**Example SPSS Output for Crosstabulation**

The table below looks at the relationship between the explanatory (independent) variable age group (zragecat) and the dependent variables attitudes to the statement ‘The law should always be obeyed even if a particular law is wrong’ (zwronglaw).

Independent variable goes in columns.



% within zwronglaw is ROW PERCENTAGE. Usually we do not want to work with these, as in most circumstances we want the groups of the independent variable broken down over the categories of the dependent variable

% within zragecat is the COLUMN PERCENTAGE. These are what we will use for making up tables for presentation purposes.

Dependent variable goes in rows

Interpretation:

* 30.9% of those who agree that the law should always be obeyed are aged 18-39 (NOTE: This is not an especially useful piece of information as the percentage in each age group depends on how wide we make our age bands.)
* 36.0% of those aged 18-39 agree that the law should always be obeyed, this contrasts with 53.8% of those aged 60 and over. Clearly older people are more likely to believe that the law should always be obeyed while younger people are more likely to believe that there are some cases when it should not be.
* 1511 respondents agreed with the statement
* 1297 respondents are aged 18-39
* 467 respondents both agreed with the statement and are aged 18-39

**Understanding the chi-square test: Expected Values**

The chi-square test compares the actual table of data with a hypothetical table with no relationship between the two variables. This handout described how the hypothetical table (or expected values) are calculated.



 **Calculating expected values**

To calculate the hypothetical table we distribute the respondents according to the overall distributions for each of the variables. We can see that we have 1511/3500 respondents or 43.2% who agree with the statement. We can also see that we have 1297 respondents aged 18-39. If there is no relationship between the variables we would expect 43.2% of these 1297 to agree with the statement. This is 559.9 people.

There are 467 people aged 18-39 who agree with the statement. We would expect there to be 559.9 people in this category. Therefore, those aged 18-39 are less likely to agree with the statement than we would expect.

To calculate the chi-square test statistic, the expected value is calculated for each cell in the table. These are then compared with the real or observed value for each cell. The larger the difference between the two the larger is the chi-square statistic and the more likely it is that we will reject the hypothesis of no relationship between the two variables.

**Understanding the chi-square test: SPSS Output and assumptions**

All statistical tests have certain assumptions, the chi-square test has two assumptions:

* No cell has an expected value less than 1
* No more than 20% of cells have expected values less than 5

These assumptions are necessary because the expected values are used as a denominator in the chi-square calculation, if the values are small we would artificially inflate the value of the chi-square statistic.

SPSS checks these assumptions for you:



This tells you the minimum expected count. It must be greater than 1

This tell you the percentage of cells with expected values less than 5 – it must be lower than 20%

This information tells you about the validity of the test. **ALL** tests must be valid. It does not tell you anything at all about the data itself.

**Understanding the chi-square test: SPSS Output and interpretation**

Once you have checked the validity of the test you can go on to interpret it (if it is not valid you must look at the data for ways to recode.)



You are interested in the Pearson Chi-square test.

For a relationship to exist between the two variables this value must be LESS THAN 0.05 (for a 95% significance level).

In this example the significance is very small (much less than 0.05). We are therefore very confident that a relationship exists between age and attitudes to obeying the law. However, we must return to the table of data to look at the direction and strength of the relationship and to provide an interpretation of it.

Paula Surridge, School of Sociology, Politics and International Studies, University of Bristol