Pearson

Chapter 15: Chi-square tests

Learning objectives
After studying this chapter you should be able to:
1. identify how and when to use the chi-square test for contingency tables
2. use the Marascuilo procedure for determining pairwise differences when evaluating more than two proportions
3. apply the chi-square test of independence
4. use the chi-square test to evaluate the goodness of fit of a set of data to a specific probability distribution
5. use the chi-square distribution to test for the population variance or standard deviation

15.1 (a) For df = 12 and α = 0.01, χ^2 = 26.217
(b) For df = 23 and α = 0.025, χ^2 = 38.076
(c) For df = 24 and α = 0.05, χ^2 = 36.415

15.2 (a) For df = 4 and α = 0.95, χ^2 = 0.711
(b) For df = 26 and α = 0.975, χ^2 = 13.844
(c) For df = 15 and α = 0.99, χ^2 = 5.229

15.3 (a), (b)

	A	**B**	**Total**
1	o_i = 20 e_i = 20 Chi-sq contrib= 0	o_i = 30 e_i = 30 Chi-sq contrib= 0	50
2	o_i = 30 e_i = 30 Chi-sq contrib= 0	o_i = 45 e_i = 45 Chi-sq contrib= 0	75
Total	50	75	125

(c) $\chi^2_{STAT} = \sum_{\text{All Cells}} \frac{(f_0 - f_e)^2}{f_e}$ = 0 + 0 + 0 + 0 = 0. Since χ^2_{STAT} < 3.841, it is not significant at the 5% level of significance with 1 df.

15.4 (a)

	A	**B**	**Total**
1	o_i = 15 e_i = 13.86	o_i = 24 e_i = 25.14	39
2	o_i = 12 e_i = 13.14	o_i = 25 e_i = 23.86	37
Total	27	49	76

(b) Decision rule: If χ^2 > 6.635, reject H_0 at the 1% significance level with 1 *df.*

Test statistic: $\chi^2_{STAT} = \sum_{\text{All Cells}} \frac{(f_0 - f_e)^2}{f_e}$ = 0.298833

Decision: Since $\chi^2_{S \quad T}$ = 0.29883 is not greater than the critical value of 6.635, it is not significant at the 1% level of significance.

15.5 (a) Excel output:

Observed Frequencies			
	Gender		
Prefer black cars?	**Men**	**Women**	**Total**
Yes	**13**	**5**	**18**
No	**37**	**45**	**82**
Total	**50**	**50**	**100**

Expected Frequencies			
	Gender		
Prefer black cars?	Men	Women	Total
Yes	9	9	18
No	41	41	82
Total	50	50	100

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.841459**
Chi-Square Test Statistic	**4.336043**
***p*-Value**	**0.037314**
Reject the null hypothesis	

Expected frequency assumption is met.

H_0: $\pi_1 = \pi_2$ H_1: $\pi_1 \neq \pi_2$ where population: 1 = males, 2 = females

Decision rule: df = 1. If $\chi^2_{STAT} > 3.8415$, reject H_0 at 5% significance level.

Test statistic: χ^2_{STAT} = 4.3360

Conclusion: Since χ^2_{STAT} = 4.3360 is greater than the upper critical bound of 3.8415, reject H_0. There is evidence to conclude that there is significant difference between the proportions of males and females who prefer black cars at the 5% level of significance.

(b) Excel output:

Observed Frequencies			
	Gender		
Prefer black cars?	**Men**	**Women**	**Total**
Yes	**52**	**20**	**72**
No	**148**	**180**	**328**
Total	**200**	**200**	**400**

Expected Frequencies			
	Gender		
Prefer black cars?	Men	Women	Total
Yes	36	36	72
No	164	164	328
Total	200	200	400

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.841459**
Chi-Square Test Statistic	**17.34417**
***p*-Value**	**3.12E-05**
Reject the null hypothesis	

Expected frequency assumption is met.

H_0: $\pi_1 = \pi_2$ H_1: $\pi_1 \neq \pi_2$ where population 1 = males, 2 = females

Decision rule: df = 1. If χ^2_{STAT} > 3.8415, reject H_0 at 5% significance level.

Test statistic: χ^2_{STAT} = 17.3442

Conclusion: Since χ^2_{STAT} = 17.3442 is greater than the upper critical bound of 3.8415, reject H_0. There is sufficient evidence to conclude that there is significant difference between the proportions of males and females who prefer black cars at the 5% level of significance.

(c) The chi-square test statistic is more likely to be significant for larger sample sizes, as demonstrated in the increase in the test statistic value between (a) and (b).

15.6 (a) Excel output:

Observed Frequencies			
	Country of Origin		
Own shares?	**China**	**Japan**	**Total**
Yes	**14**	**36**	**50**
No	**320**	**80**	**400**
Total	**334**	**116**	**450**

Expected Frequencies			
	Country of Origin		
Own shares?	China	Japan	Total
Yes	37.1111	12.8889	50
No	296.8889	103.1111	400
Total	334	116	450

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.8415**
Chi-Square Test Statistic	**62.8123**
***p*-value**	**0.0000**
Reject the null hypothesis	

H_0: $\pi_1 = \pi_2$ H_1: $\pi_1 \neq \pi_2$ where population: 1 = China, 2 = Japan

Decision rule: *df* = 1. If χ^2_{STAT} > 3.8415, reject H_0 at 5% significance level.

Test statistic: χ^2_{STAT} = 62.8123

Conclusion: Since χ^2_{STAT} = 62.8123 is greater than the upper critical bound of 3.8415, reject H_0. There is sufficient evidence to conclude that there is significant difference in share ownership between business executives in China and Japan at the 5% level of significance.

(b) The *p*-value is 0.000. The probability of obtaining a test statistic of 62.8123 or larger when the null hypothesis is true is about 0.000.

(c) We can conclude that there is overwhelming evidence in this sample to reject the null hypothesis, that business executives in both countries are equally likely to own shares on the stock market.

15.7 Excel output:

Observed Frequencies			
	Workers		
Mode of Negotiation	**Blue-collar**	**White-collar**	**Total**
Collective Bargaining	**112**	**23**	**135**
Individual Contracts	**53**	**25**	**78**
Total	**165**	**48**	**213**

Expected Frequencies			
	Workers		
Mode of Negotiation	Blue-collar	White-collar	Total
Collective Bargaining	104.5775	30.4225	135
Individual Contracts	60.4225	17.5775	78
Total	165	48	213

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.8415**
Chi-Square Test Statistic	**6.3840**
***p*-value**	**0.0115**
Reject the null hypothesis	

H_0: $\pi_1 = \pi_2$ H_1: $\pi_1 \neq \pi_2$ where population: 1 = blue-collar, 2 = white-collar

Decision rule: *df* = 1. If χ^2_{STAT} > 3.8415, reject H_0 at 5% significance level.

Test statistic: χ^2_{STAT} = 6.3840

Conclusion: Since χ^2_{STAT} = 6.3840 is greater than the upper critical bound of 3.8415, reject H_0. There is sufficient evidence to conclude that there is significant difference in negotiation mode preference between blue- and white-collar workers at the 5% level of significance.

15.8 (a) Excel output:

Observed Frequencies			
	Tyres		
Punctured?	**Sandtrak**	**Standard**	**Total**
Yes	**6**	**18**	**24**
No	**114**	**162**	**276**
Total	**120**	**180**	**300**

Expected Frequencies			
	Tyres		
Punctured?	Sandtrak	Standard	Total
Yes	9.6000	14.4000	24
No	110.4000	165.6000	276
Total	120	180	300

Data	
Level of Significance	**0.01**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**0.0002**
Chi-Square Test Statistic	**2.4457**
***p*-value**	**0.1179**
Reject the null hypothesis	

H_0: $\pi_1 = \pi_2$ H_1: $\pi_1 \neq \pi_2$ where population: 1 = Sandtrak, 2 = standard

Decision rule: *df* = 1. If χ^2_{STAT} > 0.0002, reject H_0 at 99% significance level.

Test statistic: χ^2_{STAT} = 2.4457

Conclusion: Since χ^2_{STAT} = 2.4457 is higher than the lower critical bound of 0.0002, reject H_0. There is sufficient evidence to conclude that there is significant difference in the puncture resistance between the SDsandtrak and standard tyre at the 1% level of significance. It can be concluded at the 1% level that the new SDsandtrak can be advertised to reduce the rate of punctures in off-road driving over the SD standard tyre.

(b) The level of significance is chosen by the practitioner according to the situation and the potential consequences of failing to reject a false null hypothesis. In this case, driver and passenger safety are considered when choosing a large significance level.

15.9 (a) Excel output:

Observed Frequencies			
	Location		
Response	**Adelaide**	**Perth**	**Total**
Yes	**73**	**98**	**171**
No	**138**	**117**	**255**
Total	**211**	**215**	**426**

Expected Frequencies			
	Location		
Response	Adelaide	Perth	Total
Yes	84.6972	86.3028	171
No	126.3028	128.6972	255
Total	211	215	426

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.8415**
Chi-Square Test Statistic	**5.3473**
***p*-value**	**0.0208**
Reject the null hypothesis	

H_0: $\pi_1 = \pi_2$ H_1: $\pi_1 \neq \pi_2$ where population: 1 = Adelaide, 2 = Perth

Decision rule: *df* = 1. If χ^2_{STAT} > 3.8415, reject H_0 at 5% significance level.

Test statistic: χ^2_{STAT} = 5.3473

Conclusion: Since χ^2_{STAT} = 5.3473 is greater than the upper critical bound of 3.8415, reject H_0. There is sufficient evidence to conclude that there is significant difference in negotiation mode preference between the two treatment groups in the proportion of responses to the survey at the 5% level of significance.

(b) The *p*-value is 0.0208. The probability of obtaining a test statistic of 5.3473 or larger when the null hypothesis is true is about 0.0208.

(c) Adelaide financial planners are incentivised to participate whereas Perth financial planners are not. This introduces a new element of bias to the study.

15.10 (a) $df = (r - 1)(c - 1) = (2 - 1)(5 - 1) = 4$

(b) $\chi^2_{STAT} = 9.488$

(c) $\chi^2_{STAT} = 13.277$

15.11 (a) Excel output for expected frequencies of each cell

	A	B	C	Total
1	16.0000	15.0857	32.9143	64
2	19.0000	17.9143	39.0857	76
Total	35	33	72	140

(b) $\chi^2_{STAT} = \sum_{\text{All Cells}} \frac{(f_0 - f_e)^2}{f_e}$ = 2.945159

This is not statistically significant at the 1% level.

(c) The Marascuilo procedure is only appropriate if we were to reject the null hypothesis.

15.12 (a) Excel output for expected frequencies of each cell

	A	B	C	Total
1	25.0000	25.0000	25.0000	75.0000
2	25.0000	25.0000	25.0000	75.0000
Total	50	50	50	150

(b) $\chi^2_{STAT} = \sum_{\text{All Cells}} \frac{(f_0 - f_e)^2}{f_e}$ = 4.000

This is not statistically significant at the 5% level.

15.13 (a) H_0: $\pi_1 = \pi_2 = \pi_3 = \pi_4 = \pi_5$

H_1: $\pi_1 \neq \pi_2 \neq \pi_3 \neq \pi_4 \neq \pi_5$

where population: 1 = Australia, 2 = New Zealand, 3= China, 4= Japan, 5=South Korea

Decision rule: $df = 5 - 1 = 4$. If $\chi^2_{STAT} > 9.488$, reject H_0 at 5% significance level.

Test statistic:

$$\chi^2_{STAT} = \sum_i \frac{(f_{0i} - f_{ei})^2}{f_{ei}} = \frac{(265-100)^2}{100} + \frac{(240-100)^2}{100} + \frac{(190-100)^2}{100} + \frac{(270-100)^2}{100} + \frac{(245-100)^2}{100} = 1048.5$$

Conclusion: Since χ^2_{STAT} = 1,049 is greater than the upper critical bound of 9.488, reject H_0. There is sufficient evidence to conclude that there is significant difference in the proportion of households who own more than 1 laptop computer at the 5% level of significance.

(b) The p-value is 0.000. The probability of obtaining a test statistic of 62.8123 or larger when the null hypothesis is true is about 0.000.

(c) Excel output of the Marascuilo procedure:

Marascuilo Procedure

Group	Sample Proportions
1	0.53
2	0.48
3	0.38
4	0.54
5	0.49

MARASCUILO TABLE

Proportions	Absolute Differences	Critical Range	Statistically Significant?
|Group 1 - Group 2|	0.05	0.137062905	No
|Group 1 - Group 3|	0.15	0.143818775	Yes
|Group 1 - Group 4|	0.01	0.132844571	No
|Group 1 - Group 5|	0.04	0.136368911	No
|Group 2 - Group 3|	0.1	0.147080386	No
|Group 2 - Group 4|	0.06	0.136368911	No
|Group 2 - Group 5|	0.97	0.139804435	Yes
|Group 3 - Group 4|	0.16	0.143157536	Yes
|Group 3 - Group 5|	0.11	0.146433876	No
|Group 4 - Group 5|	0.05	0.135671368	No

Australia is significantly different from China; New Zealand is significantly different from South Korea; and China is significantly different from Japan.

15.14 (a) Excel output:

Observed Frequencies				
	Age group			
Major shopping day	**Under 35**	**35-54**	**Over 54**	**Total**
Saturday	**48**	**56**	**24**	**128**
A day other than Saturday	**152**	**144**	**176**	**472**
Total	**200**	**200**	200	**600**

Expected Frequencies				
	Age group			
Major shopping day	Under 35	35-54	Over 54	Total
Saturday	42.6667	42.6667	42.6667	128.0000
A day other than Saturday	157.3333	157.3333	157.3333	472.0000
Total	200	200	200	600

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	3
Degrees of Freedom	2

Results	
Critical Value	**5.991465**
Chi-Square Test Statistic	**16.525424**
***p*-value**	**2.5796E-04**
Reject the null hypothesis	

There is evidence of a significant difference between the age groups with respect to major grocery shopping day at the 5% significance level.

(b) The *p*-value is 0.0003. The probability of obtaining a sample that gives rise to a test statistic that is equal to or more than 16.5254 is 0.0003 if the null hypothesis is true.

(c)

Pairwise Comparisons	Critical Range	$\|p_j - p_{j'}\|$
1 to 2	0.1073	0.04
2 to 3	0.0959	0.16*
1 to 3	0.0929	0.12*

There is a significant difference between the 35–54 and over 54 groups, and between the under 35 and over 54 groups.

(d) The stores can use this information to target their marketing to the specific group of shoppers on Saturday and the days other than Saturday.

15.15 (a) $H_0: \pi_1 = \pi_2 = \pi_3$ H_1: at least one proportion differs

where population 1 = under 35, 2 = 35–54, 3 = over 54

Decision rule: *df* = (*c* – 1) = (3 – 1) = 2. If χ^2_{STAT} > 5.9915, reject H_0.

Test statistic: χ^2_{STAT} = 4.1314

Decision: Since χ^2_{STAT} = 4.1314 is less than the upper critical bound of 5.9915, do not reject H_0. There is not enough evidence to show that there is a significant relationship between age and major grocery shopping day.

(b) *p*-value = 0.1267. The probability of obtaining a sample that gives rise to a test statistic that is equal to or more than 4.1314 is 0.12.67 if the null hypothesis is true.

The larger the sample size, the more power the χ^2 test has and, hence, there is a higher likelihood of rejecting a false null hypothesis.

15.16 (a) Excel output:

Observed Frequencies					
	Age group (yrs)				
Prefer low-carb beer	**18-25**	**26-45**	**46-65**	**>65**	**Total**
Yes	**45**	**45**	**23**	**12**	**125**
No	**32**	**43**	**31**	**18**	**124**
Total	**77**	**88**	**54**	30	**249**

Expected Frequencies					
	Age group (yrs)				
Prefer low-carb beer	18-25	26-45	46-65	>65	Total
Yes	38.6546	44.1767	27.1084	15.0602	125.0000
No	38.3454	43.8233	26.8916	14.9398	124.0000
Total	77	88	54	30	249

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	4
Degrees of Freedom	3

Results	
Critical Value	**7.814728**
Chi-Square Test Statistic	**4.621503**
***p*-value**	**2.0171E-01**
Do not reject the null hypothesis	

At the 5% level of significance there is not a statistically significant difference in the proportion of people of different ages who prefer low-carbohydrate beer.

(b) Since we were not able to reject the null hypothesis at this level the Marascuilo procedure is not relevant.

15.17 (a) Excel output:

Observed Frequencies					
	University Course				
	Arts	**Business**	**Engineering**	**Science**	**Total**
ATAR < 90	164	172	68	120	**524**
ATAR > 90	172	195	73	139	**579**
Total	**336**	**367**	**141**	**259**	**1103**

Expected Values					
	University Course				
	Arts	**Business**	**Engineering**	**Science**	**Total**
ATAR < 90	159.62	174.35	66.98	123.04	524
ATAR > 90	176.38	192.65	74.02	135.96	579
Total	336	367	141	259	1103

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	4
Degrees of Freedom	4

Results	
Critical Value	**7.8147**
Chi-Square Test Statistic	**0.4619682 69**
***p*-Value**	**0.2152330 8**
Do not reject the null hypothesis	

(b) The *p*-value is greater than 5%. The probability of obtaining a test statistic of 0.462 or larger when the null hypothesis is true is about 21.52%.

15.18

Observed Frequencies						
	Country					
Free WiFi available?	**UK**	**Australia**	**US**	**China**	**Malaysia**	**Total**
No	**38**	**43**	**59**	**52**	**60**	**252**
Yes	**62**	**57**	**41**	**48**	**40**	**248**
Total	**100**	**100**	**100**	**100**	**100**	**500**

Expected Frequencies						
	Country					
Free WiFi available?	UK	Australia	US	China	Malaysia	Total
Yes	50.4	50.4	50.4	50.4	50.4	252
No	49.6	49.6	49.6	49.6	49.6	248
Total	100	100	100	100	100	500

Data	
Level of Significance	**0.025**
Number of Rows	2
Number of Columns	5
Degrees of Freedom	4

Results	
Critical Value	**11.14329**
Chi-Square Test Statistic	**15.08897**
***p*-Value**	**0.00452**
Reject the null hypothesis	

(b) The *p*-value is 0.00452. The probability of obtaining a test statistic of 15.08897 or larger when the null hypothesis is true is about 0.00452.

15.19 Excel output of the Marascuilo procedure:

Group	**Sample Proportions**
1	0.3
2	0.3
3	0.4

MARASCUILO TABLE

Proportions	Absolute Differences	Critical Range	Statistically Significant?
|Group 1 - Group 2|	0	0.021887662	No
|Group 1 - Group 3|	0.1	0.021091479	Yes
|Group 2 - Group 3|	0.1	0.021091479	Yes

Travellers to North Queensland appear to be significantly different from travellers to Bali and Fiji at the 95% significance level.

15.20 Excel output:

Observed Frequencies			
	Country of Origin		
Political voting intention	**Australia**	**Overseas**	**Total**
Liberal	**12**	**23**	**35**
Labor	**23**	**28**	**51**
National	**8**	**12**	**20**
Other	**4**	**8**	**12**
Total	**47**	**71**	**118**

Expected Frequencies			
	Country of Origin		
Political voting intention	Australia	Overseas	Total
Liberal	13.9407	21.0593	35
Labor	20.3136	30.6864	51
National	7.9661	12.0339	20
Other	4.7797	7.2203	12
Total	47	71	118

Data	
Level of Significance	**0.05**
Number of Rows	4
Number of Columns	2
Degrees of Freedom	3

Results	
Critical Value	**7.814728**
Chi-Square Test Statistic	**1.251070**
***p*-value**	**7.4078E-01**
Do not reject the null hypothesis	

There is not enough evidence to conclude that there is a statistically significant difference between political party voting intentions, at the 5% significance level.

15.21 $df = (r - 1)(c - 1) = (5 - 1)(6 - 1) = 4*5 = 20$

15.22 (a) χ^2 = 21.026
(b) χ^2 = 18.549
(c) χ^2 = 26.217
(d) χ^2 = 18.307
(e) χ^2 = 13.277

15.23 Excel output:

Observed Frequencies					
	Age group (yrs)				
Prefer low-carb beer	**18-25**	**26-45**	**46-65**	**>65**	**Total**
Yes	**45**	**45**	**23**	**12**	**125**
No	**32**	**43**	**31**	**18**	**124**
Total	**77**	**88**	**54**	30	**249**

Expected Frequencies					
	Age group (yrs)				
Prefer low-carb beer	18-25	26-45	46-65	>65	Total
Yes	38.6546	44.1767	27.1084	15.0602	125.0000
No	38.3454	43.8233	26.8916	14.9398	124.0000
Total	77	88	54	30	249

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	4
Degrees of Freedom	3

Results	
Critical Value	**7.814728**
Chi-Square Test Statistic	**4.621503**
***p*-value**	**2.0171E-01**
Do not reject the null hypothesis	

H_0 : There is no relationship between age group and preference for low-carb beer.

H_1 : There is a relationship between age group and preference for low-carb beer.

Decision rule: *df* = (*r* − 1)(*c* − 1) = (2 − 1)(4 − 1) = 3. If χ^2_{STAT} > 7.815, reject H_0.

Test statistic: χ^2_{STAT} = 4.622

Decision: Since χ^2_{STAT} = 4.622 is less than critical level of 7.815, do not reject H_0 at 5% significance level. There is not enough evidence to support the alternative hypothesis that states that age group and preference for low-carbohydrate beer are dependent events.

15.24 (a) Excel output:

Observed Frequencies				
	Stress			
Commuting time	**High**	**Moderate**	**Low**	**Total**
Under 15 min	**9**	**5**	**18**	**32**
15-45 min	**17**	**8**	**28**	**53**
Over 45 min	**18**	**6**	**7**	**31**
Total	**44**	**19**	53	**116**

Expected Frequencies				
	Stress			
Commuting time	High	Moderate	Low	Total
Under 15 min	12.138	5.241	14.621	32
15-45 min	20.103	8.681	24.216	53
Over 45 min	11.759	5.078	14.164	31
Total	44	19	53	116

Data	
Level of Significance	**0.01**
Number of Rows	3
Number of Columns	3
Degrees of Freedom	4

Results	
Critical Value	**13.276704**
Chi-Square Test Statistic	**9.831141**
***p*-value**	**4.3370E-02**
Do not reject the null hypothesis	

H_0 : There is no relationship between stress and commuting time.

H_1 : There is a relationship between stress and commuting time.

Decision rule: $df = (r - 1)(c - 1) = (3 - 1)(3 - 1) = 4$ and α=1%. If $\chi^2_{STAT} >$ 13.277, reject H_0.

Test statistic: $\chi^2_{STAT} = 9.831$

Conclusion: Since $\chi^2_{STAT} = 9.831$ is less than the upper critical bound of 13.277, do not reject H_0 at 1% significance level. There is not enough evidence to support the alternative hypothesis, which states that stress and commuting time are dependent events.

(b) Using a significance level of 0.05 would alter our decision rule and conclusion in the following way:

Decision rule: $df = (r - 1)(c - 1) = (3 - 1)(3 - 1) = 4$ and α=5%. If $\chi^2_{STAT} >$ 9.488, reject H_0.

Conclusion: Since χ^2_{STAT} = 9.831 is greater than the upper critical bound of 9.488, reject H_0 at 5% significance level. There is not enough evidence to say that stress and commuting time are independent events.

15.25 Excel output:

Observed Frequencies			
	Country of Origin		
Political voting intention	**Australia**	**Overseas**	**Total**
Liberal	**12**	**23**	**35**
Labor	**23**	**28**	**51**
National	**8**	**12**	**20**
Other	**4**	**8**	**12**
Total	**47**	**71**	**118**

Expected Frequencies			
	Country of Origin		
Political voting intention	Australia	Overseas	Total
Liberal	13.9407	21.0593	35
Labor	20.3136	30.6864	51
National	7.9661	12.0339	20
Other	4.7797	7.2203	12
Total	47	71	118

Data	
Level of Significance	**0.01**
Number of Rows	4
Number of Columns	2
Degrees of Freedom	3

Results	
Critical Value	**11.344867**
Chi-Square Test Statistic	**1.251070**
***p*-value**	**7.4078E-01**
Do not reject the null hypothesis	

H_0 : There is no relationship between country of origin and voting intention.

H_1 : There is a relationship between country of origin and voting intention.

Decision rule: $df = (r - 1)(c - 1) = (4 - 1)(2 - 1) = 3$ and α=1%. If $\chi^2_{STAT} > 11.345$, reject H_0.

Test statistic: χ^2_{STAT} = 1.251

Conclusion: Since χ^2_{STAT} = 1.251 is less than the upper critical bound of 11.345, do not reject H_0. At the 1% level of significance, there appears to be no relationship between country of origin and voting intention.

15.26 Excel output:

Observed Frequencies					
	Country				
Preferred method	**UK**	**USA**	**Australia**	**China**	**Total**
Face to face	**824**	**509**	**667**	**545**	**2545**
Phone, talk	**70**	**111**	**127**	**191**	**499**
Phone, text	**94**	**154**	**75**	**91**	**414**
Email	**118**	**120**	**142**	**20**	**400**
Total	**1106**	**894**	**1011**	**847**	**3858**

Expected Frequencies					
	Country				
Preferred method	UK	USA	Australia	China	Total
Face to face	729.5931	589.7434	666.9246	558.739	2545
Phone, talk	143.0518	115.6314	130.7644	109.5524	499
Phone, text	118.6843	95.93468	108.4899	90.89114	414
Email	114.6708	92.69051	104.8212	87.81752	400
Total	1106	894	1011	847	3858

Data	
Level of Significance	**0.01**
Number of Rows	4
Number of Columns	4
Degrees of Freedom	9

Results	
Critical Value	**21.66599**
Chi-Square Test Statistic	**246.0794**
***p*-Value**	**6.71E-48**
Reject the null hypothesis	

H_0: There is no relationship between country and type of preferred communication method.
H_1: There is a relationship between country and type of preferred communication method.

Decision rule: $df = (r - 1)(c - 1) = (4 - 1)(4 - 1) = 9$ and α=1%. If $\chi^2_{STAT} > 21.666$, reject H_0.

Test statistic: χ^2_{STAT} = 246.0794

Conclusion: Since χ^2_{STAT} = 246.0794 is greater than the upper critical bound of 21.666, reject H_0. There appears to be a highly significant relationship between country and type of preferred communication method (also supported by a *p*-value that is virtually zero).

15.27 (a) Excel output:

Observed Frequencies				
	Level of return			
Managed fund type	**High**	**Medium**	**Low**	**Total**
Maximum capital gain	**25**	**41**	**52**	**118**
Long-term growth	**22**	**31**	**42**	**95**
Growth and current income	**33**	**41**	**53**	**127**
Balanced income	**35**	**39**	**42**	**116**
Common stock	**28**	**15**	**10**	**53**
Total	**143**	**167**	199	**509**

Expected Frequencies				
	Level of return			
Managed fund type	High	Medium	Low	Total
Maximum capital gain	33.151	38.715	46.134	118
Long-term growth	26.690	31.169	37.141	95
Growth and current income	35.680	41.668	49.652	127
Balanced income	32.589	38.059	45.352	116
Common stock	14.890	17.389	20.721	53
Total	143.0000	167	199	509

Data	
Level of Significance	**0.05**
Number of Rows	5
Number of Columns	3
Degrees of Freedom	8

Results	
Critical Value	**15.507313**
Chi-Square Test Statistic	**22.650608**
***p*-value**	**3.8430E-03**
Reject the null hypothesis	

H_0: There is no relationship between level of return and type of managed fund.
H_1: There is a relationship between level of return and type of managed fund.

Decision rule: *df* = (*r* – 1)(*c* – 1) = (5 – 1)(3 – 1) = 8 and α=5%. If χ^2_{STAT} > 15.507, reject H_0.
Test statistic: χ^2_{STAT} = 22.651

Conclusion: Since χ^2_{STAT} = 22.651 is greater than the upper critical bound of 15.507, reject H_0. There appears to be a significant relationship between level of return and type of managed fund.

(b) This finding tells us that it matters (in terms of level of return) as to which type of managed fund an investor chooses.

15.28 H_0: The number of sick days in the past 12 months follows a Poisson distribution.
H_1: The number of sick days in the past 12 months does not follow a Poisson distribution.

$$\bar{X} = \frac{\sum_{j=1}^{c} m_i f_i}{n} = \frac{810}{484} = 1.7$$

Our estimate of λ is 1.7 and this enables us to determine the expected probability distribution of the data.

Number of sick days	**Expected Probabilities**	**Expected Frequencies**
0	0.19	91
1	0.31	152
2	0.26	127
3	0.15	71
4	0.06	30
5	0.02	10
6	0.01	3
	1.00	**483**

Decision rule: *df* = *k* – *p* - 1 = 7 – 1 - 1 = 5 and α = 1%. If χ^2 > 11.071, reject H_0.

Test statistic: $\chi^2_{STAT} = \sum_{\text{All Cells}} \frac{(f_0 - f_e)^2}{f_e}$ = 53.321

Conclusion: Since χ^2 = 53.321 is greater than the upper critical bound of 11.071, reject H_0. The distribution of sick days does not appear to follow the Poisson distribution.

15.29 With a significance level of 0.01 and a population mean of 1.5, the distribution of sick days does not follow the Poisson distribution.

15.30 H_0: The distribution of commercial mortgages approved follows a Poisson distribution.
H_1: The distribution of commercial mortgages approved does not follow a Poisson distribution.

$$\bar{X} = \frac{\sum_{j=1}^{c} m_i f_i}{n} = \frac{219}{104} = 2.1$$

Our estimate of λ is 2.1 and this enables us to determine the expected probability distribution of the data.

No. of commercial mortgages approved	Expected Probabilities	Expected Frequencies
0	0.12	13
1	0.26	27
2	0.27	28
3	0.19	20
4	0.10	10
5	0.04	4
6	0.01	2
7	0.00	0
	1.00	**104**

Decision rule: $df = k - p - 1 = 8 - 1 - 1 = 6$ and $\alpha = 1\%$. If $\chi^2 > 16.812$, reject H_0.

Test statistic: $\chi^2_{STAT} = \sum_{\text{All Cells}} \frac{(f_0 - f_e)^2}{f_e} = 2.539$

Conclusion: Since $\chi^2 = 2.539$ is lower than the upper critical bound of 16.812, do not reject H_0.
The distribution of commercial mortgages approved appears to follow the Poisson distribution.

15.31 H_0: The distribution of battery lives follows a normal distribution.
H_1: The distribution of battery lives does not follow a normal distribution.

Life (in years)	Expected Probabilities	Expected Frequencies
0-under 1	0.07	37
1-under 2	0.29	146
2-under 3	0.40	201
3-under 4	0.19	96
4-under 5	0.03	16
5-under 6	0.00	1

Decision rule: $df = k - p - 1 = 6 - 2 - 1 = 3$ and $\alpha = 5\%$. If $\chi^2 > 7.815$, reject H_0.

Test statistic: $\chi^2_{STAT} = \sum_{\text{All Cells}} \frac{(f_0 - f_e)^2}{f_e} = 195.73$

Conclusion: Since $\chi^2 = 195.73$ is greater than the upper critical bound of 7.815, reject H_0.
The distribution of battery lives does not appear to follow a normal distribution at the 5% level of significance.

15.32 (a) $\bar{X} = \frac{\sum_{j-1}^{c} m_j f_j}{n} = \frac{6900}{500} = 13.8$

$S^2 = \frac{\sum_{j=1}^{c}(m_j - \bar{X})^2 f_j}{n-1} = \frac{20{,}505}{499} = 41.09$

$S = 6.41$

(b) H_0: The distribution of call length follows a normal distribution.
H_1: The distribution of call length does not follow a normal distribution.

Length (in minutes)	Expected Probabilities	Expected Frequencies
0-under 5	0.0853	42.65
5-under 10	0.1923	96.15
10-under 15	0.2977	148.85
15-under 20	0.2587	129.35
20-under 25	0.1259	62.95
25-under 30	0.0344	17.20

Decision rule: $df = k - p - 1 = 6 - 2 - 1 = 3$ and $\alpha = 5\%$. If $\chi^2 > 7.815$, reject H_0.

Test statistic: $\chi^2_{STAT} = \sum_{\text{All Cells}} \frac{(f_0 - f_e)^2}{f_e} = 13.28$

Conclusion: Since $\chi^2 = 13.28$ is greater than the upper critical bound of 7.815, reject H_0. The distribution of call lengths does not appear to follow a normal distribution at the 5% level of significance.

15.33 (a) For $df = 25$ and $\alpha = 0.01$, $\chi^2_{\alpha/2} = 10.520$ and $\chi^2_{1-\alpha/2} = 46.928$.

(b) For $df = 16$ and $\alpha = 0.05$, $\chi^2_{\alpha/2} = 6.908$ and $\chi^2_{1-\alpha/2} = 28.845$.

(c) For $df = 13$ and $\alpha = 0.10$, $\chi^2_{\alpha/2} = 5.892$ and $\chi^2_{1-\alpha/2} = 22.362$.

15.34 (a) For $df = 23$ and $\alpha = 0.01$, $\chi^2_{\alpha/2} = 9.2604$ and $\chi^2_{1-\alpha/2} = 44.1814$.

(b) For $df = 19$ and $\alpha = 0.05$, $\chi^2_{\alpha/2} = 8.9065$ and $\chi^2_{1-\alpha/2} = 32.8523$.

(c) For $df = 15$ and $\alpha = 0.10$, $\chi^2_{\alpha/2} = 7.2609$ and $\chi^2_{1-\alpha/2} = 24.9958$.

15.35 $\chi^2_{STAT} = \frac{(n-1) \cdot S^2}{\sigma^2} = \frac{24 \cdot 150^2}{100^2} = 54$

15.36 $\chi^2_{STAT} = \frac{(n-1)\cdot S^2}{\sigma^2} = \frac{15\cdot 10^2}{12^2}$ = 10.417

15.37 $df = n - 1 = 16 - 1 = 15$

15.38 (a) For df = 15 and α = 0.05, $\chi^2_{\alpha/2}$ = 6.262 and $\chi^2_{1-\alpha/2}$ = 27.488.

(b) For df = 15 and α = 0.05, $\chi^2_{\alpha/2}$ = 7.261.

15.39 (a) If $H_1: \sigma \neq 12$, do not reject H_0 since the test statistic χ^2 = 10.417 falls between the two critical bounds, $\chi^2_{\alpha/2}$ = 6.262 and $\chi^2_{1-\alpha/2}$ = 27.488.

(b) If $H_1: \sigma < 12$, do not reject H_0 since the test statistic χ^2 = 10.417 is greater than the critical bound 7.261.

15.40 You must assume that the data in the population are normally distributed to be able to use the chi-square test of a population variance or standard deviation. If the data selected do not come from an approximately normally distributed population, particularly for small sample sizes, the accuracy of the test can be seriously affected.

15.41 (a) H_0: $\sigma \leq 1.2^0C$. The standard deviation of the oven temperature has not increased above 1.2^0C.

H_1: $\sigma > 1.2^0C$. The standard deviation of the oven temperature has increased above 1.2^0C.

Decision rule: df = 29. If χ^2_{STAT} > 42.557, reject H_0.

Test statistic: $\chi^2_{STAT} = \frac{(n-1)\cdot S^2}{\sigma^2} = \frac{29\cdot 2.1^2}{1.2^2}$ = 88.813

Conclusion: Since the test statistic of χ^2_{STAT} = 88.813 is greater than the critical bound of 42.557, reject H_0. There is sufficient evidence to conclude that the standard deviation of the oven temperature has increased above 1.2^0C.

(b) You must assume that the data in the population are normally distributed to be able to use the chi-square test of a population variance or standard deviation.

(c) p-value = 5.53 x 10^{-8} or 0.00000005. The probability that a sample is obtained whose standard deviation is equal to or larger than 2.1^0C when the null hypothesis is true is 5.53 x 10^{-8}, a very small probability.
Note: The p-value was found using Excel.

15.42 (a) H_0: σ = \$200. The standard deviation of the amount of auto repairs is equal to \$200.

H_1: $\sigma \neq$ \$200. The standard deviation of the amount of auto repairs is not equal to \$200.

Decision rule: df = 24. If χ^2_{STAT} < 12.401 or χ^2_{STAT} > 39.364, reject H_0.

Test statistic: $\chi^2_{STAT} = \frac{(n-1)\cdot S^2}{\sigma^2} = \frac{24\cdot 237.52^2}{200^2}$ = 33.849

Decision: Since the test statistic of χ^2_{STAT} = 33.849 is between the critical bounds of 12.401 and 39.364, do not reject H_0. There is insufficient evidence to conclude that the standard deviation of the amount of auto repairs is not equal to $200.

(b) You must assume that the data in the population are normally distributed to be able to use the chi-square test of a population variance or standard deviation.

(c) *p*-value = 2(0.0874) = 0.1748. The probability of obtaining a sample whose standard deviation will give rise to a test statistic equal to or more extreme than 33.849 is 0.1748 when the null hypothesis is true.
Note: The *p*-value was found using Excel.

15.43 (a) H_0: σ = $32. The standard deviation of the monthly cost of gas usage within the local region is $32.
H_1: $\sigma \neq$ $32. The standard deviation of the monthly cost of gas usage within the local region differs from $32.
Decision rule: *df* = 14. If χ^2_{STAT} < 6.571 or χ^2_{STAT} > 23.685, reject H_0.

Test statistic: $\chi^2_{stat} = \dfrac{(n-1).S^2}{\sigma^2} = \dfrac{14(29.25)^2}{32^2} = 11.6971$

Conclusion: Since the test statistic of χ^2_{STAT} = 11.6971 is between the critical bounds of 6.571 and 23.685, do not reject H_0. There is insufficient evidence to conclude that the standard deviation of the monthly cost of calls within the local calling region differs from $32.

(b) You must assume that the data in the population are normally distributed to be able to use the chi-square test of a population variance or standard deviation.

(c) *p*-value = 2(1 – 0.6306) = 0.7388. The probability of obtaining a test statistic equal to or more extreme than the result obtained from this sample data is 0.7388 if the standard deviation of the monthly cost of gas usage within the local region is $32.

Note: Excel returns an upper-tail area of 0.6306 for χ^2_{STAT} = 11.6971. But since the sample standard deviation is smaller than the hypothesised value, the amount of area in the lower tail is (1 – 0.6306). That value is doubled to accommodate the two-tail hypotheses.

15.44 (a) H_0: σ = 0.135 cm. The standard deviation of the diameter of doorknobs is greater than or equal to 0.135 cm in the redesigned production process.
H_1: σ < 0.135 cm. The standard deviation of the diameter of doorknobs is less than 0.135 cm in the redesigned production process.
Decision rule: *df* = 24. If χ^2_{STAT} < 13.848, reject H_0.

Test statistic: $\chi^2_{STAT} = \dfrac{(n-1)\cdot S^2}{\sigma^2} = \dfrac{24\cdot 0.125^2}{0.135^2}$ = 20.576

Conclusion: Since the test statistic of χ^2_{STAT} = 12.245 is greater than the critical bound of 13.848, do not reject H_0. There is insufficient evidence to conclude that the standard deviation of the diameter of doorknobs is less than 0.135 cm in the redesigned production process.

(b) You must assume that the data in the population are normally distributed to be able to use the chi-square test of a population variance or standard deviation.

(c) p-value = 0.0664. The probability of obtaining a test statistic equal to or more extreme than the result obtained from this sample data is 6.64% if the population standard deviation is indeed no less than 0.135 cm.

15.45 (a) H_0: σ = 5g. The standard deviation of the weight of smoked salmon is 5 g.
H_1: $\sigma \neq$ \$12. The standard deviation of the weight of smoked salmon differs from 5 g.

Decision rule: df = 29. If χ^2_{STAT} < 16.047 or χ^2_{STAT} > 45.722, reject H_0.

Test statistic: $\chi^2_{STAT} = \frac{(n-1)\cdot S^2}{\sigma^2} = \frac{29\cdot 6.69^2}{5^2} = 51.92$

Conclusion: Since the test statistic of χ^2_{STAT} = 51.92 is greater than the critical boundary of 45.722, reject H_0. There is sufficient evidence to conclude that the standard deviation of salmon weights differs from 5 g at the 5% level of significance.

(b) You must assume that the data in the population are normally distributed to be able to use the chi-square test of a population variance or standard deviation.

(c) p-value = 2(1 – 0.994436) = 0.011129. The probability of obtaining a test statistic equal to or more extreme than the result obtained from this sample data is 0.011129 if the standard deviation of the weight of smoked salmon is 5 g.

15.46 (a) H_0: $\sigma \leq$ 2.5 ampere-hours. The standard deviation in the capacity of the battery is equal to 2.5 ampere-hours.
H_1: σ > 2.5 ampere-hours. The standard deviation in the capacity of the battery differs from 2.5 ampere-hours.

Decision rule: df = 19. If χ^2_{STAT} > 30.144, reject H_0.

Test statistic: $\chi^2_{STAT} = \frac{(n-1)\cdot S^2}{\sigma^2} = \frac{19\cdot 2.6589^2}{2.5^2}$ = 21.492

Conclusion: Since the test statistic of χ^2_{STAT} = 21.492 is less than the critical bound of 30.144, do not reject H_0. There is not sufficient evidence to conclude that the standard deviation in the capacity of a certain type of battery differs from 2.5 ampere-hours.

(b) You must assume that the data in the population are normally distributed to be able to use the chi-square test of a population variance or standard deviation.

(c) p-value = 0.3103. The probability of obtaining a test statistic equal to or more extreme than the result obtained from this sample data is 0.3103 if the population standard deviation is indeed no greater than 2.5 ampere-hours.

15.47 The chi-square test for the difference between two proportions can be used only when the alternative hypothesis is two-tailed.

15.48 The chi-square test can be used for *c* populations as long as all expected frequencies are at least one.

15.49 The chi-square test for independence can be used as long as all expected frequencies are at least one.

15.50 (a) Excel output:

Observed Frequencies				
	Bank			
Ease of access	**Comm.**	**ANZ**	**Westpac**	**Total**
Easy to find	**77**	**70**	**65**	**212**
Difficult to find	**22**	**30**	**25**	**77**
Very difficult to find	**11**	**10**	**8**	**29**
Total	**110**	**110**	98	**318**

Expected Frequencies				
	Bank			
Ease of access	Comm.	ANZ	Westpac	Total
Easy to find	73.333	73.333	65.333	212
Difficult to find	26.635	26.635	23.730	77
Very difficult to find	10.031	10.031	8.937	29
Total	110	110	98	318

Data	
Level of Significance	**0.05**
Number of Rows	3
Number of Columns	3
Degrees of Freedom	4

Results	
Critical Value	**9.487729**
Chi-Square Test Statistic	**1.828157**
***p*-value**	**7.6733E-01**
Do not reject the null hypothesis	

H_0: There is no difference between the public perceptions of ease of access to these three banks.
H_1: There is a difference between the public perception of ease of access to these three banks.

Conclusion: Given the Excel output above, there isn't a significant difference between the public perception of ease of access to these three banks at the 5% significance level.

(b) The comment in the question states that there should not be a difference in the perception of people about the ease of accessing different main brand service suppliers. Our findings support this.

15.51 (a) Excel output:

Observed Frequencies			
	Gender		
Intend to Buy	**Male**	**Female**	**Total**
Yes	**122**	**91**	**213**
No	**79**	**49**	**128**
Total	**201**	**140**	**341**

Expected Frequencies			
	Gender		
Intend to Buy	Male	Female	Total
Yes	125.5513	87.44868	213
No	75.44868	52.55132	128
Total	201	140	341

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.841459**
Chi-Square Test Statistic	**0.651822**
***p*-value**	**0.419462**
Do not reject the null hypothesis	

H_0: $\pi_1 = \pi_2$ H_1: $\pi_1 \neq \pi_2$ where population: 1 = males, 2 = females

Conclusion: Since χ^2_{STAT} = 0.651822 is lower than the upper critical bound of 3.8415, do not reject H_0. There is insufficient evidence to conclude that there is significant difference between the proportions of males and females who intend to buy the carpet cleaning product at the 5% significance level.

(b) The *p*-value is 0.419462. The probability of obtaining a test statistic of 0.651822 or larger when the null hypothesis is true is about 41.95%.

15.52 (a) Excel output:

Observed Frequencies			
	Island		
Number of days of use	**North**	**South**	**Total**

<10	**75**	**86**	**161**
10 or more	**44**	**45**	**89**
Total	**119**	**131**	**250**

Expected Frequencies			
	Island		
Number of days of use	North	South	Total
<10	76.636	84.364	161
10 or more	42.364	46.636	89
Total	119	131	250

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.8415**
Chi-Square Test Statistic	**0.18722**
***p*-value**	**0.66524**
Do not reject the null hypothesis	

H_0: $\pi_1 = \pi_2$ H_1: $\pi_1 \neq \pi_2$ where population: 1 = North, 2 = South

Conclusion: Since χ^2_{STAT} = 0.18722 is lower than the upper critical bound of 3.8415, do not reject H_0. There is insufficient evidence to conclude that there is significant difference in the number of days of use between North and South islands at the 5% significance level.

(b) The *p*-value is 0.66524. The probability of obtaining a test statistic of 0.18722 or larger when the null hypothesis is true is about 66.52%.

15.53 Excel output:

Observed Frequencies			
	Country		
Windscreen damage	**Australia**	**New Zealand**	**Total**
Yes	**23**	**32**	**55**
No	**387**	**448**	**835**
Total	**410**	**480**	**890**

Expected Frequencies			
	Country		
Windscreen damage	Australia	New Zealand	Total
Yes	25.3371	29.6629	55
No	384.6629	450.3371	835
Total	410	480	890

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.8415**
Chi-Square Test Statistic	**0.4260**
***p*-value**	**0.5139**
Do not reject the null hypothesis	

H_0: $\pi_1 = \pi_2$ H_1: $\pi_1 \neq \pi_2$ where population: 1 = Australia, 2 = New Zealand

Conclusion: Since χ^2_{STAT} = 0.4260 is lower than the upper critical bound of 3.8415, do not reject H_0. There is insufficient evidence to conclude that there is significant difference in the proportion of cars returned with windscreen damage between Australia and New Zealand at the 5% significance level.

15.54 (a) Excel output:

Observed Frequencies				
	Age group			
Support increase in age pension	**18-34 yrs**	**35-54 yrs**	**>54 yrs**	**Total**
Yes	**34**	**45**	**55**	**134**
No	**78**	**51**	**24**	**153**
Total	**112**	**96**	79	**287**

Expected Frequencies				
	Age group			
Support increase in age pension	18-34 yrs	35-54 yrs	>54 yrs	Total
Yes	52.2927	44.8223	36.8850	134.0000
No	59.7073	51.1777	42.1150	153.0000
Total	112	96	79	287

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	3
Degrees of Freedom	2

Results	
Critical Value	**5.991465**
Chi-Square Test Statistic	**28.693186**
***p*-value**	**5.8797E-07**
Reject the null hypothesis	

H_0: $\pi_1 = \pi_2 = \pi_3$ H_1: $\pi_1 \neq \pi_2 \neq \pi_3$

where population: 1 = 18–34 yrs, 2 = 35–54 yrs, 3 = > 54 yrs

Conclusion: Since χ^2_{STAT} = 28.693 is greater than the upper critical bound of 5.991, reject H_0. There is sufficient evidence to conclude that there is a significant difference between the age groups in the proportion of support at the 5% significance level.

(b) The *p*-value is virtually zero. The probability of obtaining a test statistic of 28.693 or larger when the null hypothesis is true is about 0.000.

(c) Excel output of the Marascuilo procedure:

Group	Sample Proportions
1 (18-34 yrs)	0.39
2 (35-54 yrs)	0.33
3 (>54 yrs)	0.28

MARASCUILO TABLE

Proportions	Absolute Differences	Critical Range	Statistically Significant ?
|Group 1 - Group 2|	0.06	0.163460829	No
|Group 1 - Group 3|	0.11	0.16662284	No
|Group 2 - Group 3|	0.05	0.170339792	No

Even though we rejected the null hypothesis in part (a), the results from the Marascuilo procedure displayed in the table above indicate that there is not enough data to conclude any particular difference is significant. However, note that the comparison between age groups 1 and 3 comes closest to significance. Further tests with larger samples might actually show a statistically significant difference between these two age groups.

15.55 (a) Excel output:

Observed Frequencies				
	Attitude			
Type of job	**Favour**	**Neutral**	**Oppose**	**Total**
Hourly worker	**108**	**46**	**71**	**225**
Supervisor	**18**	**12**	**30**	**60**
Middle Management	**35**	**14**	**26**	**75**
Upper Management	**24**	**7**	**9**	**40**
Total	**185**	**79**	136	**400**

Expected Frequencies				
	Attitude			
Type of job	Favour	Neutral	Oppose	Total
Hourly worker	104.0625	44.4375	76.5000	225
Supervisor	27.7500	11.8500	20.4000	60
Middle Management	34.6875	14.8125	25.5000	75
Upper Management	18.5000	7.9000	13.6000	40
Total	185	79	136	400

Data	
Level of Significance	**0.05**
Number of Rows	4
Number of Columns	3
Degrees of Freedom	6

Results	
Critical Value	**12.591587**
Chi-Square Test Statistic	**11.895309**
***p*-value**	**6.4346E-02**
Do not reject the null hypothesis	

H_0 : There is no relationship between attitude and type of job.

H_1 : There is a relationship between attitude and type of job.

Decision rule: *df* = (*r* − 1)(*c* − 1) = (4 − 1)(3 − 1) = 6. If χ^2_{STAT} > 12.592, reject H_0.

Test statistic: χ^2_{STAT} = 11.895

Conclusion: Since χ^2_{STAT} = 11.895 is less than the upper critical bound of 12.592, do not reject H_0 at 5% significance level. The evidence suggests that there is no relationship between attitude towards self-managed work teams and type of job.

(b) Excel output:

Observed Frequencies				
	Attitude			
Type of job	**Favour**	**Neutral**	**Oppose**	**Total**
Hourly worker	**135**	**23**	**67**	**225**
Supervisor	**39**	**7**	**14**	**60**
Middle Management	**47**	**6**	**22**	**75**
Upper Management	**26**	**6**	**8**	**40**
Total	**247**	**42**	111	**400**

Expected Frequencies				
	Attitude			
Type of job	Favour	Neutral	Oppose	Total
Hourly worker	138.9375	23.6250	62.4375	225
Supervisor	37.0500	6.3000	16.6500	60
Middle Management	46.3125	7.8750	20.8125	75
Upper Management	24.7000	4.2000	11.1000	40
Total	247	42	111	400

Data	
Level of Significance	**0.05**
Number of Rows	4
Number of Columns	3
Degrees of Freedom	6

Results	
Critical Value	**12.591587**
Chi-Square Test Statistic	**3.293706**
***p*-value**	**7.7118E-01**
Do not reject the null hypothesis	

(b) Conclusion: Since χ^2_{STAT} = 3.294 is less than the upper critical bound of 12.592, do not reject H_0 at 5% significance level. The evidence suggests that there is no relationship between attitude towards time off without pay and type of job.

15.56 (a) Excel output:

Observed Frequencies				
	Country			
Emphasis on banking?	**China**	**India**	**Australia**	**Total**
Yes	**34**	**45**	**48**	**127**
No	**55**	**50**	**51**	**156**
Total	**89**	**95**	99	**283**

Expected Frequencies				
	Country			
Emphasis on banking?	China	India	Australia	Total
Yes	39.9399	42.6325	44.4276	127.0000
No	49.0601	52.3675	54.5724	156.0000
Total	89	95	99	283

Data	
Level of Significance	**0.95**
Number of Rows	2
Number of Columns	3
Degrees of Freedom	2

Results	
Critical Value	**0.102587**
Chi-Square Test Statistic	**2.362197**
***p*-value**	**3.0694E-01**
Reject the null hypothesis	

$H_0 : \pi_1 = \pi_2 = \pi_3$

$H_1 : \pi_1 \neq \pi_2 \neq \pi_3$ where population 1 = China, 2 = India, 3 = Australia

Decision rule: $df = (r - 1)(c - 1) = (2 - 1)(3 - 1) = 2$ and α = 95%. If χ^2_{STAT} > 0.103, reject H_0.

Test statistic: χ^2_{STAT} = 2.362

Conclusion: Since χ^2_{STAT} = 2.362 is greater than the upper critical bound of 0.103, reject H_0 at 95% significance level. The evidence suggests that there is a difference in the emphasis on an understanding of the banking system between these three countries.

(b) Excel output of the Marascuilo procedure:

Group	Sample Proportions
1	0.314487633
2	0.335689046
3	0.349823322

MARASCUILO TABLE

Proportions	Absolute Differences	Critical Range	Statistically Significant?
|Group 1 - Group 2|	0.021201413	0.16904211	No
|Group 1 - Group 3|	0.035335689	0.168154744	No
|Group 2 - Group 3|	0.014134276	0.166814844	No

None of the differences between countries are statistically significant.

(c) The *p*-value is 0.03069. The probability of obtaining a test statistic of 2.362 or larger when the null hypothesis is true is about 3.069%.

(d) Excel output:

Observed Frequencies				
	Country			
Corporate Accounting?	**China**	**India**	**Australia**	**Total**
Less than 20%	**29**	**35**	**45**	**109**
21-40%	**50**	**47**	**41**	**138**
Greater than 40%	**10**	**13**	**13**	**36**
Total	**89**	**95**	99	**283**

Expected Frequencies				
	Country			
Corporate Accounting?	China	India	Australia	Total
Less than 20%	34.279	36.590	38.131	109
21-40%	43.399	46.325	48.276	138
Greater than 40%	11.322	12.085	12.594	36
Total	89	95	99	283

Data	
Level of Significance	**0.05**
Number of Rows	3
Number of Columns	3
Degrees of Freedom	4

Results	
Critical Value	**9.487729**
Chi-Square Test Statistic	**4.466557**
***p*-value**	**3.4653E-01**
Do not reject the null hypothesis	

(d) H_0: There is no relationship between the level of corporate accounting desired in each country.
H_1: There is a relationship between the level of corporate accounting desired in each country.

Decision rule: *df* = (*r* − 1)(*c* − 1) = (3 − 1)(3 − 1) = 4 and α = 5%. If χ^2_{STAT} > 9.488, reject H_0.
Test statistic: χ^2_{STAT} = 4.467
Conclusion: Since χ^2_{STAT} = 4.467 is less than the upper critical bound of 9.488, do not reject H_0 at 5% significance level. The evidence suggests that there is no relationship between the level of corporate accounting desired in each country.

(e) The *p*-value is 0.03465. The probability of obtaining a test statistic of 4.467 or larger when the null hypothesis is true is about 3.465%.

(f) Excel output:

Observed Frequencies				
	Country			
Market systems?	**China**	**India**	**Australia**	**Total**
Yes	**59**	**50**	**45**	**154**
No	**30**	**45**	**54**	**129**
Total	**89**	**95**	99	**283**

Expected Frequencies				
	Country			
Market systems?	China	India	Australia	Total
Yes	48.4311	51.6961	53.8728	154.0000
No	40.5689	43.3039	45.1272	129.0000
Total	89	95	99	283

Data	
Level of Significance	**0.025**
Number of Rows	2
Number of Columns	3
Degrees of Freedom	2

Results	
Critical Value	**7.377759**
Chi-Square Test Statistic	**8.387754**
***p*-value**	**1.5088E-02**
Reject the null hypothesis	

(f) $H_0 : \pi_1 = \pi_2 = \pi_3$

$H_1 : \pi_1 \neq \pi_2 \neq \pi_3$ where population 1 = China, 2 = India, 3 = Australia

Decision rule: *df* = (*r* – 1)(*c* – 1) = (2 – 1)(3 – 1) = 2 and α = 2.5%. If χ^2_{STAT} > 7.478, reject H_0.

Test statistic: χ^2_{STAT} = 8.388

Conclusion: Since χ^2_{STAT} = 8.388 is greater than the upper critical bound of 7.478, reject H_0 at 2.5% significance level. The findings suggest that there is evidence of a significant difference between the countries on whether comparative market systems should be a significant part of an economics senior subject.

(g) See part (b). None of the differences between countries are statistically significant. Further investigation is required.

15.57 (a) Excel output:

Observed Frequencies			
	Accommodation		
City	**Pool**	**No pool**	**Total**
Barcelona	130	5887	6017
Lyon	10	912	922
Melbourne	143	1685	1828
London	79	7644	7723
Total	362	16128	16490

Expected Frequencies			
	Accommodation		
City	Pool	No pool	Total
Barcelona	132.0894	5884.911	6017
Lyon	20.24039	901.7596	922
Melbourne	40.12953	1787.87	1828
London	169.5407	7553.459	7723
Total	362	16128	16490

Data

Level of Significance	0.01
Number of Rows	4
Number of Columns	2
Degrees of Freedom	3

Results	
Critical Value	11.34487
Chi-Square Test Statistic	324.3916
p-Value	5.22476E-70
Reject the null hypothesis	

(a) $H_0: \pi_1 = \pi_2$

$H_1: \pi_1 \neq \pi_2$ where population 1 = swimming pool, 2 = no swimming pool

Decision rule: $df = (r - 1)(c - 1) = (4 - 1)(2 - 1) = 3$ and $\alpha = 1\%$. If $\chi^2_{STAT} > 11.3449$, reject H_0.

Test statistic: $\chi^2_{STAT} = 324.3916$

Conclusion: Since $\chi^2_{STAT} = 324.3916$ is more than the critical bound of 11.3449, reject H_0 at 1% significance level. There is a significant difference between accommodation with swimming pool proportions in different countries.

(b) Marascuilo procedure
Level of significance =0.01
Square root of critical value = 3.368214

Sample proportions			
P1	0.021605		
P2	0.010846		
P3	0.078228		
P4	0.010229		
Marascuilo table			
Proportions	Absolute differences	Critical Range	
|City1-City2|	0.010759	0.01311	Not significant
|City1-City3|	0.056622	0.022076	Significant
|City1-City4|	0.011376	0.007398	Significant
|City2-City3|	0.067382	3.233808	Not significant
|City2-City4|	0.000617	0.01211	Not significant
|City3-City4|	0.067998	0.021503	Significant

15.58 (a) Excel output:

Observed Frequencies			
	Preference after ads		
Preference before ads	**Toyota**	**Mazda**	**Total**
Toyota	**79**	**13**	**92**
Mazda	**11**	**97**	**108**
Total	**90**	**110**	**200**

Expected Frequencies			
	Preference after ads		
Preference before ads	Toyota	Mazda	Total
Toyota	41.4	50.6	92
Mazda	48.6	59.4	108
Total	90	110	200

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.8415**
Chi-Square Test Statistic	**114.9791**
***p*-value**	**7.95E-27**
Reject the null hypothesis	

$H_0 : \pi_1 = \pi_2$

$H_1 : \pi_1 \neq \pi_2$ where population 1 = Toyota, 2 = Mazda

Decision rule: $df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1$ and $\alpha = 5\%$. If $\chi^2_{STAT} >$ 3.8415, reject H_0.

Test statistic: $\chi^2_{STAT} = 114.9791$

Conclusion: Since $\chi^2_{STAT} = 114.9791$ is greater than the upper critical bound of 3.8415, reject H_0 at 5% significance level. There is a significant difference in the proportion of respondents who prefer Toyota before and after viewing the ads.

(b) The *p*-value is approximately 0.000. The probability of obtaining a test statistic of 114.9791 or larger when the null hypothesis is true is about 0.000.

(c) Table is derived from the margins of the first table to form the variable preference, 'Before ad' and 'After ad'.

(d) Excel output:

Observed Frequencies

	Preference after ads		
Preference	**Toyota**	**Mazda**	**Total**
Before	**92**	**108**	**200**
After	**90**	**110**	**200**
Total	**182**	**218**	**400**

Expected Frequencies			
	Preference after ads		
Preference	Toyota	Mazda	Total
Before	91	109	200
After	91	109	200
Total	208	192	400

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.841459**
Chi-Square Test Statistic	**0.040327**
***p*-value**	**0.840843**
Do not reject the null hypothesis	

According to the above information, there is not a significant difference in the proportion of respondents who prefer Toyota before and after viewing the ads.

(e) The *p*-value has no meaning because the null hypothesis was not rejected.

(f) The difference in the results of (a) and (d) is due to the layout of the data in each table. Use the method in part (a) since the contingency table in part (d) does not make sense, given our research objectives.

15.59 (a) Excel output:

Observed Frequencies			
	Income Level		
Subscriber Option	**High**	**Low**	**Total**
TV episodes	**352**	**769**	**1121**
Movies	**198**	**498**	**696**
Total	**550**	**1267**	**1817**

Expected Frequencies			
	Income Level		
Subscriber Option	High	Low	Total

TV episodes	339.3231	781.6769	1121
Movies	210.6769	485.3231	696
Total	550	1267	1817

Data	
Level of Significance	**0.1**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**2.705543**
Chi-Square Test Statistic	**1.773126**
***p*-value**	**0.182996**
Do not reject the null hypothesis	

H_0 : There is no relationship between income level and subscriber option.

H_1 : There is a relationship between income level and subscriber option.

Decision rule: *df* = (*r* – 1)(*c* – 1) = (2 – 1)(2 – 1) = 4 and α = 10%. If χ^2_{STAT} > 2.706, reject H_0.

Test statistic: χ^2_{STAT} = 1.773

Conclusion: Since χ^2_{STAT} = 1.773 is less than the upper critical bound of 2.706, do not reject H_0 at 10% significance level. The evidence suggests that there is no relationship between the income level and subscriber download choice.

(b) The *p*-value is 0.182996. The *p*-value has no meaning because the null hypothesis was not rejected.

15.60 (a) Excel output:

Observed Frequencies			
	Gender		
Duration of current job	**Male**	**Female**	**Total**
Under 1 year	**20.7**	**31.2**	**51.9**
1 year and under 2 years	**9.9**	**15.3**	**25.2**
2 years and under 3 years	**9.5**	**13**	**22.5**
3 years and under 4 years	**7.4**	**9.6**	**17**
4 years and under 5 years	**5.7**	**6**	**11.7**
5 years and under 10 years	**18.8**	**15.6**	**34.4**
10 years and under 15 years	**11.5**	**4.8**	**16.3**
15 years and under 20 years	**6**	**2.3**	**8.3**
Over 20 years	**10.5**	**2.2**	**12.7**
Total	**100**	**100**	**200**

Expected Frequencies			
	Gender		
Duration of current job	Male	Female	Total
Under 1 year	25.95	25.96	51.9
1 year and under 2 years	12.6	12.6	25.2
2 years and under 3 years	11.25	11.25	22.5
3 years and under 4 years	8.5	8.5	17
4 years and under 5 years	5.85	5.85	11.7
5 years and under 10 years	17.2	17.2	34.4
10 years and under 15 years	8.15	8.15	16.3
15 years and under 20 years	4.15	4.15	8.3
Over 20 years	6.35	6.35	12.7
Total	100	100	200

Results	
Chi-Square Test Statistic	14.24373
***p*-value**	**0.07563**

(b) Excel output:

Observed Frequencies			
	Gender		
Multiple job holders	**Male**	**Female**	**Total**
Multiple jobs	**3.3**	**97.5**	**100.8**
Single jobs	**96.7**	**2.5**	**99.2**
Total	**100**	**100**	**200**

Expected Frequencies			
	Gender		
Multiple job holders	Male	Female	Total
Multiple jobs	50.4000	50.4000	100.8
Single jobs	49.6000	49.6000	99.2
Total	100	100	200

Results	
Chi-Square Test Statistic	**177.4842**
***p*-value**	**0.0000**

(c) Excel output:

Observed Frequencies			
	Gender		
Difficulties reported	**Male**	**Female**	**Total**
Too young or too old	**12.74**	**20.1**	**32.84**
Unsuitable hours	**3.94**	**5.1**	**9.04**
Transport problems	**9.74**	**3.2**	**12.94**
Lacked training/education	**13.74**	**2.2**	**15.94**

Insufficient work experience	**15.54**	**18**	**33.54**
No vacancies in line of work	**19.84**	**25.8**	**45.64**
No vacancies at all	**24.63**	**25.6**	**50.23**
Total	**100**	**100**	**200.17**

Expected Frequencies			
	Gender		
Difficulties reported	Male	Female	Total
Too young or too old	16.434	16.406	33
Unsuitable hours	4.524	4.516	9
Transport problems	6.475	6.465	13
Lacked training/education	7.977	7.963	16
Insufficient work experience	16.784	16.756	34
No vacancies in line of work	22.839	22.801	46
No vacancies at all	25.136	25.094	50
Total	100.1700	100	200.17

Results	
Chi-Square Test Statistic	**14.435610**
***p*-value**	**7.1094E-02**

(d) Excel output:

Observed Frequencies			
	Gender		
Duration of unemployment	**Male**	**Female**	**Total**
1-4 weeks	**11.5**	**12.9**	**24.4**
5-13 weeks	**7.5**	**17.5**	**25**
14-26 weeks	**9.5**	**23.2**	**32.7**
Total	**28.5**	**53.6**	**82.1**

Expected Frequencies			
	Gender		
Duration of unemployment	Male	Female	Total
1-4 weeks	8.470	15.930	24
5-13 weeks	8.678	16.322	25
14-26 weeks	11.351	21.349	33
Total	28.5	53.6	82.1

Results	
Chi-Square Test Statistic	**2.367695**
***p*-value**	**6.6847E-01**

(e) Excel output:

Observed Frequencies			
	Gender		
Birthplace of unemployed	**Male**	**Female**	**Total**
Australia	**34.9**	**34.6**	**69.5**
Outside of Australia	**15.5**	**15.2**	**30.7**

Total	**50.4**	**49.8**	**100.2**

Expected Frequencies			
	Gender		
Birthplace of unemployed	Male	Female	Total
Australia	34.9581	34.5419	69.5
Outside of Australia	15.4419	15.2581	30.7
Total	50.4	49.8	100.2

Results	
Chi-Square Test Statistic	**0.0006**
***p*-value**	**0.9799**

(f) Ranked order of importance of each problem area analysed (according to *p*-values shown above) are:

1. Birthplace of unemployed *p*-value: 0.9799
2. Duration of current job *p*-value: 0.07563
3. Duration of unemployment *p*-value: 0.0669
4. Difficulties reported *p*-value: 0.0071
5. Multiple job holder *p*-value: 0.0000

15.61 Numbers of persons using each social media tool have been rounded to the nearest integer.

Excel output: Facebook

Observed Frequencies			
	Column variable		
Row variable	**Male**	**Female**	**Total**
Facebook user	**536**	**571**	**1107**
Non Facebook user	**53**	**18**	**71**
Total	**589**	**589**	**1178**

Expected Frequencies			
	Column variable		
Row variable	Male	Female	Total
Facebook user	553.5	553.5	1107
Non Facebook user	35.5	35.5	71
Total	589	589	1178

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of	2

Columns	
Degrees of Freedom	1

Results	
Critical Value	**3.841459**
Chi-Square Test Statistic	**18.36012**
***p*-Value**	**1.83E-05**
Reject the null hypothesis	

Excel output: Instagram

Observed Frequencies			
	Column variable		
Row variable	**Male**	**Female**	**Total**
Instagram user	**295**	**241**	**536**
Non Instagram user	**294**	**348**	**642**
Total	**589**	**589**	**1178**

Expected Frequencies			
	Column variable		
Row variable	Male	Female	Total
Instagram user	268	268	536
Non Instagram user	321	321	642
Total	589	589	1178

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.841459**
Chi-Square Test Statistic	**9.982355**
***p*-Value**	**0.00158**

Reject the null hypothesis

Excel output:LinkedIn

Observed Frequencies			
	Column variable		
Row variable	Male	Female	Total
LinkedIn user	130	82	212
Non LinkedIn user	459	507	966
Total	589	589	1178

Expected Frequencies			
	Column variable		
Row variable	Male	Female	Total
LinkedIn user	106	106	212
Non LinkedIn user	483	483	966
Total	589	589	1178

Data	
Level of Significance	0.05
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	3.841459
Chi-Square Test Statistic	13.25302
p-Value	0.000272
Reject the null hypothesis	

Excel output: Twitter

Observed Frequencies			
	Column variable		
Row variable	Male	Female	Total
Twitter user	206	165	371

Non Twitter user	**383**	**424**	**807**
Total	**589**	**589**	**1178**

Expected Frequencies			
	Column variable		
Row variable	Male	Female	Total
Twitter user	185.5	185.5	371
Non Twitter user	403.5	403.5	807
Total	589	589	1178

Data	
Level of Significance	**0.05**
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	**3.841459**
Chi-Square Test Statistic	**6.614021**
***p*-Value**	**0.010118**
Reject the null hypothesis	

15.62–15.66 Team project.

15.62 (a)

	Objective		
Category	**Negative**	**Non-negative**	**Grand Total**
Australian	8	25	33
International	10	5	15
Grand Total	18	30	48

(b) H_0: There is no relationship between the category of investment fund and whether or not there is a negative monthly return.
H_1: There is a relationship between the category of an investment fund and whether or not there is a negative monthly return.

Decision rule: If $\chi^2 > 3.8415$, reject H_0.

Test statistic: $\chi^2 = 7.9192$

Decision: Since $\chi^2 =$ 7.9192 is greater than the critical bound of 3.8415, reject H_0. There is enough evidence to conclude at the 5% level of significance that there is a relationship between the category of an investment and whether or not there is a negative monthly return.

15.63 (a)

	Market Capitalisation			
Fee	**High**	**Medium**	**Low**	**Grand Total**
No	8	11	2	20
Yes	2	13	5	21
Grand Total	10	24	7	41

(b) H_0: There is no relationship between the market capitalisation of an investment fund and whether or not there is an outperformance fee.
H_1: There is a relationship between the market capitalisation of an investment fund and whether or not there is an outperformance fee.

Decision rule: If $\chi^2 > 5.9915$, reject H_0.

Test statistic: $\chi^2 = 5.0310$

Decision: Since $\chi^2 = 5.0310$ is less than the critical bound of 5.9915, do not reject H_0. There is not enough evidence to conclude at the 5% level of significance that there is a relationship between the market capitalisation of an investment fund and whether or not there is an outperformance fee.
(*Note:* in this calculation three cells have expected frequencies below 5 but all expected frequencies are greater than 1.)

15.64 (a)

	a			
Traded volume	**High**	**Medium**	**Low**	**Grand Total**
High	6	17	2	25
Low	4	8	11	23
Grand Total	10	25	23	48

(b) H_0: There is no relationship between the market capitalisation of an investment fund and its traded volume.
H_1: There is a relationship between the market capitalisation of an investment fund and its traded volume.

Decision rule: If $\chi^2 > 5.9915$, reject H_0.

Test statistic: $\chi^2 = 9.8044$

Decision: Since $\chi^2 = 9.8044$ is greater than the critical bound of 5.9915, reject H_0. There is enough evidence to conclude at the 5% level of significance that there is a relationship between the market capitalisation and traded volume.

15.65 $$\overline{X} = \frac{\sum X}{n} = \frac{4.8441}{39} = 0.1242$$

$$S = \sqrt{\frac{\sum X^2 - n\overline{X}^2}{n-1}} = \sqrt{\frac{1.1445 - 39(0.1242)^2}{38}} = 0.1195$$

H_0: Three-year return follows a normal distribution.
H_1: Three-year return does not follow a normal distribution.

$$\chi^2_{k-p-1} = \sum_k \frac{\left(f_0 - f_e\right)^2}{f_e} = 10.9213$$

$$\chi^2_{crit} = \chi^2_{7-2-1,0.05} = \chi^2_{4,0.05} = 9.4877$$

Since 10.9213 > 9.4877, reject H_0. There is sufficient evidence to conclude that the distribution of three-year return does not follow a normal distribution. Note that variations in answers are possible according to the categories chosen.

15.66 $$\overline{X} = \frac{\sum X}{n} = \frac{65.4866}{37} = 1.7699$$

$$S = \sqrt{\frac{\sum X^2 - n\overline{X}^2}{n-1}} = \sqrt{\frac{3845.016 - 37(1.7699)^2}{36}} = 10.1777$$

H_0: Five-year return follows a normal distribution.
H_1: Five-year return does not follow a normal distribution.

$$\chi^2_{k-p-1} = \sum_k \frac{\left(f_0 - f_e\right)^2}{f_e} = 3690.68$$

$$\chi^2_{crit} = \chi^2_{7-2-1,0.05} = \chi^2_{4,0.05} = 9.4877$$

Since 3690.68 > 9.48771, reject H_0. There is sufficient evidence to conclude that the distribution of five-year return does not follow a normal distribution. Note the effect of the large outlier on expected frequencies. Variations in answers are possible according to the categories chosen.