The Chi-Square Distribution: Homework*

Susan Dean
Barbara Illowsky, Ph.D.

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 3.0†

Abstract

This module provides homework on Chi-Square Distribution as a part of Collaborative Statistics collection (col10522) by Barbara Illowsky and Susan Dean.

Exercise 1

a. Explain why the "goodness of fit" test and the "test for independence" are generally right tailed tests.
b. If you did a left-tailed test, what would you be testing?

1 Word Problems

For each word problem, use a solution sheet to solve the hypothesis test problem. Go to The Table of Contents 14. Appendix for the chi-square solution sheet. Round expected frequency to two decimal places.

Exercise 2

A 6-sided die is rolled 120 times. Fill in the expected frequency column. Then, conduct a hypothesis test to determine if the die is fair. The data below are the result of the 120 rolls.

Face Value	Frequency	Expected Frequency
1	15	
2	29	
3	16	
4	15	
5	30	
6	15	

*Version 1.20: Jun 20, 2012 1:41 am -0500
†http://creativecommons.org/licenses/by/3.0/

Table 1

Exercise 3 *(Solution on p. 11.)*

The marital status distribution of the U.S. male population, age 15 and older, is as shown below. (*Source: U.S. Census Bureau, Current Population Reports*)

Marital Status	Percent	Expected Frequency
never married	31.3	
married	56.1	
widowed	2.5	
divorced/separated	10.1	

Table 2

Suppose that a random sample of 400 U.S. young adult males, 18 – 24 years old, yielded the following frequency distribution. We are interested in whether this age group of males fits the distribution of the U.S. adult population. Calculate the frequency one would expect when surveying 400 people. Fill in the above table, rounding to two decimal places.

Marital Status	Frequency
never married	140
married	238
widowed	2
divorced/separated	20

Table 3

The next two questions refer to the following information. The columns in the chart below contain the Race/Ethnicity of U.S. Public Schools for a recent year, the percentages for the Advanced Placement Examinee Population for that class and the Overall Student Population. (*Source: http://www.collegeboard.com*). Suppose the right column contains the result of a survey of 1000 local students from that year who took an AP Exam.

Race/Ethnicity	AP Examinee Population	Overall Student Population	Survey Frequency
Asian, Asian American or Pacific Islander	10.2%	5.4%	113

continued on next page

Black or African American	8.2%	14.5%	94
Hispanic or Latino	15.5%	15.9%	136
American Indian or Alaska Native	0.6%	1.2%	10
White	59.4%	61.6%	604
Not reported/other	6.1%	1.4%	43

Table 4

Exercise 4
Perform a goodness-of-fit test to determine whether the local results follow the distribution of the U. S. Overall Student Population based on ethnicity.

Exercise 5 ***(Solution on p. 11.)***
Perform a goodness-of-fit test to determine whether the local results follow the distribution of U. S. AP Examinee Population, based on ethnicity.

Exercise 6
The City of South Lake Tahoe, CA, has an Asian population of 1419 people, out of a total population of 23,609 (*Source: U.S. Census Bureau*). Suppose that a survey of 1419 self-reported Asians in Manhattan, NY, area yielded the data in the table below. Conduct a goodness of fit test to determine if the self-reported sub-groups of Asians in the Manhattan area fit that of the Lake Tahoe area.

Race	**Lake Tahoe Frequency**	**Manhattan Frequency**	
Asian Indian	131	174	
Chinese	118	557	
Filipino	1045	518	
Japanese	80	54	
Korean	12	29	
Vietnamese	9	21	
Other	24	66	

Table 5

The next two questions refer to the following information: UCLA conducted a survey of more than 263,000 college freshmen from 385 colleges in fall 2005. The results of student expected majors by gender were reported in *The Chronicle of Higher Education (2/2/2006)*. Suppose a survey of 5000 graduating females and 5000 graduating males was done as a follow-up last year to determine what their actual major was. The results are shown in the tables for Exercises 7 and 8. The second column in each table does not add to 100% because of rounding.

Exercise 7 ***(Solution on p. 11.)***
Conduct a hypothesis test to determine if the actual college major of graduating females fits the distribution of their expected majors.

Major	Women - Expected Major	Women - Actual Major
Arts & Humanities	14.0%	670
Biological Sciences	8.4%	410
Business	13.1%	685
Education	13.0%	650
Engineering	2.6%	145
Physical Sciences	2.6%	125
Professional	18.9%	975
Social Sciences	13.0%	605
Technical	0.4%	15
Other	5.8%	300
Undecided	8.0%	420

Table 6

Exercise 8

Conduct a hypothesis test to determine if the actual college major of graduating males fits the distribution of their expected majors.

Major	Men - Expected Major	Men - Actual Major
Arts & Humanities	11.0%	600
Biological Sciences	6.7%	330
Business	22.7%	1130
Education	5.8%	305
Engineering	15.6%	800
Physical Sciences	3.6%	175
Professional	9.3%	460
Social Sciences	7.6%	370
Technical	1.8%	90
Other	8.2%	400
Undecided	6.6%	340

Table 7

Exercise 9 ***(Solution on p. 11.)***

A recent debate about where in the United States skiers believe the skiing is best prompted the following survey. Test to see if the best ski area is independent of the level of the skier.

U.S. Ski Area	Beginner	Intermediate	Advanced
Tahoe	20	30	40
Utah	10	30	60
Colorado	10	40	50

Table 8

Exercise 10

Car manufacturers are interested in whether there is a relationship between the size of car an individual drives and the number of people in the driver's family (that is, whether car size and family size are independent). To test this, suppose that 800 car owners were randomly surveyed with the following results. Conduct a test for independence.

Family Size	Sub & Compact	Mid-size	Full-size	Van & Truck
1	20	35	40	35
2	20	50	70	80
3 - 4	20	50	100	90
5+	20	30	70	70

Table 9

Exercise 11 *(Solution on p. 11.)*

College students may be interested in whether or not their majors have any effect on starting salaries after graduation. Suppose that 300 recent graduates were surveyed as to their majors in college and their starting salaries after graduation. Below are the data. Conduct a test for independence.

Major	< $50,000	$50,000 - $68,999	$69,000 +
English	5	20	5
Engineering	10	30	60
Nursing	10	15	15
Business	10	20	30
Psychology	20	30	20

Table 10

Exercise 12

Some travel agents claim that honeymoon hot spots vary according to age of the bride and groom. Suppose that 280 East Coast recent brides were interviewed as to where they spent their honeymoons. The information is given below. Conduct a test for independence.

Location	20 - 29	30 - 39	40 - 49	50 and over
Niagara Falls	15	25	25	20
Poconos	15	25	25	10
Europe	10	25	15	5
Virgin Islands	20	25	15	5

Table 11

Exercise 13 *(Solution on p. 11.)*

A manager of a sports club keeps information concerning the main sport in which members participate and their ages. To test whether there is a relationship between the age of a member and his or her choice of sport, 643 members of the sports club are randomly selected. Conduct a test for independence.

Sport	18 - 25	26 - 30	31 - 40	41 and over
racquetball	42	58	30	46
tennis	58	76	38	65
swimming	72	60	65	33

Table 12

Exercise 14

A major food manufacturer is concerned that the sales for its skinny French fries have been decreasing. As a part of a feasibility study, the company conducts research into the types of fries sold across the country to determine if the type of fries sold is independent of the area of the country. The results of the study are below. Conduct a test for independence.

Type of Fries	Northeast	South	Central	West
skinny fries	70	50	20	25
curly fries	100	60	15	30
steak fries	20	40	10	10

Table 13

Exercise 15 *(Solution on p. 11.)*

According to Dan Lenard, an independent insurance agent in the Buffalo, N.Y. area, the following is a breakdown of the amount of life insurance purchased by males in the following age groups. He is interested in whether the age of the male and the amount of life insurance purchased are independent events. Conduct a test for independence.

Age of Males	None	< $200,000	$200,000 - $400,000	$401,001 - $1,000,000	$1,000,000 +
20 - 29	40	15	40	0	5
30 - 39	35	5	20	20	10
40 - 49	20	0	30	0	30
50 +	40	30	15	15	10

Table 14

Exercise 16

Suppose that 600 thirty–year–olds were surveyed to determine whether or not there is a relationship between the level of education an individual has and salary. Conduct a test for independence.

Annual Salary	Not a high school graduate	High school graduate	College graduate	Masters or doctorate
< $30,000	15	25	10	5
$30,000 - $40,000	20	40	70	30
$40,000 - $50,000	10	20	40	55
$50,000 - $60,000	5	10	20	60
$60,000 +	0	5	10	150

Table 15

Exercise 17 *(Solution on p. 12.)*

A Psychologist is interested in testing whether there is a difference in the distribution of personality types for business majors and social science majors. The results of the study are shown below. Conduct a Test of Homogeneity. Test at a 5% level of significance.

	Open	Conscientious	Extrovert	Agreeable Neurotic	
Business	41	52	46	61 58	
Social Science	72	75	63	80 65	

Table 16

Exercise 18 *(Solution on p. 12.)*

Do men and women select different breakfasts? The breakfast ordered by randomly selected men and women at a popular breakfast place is shown below. Conduct a test of homogeneity. Test at a 5% level of significance

	French Toast	Pancakes	Waffles	Omelettes
Men	47	35	28	53
Women	65	59	55	60

Table 17

Exercise 19 *(Solution on p. 12.)*

Is there a difference between the distribution of community college statistics students and the distribution of university statistics students in what technology they use on their homework? Of the randomly selected community college students 43 used a computer, 102 used a calculator with built in statistics functions, and 65 used a table from the textbook. Of the randomly selected university students 28 used a computer, 33 used a calculator with built in statistics functions, and 40 used a table from the textbook. Conduct an appropriate hypothesis test using a 0.05 level of significance.

Exercise 20 *(Solution on p. 12.)*

A fisherman is interested in whether the distribution of fish caught in Green Valley Lake is the same as the distribution of fish caught in Echo Lake. Of the 191 randomly selected fish caught in Green Valley Lake, 105 were rainbow trout, 27 were other trout, 35 were bass, and 24 were catfish. Of the 293 randomly selected fish caught in Echo Lake, 115 were rainbow trout, 58 were other trout, 67 were bass, and 53 were catfish. Perform the hypothesis test at a 5% level of significance.

Exercise 21 *(Solution on p. 12.)*

A plant manager is concerned her equipment may need recalibrating. It seems that the actual weight of the 15 oz. cereal boxes it fills has been fluctuating. The standard deviation should be at most $\frac{1}{2}$ oz. In order to determine if the machine needs to be recalibrated, 84 randomly selected boxes of cereal from the next day's production were weighed. The standard deviation of the 84 boxes was 0.54. Does the machine need to be recalibrated?

Exercise 22

Consumers may be interested in whether the cost of a particular calculator varies from store to store. Based on surveying 43 stores, which yielded a sample mean of \$84 and a sample standard deviation of \$12, test the claim that the standard deviation is greater than \$15.

Exercise 23 *(Solution on p. 12.)*

Isabella, an accomplished **Bay to Breakers** runner, claims that the standard deviation for her time to run the 7 $\frac{1}{2}$ mile race is at most 3 minutes. To test her claim, Rupinder looks up 5 of her race times. They are 55 minutes, 61 minutes, 58 minutes, 63 minutes, and 57 minutes.

Exercise 24

Airline companies are interested in the consistency of the number of babies on each flight, so that they have adequate safety equipment. They are also interested in the variation of the number of babies. Suppose that an airline executive believes the average number of babies on flights is 6 with a variance of 9 at most. The airline conducts a survey. The results of the 18 flights surveyed give a sample average of 6.4 with a sample standard deviation of 3.9. Conduct a hypothesis test of the airline executive's belief.

Exercise 25 *(Solution on p. 13.)*

The number of births per woman in China is 1.6 down from 5.91 in 1966 (Source*World Bank, 6/5/12*). This fertility rate has been attributed to the law passed in 1979 restricting births to one per woman. Suppose that a group of students studied whether or not the standard deviation of births per woman was greater than 0.75. They asked 50 women across China the number of births they had. Below are the results. Does the students' survey indicate that the standard deviation is greater than 0.75?

# of births	Frequency
0	5
1	30
2	10
3	5

Table 18

Exercise 26
According to an avid aquariest, the average number of fish in a 20–gallon tank is 10, with a standard deviation of 2. His friend, also an aquariest, does not believe that the standard deviation is 2. She counts the number of fish in 15 other 20–gallon tanks. Based on the results that follow, do you think that the standard deviation is different from 2? Data: 11; 10; 9; 10; 10; 11; 11; 10; 12; 9; 7; 9; 11; 10; 11

Exercise 27 *(Solution on p. 13.)*
The manager of "Frenchies" is concerned that patrons are not consistently receiving the same amount of French fries with each order. The chef claims that the standard deviation for a 10–ounce order of fries is at most 1.5 oz., but the manager thinks that it may be higher. He randomly weighs 49 orders of fries, which yields a mean of 11 oz. and a standard deviation of 2 oz.

2 Try these true/false questions.

Exercise 28 *(Solution on p. 13.)*
As the degrees of freedom increase, the graph of the chi-square distribution looks more and more symmetrical.

Exercise 29 *(Solution on p. 13.)*
The standard deviation of the chi-square distribution is twice the mean.

Exercise 30 *(Solution on p. 13.)*
The mean and the median of the chi-square distribution are the same if df = 24.

Exercise 31 *(Solution on p. 13.)*
In a Goodness-of-Fit test, the expected values are the values we would expect if the null hypothesis were true.

Exercise 32 *(Solution on p. 13.)*
In general, if the observed values and expected values of a Goodness-of-Fit test are not close together, then the test statistic can get very large and on a graph will be way out in the right tail.

Exercise 33 *(Solution on p. 13.)*
The degrees of freedom for a Test for Independence are equal to the sample size minus 1.

Exercise 34 *(Solution on p. 13.)*
Use a Goodness-of-Fit test to determine if high school principals believe that students are absent equally during the week or not.

Exercise 35 *(Solution on p. 13.)*
The Test for Independence uses tables of observed and expected data values.

Exercise 36 *(Solution on p. 13.)*
The test to use when determining if the college or university a student chooses to attend is related to his/her socioeconomic status is a Test for Independence.

Exercise 37 *(Solution on p. 13.)*

The test to use to determine if a six-sided die is fair is a Goodness-of-Fit test.

Exercise 38 *(Solution on p. 13.)*

In a Test of Independence, the expected number is equal to the row total multiplied by the column total divided by the total surveyed.

Exercise 39 *(Solution on p. 13.)*

In a Goodness-of Fit test, if the p-value is 0.0113, in general, do not reject the null hypothesis.

Exercise 40 *(Solution on p. 13.)*

For a Chi-Square distribution with degrees of freedom of 17, the probability that a value is greater than 20 is 0.7258.

Exercise 41 *(Solution on p. 13.)*

If df = 2, the chi-square distribution has a shape that reminds us of the exponential.

Solutions to Exercises in this Module

Solution to Exercise (p. 2)

a. The data fits the distribution
b. The data does not fit the distribution
c. 3
e. 19.27
f. 0.0002
h. Decision: Reject Null; Conclusion: Data does not fit the distribution.

Solution to Exercise (p. 3)

c. 5
e. 13.4
f. 0.0199
g. Decision: Reject null when $a = 0.05$; Conclusion: Local data do not fit the AP Examinee Distribution. Decision: Do not reject null when $a = 0.01$; Conclusion: There is insufficient evidence to conclude that Local data do not fit the AP Examinee Distribution.

Solution to Exercise (p. 3)

c. 10
e. 11.48
f. 0.3214
h. Decision: Do not reject null when $a = 0.05$ and $a = 0.01$; Conclusion: There is insufficient evidence to conclude that the distribution of majors by graduating females does not fit the distribution of expected majors.

Solution to Exercise (p. 4)

c. 4
e. 10.53
f. 0.0324
h. Decision: Reject null; Conclusion: Best ski area and level of skier are not independent.

Solution to Exercise (p. 5)

c. 8
e. 33.55
f. 0
h. Decision: Reject null; Conclusion: Major and starting salary are not independent events.

Solution to Exercise (p. 6)

c. 6
e. 25.21
f. 0.0003
h. Decision: Reject null

Solution to Exercise (p. 6)

c. 12
e. 125.74
f. 0
h. Decision: Reject null

Solution to Exercise (p. 7)

c: 4
d: Chi-Square with df = 4
e: 3.01
f: p-value = 0.5568
h: ii. Do not reject the null hypothesis.
iv. There is insufficient evidence to conclude that the distribution of personality types is different for business and social science majors.

Solution to Exercise (p. 7)

c: 3
e: 4.01
f: p-value = 0.2601
h: ii. Do not reject the null hypothesis.
iv. There is insufficient evidence to conclude that the distribution of breakfast ordered is different for men and women.

Solution to Exercise (p. 8)

c: 2
e: 7.05
f: p-value = 0.0294
h: ii. Reject the null hypothesis.
iv. There is sufficient evidence to conclude that the distribution of technology use for statistics homework is not the same for statistics students at community colleges and at universities.

Solution to Exercise (p. 8)

c: 3
d: Chi-Square with df = 3
e: 11.75
f: p-value = 0.0083
h: ii. Reject the null hypothesis.
iv. There is sufficient evidence to conclude that the distribution of fish in Green Valley Lake is not the same as the distribution of fish in Echo Lake.

Solution to Exercise (p. 8)

c. 83
d. Chi-Square with df = 83
e. 96.81
f. p-value = 0.1426; There is a 0.1426 probability that the sample standard deviation is 0.54 or more.
h. Decision: Do not reject null; Conclusion: There is insufficient evidence to conclude that the standard deviation is more than 0.5 oz. It cannot be determined whether the equipment needs to be recalibrated or not.

Solution to Exercise (p. 8)

c. 4
d. Chi-Square with df = 4
e. 4.52
f. 0.3402
h. Decision: Do not reject null.

Solution to Exercise (p. 8)

c. 49
d. Chi-Square with df = 49
e. 54.37
f. p-value = 0.2774; If the null hypothesis is true, there is a 0.2774 probability that the sample standard deviation is 0.79 or more.
h. Decision: Do not reject null; Conclusion: There is insufficient evidence to conclude that the standard deviation is more than 0.75. It cannot be determined if the standard deviation is greater than 0.75 or not.

Solution to Exercise (p. 9)

a. $\sigma^2 \leq (1.5)^2$
c. 48
d. Chi-Square with df = 48
e. 85.33
f. 0.0007
h. Decision: Reject null.

Solution to Exercise (p. 9)
True
Solution to Exercise (p. 9)
False
Solution to Exercise (p. 9)
False
Solution to Exercise (p. 9)
True
Solution to Exercise (p. 9)
True
Solution to Exercise (p. 9)
False
Solution to Exercise (p. 9)
True
Solution to Exercise (p. 9)
True
Solution to Exercise (p. 9)
True
Solution to Exercise (p. 10)
True
Solution to Exercise (p. 10)
True
Solution to Exercise (p. 10)
False
Solution to Exercise (p. 10)
False
Solution to Exercise (p. 10)
True