CHAPTER 21 Appendix

The Chi-Square Test for Goodness of Fit

The chi-square test essentially extends the one-proportion test with H_0: $p(\text{success}) = p_0$ and its unstated companion $p(\text{failure}) = 1 - p_0$ to more than two categories. The null hypothesis specifies the proportions in the population for all categories of the variable of interest, subject to the constraint that they add to 1. We ask whether the observed counts are likely to have been observed if the null hypothesis is true.

The chi-square test for goodness of fit is a "many-sided" test, in that rejecting the null hypothesis does not say which observed proportions are different from what was specified, nor does it indicate how they differ. However, once the null hypothesis is rejected, we can compare the observed counts to the expected counts and examine the contribution to the chi-square test statistic from each cell. The cell with the largest contribution differs most from the null hypothesis; the comparison of the counts will inform you about the direction of the difference.

Goodness-of-Fit Tests

Excel

There is no command or dialog in Excel to perform this test. However, the chi-square statistic can be calculated easily using spreadsheet commands.

1. Enter the observed counts in a column. Enter the corresponding hypothesized proportions (in decimal form) in another column. An example using the data on seed coat colors is shown here.

Observed	Hypothesized Proportion
321	0.7500
77	0.1875
31	0.0625

2. If needed, click in an empty cell and calculate the sum of the counts (n) using the **=sum(** command. Here, with counts in column a, that command would be **=sum(a2..a4)**. In our example, that sum is placed in cell a5.
3. Click in the cell next to the first hypothesized proportion. Calculate its expected count as $n*p$. Here, that formula would go in cell c2 as **=a$5*b2**. Use the cursor to copy that command down column c as far as needed.
4. Calculate the first component $(O - E)^2/E$. Using our example, that would be entered (in cell d2) as **=(a2-c2)^2/(c2)**.
5. Use the cursor to drag and copy that formula down the column.

6. Click in an empty cell and use the sum command to find the value of X^2. Here, that command would be **=sum(d2..d4)**. At this point, our spread sheet includes the following data:

	A	B	C	D
1	Observed	Hypothesized Proportion	Expected Count	Contribution
2	321	0.75	321.75	0.001748252
3	77	0.1875	80.4375	0.146901709
4	31	0.0625	26.8125	0.653991841
5	429			0.802641803

7. Find the P-value of the test. Click in an empty cell, then select **Formulas ➔ More Formulas ➔ Statistical ➔ CHISQ.DIST.RT**. Enter the value of the statistic just found and (categories – 1) for the degrees of freedom. The P-value here is 0.6694.

---

1. Enter the category labels in one column and the observed counts in another column.
2. **Analyze ➔ Distribution**
3. Select the category labels as **Y, Columns** and the observed counts as **Freq**.
4. **OK**
5. Click on the red triangle beside the variable name; select **Test Probabilities**.
6. Enter the hypothesized probability for each category.
7. **Done**
8. If the null hypothesis is rejected (the test has a small P-value), insert a new variable as a formula, with the formula $n*p$. This will allow you to compare the observed and expected counts.

JMP gives two different statistics and P-values. You want the one labeled Pearson.

The JMP Video Technology Manual: *Chi-Square GOF Test* offers more explanation and a worked example.

Minitab

1. Enter the observed counts in one column and the corresponding hypothesized proportions in another column.
2. **Stat ➔ Tables ➔ Chi-Square Goodness-of-Fit (One Variable)**
3. Click to **Select** and enter the column of observed counts in the labeled box.
4. Move the radio button to **Specific proportions**; click to select and enter the column of hypothesized proportions.
5. **OK**

By default, Minitab produces two graphs: one displaying the contributions to the chi-square statistic and another that compares the observed and expected counts. You can turn these off by clicking **Graphs** and unchecking the boxes.

The Minitab Video Technology Manual *Chi-Square GOF Test* offers additional explanation and a worked example.

SPSS

1. Enter a *numeric* category identifier in one column and the observed counts in another column.
2. **Data ➔ Weight Cases**
3. Change the radio button to **Weight cases by**. Click to highlight and select the observed counts variable into the box.
4. **OK**
5. **Analyze ➔ Nonparametric Tests ➔ Legacy Dialogs ➔ Chi-square**
6. Click to select the numeric category variable into the **Test Variable** list box.
7. Calculate the expected counts using **Transform ➔ Compute Variable**. Name a new column and enter the formula *n**p*, using whatever names you gave those columns.
8. Move the radio button in the Expected Values block to **Values**. Enter the expected counts you just calculated; click **Add** after each.
9. **OK**

CRUNCHIT!

1. Enter the observed counts in one column and the hypothesized proportions in another column.
2. **Statistics ➔ Chi-squared Goodness-of-Fit**
3. Click to select the columns into the labeled boxes.
4. **Calculate**

The CrunchIt! Help Video: *Chi-Square Goodness of Fit Test* provides an example.

TI-83/-84

If you have a TI-84:

1. Enter the observed counts in one list and the expected counts (calculate these as np_i) in another list.
2. Press STAT ▶ ▶ for Tests.
3. Select option **D:X²GOF-Test**.
4. Enter the list names for the observed and expected counts, and the degrees of freedom for the test (categories – 1).
5. Press ▼ to Calculate.
6. Press ENTER.

If you have a TI-83:

1. Enter the observed counts in one list and the expected counts (calculate these as np_i) in another list.
2. Place the cursor in the list name of an empty list. Calculate the components of the chi-square statistic as $(O - E)^2/E$. Assuming the observed counts are in L1 and the expected counts are in L2, the entry will look like this:

```
=(L1-L2)^2/L2
```

3. Sum the components. Press 2nd STAT ▶ ▶ for List Math. Select option **5:sum(**, then enter the name of the list where the components were stored. Press ENTER.
4. Find the *P*-value of the test. Press 2nd VARS and select option **8:X^{2}cdf(**. Enter the calculated test statistic, then 1 2nd , 9 9 (for +∞, or a large positive number), and the degrees of freedom, categories – 1.
5. Press ENTER.

The TI Video Technology Manuals video *Chi-Square GOF Test* provides additional explanation and worked examples.

Enter a variable for observed counts (here, **obs**) and another variable for expected proportions (here, **prop0**). The command for the test is

```
> chisq.test(obs,p=prop0)
```

The R Video Technology Manuals video *Chi-Square Goodness of Fit Test* provides additional explanation and worked examples.