Crosstabs (with Chi-Square)

A crosstabulation displays the number of cases in each category defined by two or more grouping variables.

```
PROC FREQ;
TABLES happy * freqdum / CHISQ;
RUN;
```

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of HAPPY by FREQDUM

HAPPY(GENERAL HAPPINESS)	FREQDUM(SEXFREQ DUMMY) 0	1	Total
1	103 9.72 33.23 22.79	207 19.53 66.77 34.05	310 29.25
2	286 26.98 46.20 63.27	333 31.42 53.80 54.77	619 58.40
3	63 5.94 48.09 13.94	68 6.42 51.91 11.18	131 12.36
Total	452 42.64	608 57.36	1060 100.00

Frequency Missing = 1705

For example, we see that there are 207 cases reporting "very happy" for general happiness and "more than once per month" for frequency of sex.

Statistics for Table of HAPPY by FREQDUM

Statistic	DF	Value	Prob
Chi-Square	2	16.0387	0.0003
Likelihood Ratio Chi-Square	2	16.2971	0.0003
Mantel-Haenszel Chi-Square	1	13.1235	0.0003
Phi Coefficient		0.1230	
Contingency Coefficient		0.1221	
Cramer's V		0.1230	

Effective Sample Size = 1060
Frequency Missing = 1705

WARNING: 62% of the data are missing.

The chi-square measures test the hypothesis that the row and column variables in a crosstabulation are independent. While the chi-square measures may indicate that there is a relationship between two variables, they do not indicate the strength or direction of the relationship.

Note: In this particular dataset, recall that there is 1396 missing cases on the general happiness variable. SAS creates a “warning” when more than 50% of the cases on this test, which may create skewed results.

Chi-Square

The Chi-Square Goodness of Fit Test determines if the observed frequencies are different from what we would expect to find (we expect equal numbers in each group within a variable). Use a Chi-Square Test when you want to know if there is a significant relationship between two categorical variables. In this example, we use "frequency of sex" and "happiness."

Null Hypothesis: There are approximately equal numbers of cases in each group.
Alternate Hypothesis: There are not equal numbers of cases in each group.

```
PROC FREQ;
TABLES happy / CHISQ;
RUN;
```

The FREQ Procedure

GENERAL HAPPINESS				
HAPPY	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
1	415	30.31	415	30.31
2	784	57.27	1199	87.58
3	170	12.42	1369	100.00

Frequency Missing = 1396

Chi-Square Test for Equal Proportions	
Chi-Square	418.6866
DF	2
Pr > ChiSq	<.0001

We have a Chi-Square value of 418, which is large. Our significance level is .000. We can conclude that there are not equal numbers of cases in each happiness category.

Effective Sample Size = 1369
Frequency Missing = 1396

WARNING: 50% of the data are missing.