EXERCISE 19

The analysis of nominal data

Before you start

Before proceeding with this practical, we strongly recommend you to read Section 11.5 (measures of association strength for nominal data) in Chapter 11.

THE CHI-SQUARE TEST OF GOODNESS-OF-FIT

Some nominal data on one qualitative variable

Suppose that a researcher, interested in children's preferences, expects a spatial response bias towards the right-hand side. Thirty children enter a room containing three identically-marked doors: one to the right; another to the left; and a third straight ahead. They are told they can go through any of the three doors. Their choices are shown in Table 1.

Table 1. The choices of one of three exit doors by thirty children

	Door	
Left	Centre	Right
5	8	17

It looks as if there is indeed a preference for the rightmost door, at least among the children sampled. Had the children been choosing at random, we should have expected about 10 in each category: that is, the theoretical, or expected distribution (E), of the tallies is **uniform**. The observed frequencies (O), on the other hand, have a distribution which is far from uniform.

Pearson's chi-square test can be used to test the goodness-of-fit of the expected to the observed distribution. Its rationale is lucidly discussed in any good statistics textbook (e.g. Howell, 2007). Here, we shall merely describe the SPSS procedure.

Preparing the data set

In **Variable View**, name the grouping variable *Position* for the three positional categories and a second variable *Count* for the numbers of children in the different categories. To the three categories, assign the values *1*, *2*, and *3* and in the **Values** column, enter the respective labels *Left*, *Centre,* and *Right*. Check that the values in the **Decimals** column are *0*. Click the **Data View** tab and enter the data.

Weight cases

To ensure that SPSS treats the entries in *Count* as frequencies rather than scores, follow the procedure described in Section 11.5.3.

Run the Chi-square test

To obtain the correct dialog box, select
Analyze➔Nonparametric Tests➔Chi-Square…
to open the **Chi-Square Test** dialog box. Click *Position* (not on *Count*) and then ► to transfer *Position* to the **Test Variable List:** box. Click **OK** to run the procedure.

- **Write down the value of the chi-square statistic and its p-value. Is chi-square significant?**
- **Write down the implications for the experimenter's research hypothesis. When considering the implications, be clear about the precise null hypothesis being tested. Is the experimental hypothesis the simple negation of the null hypothesis? Can you suggest any further tests that might be useful here?**

Running the goodness-of-fit test on a set of raw data

When the researcher carried out the experiment, the door that each child chose was noted at the time. In terms of the code numbers, their choices might have been recorded as:

1, 1, 3, 2, 1, 1, 3, 3, 3 , ..., and so on.

If the user defines the variable *Position*, and enters the 30 (coded) choices that the children made, the chi-square test is then run directly: there is no need to use the **Weight Cases** procedure since there is no variable for count or frequency.

THE CHI-SQUARE TEST OF ASSOCIATION BETWEEN TWO QUALITATIVE VARIABLES

An experiment on children's choices

Suppose that a researcher, having watched a number of children enter a room and recorded each child's choice between two objects, wants to know whether there is a tendency for boys and girls to choose different objects. This question concerns two variables: *Sex* and *Choice*. In statistical terms, the researcher is asking whether they are associated: do more girls than boys choose one of the objects and more boys than girls choose the other object? Suppose that the children's choices are as in Table 2.

Table 2. Choices by 50 children of one of two objects

Object	Boys	Girls
A	20	5
B	6	19

Procedure for the chi-square test of association between two variables

Prepare a new data set from Table 2. In **Variable View**, create the variables *Object* and *Sex*, assigning code numbers and explanatory labels in the usual way. Create a third variable *Count*. The use of the **Crosstabs** procedure is fully described in Section 11.5.3. Click **Statistics…**, select **Chi-square** within the **Crosstabs: Statistics** dialog box and click **Continue**. We recommend the inclusion of expected frequencies so that you can check for the presence of cells with unacceptably low expected frequencies. Click **Cells…**, select **Expected** within the **Crosstabs: Cell Display** dialog box and click **Continue**.

Output for the chi-square test of association

The output is discussed in Section 11.5.3. Three tables are presented: the first is a **Case Processing Summary** table showing how many valid cases have been processed; the second is a **Crosstabulation** table with the observed and expected frequencies in each cell, along with row and column totals; and the third is a table (headed **Chi-square Tests**) listing various statistics, together with their associated significance levels.

- **Write down the value of the Pearson chi-square and its associated tail probability (p-value). Is it significant?**
- **In terms of the experimental hypothesis, what has this test shown?**

MEASURES OF ASSOCIATION STRENGTH FOR NOMINAL DATA

So far we have considered the use of the **chi-square statistic** to test for the presence of an association between two qualitative variables. Recall that, provided that the data are suitable, the **Pearson correlation** measures the strength of a linear association between two interval variables. In that case, therefore, the same statistic serves both as a test for the presence of an association and as a measure of associative strength. It might be thought that, with nominal data, the chi-square statistic would serve the same dual function. The chi-square statistic, however, cannot serve as a satisfactory measure of associative strength, because its value depends partly upon the total frequency.

To illustrate the calculation of measures of association for two-way contingency tables, we shall use again the data of choice of objects by children. Run the **Crosstabs** procedure again but this time deselect **Chi-square** and ensure that **Phi and Cramér's V** are selected within the **Nominal** box of the **Crosstabs: Statistics** dialog box.

The output consists of three tables: the first is a **Case Processing Summary** table, the second is a **Crosstabulation** table, and the third is a table called **Symmetric Measures** listing the values of **Phi** and **Cramér's V** together with their associated significance levels.

- **Write down the value of Phi for the strength of the association between the qualitative variables of Gender and Object. Has a strong association been demonstrated?**

Finishing the session

Close down SPSS and any other windows before logging out of the computer.