Article 8. An introduction to hypothesis testing. Non-parametric comparison of two groups—1

P Driscoll, F Lecky

Objectives
- Dealing with unpaired non-parametric data
- Dealing with small samples of nominal data

In covering these objectives the following terms will be introduced:
- $\chi^2$ test
- Fisher’s exact test
- Yates’s correction

In the previous article the basic principles behind comparing two groups was discussed. The $t$ test was also shown to play an important part in this process when dealing with parametric data. As it is a versatile test, it can be used to compare independent groups and paired groups as well as give an estimate of a population mean when a sample is known. In contrast, when dealing with non-parametric data, several tests have to be used. Fortunately though the same principles apply.

As these non-parametric tests do not assume the distribution of the data is normal, they can be used on a much wider spectrum of results. The cost of this is they lack power and are mainly tests of significance. Consequently they will tell if a difference exists but not how big it is (table 1).

Table 1: Contrast of parametric and non-parametric tests in two group comparisons

<table>
<thead>
<tr>
<th>Parametric</th>
<th>Non-parametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumptions</td>
<td>Normal distribution</td>
</tr>
<tr>
<td>Data</td>
<td>Uniform variance</td>
</tr>
<tr>
<td>Significance</td>
<td>Paired and unpaired data</td>
</tr>
<tr>
<td>Small differences</td>
<td>Detected</td>
</tr>
<tr>
<td>Size of the difference</td>
<td>Can be calculated</td>
</tr>
</tbody>
</table>

$\chi^2$ test

OVERVIEW

In the medical literature there are many examples of studies that have used nominal data. As described in article 1, this is where the data are divided into categories that do not have any inherent order. It is possible to see how the proportions of observations in different categories from a sample compare (“fit”) with that found in a population. Consequently this is known as “A goodness of fit” test. It is also possible to compare the distribution between samples as well as to test the association between variables. In all cases the systematic approach described already in this series is used (box 1). However, the $z$ and $t$ test is replaced with the $\chi^2$ test.

Goodness of fit

To demonstrate this consider the following example. Dr Egbert Everard, SpR in Emergency Department, Hope Hospital, Salford M6 8HD, UK

Accident and Emergency Department, Hope Hospital, Salford M6 8HD, UK

Correspondence to: Mr Driscoll, Consultant in Accident and Emergency Medicine (pdriscoll@hope.srht.nwest.nhs.uk)

Table 2: Table of non-parametric tests for two group comparisons

<table>
<thead>
<tr>
<th>Nominal data</th>
<th>Ordinal data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaired:</td>
<td>Ranked data:</td>
</tr>
<tr>
<td>$\chi^2$ test</td>
<td>Paired—Wilcoxon</td>
</tr>
<tr>
<td>Fisher exact test</td>
<td>Unpaired—Mann-Whitney</td>
</tr>
<tr>
<td>McNemar’s test</td>
<td></td>
</tr>
</tbody>
</table>

Key point 1
The $\chi^2$ test is a non-parametric test of the null hypothesis. It can be used on unpaired nominal data to determine:
- A goodness of fit between a sample and a population
- Comparison between two groups
- Association between two variables
Medicine, is asked to determine if the type of complaints received by the Emergency Department of Deathstar General is similar to the national picture (table 3). To answer this he decides to use a $\chi^2$ "goodness of fit" test.

1 STATE THE NULL HYPOTHESIS AND THE ALTERNATIVE HYPOTHESIS OF THE STUDY

Having considered the problem, Egbert writes down the null hypothesis as:

“There is no significant difference between the type of complaints made by patients attending the Emergency Department of Deathstar General and those attending emergency departments nationally”.

The alternative hypothesis is the logical opposite of this—that is:

“There is a significant difference between the complaints made by patients attending the Emergency Department of Deathstar General and those attending emergency departments nationally”.

2 SELECT THE LEVEL OF SIGNIFICANCE

If the null hypothesis is correct, the distribution of complaints found in the Emergency Department of Deathstar General should be the same as that found nationally. Therefore the difference between them would be zero. However, there is bound to be some small variation simply due to chance. The question is how big a difference are we going to allow before we reject the idea that the two are all part of the same population?

We can answer this because the difference between a sample and population of nominal data has a $\chi^2$ distribution (fig 1). Egbert picks a significance level of 0.05 because, by convention, the outer 5% of the area under the curve is considered to be sufficiently away from the population mean as to represent values that cannot be attributed to chance variation. He now needs to determine the critical value for the $\chi^2$ statistic that demarcates this point ($\chi^2_{crit}$).

3 ESTABLISH THE CRITICAL VALUES

As with the $t$ distribution, the $\chi^2$ distribution changes shape depending upon the size of the sample. For mathematical reasons this is measured as the degrees of freedom rather than the actual size. When comparing a sample with a population, the degrees of freedom is one less than the number of categories in the sample. Therefore in this case, Egbert works out the degrees of freedom to be $4-1 = 3$.

Using the table of $\chi^2$ statistics, a value of 7.82 or greater for $\chi^2_{crit}$ demarcates a right tail that has, at most, 5% of the area under the curve is considered to be sufficiently away from the population mean as to represent values that cannot be attributed to chance variation. He now needs to determine the critical value for the $\chi^2$ statistic that demarcates this point ($\chi^2_{crit}$).

4 SELECT THE SAMPLE

Using records for the previous year, Egbert finds out that the type of complaints made (table 4).

5 CHOOSE AND CALCULATE THE TEST STATISTIC

If the null hypothesis applies there is no difference between the national picture and Deathstar General. Consequently the expected number of complaints in each category (E) should be the same as those actually observed.

Table 3 Complaints made by patients attending UK emergency departments

<table>
<thead>
<tr>
<th></th>
<th>Misdiagnosis</th>
<th>Waiting times</th>
<th>Communication</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>National proportions</td>
<td>15%</td>
<td>60%</td>
<td>10%</td>
<td>15%</td>
</tr>
</tbody>
</table>

Figure 1 Graph of $\chi^2$ values (horizontal axis) against a measure of probability (vertical axis). The curves are theoretical $\chi^2$ distributions for 5, 10 and 15 degrees of freedom (df). Each are skewed to the right, particularly when the df is small. As the df increases the curve becomes more normal in distribution. The total area under each curve is equal to 1 (that is, full probability).

Figure 2 Part of the $\chi^2$ table showing probability (p) for various degrees of freedom (df).
The value for E in each category can be determined from the percentages provided for the population’s categories. For example, in the national picture 15% of complaints are attributable to a misdiagnosis being made. Therefore if the null hypothesis is applied, 15% of patients attending the Emergency Department of Deathstar General should also make this type of complaint. The expected number would therefore be:

\[15 \times 60]/100 = 9\]

Carrying out the same process for each of the other categories, Egbert draws up a table of expected numbers of complaints if the null hypothesis applied (table 5).

Table 5  Expected complaints from patients attending Deathstar’s Emergency Departments

<table>
<thead>
<tr>
<th></th>
<th>Misdiagnosis</th>
<th>Waiting times</th>
<th>Communication</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected number</td>
<td>9</td>
<td>36</td>
<td>6</td>
<td>9</td>
<td>60</td>
</tr>
</tbody>
</table>

To be able to use the \(\chi^2\) test, the data need to have the following properties:

- Only frequency data are used in the categories.
- Events are independent within a sample group. This means that paired data cannot be used and individual subjects only appear once in the table.
- No expected value in the table is less than 1 and 80% have values over 5. The reason for this is small expected values can have a disproportionate effect on the overall \(\chi^2\) test statistic, irrespective of the values in other cells.
- There is a logical basis for the group classification.
- As these assumptions are valid in this study, Egbert can proceed with determining \(\chi^2_{calc}\).

An indicator of the validity of the null hypothesis would be the total differences between the observed (O) and expected (E) values in each category. However, if we simply added up each of these values, the overall result would be zero. This is because half the differences will be positive and the other half negative. To overcome this they are squared first. The \(\chi^2\) statistic for each category is then derived by dividing this figure by E. Therefore \((O−E)^2/E\) represents the test statistic for each category. The overall test statistic \((\chi^2_{calc})\) is the sum of these separate category test statistics. This can be represented by the equation:

\[\chi^2_{calc} = \sum [(O−E)^2/E]\]

Egbert therefore determines the \(\chi^2\) value for each category and derives the overall \(\chi^2\) statistic (table 6).

\[\chi^2_{calc} = \sum [(O−E)^2/E] = 43.43\]

When carrying out this calculation by hand there are two checks to ensure a simple mistake has not been made:

- The sum of the expected frequencies for all the categories should equal the total observed frequencies for all the categories (that is, 60 in this case).
- In each column, the sum of all observed frequencies minus the sum of all expected frequencies must equal zero.

6 COMPARE THE CALCULATED TEST STATISTIC WITH THE CRITICAL VALUE

The calculated value of 43.43 lies beyond the critical value of 7.82. It therefore falls outside the area of accepting the null hypothesis.

7 EXPRESS THE CHANCES THAT THE NULL HYPOTHESIS IS IN KEEPING WITH THE DATA

As described previously, the p value is the probability of getting a difference equal or greater than that found in the study (that is, 43.43) if the null hypothesis was correct.2

In contrast with the \(t\) and \(z\) statistics, only the right side of the \(\chi^2\) distribution is used. This is because only large values of \(\chi^2\) can reject the null hypothesis.

Key point 3

In other words \(\chi^2\) tests are always one sided.3

From the \(\chi^2\) tables, it can be seen that the size of the tail from 43.43 to the right tip is less than 0.001. In other words, there is less than a 0.1% chance that a difference with a magnitude of 43.43, or larger, could have resulted if the null hypothesis was valid. Therefore the p value is <0.001.

When presenting the results it is important that they are interpreted in the light of the data. Consequently Egbert concludes that there is a significant difference between the types of complaints made by Deathstar’s Emergency Department attendees and those found nationally, \(\chi^2 = 43.43\), df 3, p <0.001. The difference is most marked in the complaints made regarding misdiagnoses (greater than expected) and waiting times (less than expected).
An introduction to hypothesis testing

1. STATE THE NULL HYPOTHESIS AND THE ALTERNATIVE HYPOTHESIS OF THE STUDY

Having considered the problem, they write down the null hypothesis as:

“There is no significant difference between the type of complaints made by patients attending the Emergency Departments of Deathstar General and St Heartsinc”.

The alternative hypothesis is the logical opposite of this—that is:

“There is a significant difference between the complaints made by patients attending the Emergency Departments of Deathstar General and St Heartsinc”.

2. SELECT THE LEVEL OF SIGNIFICANCE

Following convention they pick a significance level of 0.05. They now need to determine the critical value for the $\chi^2$ statistic that demarcates this point ($\chi^2_{\text{crit}}$). As described in the previous example, if the null hypothesis is correct, the distribution of complaints found in the two departments should be the same. However, there is bound to be some difference between them due to random variation but this has a $\chi^2$ distribution. This distribution can therefore be used to find the value of $\chi^2_{\text{crit}}$.

3. ESTABLISH THE CRITICAL VALUES

With this type of comparison, the degrees of freedom is $(r-1)(c-1)$ where $r$ and $c$ are the number of rows and columns in the contingency table (table 7).

Table 7 Contingency table for the study comparing the complaints made by patients attending Deathstar’s and St Heartsinc’s Emergency Departments

<table>
<thead>
<tr>
<th></th>
<th>Deathstar General</th>
<th>Deathstar General</th>
<th>St Heartsinc</th>
<th>St Heartsinc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misdiagnosis</td>
<td>25</td>
<td>20</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Waiting</td>
<td>15</td>
<td>5</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Communication</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>40</td>
<td>40</td>
<td>30</td>
</tr>
</tbody>
</table>

In this study, they calculate the degrees of freedom to be $(2-1)(4-1) = 3$. The $\chi^2$ statistics tables indicate that this gives a $\chi^2_{\text{crit}}$ value of 7.82 or greater at the 5% level. In other words, 7.82 separates the left 95% area of acceptance of the null hypothesis from the right 5% area of rejection.

4. SELECT THE SAMPLE

Endora now checks the records from St Heartsinc to find out the type of complaints made by the emergency department patients (table 8).

Table 8 Observed complaints made by patients attending Deathstar’s and St Heartsinc’s Emergency Departments

<table>
<thead>
<tr>
<th></th>
<th>Deathstar General</th>
<th>Deathstar General</th>
<th>St Heartsinc</th>
<th>St Heartsinc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misdiagnosis</td>
<td>25</td>
<td>20</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Waiting</td>
<td>15</td>
<td>5</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Communication</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>40</td>
<td>40</td>
<td>30</td>
</tr>
</tbody>
</table>

5. CHOOSE AND CALCULATE THE TEST STATISTIC

If the null hypothesis applies there should not be a difference between the two departments. Consequently the expected number of complaints in each category (E) should be the same as those actually observed (O). The best estimate we have for the expected values comes from combining the values for two groups in each category so that an overall proportion can be calculated.

For example, the overall proportion of complaints attributable to a misdiagnosis is 45/100 (45%). Therefore if the null hypothesis applied, 45% of patients attending the Emergency Department of Deathstar General would also make this type of complaint. The expected number would therefore be:

$$\frac{45 \times 60}{100} = 27$$

Similarly the expected number of misdiagnosis complaints in St Heartsinc is:

$$\frac{45 \times 40}{100} = 18$$

Carrying out the same process for each of the other categories, they draw up a table of expected numbers of complaints if the null hypothesis applied (table 9). Again these data fulfill the assumptions made in using the $\chi^2$ test.

As the overall $\chi^2$ statistic is:

$$\chi^2_{\text{calc}} = \Sigma \left(\frac{O-E}{E}\right)^2$$

### Box 2 Summary for calculating the $p$ value using the $\chi^2$ test

- Record the observed category frequencies from the data (O)
- Calculate the expected values (E) for all category frequencies if the null hypothesis was true by $E = \text{row total} \times \text{column total/ grand total}$
- For each cell calculate $(O-E)^2/E$
- Add these values to obtain the test statistic $\chi^2$ where $\chi^2 = \Sigma (O-E)^2/E$
- Using tables of the $\chi^2$ distribution, determine the $p$ value for the null hypothesis using the test statistic value and appropriate degree of freedom.
Table 9  Expected complaints from patients attending Deathstar’s and St Heartsinc’s Emergency Departments

<table>
<thead>
<tr>
<th></th>
<th>Misdiagnosis</th>
<th>Waiting times</th>
<th>Communication</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deathstar</td>
<td>27</td>
<td>12</td>
<td>12</td>
<td>9</td>
<td>60</td>
</tr>
<tr>
<td>St Heartsinc</td>
<td>18</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>40</td>
</tr>
</tbody>
</table>

\[ \chi^2_{\text{calc}} = \sum \left( \frac{(O-E)^2}{E} \right) = \frac{(25-27)^2}{27} + \frac{(15-12)^2}{12} + \frac{(10-12)^2}{12} + \frac{(10-9)^2}{9} + \frac{(20-18)^2}{18} + \frac{(5-8)^2}{8} + \frac{(10-8)^2}{8} + \frac{(5-6)^2}{6} \]
\[ \chi^2_{\text{calc}} = 3.84 \]

6 COMPARE THE CALCULATED TEST STATISTIC WITH THE CRITICAL VALUE

This calculated value is less than the critical value of 7.82. It therefore falls within the area of accepting the null hypothesis.

7 EXPRESS THE CHANCES THAT THE NULL HYPOTHESIS IS IN KEEPING WITH THE DATA

From the \( \chi^2 \) tables, it can be seen that the size of the tail from 3.84 to the right tip is greater than 0.05. Consequently Endora and Egbert conclude that there is no significant difference between the types of complaints made by patients attending the two emergency departments, \( \chi^2 = 3.84 \), df 4, \( p > 0.05 \).

Association between variables

In the previous example we have seen how the \( \chi^2 \) test can be used to determine if the observed numbers in each category of the table differ from those expected if the null hypothesis was valid. The same process can be used to identify an association between the column and row variables. This is particularly useful when dealing with nominal data.

Egbert’s next study demonstrates this. Having seen a number of life threatening complications resulting from central line insertion, Egbert is a keen supporter of the femoral approach. To try and convince his colleges at Deathstar General he carries out a retrospective study to see if there is any association between type of approach and life threatening complications.

1 STATE THE NULL HYPOTHESIS AND THE ALTERNATIVE HYPOTHESIS OF THE STUDY

The null hypothesis for this study is:

“The incidence of life threatening complications following central line insertion is independent of the approach”.

Again the alternative hypothesis is the logical opposite of this—that is:

“The incidence of life threatening complications following central line insertion is dependent on the approach”.

2 SELECT THE LEVEL OF SIGNIFICANCE

A significance level of 0.05 is chosen. If the null hypothesis was valid, the difference between the approaches would have a \( \chi^2 \) distribution. This distribution can therefore be used to find the value of \( \chi^2_{\text{crit}} \) that corresponds to a significance level of 0.05.

3 ESTABLISH THE CRITICAL VALUES

With this type of comparison, the degrees of freedom are \( (r-1) \times (c-1) \) where \( r \) and \( c \) are, respectively, the number of rows and columns in the contingency table (table 10).

Table 10  Contingency table for the study comparing central line approach and life threatening complications

<table>
<thead>
<tr>
<th></th>
<th>Subclavian</th>
<th>Internal</th>
<th>Jugular</th>
<th>Femoral</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life threatening complications present</td>
<td>15</td>
<td>10</td>
<td>5</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>No life threatening complications</td>
<td>55</td>
<td>30</td>
<td>35</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>Column total</td>
<td>70</td>
<td>40</td>
<td>40</td>
<td>150</td>
<td></td>
</tr>
</tbody>
</table>

4 SELECT THE SAMPLE

From the emergency department records, Egbert determines the incidence of life threatening complications resulting from central line insertion in the past year (table 11).

Table 11  Contingency table for the study comparing central line approach and life threatening complications

<table>
<thead>
<tr>
<th></th>
<th>Subclavian</th>
<th>Internal</th>
<th>Jugular</th>
<th>Femoral</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life threatening complications present</td>
<td>[70 x 30]/150 = 14</td>
<td>[40 x 30]/150 = 8</td>
<td>[40 x 30]/150 = 8</td>
<td>[40 x 30]/150 = 8</td>
<td>30</td>
</tr>
<tr>
<td>No life threatening complications</td>
<td>[70 x 120]/150 = 56</td>
<td>[40 x 120]/150 = 32</td>
<td>[40 x 120]/150 = 32</td>
<td>[40 x 120]/150 = 32</td>
<td>120</td>
</tr>
<tr>
<td>Column total</td>
<td>70</td>
<td>40</td>
<td>40</td>
<td>150</td>
<td></td>
</tr>
</tbody>
</table>

5 CHOOSE AND CALCULATE THE TEST STATISTIC

The overall test statistic is calculated in the same way as described in the previous examples. Egbert therefore calculates the expected values for each category if the null hypothesis applied (table 12). Again these data fulfill the assumptions made in using the \( \chi^2 \) test.

As the overall \( \chi^2 \) statistic is therefore:

\[ \chi^2_{\text{calc}} = \sum \left( \frac{(O-E)^2}{E} \right) = \frac{(1)^2}{14} + \frac{(2)^2}{8} + \frac{(-3)^2}{8} + \frac{(-1)^2}{56} + \frac{(-2)^2}{32} + \frac{(3)^2}{32} \]
\[ \chi^2_{\text{calc}} = 2.12 \text{ (rounded up)} \]

6 COMPARE THE CALCULATED TEST STATISTIC WITH THE CRITICAL VALUE

This calculated value is less than the critical value of 5.99. It therefore falls within the area of accepting the null hypothesis.
Table 14 Contrast of parametric and non-parametric tests in two group comparisons

<table>
<thead>
<tr>
<th>Assumptions</th>
<th>Parametric</th>
<th>Non-parametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal distribution</td>
<td>Normal distribution</td>
<td>Uniform variance</td>
</tr>
<tr>
<td>Data</td>
<td>Paired and unpaired data</td>
<td>Data</td>
</tr>
<tr>
<td>Significance</td>
<td>Assessed</td>
<td>Assessed</td>
</tr>
<tr>
<td>Small differences</td>
<td>Detected</td>
<td>Detected</td>
</tr>
<tr>
<td>Size of the difference</td>
<td>Can be calculated</td>
<td>Data</td>
</tr>
</tbody>
</table>

Dealing with small samples
The $\chi^2$ test statistic complies with the $\chi^2$ distribution provided the expected values are large enough. When dealing with small samples we can no longer make this assumption. A way of dealing with this problem is to merge categories so that the expected number in each is greater than 5. Obviously the type categories combined needs to be logical.

Another way of tackling this problem is to use Fisher’s exact test.

Fisher’s exact test
The study by Rogers et al study demonstrates the use of this test. They investigated neurogenic pulmonary oedema in patients with low and high intracranial pressure (ICP).

<table>
<thead>
<tr>
<th>Verences Detected</th>
<th>Verence Can be calculated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal variance</td>
<td>Uniform variance</td>
</tr>
<tr>
<td>Non-normal variance</td>
<td>Paired and unpaired data</td>
</tr>
</tbody>
</table>

The data were not normally distributed and there are frequencies less than 5 in two of the cells. Analysis of the data was therefore carried out using Fisher’s exact test. This produced a $p$ value $<0.001$. Consequently there is less than 1 in 1000 chance that a difference this big, or greater, between these groups is attributable to chance. The null hypothesis was therefore rejected and it was concluded that patients with high ICP values are more likely to have abnormal PaO2/FiO2 ratios compared with patients low ICP.

For those who are interested, Bland describes the method of calculating this test statistic. This is rarely necessary because it is typically worked out using a computer program. More commonly though we will come across this test when reading published articles. In these cases it is important to bear in mind the following points:

- It determines the probability that the null hypothesis is correct—that is, both groups have the same proportion of outcome/conditions.
- It is used when the number in one or more of the categories of the contingency table are less than 5.

Table 13 Relation between ICP and PaO2/FiO2 ratios

<table>
<thead>
<tr>
<th>High ICP</th>
<th>Normal PaO2/FiO2</th>
<th>Abnormal PaO2/FiO2</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Low ICP</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

The distribution of $\chi^2$ is continuous, yet nominal data are not. This can give rise to the false rejection of the null hypothesis in 2x2 tables unless adjustments are made. The correction entails the following alteration:

$\left[\frac{\text{Observed frequency}}{\text{Expected frequency}} - 0.5\right]^2 / \text{expected frequency}$

This has the effect of reducing the $\chi^2$ value and, consequently, decreasing the $p$ value.

Though there is no precise rule when the Yates’s correction should be applied, Altman recommends it is always used when dealing with 2x2 tables.  In that way it has its biggest effect where the risk of bias is highest.

Key points
- In 2x2 tables with small frequencies, Yates's correction can have marked effect on the $p$ value.
- This correction is not needed for larger tables or when the $\chi^2$ does not reach statistical significance.

Summary
There are many types of non-parametric test and several are frequently used in the medical journals. In trying to choose the correct test it is important to bear in mind the type of data you are dealing with and the nature of the question.

The $\chi^2$ test is a versatile and commonly used method for comparing distribution and looking for associations between groups of unpaired nominal data. It does however have its limitations, particularly when dealing with small samples. In these cases Yates’s correction may need to be included, or the Fisher’s exact test carried out instead, depending upon the data.

Quiz
2. Harrison et al studied the association between head injuries and five types of facial injury sustained by cyclists. How many degrees of freedom (df) would there be in calculating the $\chi^2$ statistic?
3. Egbert takes time off to coordinate a defibrillation course for personnel from St Heartsinc and Deathstar General. Having dispatched several instructors to hospital...
with third degree burns, he concludes that several of the doctors appear to be dangerous (table 15). Determine if there is a difference between the hospitals using a 5% level for the null hypothesis.

4 De Vos et al studied the impact of gender on do not attempt resuscitation (DNAR) orders in hospital (table 16).\(^5\) Determine if there is there an association between these variables using a 5% level for the null hypothesis.

5 One for you to try on your own. Stancin et al conducted a study on the acute psychosocial impact of paediatric orthopaedic trauma victims.\(^6\) They wanted to compare patients with (group 1) and without (group 2) an accompanying head injury. There were 80 patients in group 1 and 28 in group 2. To determine compatibility between the groups they studied whether the patient was white or not. The results showed there was 46 white patients in group 1 and 24 in group 2. Is there an association between head injury and race?

**Answers**

1 See table 1.

2 \[ df = (\text{number of rows } - 1) \times (\text{number of columns } - 1) \]

3 Therefore \( df = (5-1) \times (2 \times 1) = 4 \)

### Table 15

<table>
<thead>
<tr>
<th>Dangerous</th>
<th>Not dangerous</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>St Heartsinc</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Deathstar General</td>
<td>18</td>
<td>32</td>
</tr>
</tbody>
</table>

### Table 16

<table>
<thead>
<tr>
<th>No DNAR</th>
<th>DNAR</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>211</td>
<td>32</td>
</tr>
<tr>
<td>Female</td>
<td>201</td>
<td>26</td>
</tr>
<tr>
<td>Column total</td>
<td>412</td>
<td>58</td>
</tr>
</tbody>
</table>

### Table 17

<table>
<thead>
<tr>
<th>No DNAR</th>
<th>Expected</th>
<th>DNAR</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>243 (\times) 412 (/) 470 (\approx) 213</td>
<td>32</td>
<td>243 (\times) 58 (/) 470 (\approx) 30</td>
</tr>
<tr>
<td>Female</td>
<td>227 (\times) 412 (/) 470 (\approx) 199</td>
<td>26</td>
<td>227 (\times) 58 (/) 470 (\approx) 28</td>
</tr>
<tr>
<td>412</td>
<td>412</td>
<td>58</td>
<td>58</td>
</tr>
</tbody>
</table>

3 There is no significant difference between the two hospitals; \(\chi^2 = 0.23, df = 1, p > 0.05.\)

4 Using the \(\chi^2\) test the expected frequencies for each cell need to be calculated (table 17). For each cell, the \(\frac{\text{observed frequency } - \text{expected frequency}}{\text{expected frequency}}\) now needs to be calculated and added together (table 18).

The test statistic is the sum of all the \(\frac{[O−E]^2}{E}\) values for each cell (table 17).

The authors would like to thank Jim Wardrope for his help and support.


### Further reading

