

IIT KHARAGPUR



NPTEL ONLINE
CERTIFICATION COURSES

Business Analytics for Management Decision

Statistical Inference

Rudra P. Pradhan
Vinod Gupta School of Management
IIT KHARAGPUR

NPTEL

Statistical Inference

- Hypothesis Testing
- One-Sample Hypothesis Tests
- Two-Sample Hypothesis Tests
- Analysis of Variance
- Chi-Square Test for Independence

NPTEL

Hypothesis Testing

- Involves drawing inferences about two contrasting propositions (each called a hypothesis) relating to the value of one or more population parameters.
- H_0 Null hypothesis: describes an existing theory
- H_1 Alternative hypothesis: the complement of H_0
- Using sample data, we either:
  - *reject* H_0 and conclude the sample data provides sufficient evidence to support H_1, or
  - *fail to reject* H_0 and conclude the sample data does not support H_1.

NPTEL

Hypothesis Testing

Example 7.1: A Legal Analogy for Hypothesis Testing

- In the U.S. legal system, a defendant is innocent until proven guilty.
- H_0: Innocent
- H_1: Guilty
- If evidence (sample data) strongly indicates the defendant is guilty, then we reject H_0.
- Note that we have not *proven* guilt or innocence!

NPTEL

Hypothesis Testing

Steps of hypothesis testing procedures

1. Identify the population parameter and formulate the hypotheses to test.
2. Select a *level of significance* (related to the risk of drawing an incorrect conclusion).
3. Determine the decision rule on which to base a conclusion.
4. Collect data and calculate a test statistic.
5. Apply the decision rule and draw a conclusion.

NPTEL

One-Sample Hypothesis Tests

Three forms:

1. H_0: parameter = constant
   H_1: parameter ≠ constant
2. H_0: parameter ≤ constant
   H_1: parameter > constant
3. H_0: parameter ≥ constant
   H_1: parameter < constant

The equality part of the hypotheses sign is always in the Null hypothesis.

NPTEL

One-Sample Hypothesis Tests

Example 7.2

Formulating a One-Sample Test of Hypothesis

- CadSoft believes the average response time for technical support requests can be reduced to less than 25 minutes.
- Set up the hypotheses.

H_0: mean response time ≥ 25

H_1: mean response time < 25

	A	B	C	D
1	Technical Support Response Times			
2				
3	Customer	Time (min)		
4	1	20		
5	2	12		
6	3	15		
7	4	11		
8	5	22		
9	6	6		
10	7	39		
11	8	19		
12	9	12		
13	10	13		

Figure 7.1

NPTEL

One-Sample Hypothesis Tests

Understanding Risk in Hypothesis Testing

We always risk drawing an incorrect conclusion.

Four outcomes are possible:

1. H_0 is true and the test correctly fails to reject H_0
2. H_0 is false and the test correctly rejects H_0
3. H_0 is true and the test incorrectly rejects H_0
4. H_0 is false and the test incorrectly fails to reject H_0

Outcome 3 is called a Type I error.

Outcome 4 is called a Type II error.

NPTEL

One-Sample Hypothesis Tests

Understanding Risk in Hypothesis Testing

- The probability of making a Type I error = α
  The probability of making a Type II error = β
- α = $P($ rejecting H_0 | H_0 is true)
  $\beta = P($not rejecting H_0 | H_0 is false)
- The value of α can be controlled.
  α is typically set to 0.01, 0.05, or 0.10.
- The value of β cannot be specified in advance and depends on the value of the (unknown) population parameter.

NPTEL

One-Sample Hypothesis Tests

Example 7.3

How β Depends on the True Population Mean

- In the CadSoft example:
  H_0: mean response time ≥ 25
  H_1: mean response time < 25
- If the true mean was 15, then the sample mean will most likely be less than 25.
- If the true mean is 24, then the sample mean may or may not be less than 25.
- The further away the true mean is from the hypothesized value, the smaller the value of β.

NPTEL

One-Sample Hypothesis Tests

Computing the test statistic

One-sample test on a mean, σ known

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

One-sample test on a mean, σ unknown

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

One-sample test on a proportion

$$z = \frac{\hat{p} - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}}$$

NPTEL

One-Sample Hypothesis Tests

Example 7.4 Computing the Test Statistic

In the CadSoft example, sample data for 44 customers revealed a mean response time of 21.91 minutes and a sample standard deviation of 19.49 minutes.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{21.91 - 25}{19.49/\sqrt{44}} = \frac{-3.09}{2.938} = -1.05$$

t = -1.05 indicates that the sample mean of 21.91 is 1.05 standard errors below the hypothesized mean of 25 minutes.

NPTEL

One-Sample Hypothesis Tests

Drawing Conclusions in a Two-tailed Test

H_0: parameter = constant

H_1: parameter ≠ constant

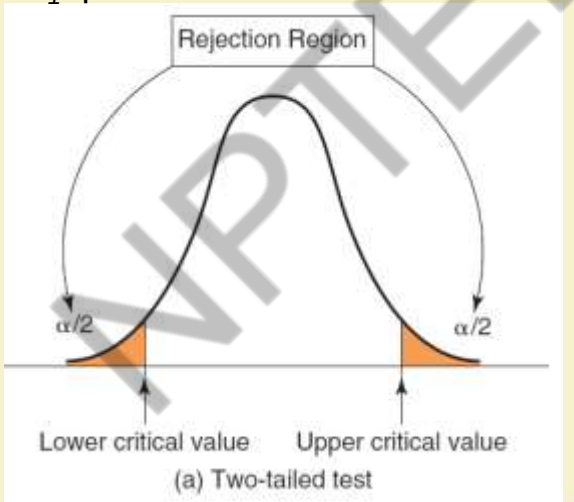


Figure 7.2a

NPTEL

One-Sample Hypothesis Tests

Drawing Conclusions in One-tailed tests

H_0: parameter ≥ constant
H_1: parameter < constant

H_0: parameter ≤ constant
H_1: parameter > constant

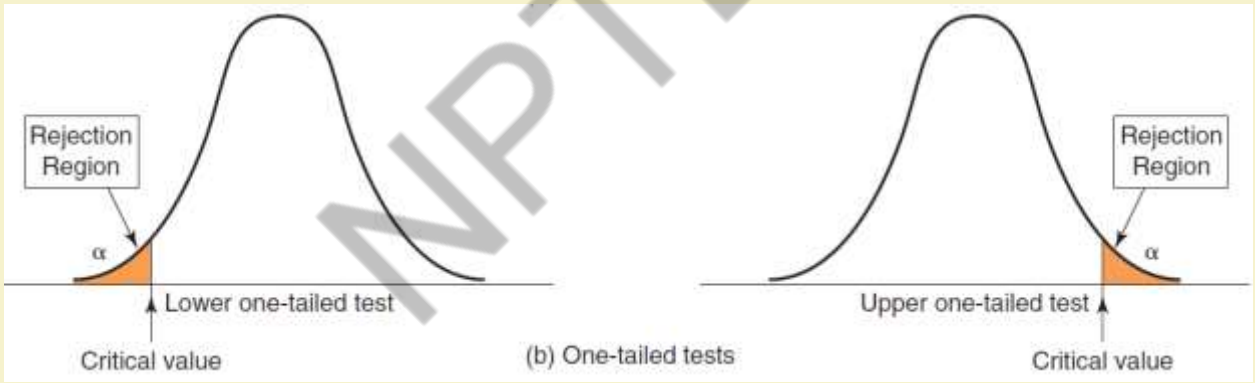


Figure 7.2b

NPTEL

One-Sample Hypothesis Tests

Example 7.5

Finding the Critical Value and Drawing a Conclusion

In the CadSoft example, suppose we set $\alpha = 0.05$.

H_0: mean response time ≥ 25
H_1: mean response time < 25

$n = 44$

$df = n - 1 = 43$

$t = -1.05$

Critical value = $t_{\alpha/2,\, n-1}$ = T.INV(1− α , n −1)

= T.INV(0.95, 43)

= 1.68

NPTEL

One-Sample Hypothesis Tests

Example 7.5 (continued)

Finding the Critical Value and Drawing a Conclusion

t = -1.05 does not fall in the rejection region.
Fail to reject H_0.

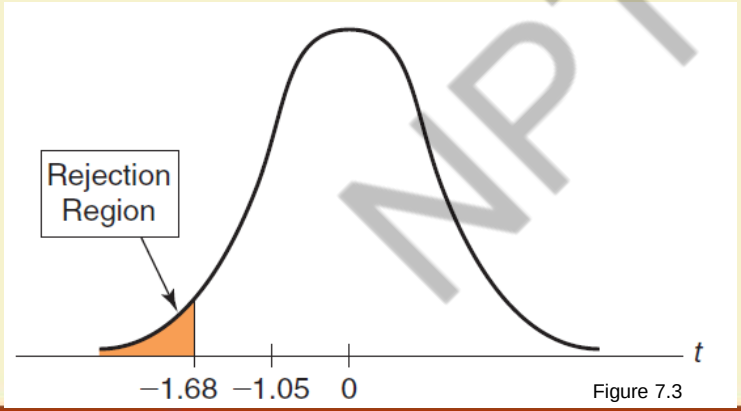


Figure 7.3

Even though the sample mean of 21.91 is well below 25, we have too much sampling error to conclude the that the true population mean is less than 25 minutes.

NPTEL

One-Sample Hypothesis Tests

p-Values

- An alternative approach to Step 3 of any hypothesis test (setting up a decision rule) uses the *p*-value rather than the critical value.
- The *p*-value is the observed significance level.
- The *p*-value decision rule is to:

  Reject H_0 if the *p*-value < α

NPTEL

One-Sample Hypothesis Tests

Example 7.6 Using *p*-Values

- In the CadSoft example, the *p*-value is the left tail area of the observed test statistic, $t = -1.05$.
- *p*-value =T.DIST(*test statistic, df, true*)

  =TDIST(-1.05, 43, true)

  = 0.1498
- Do not reject H_0 because the:

  p-value is not less than α

  0.1498 is not less than 0.05

NPTEL

One-Sample Hypothesis Tests

Example 7.7 Conducting a Two-Tailed Hypothesis Test for the Mean

Use the data set *Vacation Survey* to test whether the average age of respondents is equal to 35.

Test at a 5% significance level.

	A	B	C	D	E
1	Vacation Survey				
2					
3	Age	Gender	Relationship Status	Number of Vacations per Year	Number of Children
4	24	Male	Married	2	0
5	26	Female	Married	4	0
6	28	Male	Married	2	2
7	33	Male	Married	4	0
8	45	Male	Married	2	0

Figure 7.4

For the 34 respondents we find a sample mean of 38.677 and sample standard deviation of 7.858.

NPTEL

One-Sample Hypothesis Tests

Example 7.7 (continued) Conducting a Two-Tailed Hypothesis Test for the Mean

H_0: mean age = 35
H_1: mean age ≠ 35

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{(38.677 - 35)}{(7.858/\sqrt{34})} = 2.73$$

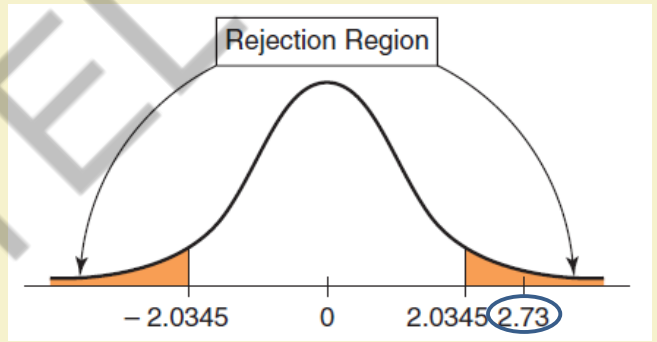


Figure 7.5

Critical value = T.INV.2T(α, *df*)
= T.INV.2T(.05, 33) = 2.0345
p-value = T.DIST.2T(2.69, 33) = 0.0111
Reject H_0.

NPTEL

One-Sample Hypothesis Tests

Example 7.8 One-Sample Test for the Proportion

- CadSoft sampled 44 customers and asked them to rate the overall quality of a software package.
- Sample data revealed that 35 respondents thought the software was very good or excellent.
- In the past, 75% of customers fell into that category.
- At the 5% significance level, test to see whether the percentage of customers in that category is now higher than 75%.

NPTEL

One-Sample Hypothesis Tests

Example 7.8 (continued)

One-Sample Test for the Proportion

H_0: proportion < 0.75

H_1: proportion ≥ 0.75

$$z = \frac{\hat{p} - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}} = \frac{0.795 - 0.75}{\sqrt{0.75(1 - 0.75)/44}} = 0.69$$

Critical value = NORM.S.INV(0.95) = 1.645

p-value = 1 − NORM.S.DIST(0.69, true) = 0.24

Do not reject H_0.

NPTEL

Two-Sample Hypothesis Tests

Two-Sample Hypothesis Tests Available in Excel's *Analysis Toolpak*

Type of Test	Excel Procedure
Two-sample test for means, σ^2 known	Excel *z*-test: Two-sample for means
Two-sample test for means, σ^2 unknown, assumed unequal	Excel *t*-test: Two-sample assuming unequal variances
Two-sample test for means, σ^2 unknown, assumed equal	Excel *t*-test: Two-sample assuming equal variances
Paired two-sample test for means	Excel *t*-test: Paired two-sample for means
Two-sample test for equality of variances	Excel *F*-test Two-sample for variances

Table 7.1

NPTEL

Two-Sample Hypothesis Tests

Example 7.9 Comparing Supplier Performance

Purchase Orders database

Set up a hypothesis test for determining if the mean lead time for Alum Sheeting (μ_1) is greater than the mean lead time for Durrable Products (μ_2).

	A	B	C	D	E	F	G	H	I	J	K
1	Purchase Orders										
2											
3	Supplier	Order No	Item No.	Item Description	Item Cost	Quantity	Cost per order	A/P Terms (Month	Order Date	Arrival Date	Lead Time
4	Alum Sheeting	A0443	1243	Airframe fasteners	$ 4.25	10,000	$ 42,500.00	30	08/08/11	08/14/11	6
5	Alum Sheeting	B0247	1243	Airframe fasteners	$ 4.25	9,000	$ 38,250.00	30	09/05/11	09/12/11	7
6	Alum Sheeting	B0567	1243	Airframe fasteners	$ 4.25	10,500	$ 44,625.00	30	10/10/11	10/17/11	7
7	Alum Sheeting	A0223	4224	Bolt-nut package	$ 3.95	4,500	$ 17,775.00	30	10/15/11	10/20/11	5
8	Alum Sheeting	A0433	5417	Control Panel	$255.00	500	$ 127,500.00	30	10/20/11	10/27/11	7
9	Alum Sheeting	A0446	5417	Control Panel	$255.00	406	$ 103,530.00	30	09/01/11	09/10/11	9
10	Alum Sheeting	B0447	5634	Side Panel	$185.00	150	$ 27,750.00	30	10/25/11	11/03/11	9
11	Alum Sheeting	B0479	5634	Side Panel	$185.00	140	$ 25,900.00	30	10/29/11	11/04/11	6

Figure 7.6

NPTEL

Two-Sample Hypothesis Tests

Example 7.9 (continued)
Comparing Supplier Performance

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_1: \mu_1 - \mu_2 > 0$$

PivotTable shows:

$$\bar{x}_1 = 7.00$$
$$\bar{x}_2 = 4.92$$

	A	B
3	**Row Labels**	**Average of Lead Time**
4	Alum Sheeting	7.00
5	Durrable Products	4.92
6	Fast-Tie Aerospace	8.47
7	Hulkey Fasteners	6.47
8	Manley Valve	6.45
9	Pylon Accessories	8.00
10	Spacetime Technologies	15.25
11	Steelpin Inc.	10.20
12	**Grand Total**	**8.41**

Figure 7.7

NPTEL

Two-Sample Hypothesis Tests

Excel's Two-sample Hypothesis Tests on Means

- If the population variances are known, use:

  z-Test: Two-Sample for Means

- If the population variances are unknown then, in most situations, use:

  t-Test: Two-Sample Assuming Unequal Variances

- If the population variances are unknown but assumed to be equal, then use:

  t-Test: Two-Sample Assuming Equal Variances

NPTEL

Two-Sample Hypothesis Tests

Cautions when Interpreting Excel's Two-sample Hypothesis Test Output

- If the test statistic is negative, Excel's one-tail *p*-value is correct for a lower tail test. For an upper tail test, subtract the reported *p*-value from 1
- The reverse is true when the test statistic is non-negative
- The critical value reported in Excel should be multiplied by -1 if $\overline{x}_1 - \overline{x}_2 < 0$

NPTEL

Copyright © 2013 Pearson Education, Inc. publishing as Prentice Hall

Two-Sample Hypothesis Tests

Example 7.10 Testing the Hypothesis for Supplier Lead-Time Performance

Data

Data Analysis

t-Test: Two-Sample Assuming Unequal Variances

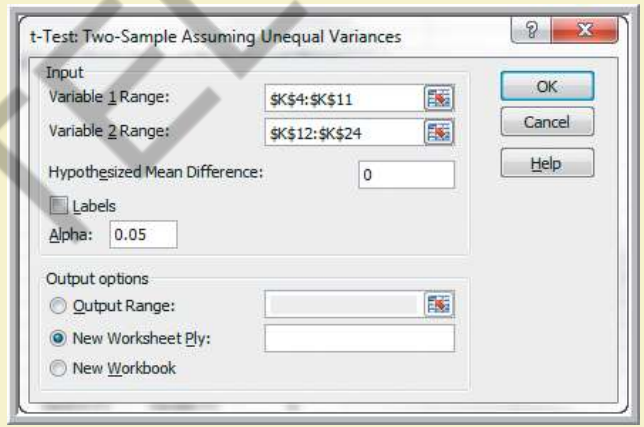


Figure 7.8

Variable 1 Range: Alum Sheeting data

Variable 2 Range: Durrable Products data

NPTEL

Two-Sample Hypothesis Tests

Example 7.10 (continued) Testing the Hypothesis for Supplier Lead-Time Performance

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_1: \mu_1 - \mu_2 > 0$$

$t = 3.83$

Critical value = 1.81

p-value = 0.00166

Reject H_0.

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		*Variable 1*	*Variable 2*
4	Mean	7.0	4.923076923
5	Variance	2.0	0.576923077
6	Observations	8.0	13
7	Hypothesized Mean Difference	0.0	
8	df	10.0	
9	t Stat	3.827958507	
10	P(T<=t) one-tail	0.001664976	
11	t Critical one-tail	1.812461123	
12	P(T<=t) two-tail	0.003329952	
13	t Critical two-tail	2.228138852	

Figure 7.9

NPTEL

Two-Sample Hypothesis Tests

Example 7.11

Using the Paired Two-Sample Test for Means

Use the *Pile Foundation* data to test for a difference in the means of the estimated and actual pile lengths.

Use a 5% significance level.

H_0: $\mu_D = 0$

H_1: $\mu_D \neq 0$

where μ_D is the difference between the 2 population means.

	A	B	C
1	Pile Foundation Data		
2			
3	Pile	Estimated	Actual
4	Number	Pile Length (ft.)	Pile Length (ft.)
5	1	10.58	18.58
6	2	10.58	18.58
7	3	10.58	18.58
8	4	10.58	18.58
9	5	10.58	28.58
10	6	10.58	26.58
11	7	10.58	17.58
12	8	10.58	27.58
13	9	10.58	27.58
14	10	10.58	37.58

Figure 7.10

NPTEL

Two-Sample Hypothesis Tests

Example 7.11 (continued)

Using the Paired Two-Sample Test for Means

t-Test: Paired Two-Sample for Means

$\overline{x}_1 = 28.18$

$\overline{x}_2 = 34.56$

$t = -10.91$

Critical value = -1.97

p-value = 1.119×10^{-23}

Reject H_0.

	A	B	C
1	t-Test: Paired Two Sample for Means		
2			
3		*Variable 1*	*Variable 2*
4	Mean	28.177556	34.55624
5	Variance	255.81004	267.0113
6	Observations	311	311
7	Pearson Correlation	0.7969284	
8	Hypothesized Mean Difference	0	
9	df	310	
10	t Stat	-10.91225	
11	P(T<=t) one-tail	5.594E-24	
12	t Critical one-tail	1.6497838	
13	P(T<=t) two-tail	1.119E-23	
14	t Critical two-tail	1.9676459	

Figure 7.11

NPTEL

Two-Sample Hypothesis Tests

Test for Equality of Variances

Understanding variation in business processes is very important.
We can test for equal variability of two groups:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad \text{or} \quad \frac{\sigma_1^2}{\sigma_2^2} = 1$$
$$H_0: \sigma_1^2 \neq \sigma_2^2$$

The ratio of two population variances follow an *F* probability distribution.

The test statistic is $F = \frac{s_1^2}{s_2^2}$

NPTEL

Two-Sample Hypothesis Tests

Example 7.12

Applying the *F*-Test for Equality of Variances

Using $\alpha = 0.05$, test for equal variances in the lead times of Alum Sheeting and Durrable Products.

$s_1^2 = 2.0$

$s_2^2 = 0.577$

$F = 3.47$

$p\text{-value} = 2(0.0286)$
$= 0.0572$

Fail to reject H_0.

	A	B	C
1	F-Test Two-Sample for Variances		
2			
3		*Variable 1*	*Variable 2*
4	Mean	7	4.923076923
5	Variance	2	0.576923077
6	Observations	8	13
7	df	7	12
8	F	3.466666667	
9	P(F<=f) one-tail	0.028595441	
10	F Critical one-tail	3.606514642	

Figure 7.12

NPTEL

Analysis of Variance

ANOVA

- Used to compare the means of two or more population groups.

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_m$$

H_1: at least one mean is different from the others

- ANOVA derives its name from the fact that we are analyzing variances in the data.
- ANOVA measures variation between groups relative to variation within groups.
- Each of the population groups is assumed to come from a normally distributed population.

NPTEL

Analysis of Variance

Example 7.13

Differences in *Insurance Survey* Data

Average health insurance satisfaction scores (1-5 scale) by education level

	College Graduate	Graduate Degree	Some College
	5	3	4
	3	4	1
	5	5	4
	3	5	2
	3	5	3
	3	4	4
	3	5	4
	4	5	
	2		
Average	3.444	4.500	3.143
Count	9	8	7

NPTEL

Analysis of Variance

Example 7.14 Applying the Excel ANOVA Tool

(*Insurance Survey* data)

Data
Data Analysis
ANOVA: Single Factor

Input Range: contiguous columns of insurance satisfaction data

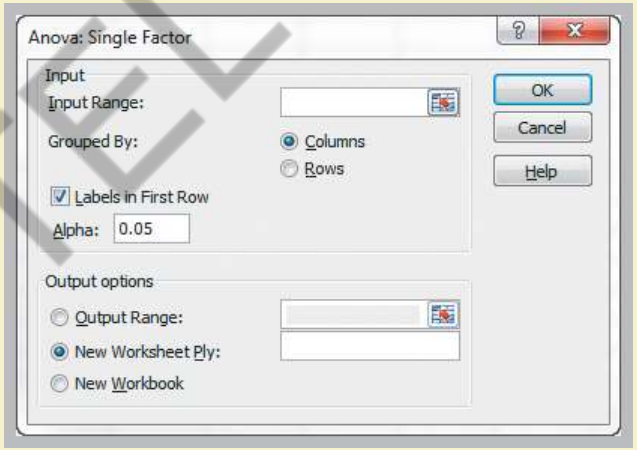


Figure 7.13

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

H_1: at least one mean is different from the others

NPTEL

Analysis of Variance

Example 7.14 (continued)

Applying the Excel ANOVA Tool

Use α = 5%.

$F > F_{crit}$

p-value
= 0.0356

Reject H_0.

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	College Graduate	9	31	3.444444444	1.027777778		
6	Graduate Degree	8	36	4.5	0.571428571		
7	Some College	7	22	3.142857143	1.476190476		
8							
9							
10	ANOVA						
11	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
12	Between Groups	7.878968254	2	3.939484127	3.924651732	0.035635398	3.466800112
13	Within Groups	21.07936508	21	1.003779289			
14							
15	Total	28.95833333	23				

Figure 7.14

NPTEL

Chi-Square Test for Independence

- Test for independence of 2 categorical variables.
- Compute expected frequencies and compare to the observed frequencies in the sample data.
- H_0: two categorical variables are independent

H_1: two categorical variables are dependent

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

where f_o are observed frequencies and and f_e are expected frequencies.

$$\text{expected frequency in row } i \text{ and column } j = \frac{(\text{grand total row } i)(\text{grand total column } j)}{\text{total number of observations}}$$

NPTEL

Chi-Square Test for Independence

Example 7.15

Independence and Marketing Strategy

Using a 5% significance level test, determine whether gender and brand preference for energy drinks can be considered independent variables.

	A	B	C	D	E	F	G	H	I
1	Energy Drink Survey								
2									
3	Respondent	Gender	Brand Preference						
4	1	Male	Brand 3		Count of Respondent	Column Labels			
5	2	Female	Brand 3		Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
6	3	Male	Brand 3		Female	9	6	22	37
7	4	Male	Brand 1		Male	25	17	21	63
8	5	Male	Brand 1		Grand Total	34	23	43	100
9	6	Female	Brand 2						
10	7	Male	Brand 2						
11	8	Female	Brand 2						
12	9	Male	Brand 1						
13	10	Female	Brand 3						

Figure 7.15

NPTEL

Chi-Square Test for Independence

Example 7.16 Computing Expected Frequencies

$$\text{expected frequency in row } i \text{ and column } j = \frac{(\text{grand total row } i)(\text{grand total column } j)}{\text{total number of observations}}$$

Cell F12:

$f_e = \frac{(37)(34)}{100}$

=16* F8/18

	E	F	G	H	I
4	Count of Respondent	Column Labels			
5	Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
6	Female	9	6	22	37
7	Male	25	17	21	63
8	Grand Total	34	23	43	100
9					
10					
11	Expected Frequency	Brand 1	Brand 2	Brand 3	Grand Total
12	Female	12.58	8.51	15.91	37
13	Male	21.42	14.49	27.09	63
14	Grand Total	34	23	43	100

Figure 7.16

NPTEL

Chi-Square Test for Independence

Example 7.17 Conducting the Chi-Square Test

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Cell F18:
$(f_o - f_e)^2/f_e$ = 1.018

Cell 120:
χ^2 = 6.49
=SUM(F18:H19)

	E	F	G	H	I
4	Count of Respondent	Column Labels			
5	Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
6	Female	9	6	22	37
7	Male	25	17	21	63
8	Grand Total	34	23	43	100
9					
10					
11	Expected Frequency	Brand 1	Brand 2	Brand 3	Grand Total
12	Female	12.58	8.51	15.91	37
13	Male	21.42	14.49	27.09	63
14	Grand Total	34	23	43	100
15					
16					
17	Chi Square Statistic	Brand 1	Brand 2	Brand 3	Grand Total
18	Female	1.018791733	0.7403	2.3311	4.0902278
19	Male	0.598338002	0.4348	1.3691	2.40219728
20	Grand Total	1.617129735	1.1751	3.7002	6.49242508
21					
22		Chi-square critical value			5.99146455
23		p-value			0.03892134

Figure 7.17

NPTEL

Chi-Square Test for Independence

Example 7.17 (continued)
Conducting the Chi-Square Test (at α = 5%)

Critical Value Approach:
df = (r - 1)(c-1)
= (2 - 1)(3 - 1) = 2

Cell 122:
= CHISQ.INV.RT(α, *df*)
= CHISQ.INV.RT(0.5, 2)
Critical value = 5.99
$\chi^2 = 6.49 > 5.99$
Reject H_0.

	E	F	G	H	I
4	Count of Respondent	Column Labels			
5	Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
6	Female	9	6	22	37
7	Male	25	17	21	63
8	Grand Total	34	23	43	100
9					
10					
11	Expected Frequency	Brand 1	Brand 2	Brand 3	Grand Total
12	Female	12.58	8.51	15.91	37
13	Male	21.42	14.49	27.09	63
14	Grand Total	34	23	43	100
15					
16					
17	Chi Square Statistic	Brand 1	Brand 2	Brand 3	Grand Total
18	Female	1.018791733	0.7403	2.3311	4.0902278
19	Male	0.598338002	0.4348	1.3691	2.40219728
20	Grand Total	1.617129735	1.1751	3.7002	6.49242508
21					
22		Chi-square critical value			5.99146455
23		p-value			0.03892134

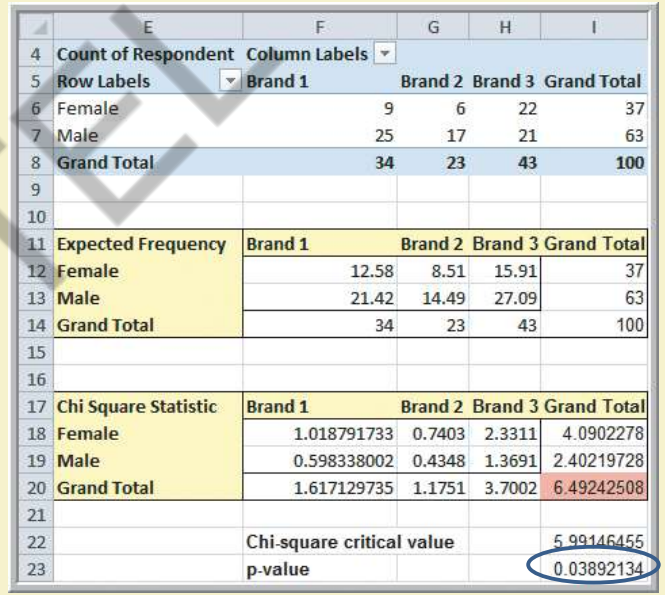
Figure 7.17

NPTEL

Chi-Square Test for Independence

Example 7.17 (continued)
Conducting the Chi-Square Test (at α = 5%)

p-Value Approach:
Cell 123:
=CHISQ.TEST(f_o, f_e)
=CHISQ.TEST(F6:H7,
F12:H13)
p-value = 0.0389
Reject H_0.



	E	F	G	H	I
4	Count of Respondent	Column Labels			
5	Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
6	Female	9	6	22	37
7	Male	25	17	21	63
8	Grand Total	34	23	43	100
9					
10					
11	Expected Frequency	Brand 1	Brand 2	Brand 3	Grand Total
12	Female	12.58	8.51	15.91	37
13	Male	21.42	14.49	27.09	63
14	Grand Total	34	23	43	100
15					
16					
17	Chi Square Statistic	Brand 1	Brand 2	Brand 3	Grand Total
18	Female	1.018791733	0.7403	2.3311	4.0902278
19	Male	0.598338002	0.4348	1.3691	2.40219728
20	Grand Total	1.617129735	1.1751	3.7002	6.49242508
21					
22		Chi-square critical value			5.99146455
23		p-value			0.03892134

Figure 7.17

NPTEL

Two-Sample Hypothesis Tests



Analytics in Practice:

Using Hypothesis Tests and Business Analytics in a Help Desk Service Improvement Project

- Schlumberger International is an international oilfield-services provider headquartered in Houston, Texas.
- Their help desk moved from Ecuador to Houston after two-sample hypothesis tests supported it.
- Use of various analytical techniques enabled their help desk service to improve dramatically.

NPTEL

Keywords

- Alternative hypothesis
- Analysis of variance (ANOVA)
- Chi-square distribution
- Chi-square statistic
- Confidence coefficient
- Factor
- Hypothesis
- Level of significance
- Null hypothesis

NPTEL

Keywords (continued)

- One-sample hypothesis tests
- One-tailed test of hypothesis
- *p*-value (observed significance level)
- Power of the test
- Statistical inference
- Two-tailed tests of hypothesis
- Type I error
- Type II error

NPTEL

Case Study: Performance Lawn Equipment (7)

- Recall that PLE produces lawnmowers and a medium size diesel power lawn tractor.
- Look into differences in customer satisfaction by region and type of satisfaction measure.
- Determine whether on-time deliveries, defectives, and process costs have improved.
- Investigate employee retention as a function of gender, college graduation, and home location.
- Write a formal report summarizing your results.

NPTEL

Books to read

- Business Analytics by J R Evans
- Business Analytics for Managers by H N Laursen

NPTEL

Thank You!!

NPTEL