Minitab Tutorial 9 for Chi- Squared

So, let's say you count some (1851) plain M&Ms, and here's what you find. Also, notice that the information from the M&M/Mars Corp. is that 30% of plain M&Ms are brown, 20% each yellow and red, and 10% each of blue, orange, and green.

In Excel, it is not too hard to type in the formulas for the Chi-Squared tests, but you can also do it in Minitab.

First, we need to enter our data into Minitab. It looks pretty similar to what we had entered in Excel.

Worksheet 1 ***

↓	C1-T	C2	C3	C
	color	p(plain)	obs1	
1	Brown	0.3	519	
2	Yellow	0.2	364	
3	Red	0.2	363	
4	Blue	0.1	200	
5	Orange	0.1	212	
6	Green	0.1	193	

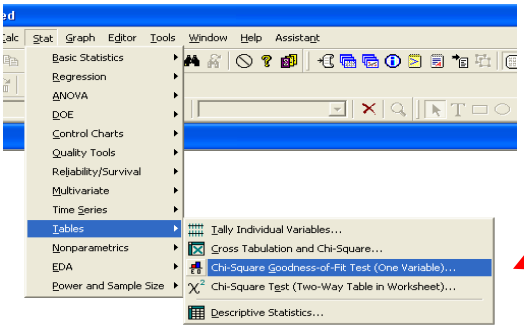
Chi-Squared Test of Goodness of Fit

So, our Null Hypothesis will be that the probabilities given to us by the corporation are correct:

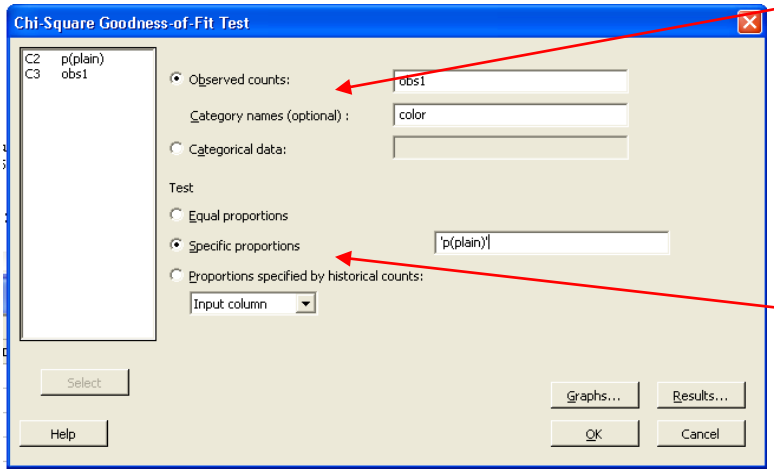
H_0: Plain M&Ms are 30% brown, 20% each yellow and red, and 10% each blue, orange, and green.

H_1: Plain M&Ms have a different distribution than that.

This is not a very specific Hypothesis. In particular, notice that if you reject the null hypothesis, you have no idea what distribution you do have (except that it isn't the one listed in H_0).



Go to **Stat: Tables: Chi-Squared Goodness-of-Fit Test.**



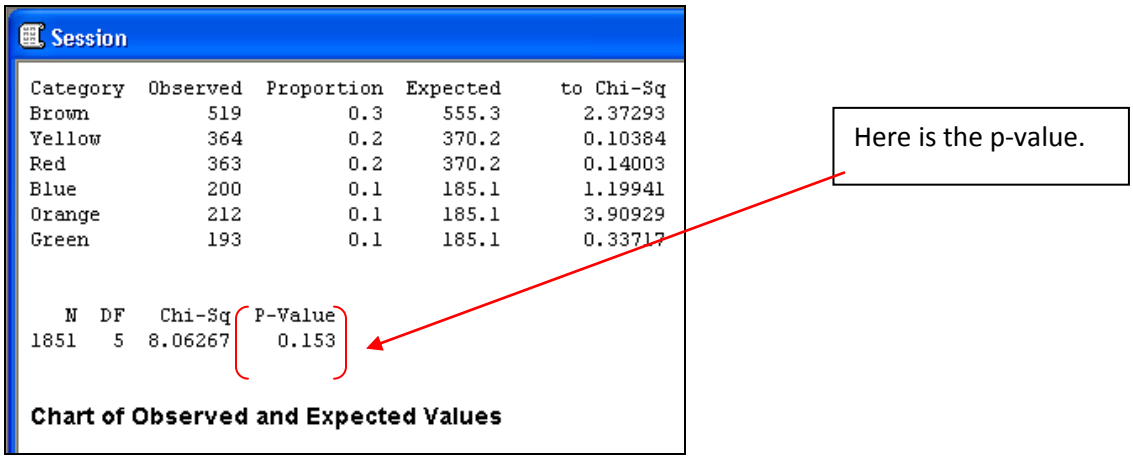
Choose observed counts, since our amounts are already totaled (if we had listed each M&M separately we would choose the second option).

Since we already know the expected values we choose the second option. But if we wanted to test if all of the proportions were equal we would choose the first option.

Minitab Tutorial 9 for Chi- Squared

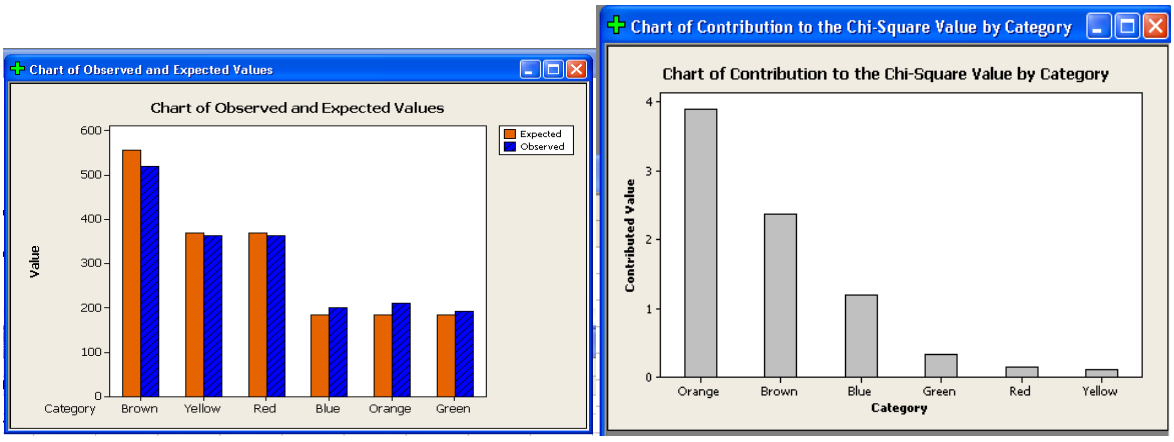
The output will show the labels furthest to the left, then the observed values, followed by the expected proportions, and then the expected number of occurrences. Following this table will be the statistics that we are most interested in.

Notice our p-value is 0.153, not significant. Therefore, we fail to reject the null hypothesis. This means that our M&Ms follow the specified distribution.



Notice that along with the output, two graphs are also generated. The first, shown here to the left, gives a comparison of the observed verse the expected values. For our data the bars are relatively close (which is to be expected since we did not reject the null hypothesis). But if we had rejected the null hypothesis we could use this chart to get an idea of what variable was least similar to the expected values.

The second chart, shown here to the right, shows the distribution of contributing values for each category, in our case the color of M&Ms. This tells us that orange was least similar to the expected value and so most of the disparity came from this variable. On the other hand, yellow was most like the expected value, and was therefore least inconsistent.



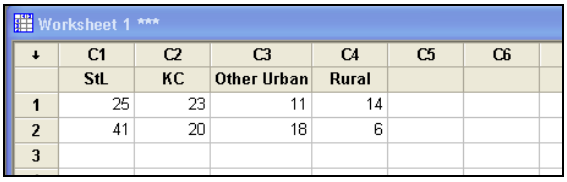
Minitab Tutorial 9 for Chi- Squared

Chi-Squared Test of Independence

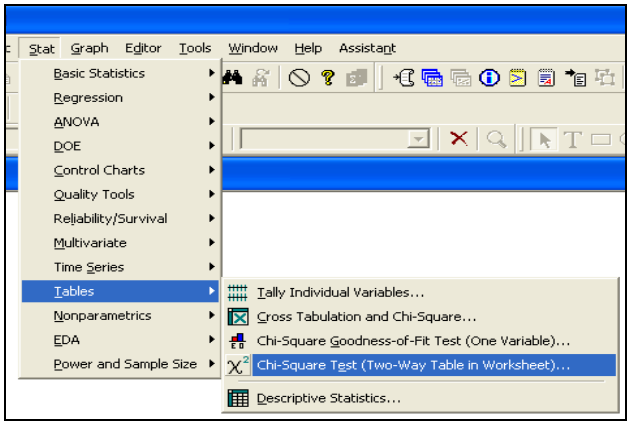
The other Chi-squared test deals with two categorical variables. The hypotheses are:
H_0: Hometown and Gender are independent.
H_1: Hometown and Gender are related.
Notice these are also not very specific hypotheses.

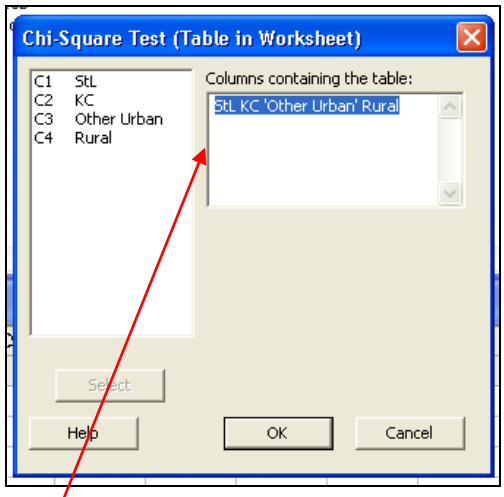
Let's say we have the following data, detailing the gender and hometown of a group of students:

	StL	KC	Other Urban	Rural	Sum
M	25	23	11	14	*73*
F	41	20	18	6	*85*
Sum	*66*	*43*	*29*	*20*	*158*



In this case we will enter data in the form of a contingency table, with each entry representing counts. It is also possible to enter data in which each row represents one observation, but would mean we would have to enter 158 rows of data for this example. Therefore, we will treat each row as separate groups that to compare. In this example row '1' represents males and row '2' represents females. The command you will want is **Stat: Tables: Chi-Squared Test (Two-Way Table in Worksheet).**

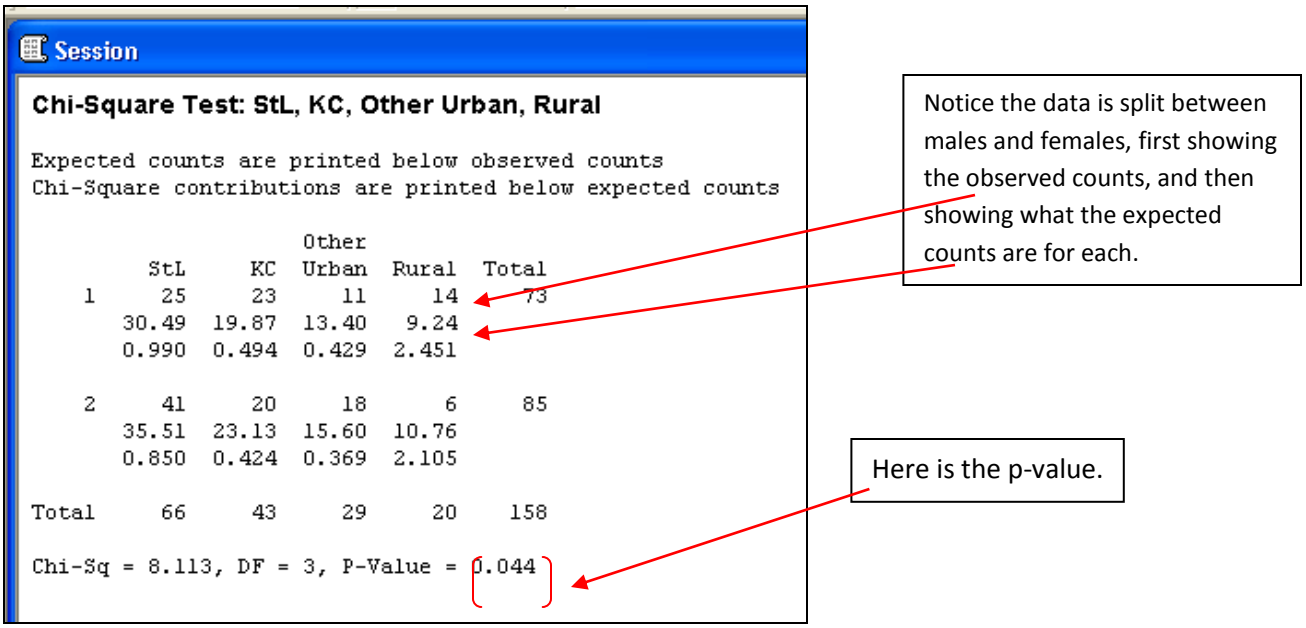




Enter all four categories here. It is also possible to only compare the distribution of more specific categories (such as comparing just the major cities).

Minitab Tutorial 9 for Chi- Squared

You get output that looks like this:



Since the p-value is less than .05 (barely), we can reject the null hypothesis, and say that gender and hometown are related.