


Categorical Data Analysis: Chi-Squared Tests

13.1 Finding Chi-Square Critical Values

To complete this section of homework watch Chapter Thirteen, Lecture Examples: [173.5](#) and [173.6](#).

1. Use the χ^2 table to find the following critical values:  [VS](#)
 - a. $\chi^2_{0.05}$ with 15 degrees of freedom
 - b. $\chi^2_{0.005}$ with 40 degrees of freedom
 - c. $\chi^2_{0.10}$ with 22 degrees of freedom
 - d. $\chi^2_{0.025}$ with 50 degrees of freedom
2. Find the rejection region for a one-dimensional χ^2 –test with the following conditions:
 - a. $k = 4$; $\alpha = 0.01$
 - b. $k = 6$; $\alpha = 0.05$
 - c. $k = 5$; $\alpha = 0.10$

13.1 Answers

1.
 - a. 24.9958,
 - b. 66.7659,
 - c. 30.8133,
 - d. 71.4202
2.
 - a. 11.3449,
 - b. 11.0705,
 - c. 7.77944


[Need more exercises?](#)



: indicates the exercise has a video devoted to it in the corresponding section of STATSprofessor.com

13.2 Checking the Assumptions for a Chi-Square Goodness-of-Fit Test

To complete this section of homework watch Chapter Thirteen, Lecture Examples: [174](#) and [175.5](#).

- Is the following experiment a multinomial experiment? One-hundred people are asked what kind of music they listen to most. They can say any genre of music they want.
- Is the following experiment a multinomial experiment? Five hundred cars are observed on the freeway and placed into one of four categories: Domestic, Japanese, European, or other.
- Is the following experiment a multinomial experiment? Researchers are planning to ask randomly chosen students which profession they find most prestigious: Doctor, Lawyer, or Teacher. They will stop the researcher once one of the three categories has one hundred people who vote it as being most prestigious.
- Researchers sample a random selection of divorced workers at a corporation and test if the percentage of divorces that involve blue collar workers is 50%; white collar workers is 45%; and executives is 5% . Does the data collected below fit the sample size requirement for a one-way χ^2 test?  [VS](#)

Blue-collar	White-collar	Executives
50	35	4

- Does the following data set based on the classification of randomly excavated pottery pieces fit the sample size requirement for a one-way χ^2 test? Assume the hypothesis to be tested is

$$\rho_B = \rho_M = \rho_P = \rho_O$$

Pottery Category	Number Found
Burnished	133
Monochrome	460
Painted	183
Other	61
Total	837



indicates the exercise has a video devoted to it in the corresponding section of STATSprofessor.com

13.2 Answers

3. No, there is not a clearly defined k , the fixed number of possible outcomes.
4. Yes, with $n = 500$ and $k = 4$
5. No, there is not fixed number of trials (i.e. – no n).
6. No, because one of the expected cell counts is below 5:


Blue-collar	White-collar	Executives
50	35	4
$89 * .5 = 44.5$	$89 * .45 = 40.05$	$89 * .05 = 4.45$

7. Yes, because $\rho_B = \rho_M = \rho_P = \rho_O$ implies the expected value will be based on $(\text{category total})/4$ and each of these expectations is greater than 5.

[Need more exercises?](#)

13.3 The Chi-Square Test Statistic

To complete this section of homework read the lecture notes for section 13.3.

8. Use the table of data below and find the χ^2 test stat that would be used to test the claim that:
 $\rho_1 = \rho_2 = \rho_3$.  [VS](#)

Group 1	Group 2	Group 3
75	74	78

Summary data to assist in the calculations:


$\frac{(O-E)^2}{E}$	0.00587	0.03671	0.07195
---------------------	---------	---------	---------

9. Use the table of data below and find the χ^2 test stat that would be used to test the claim that:
 $\rho_A = \rho_B = \rho_C = \rho_D$.

A	B	C	D
87	56	88	94



: indicates the exercise has a video devoted to it in the corresponding section of STATSprofessor.com

10. Use the table of data below and find the χ^2 test stat that would be used to test the claim that:
 $\rho_A = 0.25, \rho_B = 0.10, \rho_C = 0.60, \rho_D = 0.05$.  [VS](#)

A	B	C	D
140	80	250	30

13.3 Answers

8. $\chi^2 = 0.11453$

9. $\chi^2 = 10.8154$

$$\frac{(O-E)^2}{E} \Rightarrow$$

0.4069	7.8469	0.5608	2.0008
--------	--------	--------	--------

10. $\chi^2 = 29.1333$


$$\frac{(O-E)^2}{E} \Rightarrow$$

1.8	18	8.3333	1
-----	----	--------	---

[Need more exercises?](#)

13.4 Testing Categorical Probabilities: One-Way Table


To complete this section of homework watch Chapter Thirteen, Lecture Examples [176](#), [177](#), and [178](#).

11. M&M candies' web site claims each package of Milk Chocolate M&M's should contain 24% blue, 14% brown, 16% green, 20% orange, 13% red, and 14% yellow M&M's. Use a 1% significance level and the data (that was actually observed by a researcher named Josh Madison) to determine if M&M's really fill their bags of candy with these proportions:  [VS](#)

Blue	Brown	Green	Orange	Red	Yellow
481	371	483	544	372	369




: indicates the exercise has a video devoted to it in the corresponding section of STATSprofessor.com

12. The US government reports that the percent of uninsured in the country is 13%. They also report that the percent who purchase health insurance on their own is 20%, and the percent who purchase insurance through their employer is 67%. A random survey of Americans reveals the set of results below. Does the data provide evidence to contradict the government's claim? [VS](#) 

Uninsured	Private	Employer
315	252	1533

13. The romance novel industry claims that 90 percent of romances were read in softcover/paperback format, 7 percent listened to audiobook versions, and 3 percent read in electronic/e-book format. A recent survey of Amazon sales produced the below table of results. Use a 2.5% significance level to test the claim that romance sales on Amazon differ from the industry's claim.

Paperback	Audio	E-book
1267	108	75

14. The candy Skittles comes in five colors/flavors: green, orange, red, purple, and yellow. The company claims that each bag of Skittles has an equal number of each color. Use a 10% level of significance and the data below to test the company's claim.  [VS](#)

Green	Orange	Red	Purple	Yellow
43	50	44	44	52

15. A professor at FIU claims his grade distribution is as follows: 10% A, 15% B, 25% C, 50% D or F. Use the data below and a 5% level of significance to test the professor's claim.

A	B	C	D or F
7	30	60	103



16. The following report is from a researcher at the University of Chicago. Laumann and co-researcher Dr. Amy Derick, of the University of Chicago, surveyed 2,492 tattooed people to determine if year of birth was a predictive factor for tattoos: 986 of the group were aged 18 to 29; 657 of them were aged 30 to 40; and only 411 were aged 40 to 50. Four hundred thirty-eight had obtained their first tattoo before age 18. Use a 10% significance level to test if all age groups have the same proportion of tattooed members.

13.4 Answers

11. $H_0 : \rho_{Blue} = .24, \rho_{Brown} = .14, \rho_{Green} = .16, \rho_{Orange} = .20, \rho_{Red} = .13, \rho_{Yellow} = .14$

H_A : At least one proportion differs significantly.

Test Stat: 48.170

Critical Value: 15.0863

Reject the null, support the alternative.

The sample data allows us to reject M&M's claim. At least one color appears significantly more or less than reported.

12. $H_0 : \rho_{UI} = .13, \rho_{PR} = .20, \rho_{EP} = .67$

H_A : At least one proportion differs significantly.

Test Stat: 84.945

Critical Value: 5.99147

Reject the null, support the alternative.

The sample data allows us to reject the government's claim.

13. $H_0 : \rho_P = .90, \rho_A = .07, \rho_E = .03$

H_A : At least one proportion differs significantly.

Test Stat: 24.333

Critical Value: 7.37776

Reject the null, support the alternative.

The sample data allows us to support the claim that the industry's numbers are incorrect.



$$14. H_0 : \rho_g = \rho_o = \rho_r = \rho_p = \rho_y$$

H_A : At least one proportion differs significantly.

Test Stat: 1.4421

Critical Value: 7.77944

Do not reject the null, do not support the alternative.

The sample data does not allow us to reject Skittle's claim.

$$15. H_0 : \rho_A = .10, \rho_B = .15, \rho_C = .25, \rho_F = .50$$

H_A : At least one proportion differs significantly.

Test Stat: 10.54

Critical Value: 7.81473

Reject the null, support the alternative.

The sample data allows us to reject the professor's claim.

$$16. E_i = 623$$

$$H_0 : \rho_1 = \rho_2 = \rho_3 = \rho_4$$

H_A : At least one proportion differs significantly.

Test Stat: 340.44

Critical Value: 6.251

Reject the null, support the alternative.

[Need more exercises?](#)




indicates the exercise has a video devoted to it in the corresponding section of STATSprofessor.com


13.5 Finding Expected Cell Counts

To complete this section of homework watch *Chapter Thirteen, Lecture Example 178.5*.

The table below has data from a 2009 Canadian Journal of Human Sexuality study. The data shows gender differences in response to partner influence and social expectation questions among students who had ever had sexual intercourse. The sample for this study included only students who had ever had sexual intercourse (30.7% of the total sample) yielding 2,145 respondents after corrections. The age range for the study sample was 13 to 21 years old (mean = 15.8, standard deviation = 1.19). The vast majority were 14 to 17 years old (93%) and 45% were 16 years old. Less than 2% were aged 13 or 19 - 21 years old.

	Male	Female	Total
Did you use a condom the last time you had intercourse?			
Yes	700	744	1444
No	244	457	701
Total	944	1201	2145

17. Find E_{21} (the expected value for the cell in the second row and first column)  [VS](#)

18. Find E_{22}  [VS](#)

19. Find E_{11}

20. Find E_{12}

13.5 Answers

17. 308.51

18. 392.49

19. 635.49

20. 808.51

[Need more exercises?](#)



: indicates the exercise has a video devoted to it in the corresponding section of STATSprofessor.com


13.6 Testing Categorical Probabilities: Two-Way (Contingency) Table

To complete this section of homework watch Chapter Thirteen, Lecture Examples [179](#), [180](#), and [181](#).

21. The table below has data from a 2009 Canadian Journal of Human Sexuality study. The data shows gender differences in response to partner influence and social expectation questions among students who had ever had sexual intercourse. The sample for this study included only students who had ever had sexual intercourse (30.7% of the total sample) yielding 2,145 respondents after corrections. The age range for the study sample was 13 to 21 years old (mean = 15.8, standard deviation = 1.19). The vast majority were 14 to 17 years old (93%) and 45% were 16 years old. Less than 2% were aged 13 or 19 to 21 years old.

	Male	Female	Total
Did you use a condom the last time you had intercourse?			
Yes	700	744	1444
No	244	457	701
Total	944	1201	2145


Use a 2.5% significance level to test the claim that condom use and gender are independent.

22. The table below has data from a 2009 Canadian Journal of Human Sexuality study. The data shows gender differences in response to partner influence and social expectation questions among students who had ever had sexual intercourse. The sample for this study included only students who had ever had sexual intercourse (30.7% of the total sample) yielding 2,145 respondents after corrections. The age range for the study sample was 13 to 21 years old (mean = 15.8, standard deviation = 1.19). The vast majority were 14 to 17 years old (93%) and 45% were 16 years old. Less than 2% were aged 13 or 19 to 21 years old. Use the results and a 1% significance level to test the claim that unwanted sex and gender are independent.  [VS](#)


	Male	Female	Total
Have had sex when did not want to			
Yes	826	883	1709
No	118	318	436
Total	944	1201	2145



: indicates the exercise has a video devoted to it in the corresponding section of STATSprofessor.com

23. Use the table below, which is from a study on vitamin C intake and the common cold, to test if vitamin C intake and colds are independent.  [VS](#)

Status	Vitamin C Group	Placebo Group	Total
Children free of colds	21	11	32
Children developing colds	36	35	71
Total	57	46	103

24. Use the table below, which is from a study on heart disease and smoking, and a 0.005 significance level to test if smoking and heart disease are related.  [VS](#)

Risk Factor	Heart Disease	No Heart Disease	Total
Smoker	25	10	35
Nonsmoker	14	51	65
Total	39	61	100

25. The table below lists the marital status for a random selection of government employees along with their job status. Use the table and a 5% significance level to test if job grade and marital status are independent (note: the test statistic is $\chi^2 = 67.397$).

Job grade	Marital status				Total
	Single	Married	Divorced	Widowed	
Low-skilled	58	874	15	8	955
Blue collar	222	3927	70	20	4269
White collar	50	2396	34	10	2490
Supervisor	7	533	7	4	551
Total	337	7730	126	42	8235



26. The table below has data from a 2009 Canadian Journal of Human Sexuality study. The data shows gender differences in response to partner influence and social expectation questions among students who had ever had sexual intercourse. The sample for this study included only students who had ever had sexual intercourse (30.7% of the total sample) yielding 2,145 respondents after corrections. The age range for the study sample was 13 to 21 years old (mean = 15.8, standard deviation = 1.19). The vast majority were 14 to 17 years old (93%) and 45% were 16 years old. Less than 2% were aged 13 or 19 to 21 years old. Use the results and a 2.5% significance level to determine if the number of sexual partners and gender are related.

Number of partners	Male	Female	Total
1	428	640	1068
2	163	226	389
3	95	131	226
4+	258	204	462
Total	944	1201	2145

A table to help you calculate $\sum \frac{(O-E)^2}{E}$:

O - E	$(O - E)^2$	$\frac{(O - E)^2}{E}$
-42.0196	1765.647	3.756535
-8.1963	67.17933	0.392403
-4.4611	19.90141	0.200093
54.6769	2989.563	14.70373
42.0196	1765.647	2.952685
8.1963	67.17933	0.308445
4.4611	19.90141	0.157274
-54.6769	2989.563	11.55699



: indicates the exercise has a video devoted to it in the corresponding section of STATSprofessor.com

13.6 Answers

21. H_0 : The two categories are independent.
 H_a : The two categories are dependent.

$$\text{TestStat} : \chi^2 = 35.783$$

$$\text{CriticalValue} = \chi_{0.025,1}^2 = 5.02$$

Conclusion : The categories seem to be related.

22. H_0 : The two categories are independent.
 H_a : The two categories are dependent.
 $\text{TestStat} : \chi^2 = 63.768$
 $\text{CriticalValue} = \chi_{0.01,1}^2 = 6.63$

Conclusion: The data allows us to reject the claim of independence. The categories seem to be related.

Expected values:

752.119	956.881
191.881	244.119

23. H_0 : The two categories are independent .
 H_a : The two categories are dependent.
 $\text{TestStat} : \chi^2 = 1.987$
 $\text{CriticalValue} = \chi_{0.05,1}^2 = 3.841$

Conclusion : The data does not allow us to reject the claim of independence. The categories do not seem to be related.

Expected values:

17.709	14.291
39.291	31.709



24. H_0 : The two categories are independent .

H_a : The two categories are dependent.

$$\text{TestStat} : \chi^2 = 23.802$$

$$\text{CriticalValue} = \chi_{0.005,1}^2 = 7.879$$

Conclusion : The data supports the claim. The categories seem to be related.

Expected values:

13.65	21.35
25.35	39.65

25. H_0 : The two categories are independent.

H_a : The two categories are dependent.

$$\text{TestStat} : \chi^2 = 67.397$$

$$\text{CriticalValue} = \chi_{0.05,9}^2 = 16.919$$

Conclusion : The data allows us to reject the claim of independence.

26. H_0 : The two categories are independent.

H_a : The two categories are dependent.

$$\text{TestStat} : \chi^2 = 34.028$$

$$\text{CriticalValue} = \chi_{0.025,3}^2 = 9.348$$

Conclusion : The data allows us to support the claim of dependence.

[Need more exercises?](#)



: indicates the exercise has a video devoted to it in the corresponding section of STATSprofessor.com

Chapter 13 Mixed Review

27. The table below shows the height and area of study of 500 randomly selected men.

	Business	Humanities	Science
64 - 68 in	20	49	40
69 - 73 in	158	79	50
74 - 78 in	55	25	24

Find the expected cell count for the third row, first column (i.e. - E31).

28. Using the data below and a 0.01 significance level, calculate the **test stat** that is used to test the claim that the colors of cars in the parking lot appear with percentages of 10%, 20%, 25%, 30%, and 15% respectively. (n = 75)

Color	Grey	White	Black	Red	Other
Frequency	9	14	20	20	12

29. Suppose that you wish to perform a chi-square test of independence. The sample data is given in the contingency table below. Are the assumptions for the test met with regard to the expected value of the cells?

	0 Drinks	Less Than 7 Drinks	7 or More Drinks	
Male	20	42	38	100
Female	60	23	12	95
	80	65	50	195

30. A medical researcher is interested in determining if there is a relationship between adults over 50 who exercise regularly and low or high blood pressure. A random sample of 236 adults over 50 is selected and the results are given below. Test the claim that regular exercise and low or high blood pressure are independent. Use $\alpha = 0.01$.

	Low BP	High BP
Reg. Exercise	77	45
No Reg. Exercise	36	78

31. Using the data below and a 0.01 significance level, determine the **critical value** that is used to test the claim that the colors of cars in the parking lot appear with percentages of 10%, 20%, 25%, 30%, and 15% respectively. (n = 75)

Color	Grey	White	Black	Red	Other
Frequency	9	14	20	20	12

32. A teacher figures that final grades in the chemistry department are distributed as: A, 25%; B, 25%; C, 40%; D, 5%; and F, 5%. At the end of a randomly selected semester, the following number of grades were recorded. State the **null and alternative hypothesis** that would be used to determine if the grade distribution for the department is different than expected.

Grade	A	B	C	D	F
Frequency	36	42	60	14	8



Chapter 13 Mixed Review Answers:

27. 48.464

28. $\chi^2 = 0.778$

29. Yes, all of the expected values are at least 5.

30. Claim: Regular exercise and BP are independent.

	Low BP	High BP
Reg. Exercise	77 58.415	45 63.585
No Reg. Exercise	36 54.585	78 59.415

$$\chi^2 = 23.486$$

Critical value: 6.635

Reject the null hypothesis (claim).

31. Critical value: 13.277

32. $H_0 : \rho_A = 0.25, \rho_B = 0.25, \rho_C = 0.40, \rho_D = 0.05, \rho_F = 0.05$

H_A : At least one of the proportions are different than the claimed values above.

