and among 30 female, 6 smoke. Give two reasons why these results should not be used for a hypothesis test of the claim that the proportions of male smokers and female smokers are equal.

2. In clinical trials of the drug Zocor, some subjects were treated with Zocor and other were given a placebo. The 95% confidence interval estimate of the difference between the proportions of subjects who experienced headaches is $-0.0518 < p_1 - p_2 < 0.0194$. Write a statement interpreting that confidence interval.

3. Among 8834 malfunctioning pacemakers, in 15.8% the malfunctions were due to batteries. Find the number of successes *x*.

4. Among 129 subjects who took Chantix as an aid to stop smoking, 12.4% experienced nausea. Find the number of successes *x*.

5. Among 610 adults selected randomly from among the residents of one town, 26.1% said that they have favor stronger gun-control laws. Find the number of successes *x*.

6. A computer manufacturer randomly selects 2,410 of its computers for quality assurance and finds that 3.13% of these computer are found defective. Find the number of successes *x*.

7. Assume that you plan to use a significance level of $\alpha = 0.05$ to test the claim that $p_1 = p_2$. Use the given sample sizes and number of successes to find the pooled estimate $\overline{p}$

a) $n_1 = 288, \quad n_2 = 252, \quad x_1 = 75, \quad x_2 = 70$

b) $n_1 = 100, \quad n_2 = 100, \quad \hat{p}_1 = 0.2, \quad \hat{p}_2 = 0.18$

8. The numbers of online applications from simple random samples of college applications for 2003 and for the current year are given below.

	2003	Current Year
Number of application in sample	36	27
Number of online applications in sample	13	14

Assume that you plan to use a significance level of $\alpha = 0.05$ to test the claim that $p_1 = p_2$.

Find

a) The pooled estimate $\overline{p}$

b) The *x* test statistic

c) The critical *z* values

d) The *P*–value

Assume 95% confidence interval

e) The margin of error E
f) The 95% confidence interval.

9. Chantix is a drug used as an aid to stop smoking. The numbers of subjects experiencing insomnia for each of two treatment groups in a clinical trial of the drug Chantix are given below:

	Chantix Treatment	Placebo
Number in group	129	805
Number experiencing insomnia	19	13

Assume that you plan to use a significance level of $\alpha = 0.05$ to test the claim that $p_1 = p_2$.

Find

a) The pooled estimate $\bar{p}$

b) The *x* test statistic

c) The critical *z* values

d) The *P*–value

Assume 95% confidence interval

e) The margin of error E

f) The 95% confidence interval.

10. In a 1993 survey of 560 college students, 171 said that they used illegal drugs during the previous year. In a recent survey of 720 college students, 263 said that they used illegal drugs during the previous year. Use a 0.05 significance level to test the claim that the proportion of college students using illegal drugs in 1993 was less than it is now.

11. In a 1993 survey of 560 college students, 171 said that they used illegal drugs during the previous year. In a recent survey of 720 college students, 263 said that they used illegal drugs during the previous year. Construct the confidence interval corresponding to the hypothesis test conducted with a 0.05 significance level. What conclusion does the confidence interval suggest?

12. A simple random sample of front-seat occupants involved in car crashes is obtained. Among 2823 occupants not wearing seat belts, 31 were killed. Among 7765 occupants wearing seat belts, 16 were killed. Construct a 90% confidence interval estimate of the difference between the fatality rates for those not wearing seat belts and those wearing seat belts. What does the result suggest about the effectiveness of seat belts?

13. A Pew Research Center poll asked randomly selected subjects if they agreed with the statement that "It is morally wrong fir married people to have an affair/" Among the 386 women surveyed, 347 agrees with the statement. Among the 359 men surveyed, 305 agreed with the statement.

a) Use a 0.05 significance level to test the claim that the percentage of women who agree is difference from the percentage of men who agree. Does there appear to be a difference in the way women and men feel about this issue?

b) Construct the confidence interval corresponding to the hypothesis test conducted with a 0.05 significance level. What conclusion does the confidence interval suggest?

14. Tax returns include an option of designating $3 for presidential election campaigns, and it does not cost the taxpayer anything to make that designation. In a simple random sample of 250 tax returns from 1976, 27.6% of the returns designated the $3 for the campaign. In a simple random sample of 300 recent tax returns, 7.3% of the returns designated the $3 for the campaign. Use a 0.05 significance level to test the claim that the percentage of returns designating the $3 for the campaign was greater in 1973 than it is now.

15. In an experiment, 16% of 734 subjects treated with Viagra experienced headaches. In the same experiment, 4% of 725 subjects given a placebo experienced headaches.

a) Use a 0.01 significance level to test the claim that the proportion of headaches is greater for those treated with Viagra. Do headaches appear to be a concern for those who take Viagra?

b) Construct the confidence interval corresponding to the hypothesis test conducted with a 0.01 significance level. What conclusion does the confidence interval suggest?

16. Two different simple random samples are drawn from two different populations. The first sample consists of 20 people with 10 having a common attribute. The second sample consists of 2000 people with 1404 of them having the same common attribute. Compare the results from a hypothesis test of $p_1 = p_2$ (with a 0.05 significance level) and a 95% confidence interval estimate of $p_1 - p_2$.

17. A report on the nightly news broadcast stated that 11 out of 142 households with pet dogs were burglarized and 21 out of 217 without pet dogs were burglarized. Find the z test statistic for the hypothesis test. Assume that you plan to use a significance level of $\alpha = 0.05$ to test the claim that $p_1 = p_2$.

18. Assume that the samples are independent and that they have been randomly selected. Construct a 90% confidence interval for the difference between population proportions $p_1 = p_2$

19. The sample size needed to estimate the difference between two population proportions ti within a margin of error E with a confidence level of $1 - \alpha$ can be found as follows:

$$E = z_{\alpha/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\,.$$

In this expression, replace n_1 and n_2 by n (assuming both samples have the same size) and replace each of p_1, q_1, p_2 *and* q_2 by 0.5 (because their values are not known). Then solve for n. Use this approach to find the size pf each sample of you want to estimate the difference between the proportions of men and women who plan to vote in the next presidential election. Assume that you want 99% confidence that your error is no more than 0.05.

Section 4.2 – Inferences About Two Means: Dependent

Definitions

Two samples are ***independent*** if the sample values selected from one population are not related to or somehow paired or matched with the sample values from the other population.

Two samples are ***dependent*** if the sample values are *paired*. (That is, each pair of sample values consists of two measurements from the same subject (such as before/after data), or each pair of sample values consists of matched pairs (such as husband/wife data), where the matching is based on some inherent relationship.)

Objectives

Test a claim about the mean of the differences from dependent samples or construct a confidence interval estimate of the mean of the differences from dependent samples.

Statistical inference methods on matched-pairs data use the same methods as inference on a single population mean, except that the ***differences*** are analyzed.

Testing Hypotheses Regarding the Difference of Two Means Using a Matched-Pairs Design

To test hypotheses regarding the mean difference of matched-pairs data, the following must be satisfied:

- ✓ The sample is obtained using simple random sampling,
- ✓ The sample data are matched pairs,
- ✓ The differences are normally distributed with no outliers or the sample size, n, is large ($n \geq 30$),
- ✓ The sampled values are independent

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

Notation

d : Individual difference between the two values in a single matched pair

μ_d : Mean value of the differences d for the *population* of all pairs of data

$\bar{d}$: Mean value of the differences d for the paired *sample* data

s_d : Standard deviation of the differences d for the paired *sample* data

n : Number of all *pairs* of data

Confidence Intervals for Dependent Samples

$$\bar{d} - E < \mu_d < \bar{d} + E \qquad \textit{where} \quad E = t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

Critical values of $t_{\alpha/2}$: Use Table with $n - 1$ degrees of freedom.

***Step* 1:** Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways, where μ_d is the population mean difference of the matched-pairs data.

Two–Tailed	***Left–Tailed***	***Right–Tailed***
$H_0: \mu_d = 0$	$H_0: \mu_d = 0$	$H_0: \mu_d = 0$
$H_1: \mu_d \neq 0$	$H_1: \mu_d < 0$	$H_1: \mu_d > 0$

***Step* 2:** Select a level of significance, α, based on the seriousness of making a Type I error.

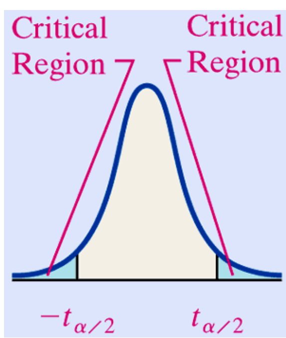
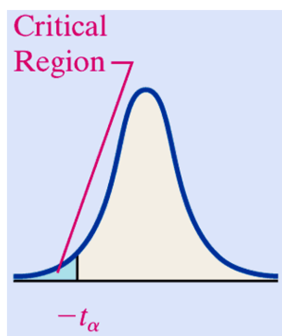
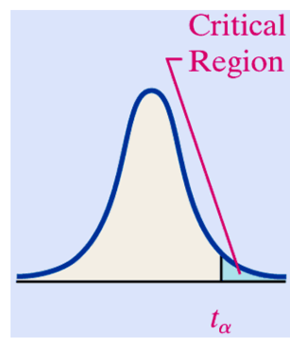
***Step* 3:** Compute the test statistic

$$t_0 = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

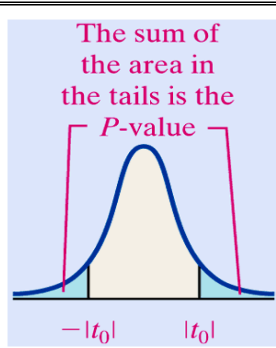
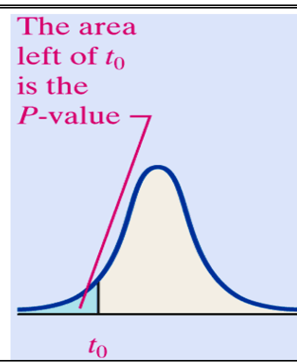
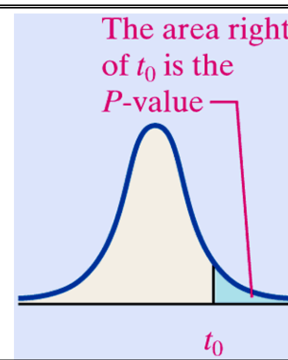
which approximately follows Student's t-distribution with $n - 1$ degrees of freedom. The values of $\bar{d}$ and s_d are the mean and standard deviation of the differenced data.

Use Table to determine the critical value using $n - 1$ degrees of freedom.

Classical Approach

<table>
<tr><th>Two–Tailed</th><th>Left–Tailed</th><th>Right–Tailed</th></tr>
<tr><td>$t_0 < -t_{\alpha/2}$ or $t_0 > t_{\alpha/2}$
Reject the null hypothesis</td><td>$t_0 < -t_{\alpha}$
Reject the null hypothesis</td><td>$t_0 > t_{\alpha}$
Reject the null hypothesis</td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

***Step* 4**: Estimate the P–value

<table>
<tr><th>Two–Tailed</th><th>Left–Tailed</th><th>Right–Tailed</th></tr>
<tr><td></td><td></td><td></td></tr>
</table>

*If **P-value** < **α**, reject the null hypothesis*

Example

Data Set below includes measured weights of college students in September and April of their freshman year. (Here we use only a small portion of the available data so that we can better illustrate the method of hypothesis testing.) Use the sample data in Table below with a 0.05 significance level to test the claim that for the population of students, the mean change in weight from September to April is equal to 0 *kg*.

Weight (kg) Measurements of Students in Their Freshman Year

April weight	66	52	68	69	71
September weight	67	53	64	71	70
Difference *d* = (April weight) – (September weight)	–1	–1	4	–2	1

Solution

Requirements are satisfied: samples are dependent, values paired from each student; although a volunteer study, we'll proceed as if simple random sample and deal with this in the interpretation.

Weight gained = April weight – Sept. weight

μ_d denotes the mean of the "April – Sept." differences in weight; the claim is $\mu_d = 0\ kg$

Step 1: claim is $\mu_d = 0\ kg$

Step 2: If original claim is not true, we have $\mu_d \neq 0\ kg$

Step 3: $H_0 : \mu_d = 0\ kg$ original claim $H_1 : \mu_d \neq 0\ kg$

Step 4: significance level is $\alpha = 0.05$

Step 5: use the student *t* distribution

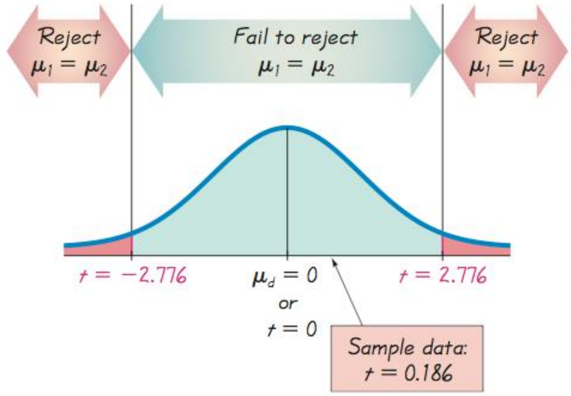
Step 6: find values of *d* and s_d differences are: –1, –1, 4, –2, 1 $d = 0.2$ and $s_d = 2.4$ now find the test statistic

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{0.2 - 0}{\frac{2.4}{\sqrt{5}}} = 0.186$$

From (*t*–Distribution Table): df = *n* – 1, area in two tails is 0.05, yields a critical value *t* = ±2.776

Degrees of Freedom	Area in Two Tails 0.01	0.02	0.05	0.10	0.20
4	**4.604**	3.747	$\boxed{\textbf{2.776}}$	2.132	1.533

Step 7: Because the test statistic does not fall in the critical region, we fail to reject the null *hypothesis*.



We conclude that there is not sufficient evidence to warrant rejection of the claim that for the population of students, the mean change in weight from September to April is equal to 0 *kg*. Based on the sample results listed in Table, there does not appear to be a significant weight gain from September to April.

The P-value method:

Using technology, we can find the *P*-value of 0.8605. (Using *t*–Distribution Table: with the test statistic of $t = 0.186$ and 4 degrees of freedom, we can determine that the *P*-value is greater than 0.20.) We again fail to reject the null hypothesis, because the *P*-value is greater than the significance level of $\alpha = 0.05$.

Confidence Interval method:

Construct a 95% confidence interval estimate of μ_d , which is the mean of the "April–September" weight differences of college students in their freshman year.

$\bar{d} = 0.2, \quad s_d = 2.4 \quad n = 5, \quad t = 2.776$

The margin error: $E = t_{\alpha/2} \dfrac{s_d}{\sqrt{n}} = 2.776 \cdot \dfrac{2.4}{\sqrt{5}} = 3.0$

The confidence interval:

$$\bar{d} - E < \mu_d < \bar{d} + E$$

$$0.2 - 3.0 < \mu_d < 0.2 + 3.0$$

$$-2.8 < \mu_d < 3.2$$

Conclusion:

We have 95% confidence that the limits of –2.8 *kg* and 3.2 *kg* contain the true value of the mean weight change from September to April. In the long run, 95% of such samples will lead to confidence interval limits that actually do contain the true population mean of the differences. Note that the confidence interval includes the value of 0 *kg*, so it is very possible that the mean of the weight changes is equal to 0 *kg*.

Exercises *Section* 4.2 – Inferences about Two Means: Dependent

1. Listed below are the time intervals (in minutes) before and after eruptions of the Old Faithful geyser. Find the values of $\bar{d}$ and s_d. In general, what does μ_d represent?

Time interval before eruption	98	92	95	87	96
Time interval after eruption	92	95	92	100	90

2. Listed below are measured fuel consumption amount (in miles/gal) from a sample of cars.

City fuel consumption	18	22	21	21
Highway fuel consumption	26	31	29	29

Assume that wou want to use a 0.05 significance level to test the claim that the paired sample data come from a population for which the mean differecne is $\mu_d = 0$. Find

a) $\bar{d}$ *b)* s_d *c)* The t test statistic *d)* The critical values.

3. Listed below are predicted high temperatures that were forecast different days.

Predicted high temperatures forecast 3 *days ahead*	79	86	79	83	80
Predicted high temperatures forecast 5 *days ahead*	80	80	79	80	79

Assume that wou want to use a 0.05 significance level to test the claim that the paired sample data come from a population for which the mean differecne is $\mu_d = 0$. Find

a) $\bar{d}$ *b)* s_d *c)* The t test statistic *d)* The critical values.

4. Listed below are body mass indices (BMI). The BMI of each student was measured in September and April of the freshman year.

a) Use a 0.05 significance level to tet the claim that the mean change in BMI for all students is equal to 0. Does BMI appear to change during freshman year?

b) Construct a 95% confidence interval estimate of the change in BMI during freshman year. Does the confidence interval include 0, and what does that suggest about BMI during freshman year?

April BMI	20.15	19.24	20.77	23.85	21.32
September BMI	20.68	19.48	19.59	24.57	20.96

5. Listed below are body temperature (in °F) of subjects measured at 8:00 AM and at 12:00 AM. Construct a 95% confidence interval estimate of the difference between the 8:00 AM temperatures and the 12:00 AM temperatures. Is body temperature basically the same at both times?

8:00 AM	97.0	96.2	97.6	96.4	97.8	99.2
12:00 AM	98.0	98.6	98.8	98.0	98.6	97.6

6. Listed below are systolic blood pressure measuremetns (mm Hg) taking rom the right and left arms of the same woman. Use a 0.05 significance level to test for a difference the measurements from the two arms. What do you conclude?

Right arm	102	101	94	79	79
Left arm	175	169	182	146	144

7. As part of the National Health and Nutrition Examination Survey, the Department of Health and Human Services obtained self-reported heights and measured heights for males ages 12 – 16. All measurement are in inches. Listed below are sample results

Reported height	68	71	63	70	71	60	65	64	54	63	66	72
Measured height	67.9	69.9	64.9	68.3	70.3	60.6	64.5	67.0	55.6	74.2	65.0	70.8

a) Is there sufficient evidence to support the claim that there is a difference between self-reported heights and measured heights of males? Use a 0.05 significance level.

b) Construct a 95% confidence interval estimate of the man difference between reported heights and measured heights. Interpret the resulting confidence interval, and comment on the implications of whether the confidence interval limits contain 0.

8. Listed below are combined city – highway fuel consumption ratings (in miles/gal) for different cars measured under both the old rating system and a new rating system introducing in 2008. The new ratings were implemented in response to complaints that the old ratings were too high. Use a 0.01 significance level to test the claim the old ratings are higher than the new ratings.

Old rating	16	18	27	17	33	28	33	18	24	19	18	27	22	18	20	29	19	27	20	21
New rating	15	16	24	15	29	25	29	16	22	17	16	24	20	16	18	26	17	25	18	19

9. Listed below are 2 tables. Construct a 95% confidence interval estimate of the mean of the differences between weights of discarded paper and weights of discarded plastic. Which seems to weigh more: discarded paper or discarded plastic?

Paper

2.41	7.57	9.55	8.82	8.72	6.96	6.83	11.42	16.08	6.38	13.05	11.36	15.09
2.80	6.44	5.86	11.08	12.43	6.05	13.61	6.98	14.33	13.31	3.27	6.67	17.65
12.73	9.83	16.39	6.33	9.19	9.41	9.45	12.32	20.12	7.72	6.16	7.98	9.64
8.08	10.99	13.11	3.26	1.65	10.00	8.96	9.46	5.88	8.26	12.45	10.58	5.87
8.78	11.03	12.29	20.58	12.56	9.92	3.45	9.09	3.69	2.61			

Plastic

0.27	1.41	2.19	2.83	2.19	1.81	0.85	3.05	3.42	2.10	2.93	2.44	2.17
1.41	2.00	0.93	2.97	2.04	0.65	2.13	0.63	1.53	4.69	0.15	1.45	2.68
3.53	1.49	2.31	0.92	0.89	0.80	0.72	2.66	4.37	0.92	1.40	1.45	1.68
1.53	1.44	1.44	1.36	0.38	1.74	2.35	2.30	1.14	2.88	2.13	5.28	1.48
3.36	2.83	2.87	2.96	1.61	1.58	1.15	1.28	0.58	0.74			

10. Suppose you wish to test the claim that μ_d, the mean value of the differences d for a population of paired data, is different from 0. Given a sample of $n = 23$ and a significance level of $\alpha = 0.05$, what criterion would be used for rejecting the null hypothesis?

11. Assume that he paired data came from a population that is normally distributed. Using a 0.05 significance level, find $\bar{d}$, s_d, the t test statistic, and the critical values to test the claim that $\mu_d = 0$

x	14	8	4	14	3	12	4	13
y	15	8	7	13	5	11	6	15

12. Assume that he paired data came from a population that is normally distributed. Using a 0.05 significance level, find $\bar{d}$, s_d, the t test statistic, and the critical values to test the claim that $\mu_d = 0$

x	12	5	1	20	3	16	12	8
y	7	10	5	15	7	14	10	13

Section 4.3 – Inferences About Two Means: Independent

Hypothesis Test for Two Means: Independent Samples

$$t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

(where $\mu_1 - \mu_2$ is ***often assumed*** to be 0)

approximately follows Student's *t*-distribution with the smaller of $n_1 - 1$ or $n_2 - 1$ degrees of freedom

Notation

μ_i = population mean

n_1 = size of the first sample

s_i = sample standard deviation

Testing Hypotheses Regarding the Difference of Two Means

To test hypotheses regarding the mean difference of matched-pairs data, the following must be satisfied:

- ✓ The sample is obtained using simple random sampling;
- ✓ The sample are independent;
- ✓ The populations from which the samples are drawn are normally distributed or the sample sizes are large $\left(n_1 \geq 30,\ n_2 \geq 30\right)$;
- ✓ For each sample, the sample size is no more than 5% of the population size.

Degrees of freedom

1. We use this simple and conservative estimate: d*f* = smaller of $n_1 - 1$ and $n_2 - 1$.
2. Statistically software typically use the more accurate but more difficult estimate formula

$$\boldsymbol{df} = \frac{(A+B)^2}{\frac{A^2}{n_1 - 1} + \frac{B^2}{n_2 - 1}} \qquad \textit{where} \quad A = \frac{s_1^2}{n_1} \quad B = \frac{s_2^2}{n_2}$$

Confidence Interval Estimate of $\mu_1 - \mu_2$: Independent Samples

The confidence interval estimate of the difference $\mu_1 - \mu_2$ is

$$\left(\bar{x}_1 - \bar{x}_2\right) - E < \left(\mu_1 - \mu_2\right) < \left(\bar{x}_1 - \bar{x}_2\right) + E \qquad \textit{Where} \qquad E = t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

***Step* 1:** Determine the null and alternative hypotheses. The hypotheses can structured in one of three ways.

Two–Tailed	***Left–Tailed***	***Right–Tailed***
$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$	$H_1: \mu_1 < \mu_2$	$H_1: \mu_1 > \mu_2$

***Step* 2:** Select a level of significance, α, based on the seriousness of making a Type I error.

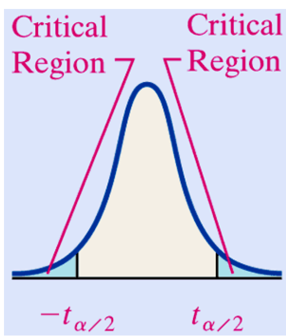
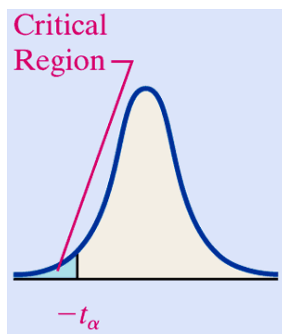
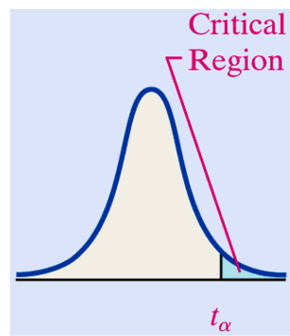
***Step* 3:** Compute the test statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

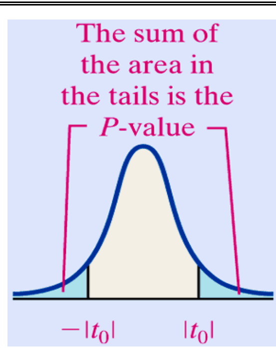
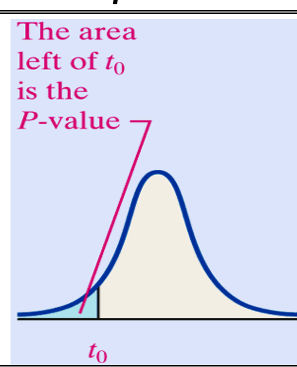
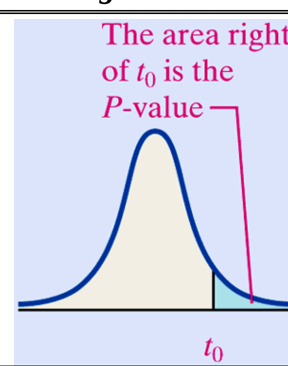
which approximately follows Student's t-distribution.

Use Table to determine the critical value using $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

Classical Approach

Two–Tailed	***Left–Tailed***	***Right–Tailed***
$t_0 < -t_{\alpha/2}$ *or* $t_0 > t_{\alpha/2}$	$t_0 < -t_{\alpha}$	$t_0 > t_{\alpha}$
Reject the null hypothesis	*Reject the null hypothesis*	*Reject the null hypothesis*
		

***Step* 4**: Estimate the *P*–value

Two–Tailed	***Left–Tailed***	***Right–Tailed***
		

If ***P*-value < *α***, *reject the null hypothesis*

Example

A headline in *USA Today* proclaimed that "Men, women are equal talkers." That headline referred to a study of the numbers of words that samples of men and women spoke in a day. Given below are the results from the study. .

Number of Words Spoken in a Day

Men	***Women***
$n_1 = 186$	$n_2 = 210$
$\bar{x}_1 = 15,668.5$	$\bar{x}_2 = 16,215.0$
$s_1 = 8632.5$	$s_2 = 7301.2$

a) Use a 0.05 significance level to test the claim that men and women speak the same mean number of words in a day. Does there appear to be a difference?

b) Construct a 95% confidence interval estimate of the difference between the mean number of words spoken by men and the mean number of words spoken by women.

Solution

a) Two population standard deviations are not known and not assumed to be equal, independent samples, simple random samples, both samples are large.

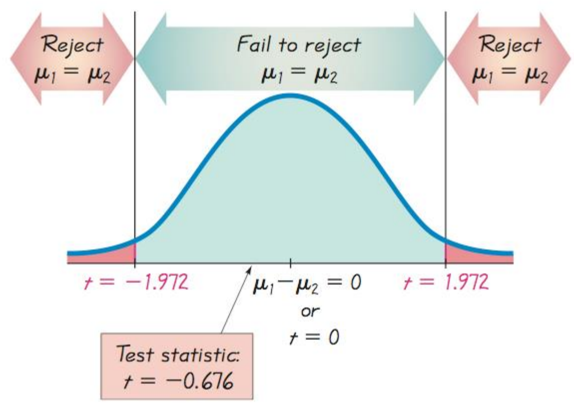
$H_0: \ \mu_1 = \mu_2$

$H_1: \ \mu_1 \neq \mu_2$

Assume: $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{(15,668.5 - 16,215.0) - 0}{\sqrt{\dfrac{8632.5^2}{186} + \dfrac{7301.2^2}{210}}}$$

$= -0.676$



Use *t* Distribution Table: area in two tails is 0.05, *df* = 185, which is not in the table, the closest value is *t* = ±1.972

Degrees of Freedom	Area in Two Tails 0.01	0.02	0.05	0.10	0.20
200	**2.601**	2.345	**1.972**	1.653	1.286

Because the test statistic does not fall within the critical region, fail to reject the null hypothesis: $\mu_1 = \mu_2$ (or $\mu_1 - \mu_2 = 0$).

Conclusion

There is not sufficient evidence to warrant rejection of the claim that men and women speak the same mean number of words in a day. There does not appear to be a significant difference between the two means.

b) $E = t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = 1.972\sqrt{\frac{8632.5^2}{186}+\frac{7301.2^2}{210}}$

$= 1595.4$

Construct the confidence interval use E = 1595.4 and

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

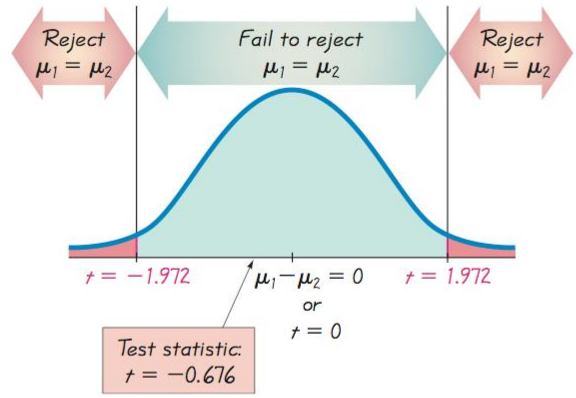
$$(15{,}668.5 - 16{,}215.0) - 1595.4 < (\mu_1 - \mu_2) < (15{,}668.5 - 16{,}215.0) + 1595.4$$

$$-2141.9 < (\mu_1 - \mu_2) < 1048.9$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

$$= \frac{(15{,}668.5 - 16{,}215.0) - 0}{\sqrt{\frac{8632.5^2}{186}+\frac{7301.2^2}{210}}}$$

$= -0.676$



Because the test statistic does not fall within the critical region, fail to reject the null hypothesis: $\mu_1 = \mu_2$ (or $\mu_1 - \mu_2 = 0$).

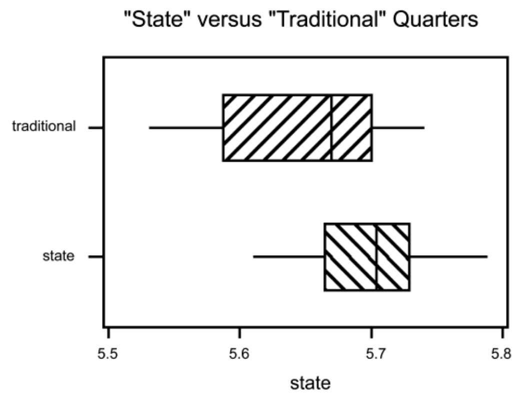
We are 95% confident that the limits of –2141.9 words and 1048.9 words actually do contain the difference between the two population means. Because those limits do contain 0, there is not sufficient evidence to warrant rejection of the claim that men and women speak the same mean number of words in a day. There does not appear to be a significant difference between the two means.

Example

A researcher wanted to know whether "state" quarters had a weight that is more than "traditional" quarters. He randomly selected 18 "state" quarters and 16 "traditional" quarters, weighed each of them and obtained the following data.

STATE (grams)		*TRADITIONAL (grams)*	
5.70	5.67	5.67	5.55
5.73	5.61	5.70	5.61
5.70	5.67	5.72	5.58
5.65	5.62	5.66	5.74
5.73	5.65	5.70	5.68
5.79	5.73	5.68	5.53
5.77	5.71	5.67	5.55
5.70	5.76	5.61	5.74
5.73	5.72		

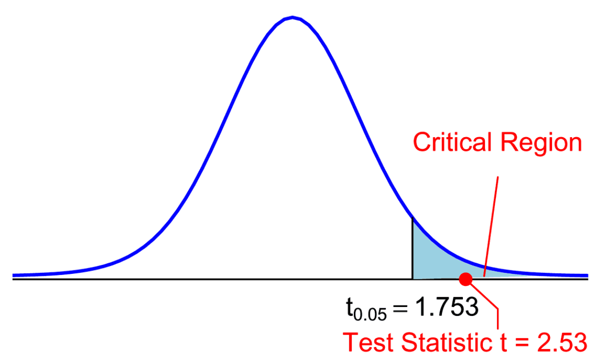
Test the claim that "state" quarters have a mean weight that is more than "traditional" quarters at the $\alpha = 0.05$ level of significance.



Solution

Step 1: We want to determine whether state quarters weigh more than traditional quarters:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$



Step 2: The level of significance is $\alpha = 0.05$.

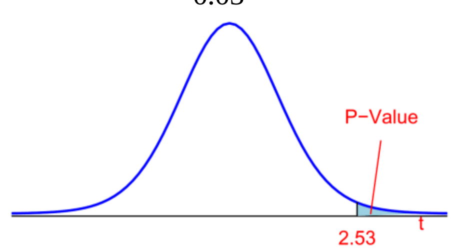
Step 3: The test statistic is

$$t_0 = \frac{5.7022 - 5.6494}{\sqrt{\frac{0.0497^2}{18} + \frac{0.0689^2}{16}}} = 2.53$$

$$t_{0.05} = 1.753$$

Step 4: Since the test statistic, $t_0 = 2.53$ is greater than the critical value $t_{0.05} = 1.753$

We reject the null hypothesis.



Because this is a right-tailed test, the *P*-value is the area under the *t*-distribution to the right of the test statistic $t_0 = 2.53$. That is, P-value = $P(t > 2.53) \approx 0.01$

Step 4: Since the P-value is less than the level of significance $\alpha = 0.05$, we reject the null hypothesis.

Step 5: There is sufficient evidence at the $\alpha = 0.05$ level to conclude that the state quarters weigh more than the traditional quarters.

Alternative Methods When σ_1 and σ_2 are Known

Requirements

1. The two population standard deviations are both known.
2. The two samples are independent.
3. Both samples are simple random samples.
4. Either or both of these conditions are satisfied: The two sample sizes are both large (with $n_1 > 30$ and $n_2 > 30$) or both samples come from populations having normal distributions.

Hypothesis Test for Two Means: Independent Samples with σ_1 and σ_2 Both Known

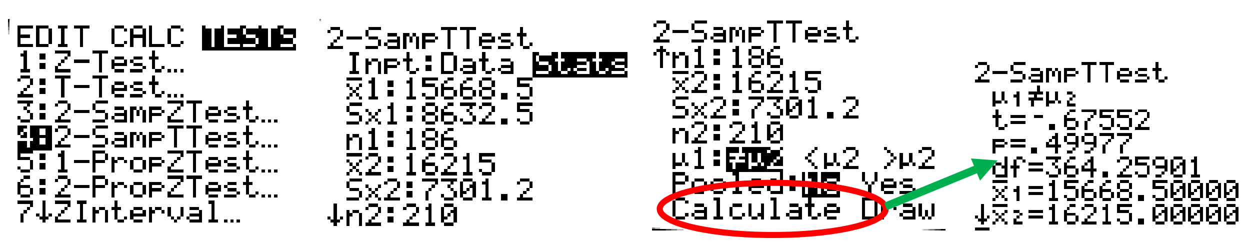
$$z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

P-values and critical values: Refer to Normal Distribution Table.

Confidence Interval: Independent Samples with σ_1 and σ_2 Both Known

$$\left(\bar{x}_1 - \bar{x}_2\right) - E < \left(\mu_1 - \mu_2\right) < \left(\bar{x}_1 - \bar{x}_2\right) + E$$

Where $E = z_{\alpha/2} \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$



Exercises Section 4.3 – Inferences about Two Means: Independent

1. If the pulse rates of men and women shown in the data below

Women:

76	72	88	60	72	68	80	64	68	68	80	76	68	72	96	72	68	72	64	80
64	80	76	76	76	80	104	88	60	76	72	72	88	80	60	72	88	88	124	64

Men:

68	64	88	72	64	72	60	88	76	60	96	72	56	64	60	64	84	76	84	88
72	56	68	64	60	68	60	60	56	84	72	84	88	56	64	56	56	60	64	72

These data are used to construct 95% confidence interval for the difference between the two population means, the result is $-12.2 < \mu_1 - \mu_2 < -1.6$, where pulse rates of men correspond to population 1 and pulse rates of women correspond to population 2. Express the confidence interval with pulse rates of women being population 1 and pulse rates of men being population 2.

2. Assume that you want to use a 0.01 significance level to test the claim that the mean pulse rate of men is less than the mean pulse rate of women. What confidence level should be used if you want to test that claim using a confidence interval?

3. To test the effectiveness of Lipitor, cholesterol levels are measured in 250 subjects before and after Lipitor treatments. Determine whether this sample is independent or dependent.

4. On each of 40 different days, you measured the voltage supplied to your home and you also measured the voltage produced by the gasoline-powered generator. One sample consists of the voltages in the house and the second sample consists of the voltages produced by the generator. Determine whether this sample is independent or dependent.

5. In a randomized controlled trial conducted with children suffering from viral croup, 46 children were treated with low humidity while 46 other children were treated with high humidity. Researchers used the Westley Croup Score to assess the results after one hour. The low humidity group had a mean score of 0.98 with standard deviation of 1.22 while the high humidity group had a mean score of 1.09 with standard deviation of 1.11.

a) Use a 0.05 significance level to test the claim that the two groups are from populations with the same mean. What does the result suggest about the common treatment of humidity?
Assume that the two samples are independent simple random samples selected from normally distributed populations.

b) Assume that $\sigma_1 = \sigma_2$, how are the results affected by this additional assumption?

6. The mean tar content of a simple random sample of 25 unfiltered king size cigarettes is 21.1 mg, with a standard deviation of 3.2 mg. The mean tar content of a simple random sample of 25 filtered 100 mm cigarettes is 13.2 mg, with a standard deviation of 3.7 mg.
Assume that the two samples are independent simple random samples selected from normally distributed populations in part a and b.

a) Construct a 90% confidence interval estimate of the difference between the mean tar content of unfiltered king size cigarettes and the mean tar content of filtered 100 mm cigarettes. Does the result suggest that 100 mm filtered cigarettes have less tar than unfiltered king size cigarettes?

b) Use a 0.05 significance level to test the claim that unfiltered king size cigarettes have a mean tar content greater than that of filtered 100 mm cigarettes. What does the result suggest about the effectiveness of cigarette filters?

c) Assume that $\sigma_1 = \sigma_2$, how are the results affected by this additional assumption?

7. The heights are measured for the simple random sample of supermodels Crawford, Bundchen, Pestova, Christenson, Hume, Moss, Campbell, Schiffer, and Taylor. They have a mean of 70.0 in. and a standard deviation of 1.5 in. 40 women who are not supermodels, listed below and they have heights with means of 63.2 in. and a standard deviation of 2.7 in.

64.3	66.4	62.3	62.3	59.6	63.6	59.8	63.3	67.9	61.4	66.7	64.8	63.1	66.7	66.8
64.7	65.1	61.9	64.3	63.4	60.7	63.4	62.6	60.6	63.5	58.6	60.2	67.6	63.4	64.1
62.7	61.3	58.2	63.2	60.5	65.0	61.8	68.0	67.0	57.0					

a) Use a 0.01 significance level to test the claim that the mean height of supermodels is greater than the mean height of women who are not supermodels

b) Construct a 98% confidence interval level for the difference between the mean height of supermodels and the mean height of women who are not supermodels. What does the result suggest about those two means?

8. Many studies have been conducted to test the effects of marijuana use on mental abilities. In one such study, groups of light and heavy users of marijuana in college were tested for memory recall, with the results given below. Use a 0.01 significance level to test the claim that the population of heavy marijuana users has a lower mean than the light users. Should marijuana use be of concern to college students?

Items sorted correctly by light marijuana users: $n = 64, \quad \bar{x} = 53.3, \quad s = 3.6$

Items sorted correctly by heavy marijuana users: $n = 65, \quad \bar{x} = 51.3, \quad s = 4.5$

9. The trend of thinner Miss America winners has generated charges that the contest encourages unhealthy diet habits among young women. Listed below are body mass indexes (BMI) for Miss America winners from two different time periods. Consider the listed values to be simple random samples selected from larger populations.

a) Use a 0.05 significance level to test the claim that recent winners have a lower mean BMI than winners from the 1920s and 1930s.

b) Construct a 90% Confidence interval for the difference between the mean BMI of recent winners and the mean BMI of winners from the 1920s and 1930s.

BMI (from recent winners):	19.5	20.3	19.6	20.2	17.8	17.9	19.1	18.8	17.6	16.8
BMI (from 1920s and 1930s):	20.4	21.9	22.1	22.3	20.3	18.8	18.9	19.4	18.4	19.1

10. Listed below are amounts of strontium-90 (in millibecquerels or mBq per gram of calcium) in a simple random sample of baby teeth obtained from Pennsylvania residents and New York residents born after 1979.

a) Use a 0.05 significance level to test the claim that the mean amount of strontium-90 from Pennsylvania residents is greater than the mean amount from New York residents.

b) Construct a 90% Confidence interval for the difference between the mean amount of strontium-90 from Pennsylvania residents and the mean amount from New York residents.

Pennsylvania:	155	142	149	130	151	163	151	142	156	133	138	161
New York:	133	140	142	131	134	129	128	140	140	140	137	143

11. Listed below are the word counts for male and female psychology students.

a) Use a 0.05 significance level to test the claim that male and female psychology students speak the same mean number of words in a day.

b) Construct a 95% Confidence interval estimate of the difference between the mean number of words spoken in a day by male and female psychology students. Do the confidence interval limits include 0, and what does that suggest about the two means?

Male	21143	17791	36571	6724	15430	11552	11748	12169	15581	23858	5269
	12384	11576	17707	15229	18160	22482	18626	1118	5319		

Female	6705	21613	11935	15790	17865	13035	24834	7747	3852	11648	25862
	17183	11010	11156	11351	25693	13383	19992	14926	14128	10345	13516
	12831	9671	17011	28575	23557	13656	8231	10601	8124		

Assume that the two samples are independent simple random samples selected from normally distributed populations. Do not assume that the population standard deviations are equal.

12. Refer to the tables below and test the claim that they contain the same amount of cola, the mean weight of cola cans of regular Coke is the same as the mean weight of cola in cans of Diet Coke. If there is a difference in the mean weights, identify the most likely explanation for that difference.

Coke	0.8192	0.815	0.8163	0.8211	0.8181	0.8247	0.8062	0.8128	0.8172	0.811
	0.8251	0.8264	0.7901	0.8244	0.8073	0.8079	0.8044	0.817	0.8161	0.8194
	0.8189	0.8194	0.8176	0.8284	0.8165	0.8143	0.8229	0.815	0.8152	0.8244
	0.8207	0.8152	0.8126	0.8295	0.8161	0.8192				

Diet	0.7773	0.7758	0.7896	0.7868	0.7844	0.7861	0.7806	0.783	0.7852	0.7879
	0.7881	0.7826	0.7923	0.7852	0.7872	0.7813	0.7885	0.776	0.7822	0.7874
	0.7822	0.7839	0.7802	0.7892	0.7874	0.7907	0.7771	0.787	0.7833	0.7822
	0.7837	0.791	0.7879	0.7923	0.7859	0.7811				

Assume that the two samples are independent simple random samples selected from normally distributed populations. Do not assume that the population standard deviations are equal.

13. An Experiment was conducted to test the effects of alcohol. Researchers measured the breath alcohol levels for a treatment group of people who drank ethanol and another group given a placebo. The results are given in the accompanying table. Use a 0.05 significance level to test the claim that the two sample groups come from populations with the same mean.

Treatment Group:		$\bar{x}_1 = 0.049$	$s_1 = 0.015$
Placebo Group:	$n_2 = 22$	$\bar{x}_2 = 0.000$	$s_2 = 0.000$

14. A researcher was interested in comparing the GPAs $n_1 = 22$ of students at two different colleges. Independent simple populations. Do samples of 8 students from college *A* and 13 students from college *B* yielding the following GPAs.

College *A*	3.7	3.2	3.0	2.5	2.7	3.6	2.8	3.4					
College *B*	3.8	3.2	3.0	3.9	3.8	2.5	3.9	2.8	4.0	3.6	2.6	4.0	3.6

Construct a 95% confidence interval for $\mu_1 - \mu_2$. The difference between the mean GPA of college *A* students and the mean GPA of college *B* students.

$\left(Note: \ \bar{x}_1 = 3.1125, \ \bar{x}_2 = 3.4385, \ s_1 = 0.4357, \ s_2 = 0.5485\right)$

15. Assume that the two samples are independent simple random samples selected from normal distributed populations. Do not assume that the population standard deviations are equal. A researcher was interested in comparing the heights of women in two different countries. Independent simple random samples of 9 women from country A and 9 women from B yielded to the following heights (in inches).

Country A	64.1	66.4	61.7	62.0	67.3	64.9	64.7	68.0	63.6
Country B	65.3	60.2	61.7	65.8	61.0	64.6	60.0	65.4	59.0

Construct a 90% confidence interval for $\mu_1 - \mu_2$ the difference between the mean height of women in country A and the mean height of women in country B. Round to two decimal places.

$\left(Note: \ \bar{x}_1 = 64.744 \ in, \ \bar{x}_2 = 62.556 \ in, \ s_1 = 2.192 \ in, \ s_2 = 2.697 \ in\right)$

Section 4.4 – Goodness-of-Fit

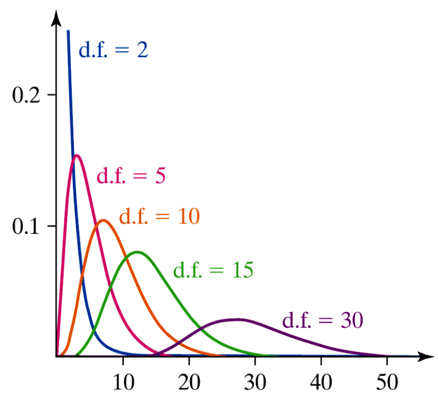
We consider sample data consisting of observed frequency counts arranged in a single row or column (called a one-way frequency table). We will use a hypothesis test for the claim that the observed frequency counts agree with some claimed distribution, so that there is a good fit of the observed data with the claimed distribution.

Definition

A ***goodness-of-fit*** test is used to test the hypothesis that an observed frequency distribution fits (or conforms to) some claimed distribution.

Characteristics of the Chi-Square Distribution

1. It is not symmetric
2. Its shape depends on the degrees of freedom, just like Student's *t*-distribution
3. As the number of degrees of freedom increases, it becomes more nearly symmetric
4. The values of χ^2 are nonnegative. That is, the values of χ^2 are greater than or equal to 0



Test Statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Notation

O represents the ***observed frequency*** of an outcome.
E represents the ***expected frequency*** of an outcome.
k represents the ***number of different categories*** or outcomes.
n represents the total ***number of trials***.

Critical Values

The chi-square distribution with $k-1$ degrees of freedom
Goodness-of-fit hypothesis tests are always *right-tailed*.

Requirements

1. The data have been randomly selected.
2. The sample data consist of frequency counts for each of the different categories.
3. All expected frequencies are greater than or equal to 1 (all $E_i \geq 1$) and
4. No more than 20% of the expected frequencies are less than 5.

CAUTION!

Goodness-of-fit tests are used to test hypotheses regarding the distribution of a variable based on a single population.

Expected Frequencies

If all expected frequencies are <u>equal</u>: $E = \frac{n}{k}$

The sum of all observed frequencies divided by the number of categories

If expected frequencies are <u>not all equal</u>: $E = np$

Each expected frequency is found by multiplying the sum of all observed frequencies by the probability for the category.

Goodness-of-Fit Test

To test the hypotheses regarding a distribution, we use the steps that follow

***Step* 1**: Determine the null and alternative hypotheses.

H_0: The random variable follows a certain distribution

H_1: The random variable does not follow the distribution in the null hypothesis

***Step* 2**: Decide on a level of significance, α, depending on the seriousness of making a Type I error.

***Step* 3:**

a) Calculate the expected counts for each of the k categories. The expected counts are $E_i = np_i$ for $i = 1, 2, \ldots, k$ where n is the number of trials and p_i is the probability of the ith category, assuming that the null hypothesis is true.

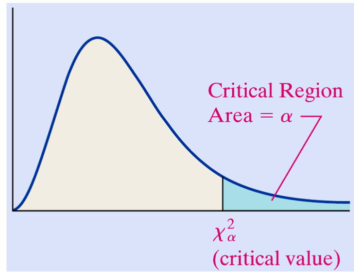
b) Verify that the requirements for the goodness-of-fit test are satisfied.

1. All expected counts are greater than or equal to 1 (all $E_i \geq 1$).
2. No more than 20% of the expected counts are less than 5.

Classical Approach

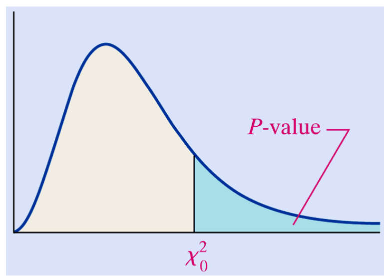
c) Compute the ***test statistic***: $\chi_0^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

***Step* 4:** Determine the critical value. All goodness-of-fit tests are right-tailed tests, so the critical value is χ^2_α with $k - 1$ degrees of freedom.



Compare the critical value to the test statistic. If $\chi^2_0 > \chi^2_\alpha$ reject the null hypothesis.

d) Use Table to obtain an approximate *P*-value by determining the area under the chi-square distribution with $k - 1$ degrees of freedom to the right of the test statistic.



***Step* 4:** If the *P*-value < α, reject the null hypothesis.

"If the *P* is low, the null must go."
(If the *P*-value is small, reject the null hypothesis that the distribution is as claimed.)

Example

Data Set includes weights from 40 randomly selected adult males and 40 randomly selected adult females. Those weights were obtained as part of the National Health Examination Survey. When obtaining weights of subjects, it is extremely important to actually weigh individuals instead of asking them to report their weights. By analyzing the last digits of weights, researchers can verify that weights were obtained through actual measurements instead of being reported. When people report weights, they typically round to a whole number, so reported weights tend to have many last digits consisting of 0. In contrast, if people are actually weighed with a scale having precision to the nearest 0.1 pound, the weights tend to have last digits that are uniformly distributed, with 0, 1, 2, … , 9 all occurring with roughly the same frequencies. Table shows the frequency distribution of the last digits from 80 weights. (For example, the weight of 201.5 lb has a last digit of 5, and this is one of the data values included in Table)

Last Digit	Frequency
0	7
1	14
2	6
3	10
4	8
5	4
6	5
7	6
8	12
9	8

Test the claim that the sample is from a population of weights in which the last digits do not occur with the same frequency. Based on the results, what can we conclude about the procedure used to obtain the weights?

Solution

Requirements are satisfied: randomly selected subjects, frequency counts, expected frequency is 8 (> 5)

At least one of the probabilities $p_0, p_1, \ldots, p_9$, is different from the others

At least one of the probabilities are the same:

$$p_0 = p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9$$

Null hypothesis contains equality

$$H_0 : p_0 = p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9$$

H_1 : At least one probability is different

No significance specified, use $\alpha = 0.05$

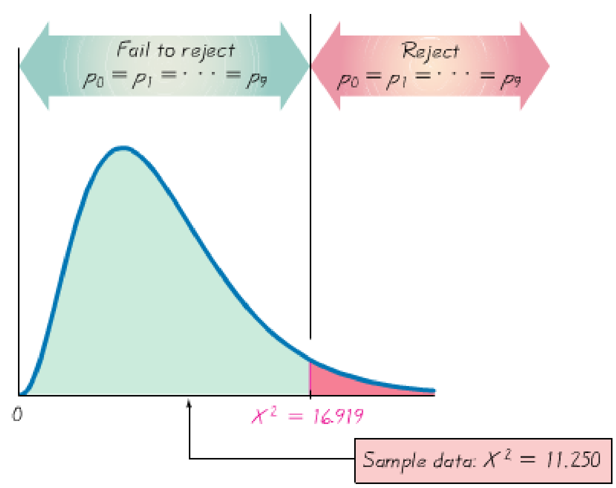
Testing whether a uniform distribution so use goodness-of-fit test: χ^2

Last Digit	*Observed Frequency O*	*Expected Frequency E*	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
0	7	8	–1	1	0.125
1	14	8	6	36	4.500
2	6	8	–2	4	0.500
3	10	8	2	4	0.500
4	8	8	0	0	0.000
5	4	8	–4	16	2.000
6	5	8	–3	9	1.125
7	6	8	–2	4	0.500
8	12	8	4	16	2.00
9	8	8	0	0	0.000
				$\chi^2 = \sum \frac{(O-E)^2}{E} = 11.250$	

The test statistic $\chi^2 = 11.250$, using $\alpha = 0.05$ and $k - 1 = 9$ degrees of freedom, the critical value is $\chi^2 = 16.919$

Because the test statistic does not fall in the critical region, there is not sufficient evidence to reject the null hypothesis.

There is not sufficient evidence to support the claim that the last digits do not occur with the same relative frequency.



Conclusion

This goodness-of-fit test suggests that the last digits provide a reasonably good fit with the claimed distribution of equally likely frequencies. Instead of asking the subjects how much they weigh, it appears that their weights were actually measured as they should have been.

Example

Table below lists the numbers of games played in the baseball World Series. That table also includes the expected proportions for the numbers of games in a World Series, assuming that in each series, both teams have about the same chance of winning. Use a 0.05 significance level to test the claim that the actual numbers of games fit the distribution indicated by the probabilities.

Games Playe4d	4	5	6	7
Actual World Series Contests	19	21	22	37
Expected Proportion	$\frac{2}{16}$	$\frac{4}{16}$	$\frac{5}{16}$	$\frac{5}{16}$

Solution

Step 1: The original claim: $p_4 = \frac{2}{16},\quad p_5 = \frac{4}{16},\quad p_6 = \frac{5}{16},\quad p_7 = \frac{5}{16}$

Step 2: If the original claim is false, then at least one of the proportions does not have the value as claimed.

Step 3: null hypothesis contains equality

$$H_0:\ p_4 = \frac{2}{16},\quad p_5 = \frac{4}{16},\quad p_6 = \frac{5}{16},\quad p_7 = \frac{5}{16}$$

H_1 : At least one probability is not equal to the given claimed value.

Step 4: The significance level is $\alpha = 0.05$

Step 5: Testing whether the distribution of numbers of games in World Series contests is as claimed, use goodness-of-fit test: χ^2

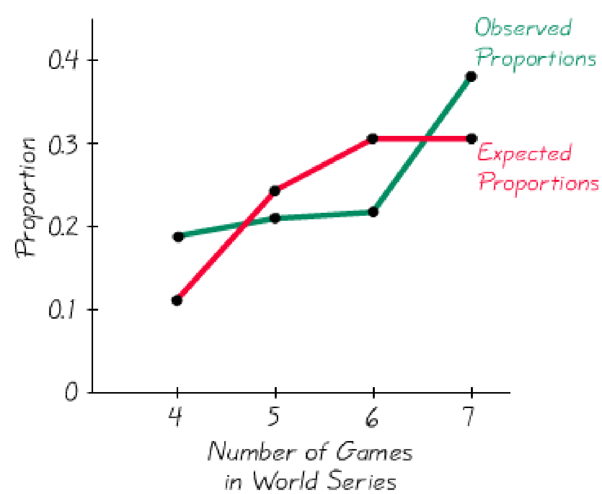
Step 6: Table above shows the computation of the χ^2 test statistic. The test statistic $\chi^2 = 7.885$, using $\alpha = 0.05$ and $k - 1 = 3$ degrees of freedom, the P-value is 0.048

Step 7: The P-value of 0.048 is less than the significance level of 0.05, so there is sufficient evidence to reject the null hypothesis. Also the test statistic of $\chi^2 = 7.885$ is in critical region bounded by the critical value of 7.185.

Step 8: There is sufficient evidence to warrant rejection of the claim that actual numbers of games in World Series contests fit the distribution indicated by the expected proportions.

Conclusion

This goodness-of-fit test suggests that the numbers of games in World Series contest do not fit the distribution expected from probability calculations



Exercises *Section* 4.4 – Goodness-of-Fit

1. A poll typically involves the selection of random digits to be used for telephone numbers. The New York Times states that "within each (telephone) exchange, random digits were added to form a complete telephone number, thus permitting access to listed and unlisted numbers. "When such digits are randomly generated, what is the distribution of those digits? Given such randomly generated digits, what is a test for "goodness-off-fit"?

2. When generating random digits, we can test the generated digits for goodness-of-fit with the distribution in which all of the digits are equally likely. What does an exceptionally large value of the χ^2 test statistic suggest about the goodness-of-fit? What does an exceptionally small value of the χ^2 test statistic (such as 0.002) suggest about the goodness-of-fit?

3. You purchased a slot machine, and tested it by playing it 1197 times. There are 10 different categories of outcome, including no win, win jackpot, win with three bells, and so on. When testing the claim the observed outcomes agree with the expected frequencies, the author obtained a test statistic of $\chi^2 = 8.185$. Use a 0.05 significance level to test the claim that the actual outcomes agree with the expected frequencies. Does the slot machine appear to be functioning as expected?
Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

4. Do "*A*" students tend to sit in a particular part of the classroom? The author recorded the locations of the students who received grades *A*, with these results: 17 sat in the front, 9 sat in the middle, and 5 sat in the back of the classroom. When testing the assumption that the "*A*" students are distributed evenly throughout the room, the author obtained the test statistic of $\chi^2 = 7.226$. If using a 0.05 significance level, is there sufficient evidence to support the claim that the "*A*" students are not evenly distributed throughout the classroom? If so, does that mean you can increase your likelihood of getting an *A* by sitting in the front of the room? Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

5. Randomly selected nonfat occupational injuries and illnesses are categorized according to the day of the week that they first occurred, and the results are listed below. Use a 0.05 significance level to test the claim that such injuries and illness occur with equal frequency on the different days of the week. Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

Day	*Mon*	*Tues*	*Wed*	*Thurs*	*Fri*
Number	23	23	21	21	19

6. Records of randomly selected births were obtained and categorized according to the day of the week that they occurred. Because babies are unfamiliar with our schedule of weekdays, a reasonable claim is that occur on the different days with equal frequency. Use a 0.01 significance level to test that claim. Can you provide an explanation for the result? Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

Day	*Sun*	*Mon*	*Tues*	*Wed*	*Thurs*	*Fri*	*Sat*
Number of births	77	110	124	122	120	123	97

7. The table below lists the frequency of wins for different post positions in the Kentucky Derby horse race. A post position of 1 is closest to the inside rail, so that horse has the shortest distance to run. (Because the number of horses varies from year to year, only the first ten post positions are included.) Use a 0.05 significance level to test the claim that the likelihood of winning is the same for the different post positions. Based on the result, should bettor consider the post position of a horse racing in the Kentucky Derby? Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

Post Position	1	2	3	4	5	6	7	8	9	10
Wins	19	14	11	14	14	7	8	11	5	11

8. The table below lists the cases of violent crimes are randomly selected and categorized by month. Use a 0.01 significance level to test the claim that the rate of violent crime is the same for each month. Can you explain the result? Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

Month	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Number	786	704	835	826	900	868	920	901	856	862	783	797

9. The table below lists the results of the Advanced Placement Biology class conducted genetics experiments with fruit flies. Use a 0.05 significance level to test the claim that the observed frequencies agree with the proportions that were expected according to principles of genetics Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

Characteristic	***Red eye / normal wing***	***Sepia eye / normal wing***	***Red eye / vestigial wing***	***Sepia eye / vestigial wing***
Frequency	59	15	2	4

10. The table below lists the claims that its M&M plain candies are distributed with the following color percentages: 16% green, 20% orange, 14% yellow, 24% blue, 13% red, and 13% brown. Use a 0.05 significance level to test the claim that the color distribution is as claimed.

Green	Orange	Yellow	Blue	Red	Brown
19	25	8	27	13	8

Section 4.5 – Comparing Three or More Means

Analysis of Variance (**ANOVA**) is an inferential method used to test the equality of three or more population means.

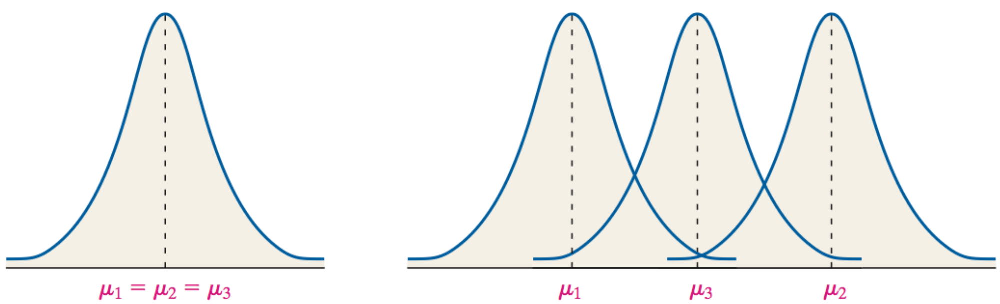
Requirements of a One-Way ANOVA Test

1. There must be k simple random samples; one from each of k populations or a randomized experiment with k treatments.
2. The k samples are independent of each other; that is, the subjects in one group cannot be related in any way to subjects in a second group.
3. The populations are normally distributed.
4. The populations must have the same variance; that is, each treatment group has the population variance σ^2

Testing a Hypothesis Regarding $k = 3$

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : At least one population mean is different from the others



Testing Using ANOVA

The methods of one-way ANOVA are ***robust***, so small departures from the normality requirement will not significantly affect the results of the procedure. In addition, the requirement of equal population variances does not need to be strictly adhered to, especially if the sample size for each treatment group is the same. Therefore, it is worthwhile to design an experiment in which the samples from the populations are roughly equal in size.

Verifying the Requirement of Equal Population Variance

The one-way ANOVA procedures may be used if the largest sample standard deviation is no more than twice the smallest sample standard deviation.

Example

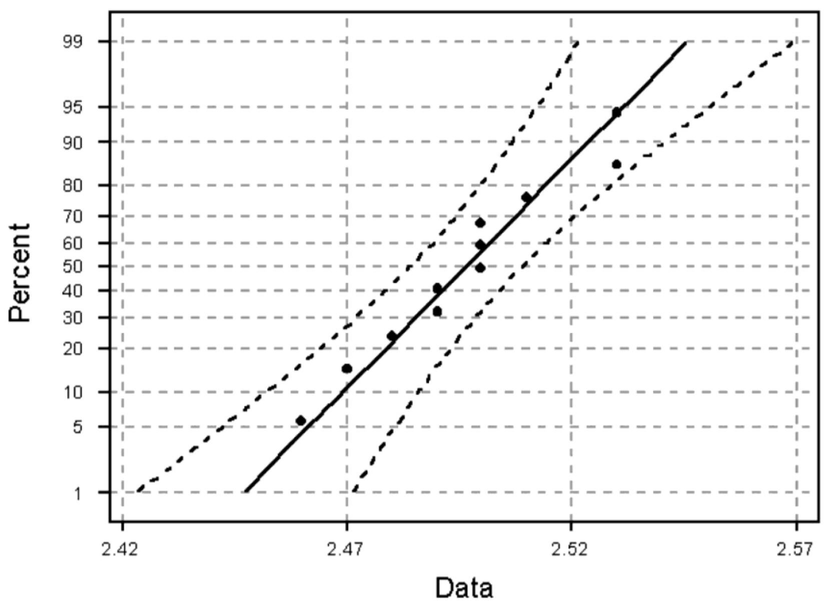
The following data represent the weight (in grams) of pennies minted at the Denver mint in 1990, 1995, and 2000. Verify that the requirements in order to perform a one-way ANOVA are satisfied.

1990	**1995**	**2000**
2.50	2.52	2.50
2.50	2.54	2.48
2.49	2.50	2.49
2.53	2.48	2.50
2.46	2.52	2.48
2.50	2.50	2.52
2.47	2.49	2.51
2.53	2.53	2.49
2.51	2.48	2.51
2.49	2.55	2.50
2.48	2.49	2.52

Solution

1. The 3 samples are simple random samples.
2. The samples were obtained independently.
3. Normal probability plots for the 3 years follow. All of the plots are roughly linear so the normality assumption is satisfied.

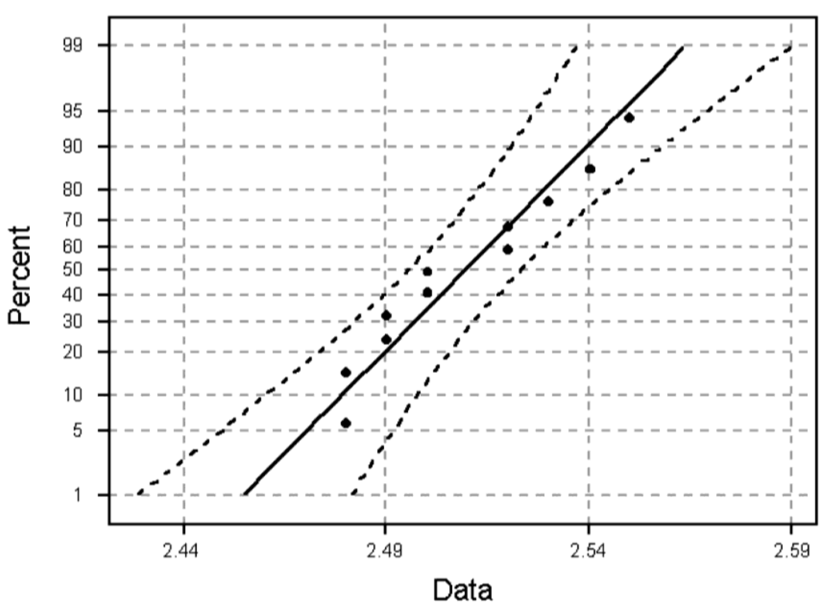
Normal Probability Plot for 1990



Mean: 2.49636

StDev: 0.0210077

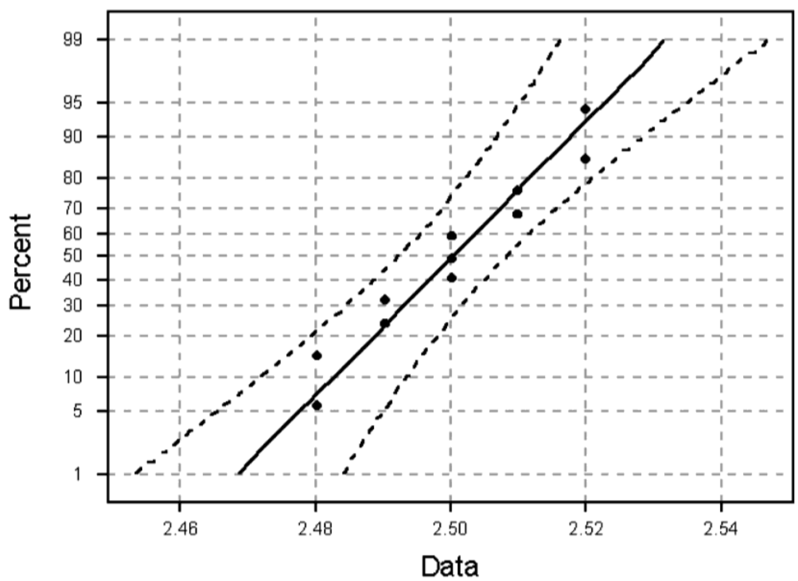
Normal Probability Plot for 1995



Mean: 2.50909

StDev: 0.0231417

Normal Probability Plot for 2000



Mean: 2.5

StDev: 0.0134840

The largest standard deviation is not more than twice the smallest standard deviation (2•0.0141 = 0.0282 > 0.02430) so the requirement of equal population variances is considered satisfied.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
1990	11	2.4964	2.5000	2.4967	0.0220	0.0066
1995	11	2.5091	2.5000	2.5078	0.0243	0.0073
2000	11	2.5000	2.5000	2.5000	0.0141	0.0043

Variable	Minimum	Maximum	Q1	Q3
1990	2.4600	2.5300	2.4800	2.5130
1995	2.4800	2.5500	2.4900	2.5300
2000	2.4800	2.5200	2.4900	2.5100

The basic idea in one-way ANOVA is to determine if the sample data could come from populations with the same mean, μ, or suggests that at least one sample comes from a population whose mean is different from the others.

To make this decision, we compare the variability among the sample means to the variability within each sample.

We call the variability among the sample means the between-sample variability, and the variability of each sample the within-sample variability.

If the ***between-sample variability*** is large relative to the ***within-sample*** variability, we have evidence to suggest that the samples come from populations with different means.

ANOVA *F*-Test Statistic

The analysis of variance *F*-test statistic is given by

$$F_0 = \frac{between-sample\ \text{variablity}}{within-sample\ \text{var}iablity}$$

$$= \frac{mean\ square\ due\ to\ treatments}{mean\ square\ due\ to\ error}$$

$$= \frac{MST}{MSE}$$

Computing the F-Test Statistic

***Step* 1**: Compute the sample mean of the combined data set by adding up all the observations and dividing by the number of observations. Call this value $\bar{x}$.

***Step* 2**: Find the sample mean for each sample (or treatment). Let $\bar{x}_1$ represent the sample mean of sample 1, $\bar{x}_2$ represent the sample mean of sample 2, and so on.

***Step* 3**: Find the sample variance for each sample (or treatment). Let s_1^2 represent the sample variance for sample 1, s_2^2 represent the sample variance for sample 2, and so on.

***Step* 4**: Compute the sum of squares due to treatments, SST, and the sum of squares due to error, SSE.

***Step* 5**: Divide each sum of squares by its corresponding degrees of freedom (k – 1 and n – k, respectively) to obtain the mean squares MST and MSE.

$$MST = \frac{\sum n_i \left(x_i - \bar{x}\right)^2}{k-1} \qquad MSE = \frac{\sum \left(n_i - 1\right) s_i^2}{n-k}$$

***Step* 6**: Compute the *F*-test statistic: $F_0 = \dfrac{mean\ square\ due\ to\ treatments}{mean\ square\ due\ to\ error} = \dfrac{MST}{MSE}$

Example

Compute the *F*-test statistic for the penny data.

1990	**1995**	**2000**
2.50	2.52	2.50
2.50	2.54	2.48
2.49	2.50	2.49
2.53	2.48	2.50
2.46	2.52	2.48
2.50	2.50	2.52
2.47	2.49	2.51
2.53	2.53	2.49
2.51	2.48	2.51
2.49	2.55	2.50
2.48	2.49	2.52

Solution

1. $\bar{x} = \dfrac{2.50+2.50+\cdots+2.50+2.52}{33} = 2.5018$
2. $\bar{x}_{1990} = 2.4964 \quad \bar{x}_{1995} = 2.5091 \quad \bar{x}_{2000} = 2.5$
3. $s_{1990}^2 = \dfrac{(2.50-2.4964)^2 + \cdots + (2.48-2.4964)^2}{11-1} = 0.0005$

   $s_{1995}^2 = \dfrac{(2.52-2.5091)^2 + \cdots + (2.49-2.5091)^2}{11-1} = 0.0006$

$$s_{200}^2 = \frac{(2.50-2.5)^2+\cdots+(2.52-2.5)^2}{11-1} \underline{=0.0002}$$

4. $SST = 11(2.4964-2.5018)^2+11(2.5091-2.5018)^2+11(2.5-2.5018)^2 \underline{=0.0009}$

$SSE = (11-1)(0.0005)+(11-1)(0.0006)+(11-1)(0.0002) \underline{=0.013}$

5. $MST = \frac{SST}{k-1} = \frac{0.0009}{3-1} \underline{=0.0005}$

$MSE = \frac{SSE}{n-k} = \frac{0.013}{33-3} \underline{=0.0004}$

6. $F_0 = \frac{MST}{MSE} = \frac{0.0005}{0.0004} \underline{=1.25}$

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Squares**	***F*-Test Statistic**
Treatment	0.0009	2	0.0005	1.25
Error	0.013	30	0.0004	
Total	0.0139	32		

Exercises Section 4.5 – Comparing Three or More Means

1. Fill in the ANOVA table

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Squares*	*F-Test Statistic*
Treatment	565	5		
Error	3560	32		
Total				

2. Fill in the ANOVA table

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Squares*	*F-Test Statistic*
Treatment	490	4		
Error	7267	21		
Total				

3. Determine the ***F***-test statistic based on the given summary statistics

Population	*Sample Size*	*Sample Mean*	*Sample Variance*
1	10	42	35
2	10	41	40
3	10	22	23

Compute $\overline{x}$, the sample mean of the combined data set, by adding up all the observations and dividing by the number pf the observations.

4. An engineer wants to know if the mean strengths of three concrete mix designs differ significantly. He randomly selects 9 cylinders that measure 6 inches in diameter and 12 inches in heights in which mixture *A* is poured, 9 cylinders of mixture *B*, and 9 cylinders of mixture *C*. After 28 days, he measures the strength (in pounds per square inch) of the cylinders. The results are presented in the table below.

Mixture A	*Mixture B*	*Mixture C*
3,980	4,070	4,130
4,040	4,340	3,820
3,760	4,620	4,020
3,870	3,730	4,150
3,990	4,870	4,190
4,090	4,120	3,840
3,820	4,640	3,750
3,940	4,180	3,990
4,080	3,850	4,320

a) Write the null and alternative hypotheses

b) Explain why the one-way ANOVA cannot be used to test these hypotheses

5. At a community college, the mathematics department has been experimenting with four different delivery mechanisms for content in their Elementary Statistics courses. One method is the traditional lecture (method I), the second is a hybrid format in which half the class time is online and the other half is face-to-face (method II), the third is online (method III), and the fourth is an emporium model from which students obtain their lectures and do their work in a lab with an instructor available for assistance (method IV). To assess the effectiveness of the four methods, students in each approach are given a final exam with the results shown in the accompanying table. Do the data suggest that any method has a different mean score from the others?

Method I	Method II	Method III	Method IV
76	88	78	89
81	52	60	90
85	77	73	79
68	73	70	62
88	64	62	83
73	38	82	75
80	57	74	54
65	63	80	70
60	83	53	80
92	65	46	94
83	78	84	76
51	64	80	78
71	87	78	81
63	92		
71			
65			

a) Write the null and alternative hypotheses

b) State the requirements that must be satisfied to use one-way ANOVA procedure

c) Assuming the requirements stated in part (b) are satisfied, use the following one-way ANOVA table to test the hypothesis of equal means at the $\alpha = 0.05$ level of significance.

d) Interpret the *P*-value.

e) Verify that the residuals are normally distributed

Section 4.6 – Testing the Significance of the Least-Squares Regression Model

Requirement 1 for Inference on the Least-Squares Regression Model

For any particular value of the explanatory variable *x*, the mean of the corresponding responses in the population depends linearly on *x*. That is,

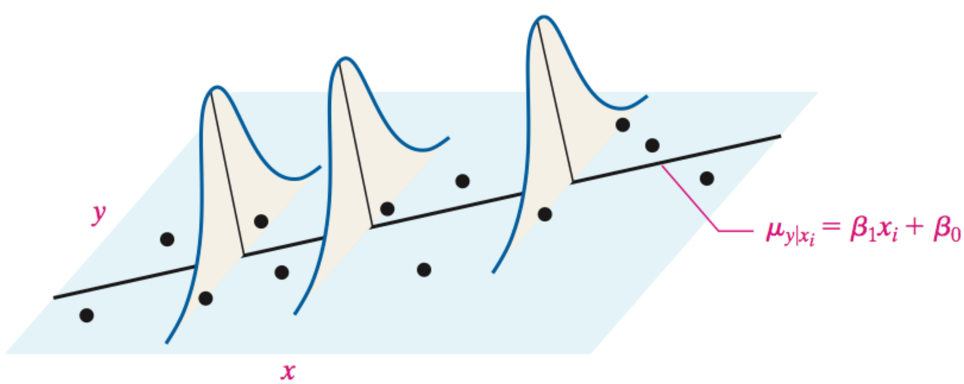
$$\mu_{y|x} = \beta_1 x + \beta_0$$

for some numbers β_0 *and* β_1 , where $\mu_{y|x}$ represents the population mean response when the value of the explanatory variable is *x*.

The response variables are normally distributed with mean $\mu_{y|x} = \beta_1 x + \beta_0$ and standard deviation *σ*.

> When doing inference on the least-squares regression model, we require (1) for any explanatory variable, *x*, the mean of the response variable, *y*, depends on the value of *x* through a linear equation, and (2) the response variable, *y*, is normally distributed with a constant standard deviation, *σ*. The mean increases/decreases at a constant rate depending on the slope, while the standard deviation remains constant.

A large value of *σ*, the population standard deviation, indicates that the data are widely dispersed about the regression line, and a small value of *σ* indicates that the data lie fairly close to the regression line



The least-squares regression model is given by $y_i = \beta_1 x_i + \beta_0 + \varepsilon_i$ where

y_i is the value of the response variable for the i^{th} individual

β_0 *and* β_1 are the parameters to be estimated based on sample data

$\beta_1 x_i$ is the value of the explanatory variable for the i^{th} individual

ε_i is a random error term with mean 0 an variance, the error terms are independent.

i = 1, …, *n*, where *n* is the sample size (number of ordered pairs in the data set)

The standard error of the estimate, s_e , is found using the formula

$$s_e = \sqrt{\frac{\sum\left(y_i - \hat{y}_i\right)^2}{n-2}} = \sqrt{\frac{\sum residuals^2}{n-2}}$$

Example

Compute the standard error of the estimate for the drilling data which is presented

Depth at Which Drilling Begins, x (in ft)	Time to Drill 5 Feet, y (in min)
35	5.88
50	5.99
75	6.74
95	6.1
120	7.47
130	6.93
145	6.42
155	7.97
160	7.92
175	7.62
185	6.89
190	7.9

Solution

***Step* 1:** The least squares regression line to be $\hat{y} = 0.116x + 5.5273$

***Step* 2, 3:** The predicted values as well as the residuals for the 12 observations

Depth, x	Time, y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$
35	5.88	5.9333	-0.0533	0.0028
50	5.99	6.1073	-0.1173	0.0138
75	6.74	6.3973	0.3427	0.1174
95	6.1	6.6293	-0.5293	0.2802
120	7.47	6.9193	0.5507	0.3033
130	6.93	7.0353	-0.1053	0.0111
145	6.42	7.2093	-0.7893	0.6230
155	7.97	7.3253	0.6447	0.4156
160	7.92	7.3833	0.5367	0.2880
175	7.62	7.5573	0.0627	0.0039
185	6.89	7.6733	-0.7833	0.6136
190	7.9	7.7313	0.1687	0.0285
				$\sum \text{residuals}^2 = 2.7012$

***Step* 4**: We find the sum of the squared residuals by summing the last column of the table:

***Step* 5**: The standard error of the estimate is then given by

$$s_e = \sqrt{\frac{\sum residuals^2}{n-2}} = \sqrt{\frac{2.7012}{10}} \underline{= 0.5197}$$

Exercises *Section* 4.6 – Testing the Significance of the Least-Squares Regression Model

Section 4.7 – Confidence and Prediction Intervals

Method for constructing a prediction interval, which is an interval estimate of a predicted value of ***y***

Unexplained, Explained, and Total Deviation

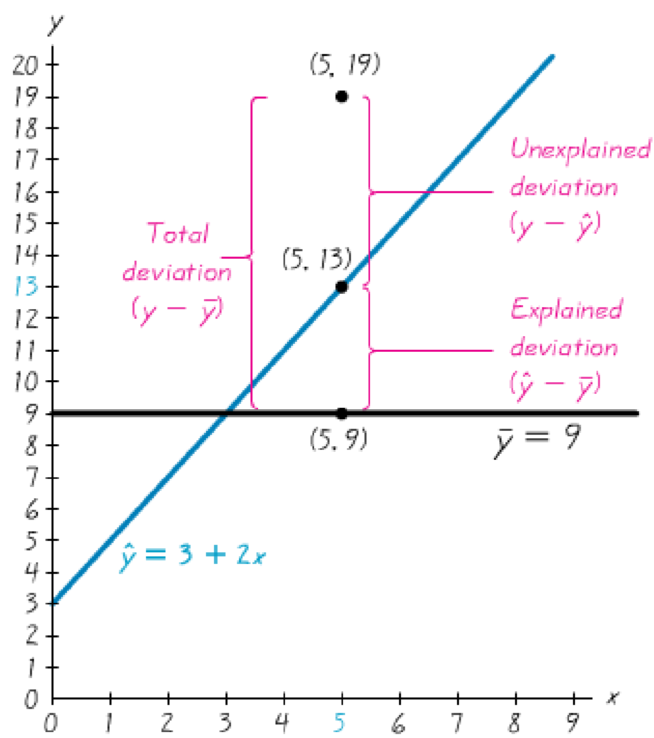
Definitions

Assume that we have a collection of paired data containing the sample point (x, y), that $\hat{y}$ is the predicted value of y (obtained by using the regression equation), and that the mean of the sample y-values is $\bar{y}$.

The ***total deviation*** of (x, y) is the vertical distance $y - \bar{y}$, which is the distance between the point (x, y) and the horizontal line passing through the sample mean $\bar{y}$.

The ***explained deviation*** is the vertical distance $\hat{y} - \bar{y}$, which is the distance between the predicted y-value and the horizontal line passing through the sample mean $\bar{y}$.

The ***unexplained deviation*** is the vertical distance $y - \hat{y}$, which is the vertical distance between the point (x, y) and the regression line. (The distance $y - \hat{y}$ is also called a ***residual***.



- ✓ The mean of the y-value is given by $\bar{y} = 9$
- ✓ One of the pairs of sample data is $x = 5 \quad and \quad y = 19$
- ✓ The point (5, 13) is one of the points on the regression line, because substituting $x = 5$ into the regression equation of $\hat{y} = 3 + 2x$ yields $\hat{y} = 3 + 2(5) = 13$

Formula

$$(\textit{total variation}) = (\textit{explained variation}) + (\textit{unexplained variation})$$

or $$\sum (y-\bar{y})^2 = \sum (\hat{y}-\bar{y})^2 + \sum (y-\hat{y})^2$$

Definition

The coefficient of determination is the amount of the variation in *y* that is explained by the regression line.

$$r^2 = \frac{\textit{explained variation}}{\textit{total variation}}$$

The value of r^2 is the proportion of the variation in *y* that is explained by the linear relationship between *x* and *y*.

Example

We used the paired pizza/subway fare costs to get $r = 0.988$. Find the coefficient of determinant. Also, find the percentage of the total variation in y (subway fare) that can be explained by the linear relationship between the cost of a slice pizza and the cost of a subway fare.

Cost of Pizza	0.15	0.35	1.00	1.25	1.75	2.00
Subway Fare	0.15	0.35	1.00	1.35	1.50	2.00

Solution

The coefficient of determinant is $r^2 = 0.988^2 = 0.976$

Because r^2 is the proportion of total variation that is explained, we conclude that 97.6% of the total variation in subway fares can be explained by the cost of a slice of pizza. This means that 2.4% of the total variation in cost of subway fares can be explained by factors other than the cost of a slice of pizza.

Definition

A ***prediction interval***, is an interval estimate of a predicted value of *y*.

Definition

The ***standard error of estimate***, denoted by s_e is a measure of the differences (or distances) between the observed sample *y*-values and the predicted values $\hat{y}$ that are obtained using the regression equation.

$$s_e = \sqrt{\frac{\sum (y-\hat{y})^2}{n-2}} \qquad (\textit{where } \hat{y} \textit{ is the predicted y-value})$$

Or

$$s_e = \sqrt{\frac{\sum y^2 - b_0 \sum y - b_1 \sum xy}{n-2}}$$

Example

Find the standard error of estimate s_e for the paired pizza/subway fare data listed below.

Cost of Pizza	0.15	0.35	1.00	1.25	1.75	2.00
Subway Fare	0.15	0.35	1.00	1.35	1.50	2.00

Solution

x	y	xy	x^2	y^2
0.15	0.15	0.0225	0.0225	0.0225
0.35	0.35	0.1225	0.1225	0.1225
1	1	1	1	1
1.25	1.35	1.6875	1.5625	1.8225
1.75	1.5	2.625	3.0625	2.25
2	2	4	4	4
6.5	6.35	9.4575	9.77	9.2175

$n = 6, \quad b_0 = 0.034560171, \quad b_1 = 0.94502138$

	Coefficients
Intercept	0.0345602
X Variable	0.9450214

$$s_e = \sqrt{\frac{\sum y^2 - b_0 \sum y - b_1 \sum xy}{n-2}}$$

$$= \sqrt{\frac{9.2175 - (0.034560171)(6.35) - (0.94502138)(9.4575)}{6-2}}$$

$\underline{\approx 0.123}$

Prediction Interval for an Individual *y*

Given the fixed value x_0 the prediction interval for an individual *y* is

$$\hat{y} - E < y < \hat{y} + E$$

Where the margin of error *E* is

$$E = t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{n\left(x_0 - \overline{x}\right)^2}{n\left(\sum x^2\right) - \left(\sum x\right)^2}}$$

And x_0 represents the given value of *x*

$t_{\alpha/2}$ has $n - 2$ degrees of freedom

s_e is given in the previous formula

Example

For the paired pizza/subway fare costs from the Chapter Problem, we have found that for a pizza cost of \$2.25, the best predicted cost of a subway fare is \$2.16. Construct a 95% prediction interval for the cost of a subway fare, given that a slice of pizza costs \$2.25 (so that $x = 2.25$).

Solution

$$s_e = 0.122987$$

$$n = 6, \quad \bar{x} = \frac{6.5}{6} = 1.083333 \quad \sum x = 6.5 \quad \sum x^2 = 9.77$$

$$\alpha = 0.05 \quad \text{(2-tails)}$$

$$t_{\alpha/2} = 2.776 \quad df = 6 - 2 = 4$$

$$E = t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{n\left(x_0 - \bar{x}\right)^2}{n\left(\sum x^2\right) - \left(\sum x\right)^2}}$$

$$= (2.776)(0.122987)\sqrt{1 + \frac{1}{6} + \frac{6(2.25 - 1.0833333)^2}{6(9.77) - (6.5)^2}}$$

$$\approx 0.441$$

With $\hat{y} = 2.16$ *and* $E = 0.441$

$$\hat{y} - E < y < \hat{y} + E$$

$$2.16 - 0.441 < y < 2.16 + 0.441$$

$$\boxed{1.72 < y < 2.60}$$

If the cost of a slice of pizza is \$2.25, we have 95% confidence that the cost of a subway fare is between \$1.72 and \$2.60. That is a fairly large range of possible values, and one major factor contributing to the large range is that the sample size is very small with $n = 6$.

TI-83/84 PLUS The TI-83/84 Plus calculator can be used to find the linear correlation coefficient r, the equation of the regression line, the standard error of estimate s_e, and the coefficient of determination (labeled r^2). Enter the paired data in lists L1 and L2, then press **STAT** and select **TESTS,** and then choose the option **LinRegTTest.** For Xlist enter L1, for Ylist enter L2, use a Freq (frequency) value of 1, and select $\neq$ 0. Scroll down to Calculate, then press the **ENTER** key.

Exercises *Section* 4.7 – Confidence and Prediction Intervals

1. A height of 70 in. is used to find the predicted weight is 180 lb. In your own words, describe a prediction interval in this situation.

2. A height of 70 in. is used to find the predicted weight is 180 lb. What is the major advantage of using a prediction interval instead of the predicted weight of 180 lb.? Why is the terminology of prediction interval used instead of confidence interval?

3. Use the value of the linear correlation $r = 0.873$ to find the coefficient of determination and the percentage of the total variation that can be explained by the linear relationship between the 2 variables

$x = tar\ in\ menthol\ cigarettes$

$y = movie\ gross$ 16	13	16	9	14	13	12		.14	13	13	16	13	13	18
9	19	2	13	14	14	15	16	6.						

1.1	0.8	1	0.9	0.8	0.8	0.8	0.8	0.9	0.8	0.8	1.2	0.8	0.8	1.3
0.7	1.4	0.2	0.8	1	0.8	0.8	1.2	0.6	0.7					

4. Use the value of the linear correlation

5. $r = 0.744$ to find the coefficient of determination and the percentage of the total variation that can be explained by the linear relationship between the 2 variables

$x = movie\ budget$

41	20	116	70	75	52	120	65	6.5	60	125	20	5	150
4.5	7	100	30	225	70	80	40	70	50	74	200	113	68
72	160	68	29	132	40								

117	5	103	66	121	116	101	100	55	104	213	34	12	290
47	10	111	100	322	19	117	48	228	47	17	373	380	118
114	120	101	120	234	209								

6. Use the value of the linear correlation $r = -0.865$ to find the coefficient of determination and the percentage of the total variation that can be explained by the linear relationship between the 2 variables

$x = car\ weight, \quad y = city\ fuel\ consumption\ in\ mi/gal$

7. Use the value of the linear correlation $r = -0.488$ to find the coefficient of determination and the percentage of the total variation that can be explained by the linear relationship between the 2 variables

$x = age\ of\ home, \quad y = home\ selling\ price$

8. Refer to the display obtained by using the paired data consisting of weights (in *lb.*) of 32 cars and their highway fuel consumption amounts (in *mi/gal*). A car weight of 4000 *lb*. to be used for predicting the highway fuel consumption amount

```
The regression equation is
Highway = 50.5 - 0.00587 Weight

Predictor        Coef     SE Coef      T      P
Constant       50.502       2.860  17.66  0.000
Weight     -0.0058685   0.0007859  -7.47  0.000

S = 2.19498   R-Sq = 65.0%   R-Sq(adj) = 63.9%

Predicted Values for New Observations

New
Obs     Fit  SE Fit        95% CI               95% PI
  1  27.028   0.497  (26.013, 28.042)  (22.431, 31.624)

Values of Predictors for New Observations
New
Obs  Weight
  1    4000
```

a) What percentage of the total variation in highway fuel consumption can be explained by the linear correlation between weight and highway fuel consumption?

b) If a car weighs 4000 *lb*., what is the single value that is the best predicted amount of highway fuel consumption? (Assume that there is a linear correlation between weight and highway fuel consumption.)

9. The paired values of the Consumer Price Index (CPI) and the cost of a slice of pizza are shown below

CPI	30.2	48.3	112.3	162.2	191.9	197.8
Cost of Pizza	0.15	0.35	1.00	1.25	1.75	2.00

a) Find the explained variation

b) Find the unexplained variation

c) Find the total variation

d) Find the coefficient of determination

e) Find the standard error of estimate s_e

f) Find the predicted cost of a slice of pizza for the year 2001, when the CPI was 187.1.

g) Find a 95% prediction interval estimate of the cost of a slice of pizza when the CPI was 187.1

In each case, there is sufficient evidence to support a claim of a linear correlation so that it is reasonable to use the regression equation when making predictions.

10. Find the best predicted temperature for a recent year in which the concentration (in parts per million) of CO_2 and temperature (in °C) for different years

CO_2	314	317	320	326	331	339	346	354	361	369
Temperature	13.9	14.0	13.9	14.1	14.0	14.3	14.1	14.5	14.5	14.4

a) Find the explained variation

b) Find the unexplained variation

c) Find the total variation

d) Find the coefficient of determination

e) Find the standard error of estimate s_e

f) Find the predicted temperature (in °C) when CO_2 concentration is 370.9 parts per million.

g) Find a 99% prediction interval estimate temperature (in °C) when CO_2 concentration is 370.9 parts per million

In each case, there is sufficient evidence to support a claim of a linear correlation so that it is reasonable to use the regression equation when making predictions.

Find a prediction interval data listed below.

Cost of Pizza	0.15	0.35	1.00	1.25	1.75	2.00
Subway Fare	0.15	0.35	1.00	1.35	1.50	2.00

11. Using: *Cost of a slice of pizza*: \$2.10; 99% *confidence*

12. Using: *Cost of a slice of pizza*: \$2.10; 90% *confidence*

13. Using: *Cost of a slice of pizza*: \$0.50; 95% *confidence*

14. Using: *Cost of a slice of pizza*: \$0.75; 99% *confidence*