SPSS: Chi Squared Tests

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.

	p(plain)	obs1
Brown	0.3	519
Yellow	0.2	364
Red	0.2	363
Blue	0.1	200
Orange	0.1	212
Green	0.1	193
Sums	1	1851

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

First, we need to enter our data into SPSS. Notice that I labeled the columns, and arbitrarily assigned numbers to the colors (I could've used any numbers, but going in order makes the most sense). I also typed in the value labels on the variable view screen (so, 1=brown, 2=yellow, etc. on any printouts).

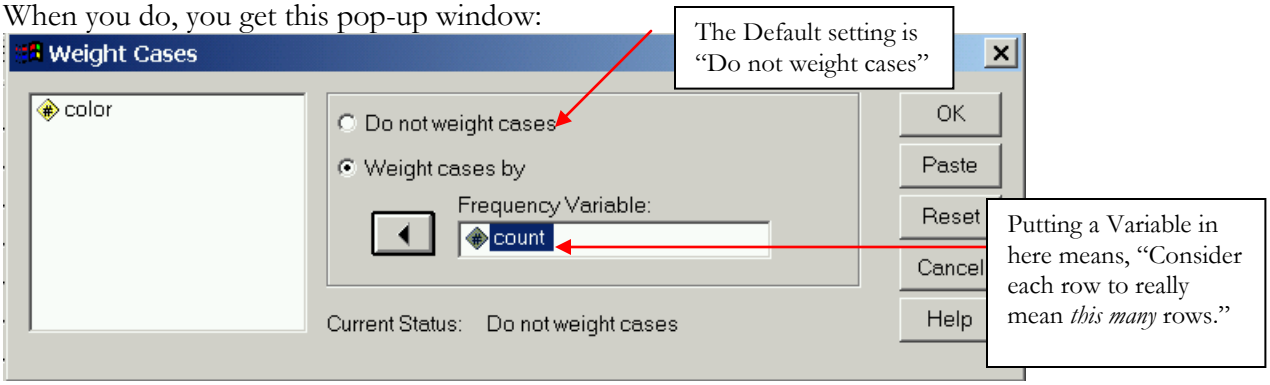
	color	count
1	1.00	519.00
2	2.00	364.00
3	3.00	363.00
4	4.00	200.00
5	5.00	212.00
6	6.00	193.00

However, SPSS typically thinks of each row as an individual. Therefore, if we ran our analysis now, it would think we had 6 M&Ms, not 1851. To run our test the "normal" way, we would need to enter 1851 rows of data! To get around this, we use a command called Weight Cases.

Weight Cases

This feature is a way of telling SPSS to pretend that a single row of data is, in fact, a whole bunch of rows. In our first row, we want to tell SPSS that we have 519 brown M&Ms. So, look for the menu item: **Data: Weight Cases**

When you do, you get this pop-up window:



Now, SPSS will pretend that the variable "color" has 519 rows with a "1," 364 with a 2, etc.

This will be in the bottom right corner of your screen, if you did it right.

You can then do the Chi-Square (or any) test like you would've before. Except now, the six rows will act like it is 1851 rows.

SPSS: Chi Squared Tests

Chi-Squared Test of Goodness of Fit

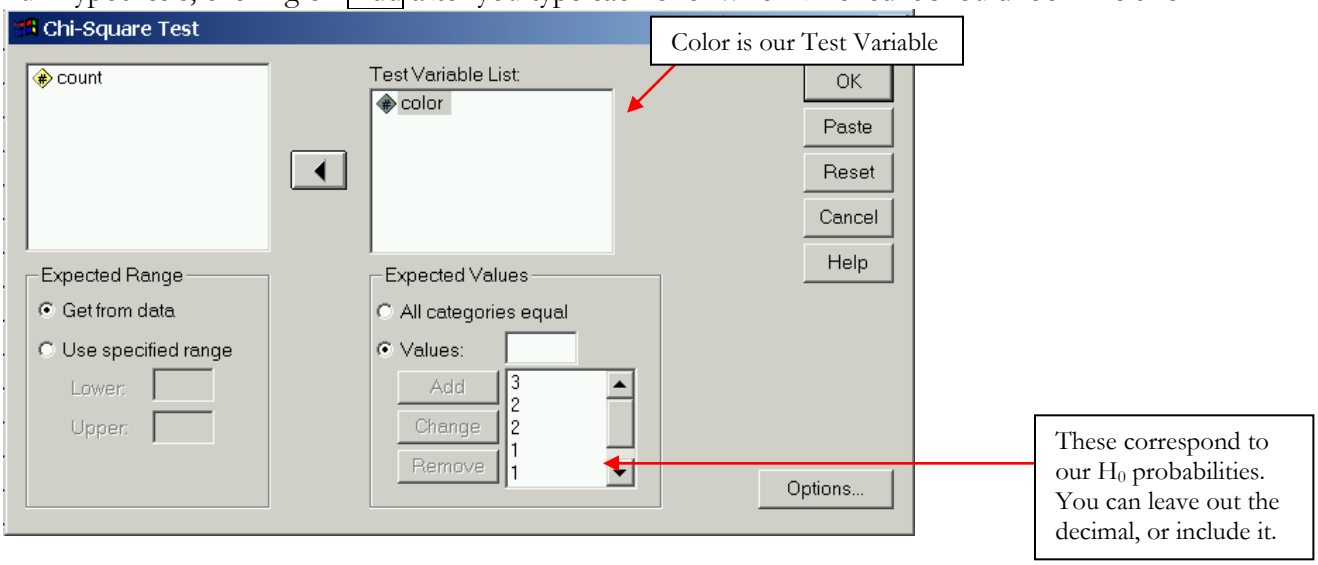
So, our Null Hypothesis will be that the probabilities given to us by the corporation are correct:

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).

Now, you click on **Analyze: Non-Parametric: Chi-Square**

Now, your test variable is color. Be sure to NOT use Count (that would be saying that you had 519 M&MS of type 519 – Oops). Then, you need to enter in your expect values. This is actually easier than Excel, because you only need the probabilities, not the computed expected values. So, click on the radio button for Expected Values: Values. Then, enter in your six relative probabilities from the null hypothesis, clicking on Add after you type each one. When finished it should look like this:



Click OK, and you get the following output:

Test Statistics

	COLOR
Chi-Square [a]	8.063
df	5
Asymp. Sig.	.153

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 185.1.

COLOR

	Observed N	Expected N	Residual
brown	519	555.3	-36.3
yellow	364	370.2	-6.2
red	363	370.2	-7.2
blue	200	185.1	14.9
orange	212	185.1	26.9
green	193	185.1	7.9
Total	1851		

Since the p-value is not less than .05, we fail to reject. Our M&Ms do follow the distribution we specified. Notice that it gives the sum, the degrees of freedom, and it also has a little footnote about the assumptions you made.

SPSS: Chi Squared Tests

Crosstabs and the 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:

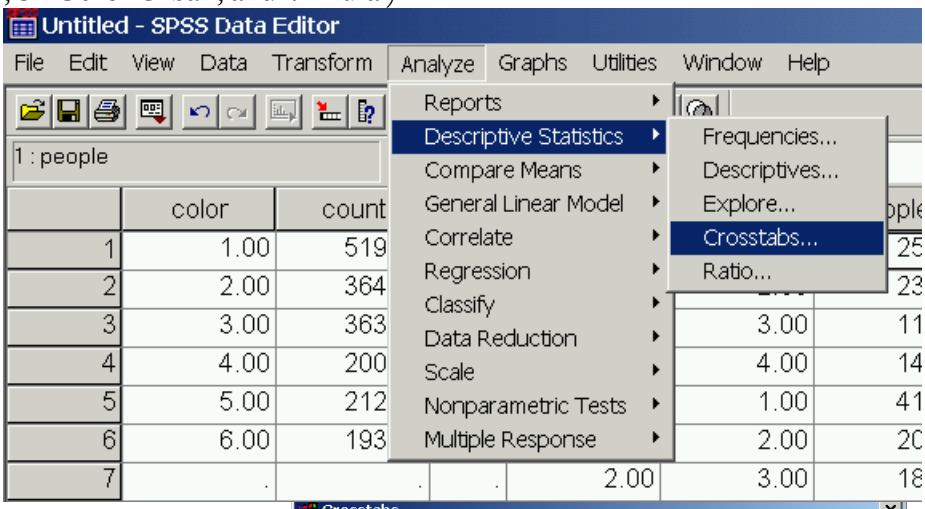
	StL	KC	Other Urban	Rural	Sum
M	25	23	11	14	*73*
F	41	20	18	6	*85*
Sum	*66*	*43*	*29*	*20*	*158*

	gender	hometown	people
1	1.00	1.00	25.00
2	1.00	2.00	23.00
3	1.00	3.00	11.00
4	1.00	4.00	14.00
5	2.00	1.00	41.00
6	2.00	2.00	20.00
7	2.00	3.00	18.00
8	2.00	4.00	6.00

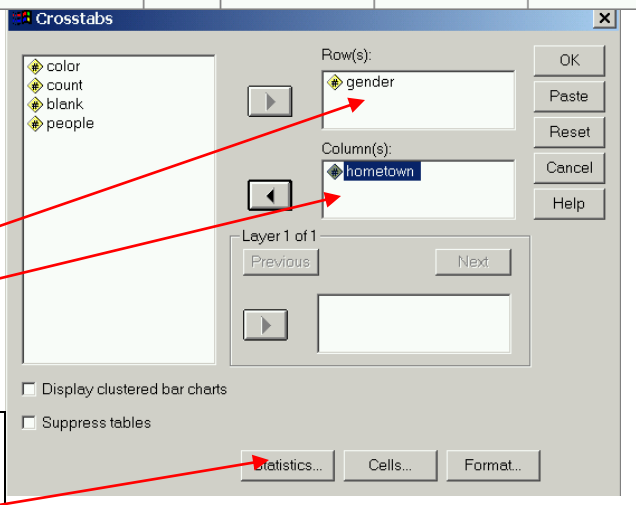
To enter it into SPSS, you have to really switch it around. It should look like this. You can see that, rather than looking like a two-way table, it gives a variable description of each cell. Again, you might want to go into the variable view, so you can label the values of each variable (For gender, make 1= male and 2 = female; for hometown, 1=St. Louis, 2 = Kansas City, 3=Other Urban, and 4=Rural).

Again, you'll want to **Weight Cases**, this time, by "people."

The Command you'll want for this kind of Chi-Square is a little bit hidden, so go to the menu for:
Analyze: Descriptive Statistics: Crosstabs



The purpose of Crosstabs is to make a two-way contingency table out of big hunks of data. You could imagine that if you hadn't weighted cases, but entered the data in the long way, you'd like to get such a two-way table. To get these, put your data in how you'd like it in the chart, like this:

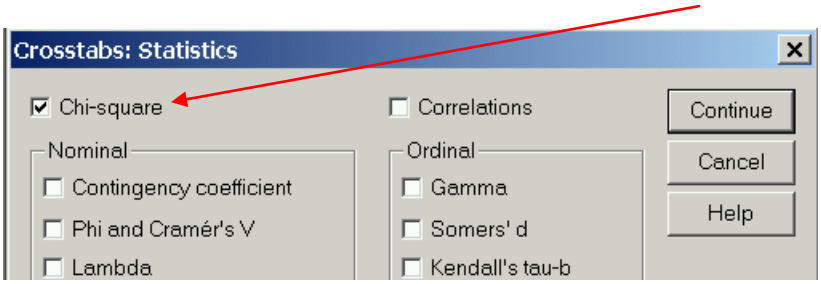


Gender and hometown are the two variables to examine. They can go either place.

Clicking on this is where you will find Chi-Square

SPSS: Chi Squared Tests

Once you click on Statistics, you can check the box for Chi-Square. It's that easy.



You get output that looks like this:

GENDER * HOMETOWN Crosstabulation

Count

		HOMETOWN				
		1.00	2.00	3.00	4.00	Total
GENDER	1.00	25	23	11	14	73
	2.00	41	20	18	6	85
Total		66	43	29	20	158

This looks like your original data chart.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8.113[a]	3	.044
Likelihood Ratio	8.213	3	.042
Linear-by-Linear Association	3.636	1	.057
N of Valid Cases	158		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 9.24.

Here is the p-value.

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