Chapter 11

Other Chi-Square Tests

Chapter 11 Overview

Introduction

- 11-1 Test for Goodness of Fit
- 11-2 Tests Using Contingency Tables

Chapter 11 Objectives

1. Test a distribution for goodness of fit, using chi-square.
2. Test two variables for independence, using chi-square.
3. Test proportions for homogeneity, using chi-square.

Characteristics of the Chi-Square Distribution

1. It is not symmetric.

Characteristics of the Chi-Square Distribution

1. It is not symmetric.

2. The shape of the chi-square distribution depends upon the degrees of freedom, just like Student's *t*-distribution.

Characteristics of the Chi-Square Distribution

1. It is not symmetric.

2. The shape of the chi-square distribution depends upon the degrees of freedom, just like Student's *t*-distribution.

3. As the number of degrees of freedom increases, the chi-square distribution becomes more symmetric as is illustrated in Figure 1.

Characteristics of the Chi-Square Distribution

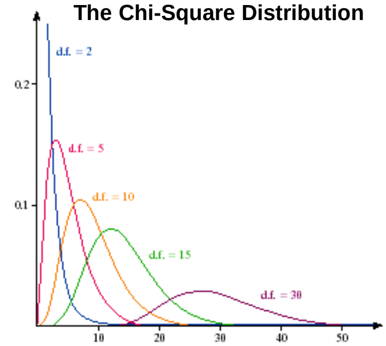
1. It is not symmetric.

2. The shape of the chi-square distribution depends upon the degrees of freedom, just like Student's *t*-distribution.

3. As the number of degrees of freedom increases, the chi-square distribution becomes more symmetric as is illustrated in Figure 1.

4. The values are non-negative. That is, the values of χ^2 are greater than or equal to 0.

The Chi-Square Distribution



11.1 Test for Goodness of Fit

- The chi-square statistic can be used to see whether a frequency distribution fits a specific pattern. This is referred to as the chi-square goodness-of-fit test.

Reading the χ^2 Table

- Find the value of χ^2 for 7 df and an area of 0.10 in the right tail of the chi-square distribution curve.

Reading the χ^2 Table

- Find the value of χ^2 for 12 df and an area of 0.05 in the left tail of the chi-square distribution curve.

Test for Goodness of Fit

Formula for the test for goodness of fit:

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

where

d.f. = k – 1 ; number of categories minus 1

O = observed frequency

E = expected frequency

Assumptions for Goodness of Fit

1. The data are obtained from a random sample.
2. The expected frequency for each category must be 5 or more.

Example 1: Fruit Soda Flavors

A market analyst wished to see whether consumers have any preference among five flavors of a new fruit soda. A sample of 100 people provided the following data. Is there enough evidence to reject the claim that there is no preference in the selection of fruit soda flavors, using the data shown previously? Let $\alpha = 0.05$.

Cherry	Strawberry	Orange	Lime	Grape
32	28	16	14	10

Step 1: State the hypotheses and identify the claim.

H_0: Consumers show no preference (claim).
(H_0: $p_1 = p_2 = p_3 = p_4 = p_5$)
H_1: Consumers show a preference.
(H_1: At least 2 of the 5 proportions are ≠ 0.20)

Example 1: Fruit Soda Flavors

	Cherry	Strawberry	Orange	Lime	Grape
Observed	32	28	16	14	10
Expected	20	20	20	20	20

Step 2: Find the critical value.

D.f. = 5 – 1 = 4, and α = 0.05. CV = 9.488.

Step 3: Compute the test value.

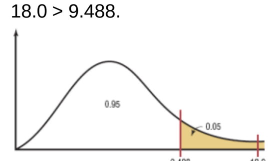
$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(32-20)^2}{20} + \frac{(28-20)^2}{20} + \frac{(16-20)^2}{20} + \frac{(14-20)^2}{20} + \frac{(10-20)^2}{20} = 18.0$$

Example 1: Fruit Soda Flavors

Step 4: Make the decision.

The decision is to reject the null hypothesis, since 18.0 > 9.488.



Step 5: Summarize the results.

There is enough evidence to reject the claim that consumers show no preference for the flavors.

Example 1: Using your graphing calc.

	Cherry	Strawberry	Orange	Lime	Grape
Observed	32	28	16	14	10
Expected	20	20	20	20	20

Step 2: Enter the observed values into L_1
Enter the expected values into L_2
calculate first and enter or enter as 100*.2

Step 3: STATS – TEST- D:χ^2GOF – Test
Observed: L_1
Expected: L_2
df: 4
Calculate

This should be displayed on your screen:
χ^2 =18
p = .001234098
df = 4 and CNTRB= (7.2 3.2…

Example 2: Retirees

The Russel Reynold Association surveyed retired senior executives who had returned to work. They found that after returning to work, 38% were employed by another organization, 32% were self-employed, 23% were either freelancing or consulting, and 7% had formed their own companies. To see if these percentages are consistent with those of Allegheny County residents, a local researcher surveyed 300 retired executives who had returned to work and found that 122 were working for another company, 85 were self-employed, 76 were either freelancing or consulting, and 17 had formed their own companies. At α = 0.10, test the claim that the percentages are the same for those people in Allegheny County.

Example 2: Retirees

	New Company	Self-Employed	Free-lancing	Owns Company
Observed	122	85	76	17

Step 1: State the hypotheses and identify the claim.

H_0: The retired executives who returned to work are distributed as follows: 38% are employed by another organization, 32% are self-employed, 23% are either freelancing or consulting, and 7% have formed their own companies (claim).

(H_0: $p_1 = .38$, $p_2 = .32$ $p_3 = .23$ $p_4 = .07$)

H_1: The distribution is not the same as stated in the null hypothesis.

(H_1: At least one of the percentages is not as stated.)

Example 2: Retirees

	New Company	Self-Employed	Free-lancing	Owns Company
Observed	122	85	76	17
Expected	.38(300)= 114	.32(300)= 96	.23(300)= 69	.07(300)= 21

Step 2: Find the critical value.

D.f. = 4 – 1 = 3, and α = 0.10. CV = 6.251.

Step 3: Compute the test value.

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

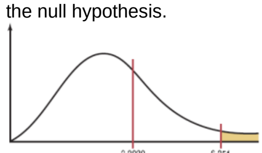
$$= \frac{(122-114)^2}{114} + \frac{(85-96)^2}{96} + \frac{(76-69)^2}{69} + \frac{(17-21)^2}{21}$$

$$= 3.2939$$

Example 2: Retirees

Step 4: Make the decision.

Since 3.2939 < 6.251, the decision is not to reject the null hypothesis.



Step 5: Summarize the results.

There is not enough evidence to reject the claim. It can be concluded that the percentages are not significantly different from those given in the null hypothesis.

Example 2: Graphing Calculator

	New Company	Self-Employed	Free-lancing	Owns Company
Observed	122	85	76	17
Expected	.38(300)=	.32(300)=	.23(300)=	.07(300)=

Step 2: Enter the observed values into L_1
Enter the expected values into L_2

Step 3: STATS – TEST- D:χ^2GOF – Test
Observed: L_1
Expected: L_2
df: 3
Calculate
This should be displayed on your screen:
χ^2 =3.293869865
p = .3484967603
df = 3 and CNTRB= {.561403…

Example 11-3: ATM Usage (p. 495)

A bank has an ATM installed inside the bank, and it is available to its customers only from 7 AM to 6 PM Mon – Fri. The manager of the bank wanted to investigate if the percentage of transactions made on the ATM is the same for each of the five days (M – F) of the week. She randomly selected one week and counted the number of transactions made on this ATM on each of the 5 days. At α = 0.01, can we reject the null hypothesis that the proportion of people who use this ATM each of the 5 days is the same? (Assume this week is typical of all weeks regarding ATM use)

DAY	Mon	Tues	Wed	Thurs	Fri
# of users	253	197	204	279	267

Example 11-3: ATM Usage

DAY	Mon	Tues	Wed	Thurs	Fri
# of users	253	197	204	279	267

Step 1: State the hypotheses and identify the claim.

H_0: The proportion of people using the ATM is the same for all 5 days of the week.

(H_0: $p_1 = p_2 = p_3 = p_4 = p_5$)

H_1: The distribution is not the same as stated in the null hypothesis.

(H_1: At least 2 of the 5 proportions are ≠ 0.20)

Example 11-3: ATM Usage

DAY	Mon	Tues	Wed	Thurs	Fri
# of users	253	197	204	279	267

Step 2: Find the critical value. Df = k - 1

D.f. = 5 – 1 = 4, and α = 0.01. CV = 13.277.

Step 3: Compute the test value.

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

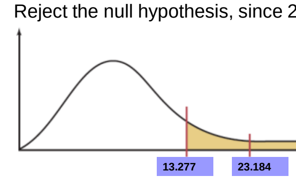
$$\frac{(253-240)^2}{240} + \frac{(197-240)^2}{240} + \frac{(204-240)^2}{240} + \frac{(297-240)^2}{240} + \frac{(267-240)^2}{240}$$

= 23.183

Example 11-3: ATM Usage

Step 4: Make the decision.

Reject the null hypothesis, since 23.184 > 13.277.



Step 5: Summarize the results.

There is enough evidence to reject the claim that the proportion of people using the ATM is the same for all 5 days of the week.

11.3 Tests Using Contingency Tables

- **When data can be tabulated in table form in terms of frequencies, several types of hypotheses can be tested by using the chi-square test.**
- **The test of independence of variables is used to determine whether two variables are independent of or related to each other when a single sample is selected.**
- **The test of homogeneity of proportions is used to determine whether the proportions for a variable are equal when several samples are selected from different populations.**

Test for Independence

- **The chi-square goodness-of-fit test can be used to test the independence of two variables.**
- **The hypotheses are:**
  - **H_0: There is no relationship between two variables.**
  - **H_1: There is a relationship between two variables.**
- **If the null hypothesis is rejected, there is some relationship between the variables.**

Test for Independence

- **In order to test the null hypothesis, one must compute the expected frequencies, assuming the null hypothesis is true.**
- **When data are arranged in table form for the independence test, the table is called a contingency table.**

Contingency Tables

	Column 1	Column 2	Column 3
Row 1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$
Row 2	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$

- **The degrees of freedom for any contingency table are d.f. = (rows – 1) (columns – 1) = (*R* – 1)(*C* – 1).**

Test for Independence

The formula for the test for independence:

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

where

d.f. = $(R-1)(C-1)$

O = observed frequency

E = expected frequency = $\dfrac{(\text{row sum})(\text{column sum})}{\text{grand total}}$

Example 5: College Education and Place of Residence

A sociologist wishes to see whether the number of years of college a person has completed is related to her or his place of residence. A sample of 88 people is selected and classified as shown. At $\alpha = 0.05$, can the sociologist conclude that a person's location is dependent on the number of years of college?

Location	No College	Four-Year Degree	Advanced Degree	Total
Urban	15	12	8	35
Suburban	8	15	9	32
Rural	6	8	7	21
Total	29	35	24	88

Example 5: College Education and Place of Residence

Step 1: State the hypotheses and identify the claim.

H_0: A person's place of residence is independent of the number of years of college completed.

H_1: A person's place of residence is dependent on the number of years of college completed (claim).

Step 2: Find the critical value.

The critical value is 9.488, since the degrees of freedom are (3 – 1)(3 – 1) = 4.

Example 5: College Education and Place of Residence

Compute the expected values.

$$E = \frac{(\text{row sum})(\text{column sum})}{\text{grand total}} \qquad E_{1,1} = \frac{(35)(29)}{88} = 11.53$$

Location	No College	Four-Year Degree	Advanced Degree	Total
Urban	15 (11.53)	12 (13.92)	8 (9.55)	35
Suburban	8 (10.55)	15 (12.73)	9 (8.73)	32
Rural	6 (6.92)	8 (8.35)	7 (5.73)	21
Total	29	35	24	88

Example 5: College Education and Place of Residence

Step 3: Compute the test value.

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

$$= \frac{(15-11.53)^2}{11.53} + \frac{(12-13.92)^2}{13.92} + \frac{(8-9.55)^2}{9.55}$$

$$+ \frac{(8-10.55)^2}{10.55} + \frac{(15-12.73)^2}{12.73} + \frac{(9-8.73)^2}{8.73}$$

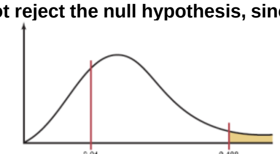
$$+ \frac{(6-6.92)^2}{6.92} + \frac{(8-8.35)^2}{8.35} + \frac{(7-5.73)^2}{5.73}$$

$$= 3.01$$

Example 5: College Education and Place of Residence

Step 4: Make the decision.

Do not reject the null hypothesis, since 3.01<9.488.



Step 5: Summarize the results.

A person's place of residence is independent of the number of years of college completed.

Graphing Calculator

- Press **2nd x-1** for **MATRIX** and move the cursor to **edit**, then press **enter**
- Enter the number of rows & columns then **enter**
- Enter the values in the matrix as they appear in the contingency table
- **Press STAT – TEST – C: χ^2 – Test** Make sure the observed matrix is [A] and the expected matrix is [B]
- **Calculate, enter**

You should see

- χ^2 – Test
- **χ^2 = 3.005538842**
- **P = .5568989167**
- **df = 4**

Example6: Alcohol and Gender

A researcher wishes to determine whether there is a relationship between the gender of an individual and the amount of alcohol consumed. A sample of 68 people is selected, and the following data are obtained. At $\alpha = 0.10$, can the researcher conclude that alcohol consumption is related to gender?

Gender	Alcohol Consumption Low	Moderate	High	Total
Male	10	9	8	27
Female	13	16	12	41
Total	23	25	20	68

Graphing Calculator

- Press **2nd x-1** for **MATRIX** and move the cursor to **edit**, then press **enter**
- Enter the number of rows & columns then **enter**
- Enter the values in the matrix as they appear in the contingency table
- **Press STAT – TEST – C: χ^2 – Test** Make sure the observed matrix is [A] and the expected matrix is [B]
- **Calculate, enter**

Example 6: Alcohol and Gender

Step 1: State the hypotheses and identify the claim.

H_0: The amount of alcohol that a person consumes is independent of the individual's gender.

H_1: The amount of alcohol that a person consumes is dependent on the individual's gender (claim).

Step 2: Find the critical value.

The critical value is 9.488, since the degrees of freedom are (3 – 1)(3 – 1) = (2)(2) = 4.

Example 6: Alcohol and Gender

Compute the expected values.

$$E = \frac{(\text{row sum})(\text{column sum})}{\text{grand total}} \qquad E_{1,1} = \frac{(27)(23)}{68} = 9.13$$

Gender	Alcohol Consumption Low	Moderate	High	Total
Male	10 (9.13)	9 (9.93)	8 (7.94)	27
Female	13 (13.87)	16 (15.07)	12 (12.06)	41
Total	23	25	20	68

Example 6: Alcohol and Gender

Step 3: Compute the test value.

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

$$= \frac{(10-9.13)^2}{9.13} + \frac{(9-9.93)^2}{9.93} + \frac{(8-7.94)^2}{7.94}$$

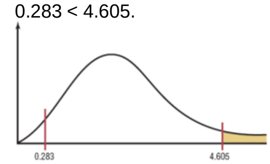
$$+ \frac{(13-13.87)^2}{13.87} + \frac{(16-15.07)^2}{15.07} + \frac{(12-12.06)^2}{12.06}$$

$$= 0.283$$

Example 6: Alcohol and Gender

Step 4: Make the decision.

Do not reject the null hypothesis, since 0.283 < 4.605.



Step 5: Summarize the results.

There is not enough evidence to support the claim that the amount of alcohol a person consumes is dependent on the individual's gender.

Test for Homogeneity of Proportions

- **Homogeneity of proportions test is used when samples are selected from several different populations and the researcher is interested in determining whether the proportions of elements that have a common characteristic are the same for each population.**

Test for Homogeneity of Proportions

- **The hypotheses are:**
  - H_0: $p_1 = p_2 = p_3 = \ldots = p_n$
  - H_1: **At least one proportion is different from the others.**
- **When the null hypothesis is rejected, it can be assumed that the proportions are not all equal.**

Assumptions for Homogeneity of Proportions

1. **The data are obtained from a random sample.**
2. **The expected frequency for each category must be 5 or more.**

Example 7: Lost Luggage

A researcher selected 100 passengers from each of 3 airlines and asked them if the airline had lost their luggage on their last flight. The data are shown in the table. At α = 0.05, test the claim that the proportion of passengers from each airline who lost luggage on the flight is the same for each airline.

	Airline 1	Airline 2	Airline 3	Total
Yes	10	7	4	21
No	90	93	96	279
Total	100	100	100	300

Graphing Calculator

- Press **2nd x-1** for **MATRIX** and move the cursor to **edit**, then press **enter**
- Enter the number of rows & columns then **enter**
- Enter the values in the matrix as they appear in the contingency table
- **Press STAT – TEST – C: χ^2 – Test** Make sure the observed matrix is [A] and the expected matrix is [B]
- **Calculate, enter**

Example 7: Lost Luggage

Step 1: State the hypotheses.

H_0: $p_1 = p_2 = p_3 = \ldots = p_n$

H_1: At least one mean differs from the other.

Step 2: Find the critical value.

The critical value is 5.991, since the degrees of freedom are (2 – 1)(3 – 1) = (1)(2) = 2.

Example 7: Lost Luggage

Compute the expected values.

$$E = \frac{(\text{row sum})(\text{column sum})}{\text{grand total}} \qquad E_{1,1} = \frac{(21)(100)}{300} = 7$$

	Airline 1	Airline 2	Airline 3	Total
Yes	10 (7)	7 (7)	4 (7)	21
No	90 (93)	93 (93)	96 (93)	279
Total	100	100	100	300

Example 7: Luggage

Step 3: Compute the test value.

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

$$= \frac{(10-7)^2}{7} + \frac{(7-7)^2}{7} + \frac{(4-7)^2}{7} + \frac{(90-93)^2}{93} + \frac{(93-93)^2}{93} + \frac{(96-93)^2}{93}$$

$$= 2.765$$

Example 7: Lost Luggage

Step 4: Make the decision.

Do not reject the null hypothesis, since 2.765 < 5.991.

Step 5: Summarize the results.

There is no difference in the proportions of the luggage lost by each airline.