

H0: There is no difference between the two therapies’ ability to cure cocaine dependence

We next calculate the Expected Values from the Observed Values and then the p-value of the chi-square statistic as we did in Example 1.

This time, however, we will use the approach employed in Example 2 of [Goodness of Fit](https://www.real-statistics.com/chi-square-and-f-distributions/goodness-of-fit/" \o "Goodness of Fit" \t "_blank), namely calculating the Pearson’s chi-square test statistic directly (using Definition 2 of [Goodness of Fit](https://www.real-statistics.com/chi-square-and-f-distributions/goodness-of-fit/" \o "Goodness of Fit" \t "_blank)).

The value of this statistic is 5.516

Since we are dealing with a 2 × 2 table of observations, df = (2 – 1)(2 – 1) = 1.

p-value = CHISQ.DIST.RT(χ2, df) = CHISQ.DIST.RT(5.516,1) = **.0188** < .05 = α

χ2-crit = CHISQ.INV.RT(α, df) = CHISQ.INV.RT(.05,1) = **3.841 < 5.516** = χ2-obs

We reject the null hypothesis and conclude there is a significant difference in the cure rate between the two therapies.

The value of this statistic is 5.725, which is not much different from the test statistic we obtained using the Pearson’s test. Since this statistic is also approximately chi-square with one degree of freedom, the analysis is quite similar:

p-value = CHISQ.DIST.RT(χ2, *df*) = CHISQ.DIST.RT(5.725,1) = .015 < .05 = *α*

χ2-crit = CHISQ.INV.RT(*α, df*) = CHISQ.INV.RT(.05,1) = 3.841 < 5.725 = χ2-obs

and so once again, we reject the null hypothesis and conclude there is significant difference in the results for the two therapies.

**Observation**: It is very important to include all observations in the test. E.g. if in Example 2 we only test Cured vs. Therapy 1 and 2, we will get erroneous results. We need to include Not Cured as well as Cured.

**Real Statistics Excel Functions**: The following supplemental functions are provided in the Real Statistics Resource Pack:

**CHI\_STAT2**(R1, R2) = Pearson’s chi-square statistic for observation values in range R1 and expectation values in range R2

**CHI\_MAX2**(R1, R2) = Maximum likelihood chi-square statistic for observation values in range R1 and expectation values in range R2

**CHI\_STAT**(R1) = Pearson’s chi-square statistic for observation values in range R1. This is CHI\_STAT2(R1, R2) where R2 is the expectation values calculated from R1.

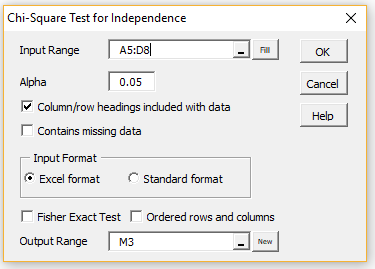
**CHI\_MAX**(R1) = Maximum likelihood chi-square statistic for observation values in range R1. This is CHI\_MAX2(R1, R2) where R2 is the expectation values calculated from R1.

**CHI\_TEST**(R1) = p-value for Pearson’s chi-square statistic for observation values in range R1. This is CHISQ.TEST(R1, R2) where R2 is the expectation values calculated from R1.

**CHI\_MAX\_TEST**(R1) = p-value for Maximum likelihood chi-square statistic for observation values in range R1

The ranges R1 and R2 must contain only numeric values.

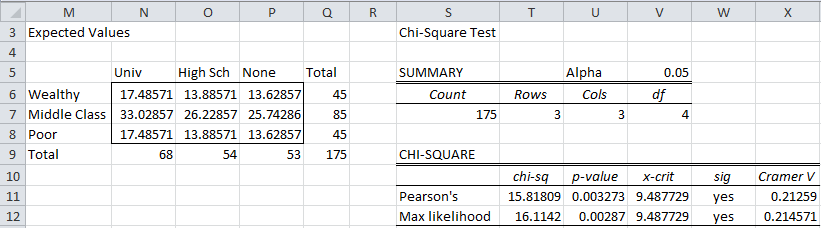
**Real Statistics Data Analysis Tool**: In addition, the Real Statistics Resource Pack provides a supplemental **Chi-Square Test**data analysis tool. To use this tool for Example 1 enter **Ctrl-m** and select the **Chi-square Test** option. A dialog box as in Figure 3 appears.

[](https://www.real-statistics.com/wp-content/uploads/2019/04/chi-square-independence-dialog.png)

**Figure 3 – Dialog box for Chi-square Test**

Insert the observation data into the **Input Range** (excluding the totals, but optionally including the row and column headings; i.e. range A5:D8), click on the **Excel format** radio button and press the **OK** button. Leave the **Fisher Exact Test** option unchecked (see [Fisher Exact Test](https://www.real-statistics.com/chi-square-and-f-distributions/fishers-exact-test/) for use of this option).

The data analysis tool builds an array with the expected values and performs both the Pearson’s and maximum likelihood chi-square tests. The Cramer effect size, and for 2 × 2 contingency tables the Odds Ratio effect size, as described in [Effect Size for Chi-square](https://www.real-statistics.com/chi-square-and-f-distributions/effect-size-chi-square/" \o "Effect Size for Chi-square" \t "_blank) are also calculated. The output from the data analysis tool for the data in Example 1 in shown in Figure 4.

[](https://www.real-statistics.com/wp-content/uploads/2012/12/chi-square-test-indep.png)

**Figure 4 – Chi-Square data analysis tool output for Example 1**

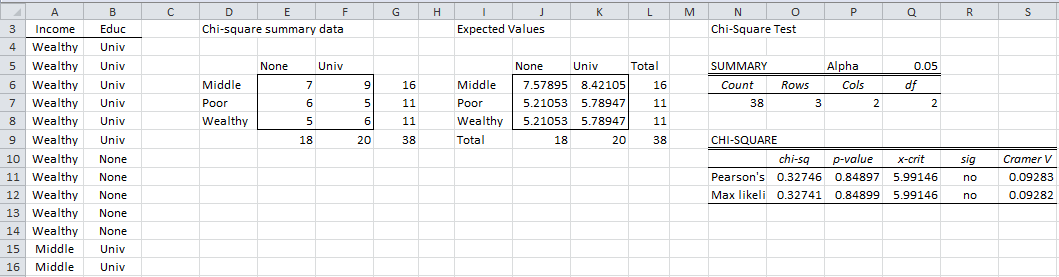
**Observation**: As described in [Goodness of Fit](https://www.real-statistics.com/chi-square-and-f-distributions/goodness-of-fit/" \o "Goodness of Fit" \t "_blank), the expected frequency for any cell in the contingency table  should generally be at least 5. With small tables (especially 2 × 2 tables), cells with expected frequencies of at least 10 would be preferable.

For large contingency tables, a small percentage of cells with expected frequency of less than 5 can be acceptable. Even for smaller contingency tables having one cell with expected frequency of less than 5 may not cause big problems, but it is probably a better choice to use [Fisher’s Exact Test](https://www.real-statistics.com/chi-square-and-f-distributions/fishers-exact-test/" \o "Fisher’s Exact Test" \t "_blank) in this case. In any event, you should avoid using the chi-square test where there is an expected frequency of less than 1 in any cell.

If the expected frequency for one or more cell is less than 5, it may be beneficial to combine one or more cells so that this condition can be met, although this must be done in such a way as to not bias the results.

**Observation**: In addition to the usual Excel input data format, the Real Statistics **Chi Square Test** data analysis tool supports another input data format called **standard format**. This format is similar to that used by SPSS and other statistical analysis programs.

**Example 3**: A survey is conducted of 38 young adults whose parents are classified either as wealthy, middle class or poor to determine whether they will graduate from university or not. The results are summarized in the table on the left side of Figure 5 (only the first 13 of 38 rows of data are shown). Based on the data collected is a person’s level of schooling independent of their parents’ wealth?

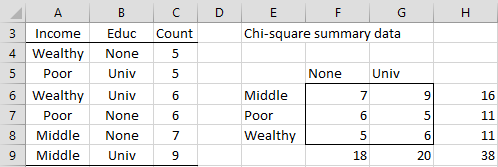
[](https://www.real-statistics.com/wp-content/uploads/2012/12/independence-test-standard-format.png)

**Figure 5 – Data and chi-square tests for Example 3**

Once again enter **Ctrl-m** and select the **Chi-square** data analysis tool. When the dialog box shown in Figure 3 appears, insert A3:B41 into the **Input Range**, click on the **Standard format** radio button and press the **OK** button.

The data analysis tool first builds a contingency table (range D5:F8 of Figure 5) and performs the same type of analysis as for Example 1 and 2. Since *sig* = no (cells R11 and R12) we cannot reject the null hypothesis that a student’s graduating from university is independent of his/her parents’ level of income.

**Observation**: Example 3 uses the two column version of the standard format. There is also a three column version, which is a frequency table version of the other standard format. This is demonstrated in Figure 6 where A4:C9 is inserted in the **Input Range**(or A3:C9 if **Column/row headings included with data** is checked). The output is identical to that shown in Figure 5.

[](https://www.real-statistics.com/wp-content/uploads/2017/12/independence-test-standard-format.png)

**Figure 6 – Standard format**

**Post-hoc Testing**

After a significant result from the chi-square test of independence, you can perform one of several follow-up tests to pinpoint the cause of the significant result. Further information about this topic can be found by clicking on the following links:

* [Post-hoc testing after chi-square independence testing](https://www.real-statistics.com/chi-square-and-f-distributions/independence-testing/independence-testing-follow-up/)
* [Real Statistic support for post-hoc testing](https://www.real-statistics.com/chi-square-and-f-distributions/independence-testing/chi-square-post-hoc-functions/)