# 11 - 1. Test for Goodness of Fit

## Objective 1. Test a Distribution for Goodness of Fit, Using Chi-Square.

The chi-square statistic can be used to determine whether a frequency distribution fits a specific pattern. Some examples include, do customers have a preference for a particular color or are all colors preferred equally, do accidents occur at certain intersections more often than at other intersections or not, or do causes of airline delays occur at different or the same rates?

Recall that chi-square distributions:

1. Are families of curves based on degrees of \_\_\_\_\_\_\_\_\_\_\_\_\_;
2. Are \_\_\_\_\_\_\_\_\_\_\_\_ skewed;
3. Are greater than or equal to zero;
4. Have a total area equal to \_\_\_\_\_ under the distribution.

### Chi-Square Goodness-of-Fit Test

The chi-square goodness-of-fit test is used to test the claim that an observed frequency distribution fits an expected distribution or a specific pattern.

Actual frequencies are called the observed frequencies.

Expected frequencies are those obtained by calculation as if there were a specific pattern.

### Calculate Expected Frequencies

1. If all the expected frequencies are \_\_\_\_\_\_\_\_\_\_\_, the expected frequency *E* can be calculated using , where *n* is the number of \_\_\_\_\_\_\_\_\_\_ and *k* is the number of \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
2. If all the expected frequencies are not \_\_\_\_\_\_\_\_\_, then the expected frequency *E* can be calculated by , where *n* is the number of observed frequencies and *p* is the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ for that category.

### Example 11-1. Finding Expected Frequencies

Find the expected frequencies for a sample of 175 truck colors when

1. the truck colors are equally popular; or
2. the proportion of popular truck colors are white 31%, black 19%, silver 11%, red 11%, gray 10%, blue 8%, and other 10%.

*Solution:*

|  | **White** | **Black** | **Silver** | **Red** | **Gray** | **Blue** | **Other** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| a. Equal  Frequencies |  |  |  |  |  |  |  |
| b. Category  Probability |  |  |  |  |  |  |  |

Observed frequencies will almost always differ from the expected frequencies due to sampling error; that is, the frequencies differ from sample to sample. We want to know if the differences are significant or due to chance.

The hypotheses are:

*H0;* There is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ in the observed frequencies and the expected frequencies.

If the null hypothesis is not rejected, it would mean that the observed frequencies have a good fit with the expected frequencies.

*H1:* There is a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ in the observed frequencies and the expected frequencies.

If the alternate hypothesis is supported, it would mean that the observed frequencies do not fit well with the expected frequencies.

The measure of discrepancy between the observed values *O* and the expected values *E* is the chi-square statistic.

### Formula for the Chi-Square Goodness-of-Fit Test

with degrees of freedom equal to the number of categories minus 1, and

where *O* = observed frequency and *E* = expected frequency.

The value of the test statistic is based on the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ between the observed and expected values. When the observed and expected frequencies \_\_\_\_\_\_\_\_\_\_, then is close to *0*. When the \_\_\_\_\_\_\_\_\_\_\_\_\_\_ is significant and the observed and expected frequencies do not fit well, then is large. Also, can never be negative.

### Assumptions for the Chi-Square Goodness-of-Fit Test

1. The data are obtained from a \_\_\_\_\_\_\_\_\_\_\_\_ sample
2. The expected frequency for each category must be \_\_\_\_\_\_\_or more.

### Procedure for the Chi-Square Goodness-of-Fit Test

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the \_\_\_\_\_\_\_\_\_\_ value from Table G. The test is always right tailed.

**Step 3** Compute the \_\_\_\_\_\_\_\_\_ value: \_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step 4** Make the decision.

**Step 5** Summarize the results.

### Example 11-2. Arrests for Crimes

In a certain city, the police department wants to determine whether there is a difference in the number of arrests for four types of crimes. A random sample of 160 arrests showed 38 for larceny thefts, 50 for property crimes, 28 for drug use and 44 for driving under the influence. Use

*Solution:*

**Step 1** State the hypotheses and identify the claim.

*H0;* There is no difference in the number of arrests for each type of crime. (claim)

*H1:* There is a difference in the number of arrests for each type of crime.

**Step 2** Find the critical value.

The critical value from Table G with degrees of freedom and is .

**Step 3** Compute the expected values: .

|  | **Larceny thefts** | **Property crimes** | **Drug**  **use** | **Driving under the influence** |
| --- | --- | --- | --- | --- |
| **Observed** | 38 | 50 | 28 | 44 |
| **Expected** | 40 | 40 | 40 | 40 |
|  |  |  |  |  |

Compute the test value.

**Step 4** Make the decision.

Do not reject the null hypothesis since the test value of which is the critical value and the test is a right tailed test.

**Step 5** Summarize the results.

There is not enough evidence to reject the claim that there is no difference in the number of arrests for each type of crime.

The test value of falls between the critical values of and , corresponding to levels of and , indicating that

### Example 11-3. Truck Colors

In a recent year, the most popular colors for light trucks were white 31%, black 19%, silver 11%, red 11%, gray 10%, blue 8%, and other 10%. A survey of randomly selected light truck owners in a particular area yielded the following results:

|  | **White** | **Black** | **Silver** | **Red** | **Gray** | **Blue** | **Other** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Survey results** | 43 | 32 | 30 | 27 | 22 | 15 | 6 |

Use .

*Solution*:

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the critical value from Table G. The test is always right tailed.

**Step 3** Compute the test value:

From Example 11.1, we know:

|  | **White** | **Black** | **Silver** | **Red** | **Gray** | **Blue** | **Other** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Expected frequencies** |  |  |  |  |  |  |  |

**Step 4** Make the decision.

**Step 5** Summarize the results.

### Using Chi-Square Goodness-of-Fit to Test for Normality

The goodness-of-fit may also be used to test if a variable is \_\_\_\_\_\_\_\_\_\_\_\_\_\_ distributed. To do so, for a frequency distribution, the expected frequency of each class is determined using the standard normal distribution.

### Procedure for Test of Normality

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the \_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_ of the variable displayed in the frequency chart.

**Step 3** Find the \_\_\_\_\_\_\_ under the standard normal distribution, for each class, using the *z* values and Table E.

**Step 4** Find the \_\_\_\_\_\_\_\_\_\_\_\_\_\_ frequencies for each class by multiplying the area by the total number of observations.

**Step 5** Find the critical value from Table G. The test is always right tailed.

**Step 6** Compute the test value:

**Step 7** Make the decision.

**Step 8** Summarize the results.

### Example 11-4. Test of Normality

Use chi-square to determine if the variable shown in the frequency distribution is normally distributed. Use .

| **Boundaries** | **Frequency** |
| --- | --- |
| 89.5-104.5 | 24 |
| 104.5-119.5 | 62 |
| 119.5-134.5 | 72 |
| 134.5-149.5 | 26 |
| 149.5-164.5 | 12 |
| 164.5-179.5 | 4 |
|  | **Total = 200** |

*Solution:*

**Step 1** State the hypotheses and identify the claim.

*H0*: The variable is normally distributed.

*H1*: The variable is not normally distributed

**Step 2** Find the mean and standard deviation of the variable displayed in the frequency chart.

| **Boundaries** | **Frequency** | ***Xm*** | ***f(Xm)*** | ***f(Xm)2*** |
| --- | --- | --- | --- | --- |
| 89.5-104.5 | 24 | 97 | 2328 | 225816 |
| 104.5-119.5 | 62 | 112 | 6944 | 777728 |
| 119.5-134.5 | 72 | 127 | 9144 | 1161288 |
| 134.5-149.5 | 26 | 142 | 3692 | 524264 |
| 149.5-164.5 | 12 | 157 | 1884 | 295788 |
| 164.5-179.5 | 4 | 172 | 688 | 118336 |
|  | **Total = 200** |  | **24680** | **3103220** |

**Step 3** Find the area under the standard normal distribution, for each class, using the *z* values and Table E.

Find the z scores for the upper class limit and use Table E to find the area for the class:

The area for \_\_\_\_\_\_\_\_\_\_\_\_\_.

The area for .

The area for \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_= \_\_\_\_\_\_\_\_\_\_\_\_\_

The area for

The area for for.

The area for

**Step 4** Find the expected frequencies for each class by multiplying the area by the total number of observations.

The expected frequencies (use two decimal places) for each class are:

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |

**Step 5** Find the critical value from Table G. The test is always right tailed.

The critical value in this test has degrees of freedom equal to the number of categories, *n*, minus *3* since *1* degree of freedom is lost for each parameter that is estimated. We have estimated the mean and standard deviation, so *2* additional degrees of freedom are needed.

The critical value with and is .

(Only 5 classes have counts of 5 or more.)

**Step 6** Compute the test value:

The last category is less than 5, so combined with the previous category.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Observed** | 24 | 62 | 72 | 26 | 16 |
| **Expected** | 26.70 | 55.10 | 66.64 | 38.96 | 12.60 |

**Step 7** Make the decision.

, so the null hypothesis is rejected.

**Step 8** Summarize the results.

Therefore, there is enough evidence to support that the distribution is not normally distributed.

# 11 – 2. Tests Using Contingency Tables

## Objective 2. Test Two Variables for Independence Using Chi-Square.

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 uses a chi-square distribution and a contingency table.

### Chi-Square Independence Test

The chi-square independence test is used to test whether two variables are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of each other.

.

with degrees of freedom equal to the product of one less than the number of rows times one less that the number of columns:

### Assumptions for Chi-Square Independence Test

1. The data are obtained from a \_\_\_\_\_\_\_\_\_\_\_\_ sample.
2. The expected value in each cell must be \_\_\_\_\_ or more.  
   If the expected values are not 5 or more, \_\_\_\_\_\_\_\_\_\_ categories.

### 

### Hypotheses of the Chi-Square Independence Test

*H0*: The variables are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of each other.

*H1*: The variables are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ upon each other.

The data for the two variables are placed into a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ table with one variable called the *\_\_\_\_\_\_\_variable* and the other called the *\_\_\_\_\_\_\_\_\_\_ variable*. The table has cells, where *R* is the number of rows and *C* is the number of columns.

For instance, the following is a contingency table:

|  | **Column 1** | **Column 2** | **Column 3** |
| --- | --- | --- | --- |
| **Row 1** | *C1,1* | *C1,2* | *C1,3* |
| **Row 2** | *C2,1* | *C2,2* | *C2,3* |

The cell values are denoted by *Cr,c*.

### The Chi-Square Independence Test

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the critical value for the right tail . Use Table G.

**Step 3** Compute the test value.  
 First find the expected values: . .  
 Compute the test value: .

**Step 4** Make the decision.

**Step 5** Summarize the results.

### Example 11-5. Procedure Preference

A researcher selects a random sample of nurses and doctors to determine whether doctors have different preferences about a new post-operative procedure that had been administered to a number of patients in a large hospital

Use . The results are displayed in the contingency table:

| **Group** | **Prefer new procedure** | **Prefer old procedure** | **No preference** | ***Totals*** |
| --- | --- | --- | --- | --- |
| **Nurses** | 100 | 80 | 20 |  |
| **Doctors** | 50 | 120 | 30 |  |
| ***Totals*** |  |  |  |  |

*Solution*:

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the critical value for the right tail . Use Table G.

**Step 3** Compute the test value.  
 First find the expected values: . .

Compute the test value: .

**Step 4** Make the decision.

**Step 5** Summarize the results.

### Example 11-6. Music CDs Sold

Are the sales of CDs (in millions) by genre related to the year in which the sales occurred? Use level of significance.

|  | **R & B** | **Country** | **Rock** | ***Totals*** |
| --- | --- | --- | --- | --- |
| **2013** | 100 | 80 | 20 |  |
| **2014** | 50 | 120 | 30 |  |
| ***Totals*** |  |  |  |  |

*Solution*:

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the critical value for the right tail . Use Table G.

**Step 3** Compute the test value.  
 First find the expected values: . .

Compute the test value: .

**Step 4** Make the decision.

**Step 5** Summarize the results.

## Objective 3. Test Proportions for Homogeneity, Using Chi-Square

The test of homogeneity of proportions is used to test the claim that different populations have the \_\_\_\_\_\_\_\_\_ proportion of subjects who have a certain attitude or characteristic. The test uses a chi-square distribution and a contingency table. The process mirrors that of the chi-square test of independence.

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 \_\_\_\_\_\_\_\_ for each population. Sample sizes are specified in advance. Thus, either the row totals or the column totals are known before samples are selected.

### Chi-Square Test of Homogeneity of Proportions

The chi-square test of homogeneity is used to test whether different populations have the \_\_\_\_\_\_\_\_\_\_ proportion of subjects having the same attitude or characteristic.

.

with degrees of freedom equal to the \_\_\_\_\_\_\_\_\_\_\_\_\_\_ of one less than the number of rows times one less that the number of columns:

### Assumptions for Chi-Square Test of Homogeneity of Proportions

1. The data are obtained from a \_\_\_\_\_\_\_\_\_\_\_\_ sample.
2. The expected value in each cell must be \_\_\_\_\_\_\_\_ or more.  
   If the expected values are not \_\_\_\_\_ or more, \_\_\_\_\_\_\_\_\_\_\_ categories.

### Hypotheses of the Chi-Square Test of Homogeneity of Proportions

*H0*: The proportions are equal. That is,

*H1*: At least one proportion is different from the others.

### The Chi-Square Test of Homogeneity of Proportions

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the critical value for the right tail . Use Table G.

**Step 3** Compute the test value.  
 First find the expected values: . .  
 Compute the test value: .

**Step 4** Make the decision.

**Step 5** Summarize the results.

### Example 11-7. Fathers in the Delivery Room

On average, 79% of American fathers are in the delivery room when their children are born. A physician’s assistant surveyed 300 randomly selected first-time fathers, 75 each from four different hospitals, to determine if they had been in the delivery room when their children were born. The results are in the table. At , is there enough evidence to reject the claim that the proportions of those who were in the delivery room at the time of birth at the four hospitals are the same?

|  | **Hospital A** | **Hospital B** | **Hospital C** | **Hospital D** | ***Total*** |
| --- | --- | --- | --- | --- | --- |
| **Present** | 66 | 60 | 57 | 56 |  |
| **Not present** | 9 | 15 | 18 | 19 |  |
| ***Total*** | 75 | 75 | 75 | 75 |  |

*Solution:*

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the critical value for the right tail . Use Table G.  
 d.f. = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

C. V. + \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step 3** Compute the test value.  
 First find the expected values: . .

|  | **Hospital A** | **Hospital B** | **Hospital C** | **Hospital D** | ***Total*** |
| --- | --- | --- | --- | --- | --- |
| **Present** | 66 ( ) | 60 ( ) | 57 ( ) | 56 ( ) |  |
| **Not present** | 9 ( ) | 15 ( ) | 18 ( ) | 19 ( ) |  |
| ***Total*** | 75 | 75 | 75 | 75 |  |

Compute the test value: .

**Step 4** Make the decision.

**Step 5** Summarize the results.