MRes in Medical Statistics
MMB8028

School of Mathematics and Statistics

Practical session on the analysis of categorical data: outline solutions

1.
The first step is to turn the information in the question into a 2×2 table. Remember that the cells of the table are mutually exclusive, so the rows are classified by treatment schedule and the columns by whether or not a patient exhibited symptoms of Raynaud's phenomenon. This gives:

	Exhibited Reynaud's phenomenon	Did not exhibit Reynaud's phenomenon
Treatment A	16	24
Treatment B	12	33

The first thing to calculate is the proportion of patients exhibiting the phenomenon on each treatment, namely:
Treatment A: 16/40 = 0.40
Treatment B: 12/45 = 0.267

To see if this sample difference provides evidence that the true proportions are different, a χ^2 test can be performed by putting the two columns of this table into two columns in Minitab and selecting the Chi-Square Test (Two-Way Table in Worksheet)… in the Tables item under the **Stat** menu. This gives the following output:

Chi-Square Test: C1, C2

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

         C1      C2    Total
   1     16      24      40
       13.18   26.82
       0.605   0.297

   2     12      33      45
       14.82   30.18
       0.538   0.264

Total    28      57      85

Chi-Sq = 1.704, DF = 1, P-Value = 0.192
```

Thus the data provide no evidence that the true proportions of patients exhibiting Reynaud's phenomenon differ between the two treatment schedules.

Note that this way of calculating chi-square starts from the 2 × 2 table. If you had had the binary outcomes on each patient, then the Cross Tabulation and Chi-Square… item would have been the one to use.

2.
As before the first thing to calculate is the proportions of children with malocclusion in the breast- and bottle-fed groups. These are:
Breast-fed: 4/20 = 0.2
Bottle-fed: 1/22 = 0.045

Applying the same procedure as above gives:

Chi-Square Test: C1, C2

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

          C1     C2   Total
    1      4     16      20
        2.38  17.62
       1.101  0.149

    2      1     21      22
        2.62  19.38
       1.001  0.135

Total      5     37      42

Chi-Sq = 2.386, DF = 1, P-Value = 0.122
2 cells with expected counts less than 5.
```

However this time there is a warning at the bottom of the output that some of the expected values are too small for a χ^2 test to be valid. As such it would be preferable to try the analysis again using Fisher's Exact test.

Unfortunately, Fisher's Exact test is not an option under the Chi-square Test (Two-Way Table in Worksheet)… item but it is under Cross Tabulation and Chi-Square…. The data are not in the natural form for this item but it can be entered without too much trouble using the Frequencies box. Suppose the table is entered in three columns, as in the hint to the question. Then the Cross Tabulation and Chi-Square… dialogue box should be filled in as shown below. Clicking on OK (twice) gives the following output:

Tabulated statistics: feed, tooth

```
Using frequencies in freq


Rows: feed    Columns: tooth

          maloccluded   normal   All

Bottle              1       21    22
Breast              4       16    20
All                 5       37    42

Cell Contents:      Count


Fisher's exact test: P-Value =  0.174484
```
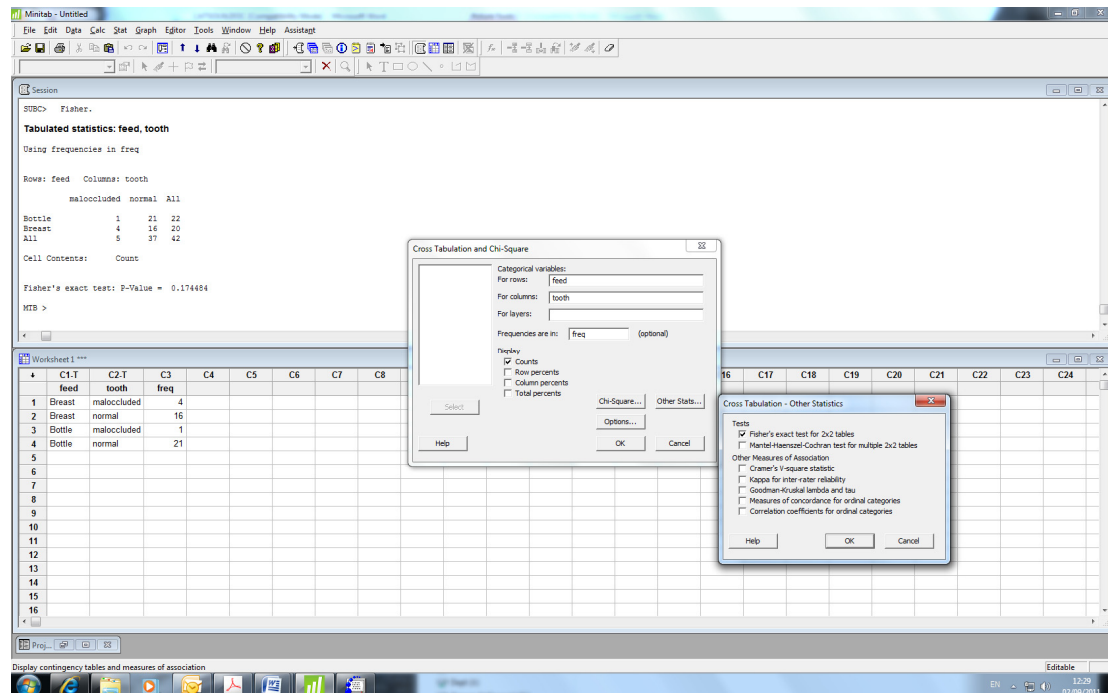
This confirms the earlier analysis, which indicates there is no evidence of any association malocclusion with type of feeding.

3.
The odds of exhibiting Reynaud's on treatment A are 16/24 = 2/3
The odds of exhibiting Reynaud's on treatment B are 12/33 = 4/11

So the odds ratio (OR) of exhibiting Reynaud's on A relative to B is $\dfrac{2/3}{4/11} =$

1.83333. To construct a 95% confidence interval for this quantity, use the result that the standard error of the *natural* log of the odds ratio is the square root of the sum of the reciprocals of the entries in the 2×2 table shown in answer to question 1.

Natural log of OR is $\log_e(1.83333) = \ln(1.83333) = 0.60613$

Entries in table are 16, 12, 24, 33. Sum of reciprocals is

$1/16 + 1/12 + 1/24 + 1/33 = 0.217803$

And this has square root 0.46669[†]

The 95% confidence interval for the natural log of the OR is then

$$0.60613 - 1.96 \times 0.46669 \text{ to } 0.60613 + 1.96 \times 0.46669,$$

[†] if you are using the Windows calculator under Windows 7 there is a square root key; for earlier versions of Windows the square root is sometimes found by selecting the Inv box then clicking on x^2, i.e. the square root is found as the inverse of squaring.

that is, from -0.30858 to 1.52084. This is the confidence interval for the *log* of the OR, not the OR itself. To get the confidence interval for the OR, take the natural antilogs of these numbers (click Inv before ln on the Windows calculator). The confidence interval required is therefore 0.734 to 4.576.

{Note two things here. First, do not forget the final step of taking antilogs. It is not too hard to remember here because one of the limits for the interval for the log OR is negative - this is not possible for the OR but is entirely legitimate for the log OR. Second, in doing this calculation, more decimal places than necessary for interpretation were carried through the intermediate steps in the computation. Rounding during the calculation can lead to a build up of error that is quite noticeable. Only round at the end of the calculation}

This means that the data estimate that the chance of exhibiting Reynaud's phenomenon (as measured by the odds) is 1.833 time greater on treatment A than B, but that the data are compatible the chance of Reynaud's being as much as 4.576 times greater on A than B, or being only 0.784 times as large on A as on B. In particular the data are compatible with an odds ratio of 1, namely that Reynaud's is equally likely under the two treatments. This is what the χ^2 test in question 1 indicated, but the confidence interval gives much more information because not only does it say the data are compatible with an odds ratio of 1 but it gives the range of odds ratios with which the data are compatible.

4.

The first task is to find the numbers of responders in each skin type. This can be done in several ways: the one shown below uses the Code item under **Data**. Choose the Numeric to Numeric… option and code from the MPD column into a new column (C3 in the screen print below). The entry 0:29.9 simply instructs Minitab to code any number in MPD which is in the *range* 0 to 29.9 as 0 in C3 and any entry of 30 in MPD as 1 in C3. In the screen below, C3 has already been named non-responder.
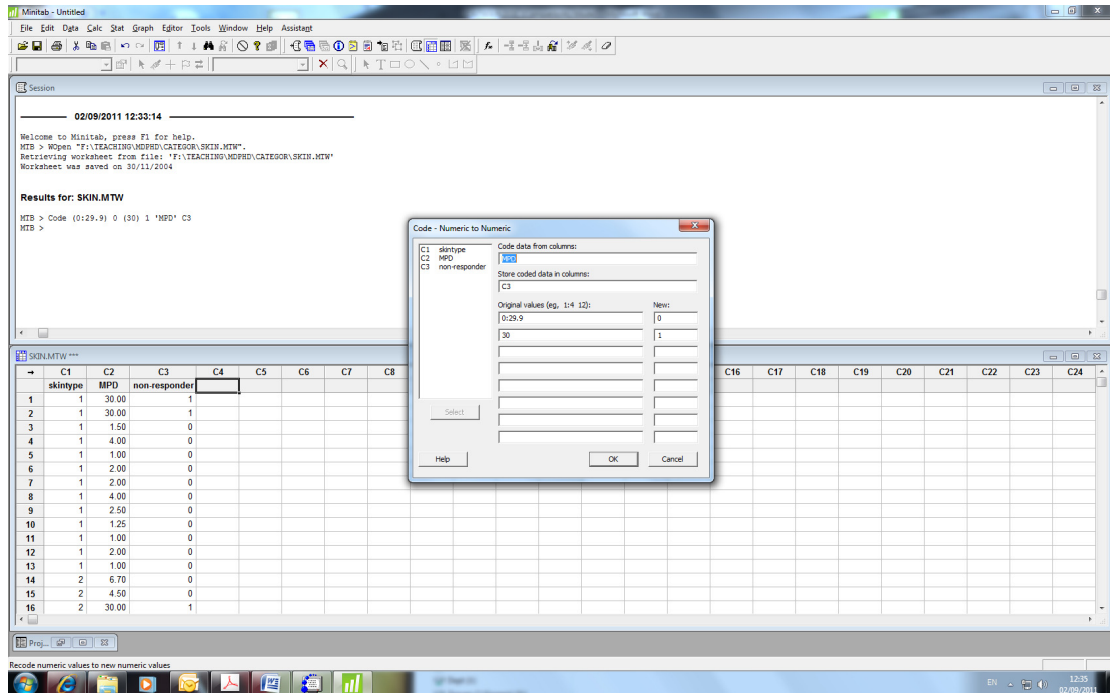
This produces a column of 0s and 1s in C3, a 0 indicating a value below 30, a 1 a value of 30, i.e. a non-responsive patient.

The analysis proceeds by forming a table, using the Cross tabulation and Chi-square... item on the Tables item under the **Stat** menu. Select skintype and non-responder as the variables classifying rows and columns.

This will give a table of how many patients are responders and non-responders in each skin type. However, by clicking on Row percents, the proportions (or at least the percentages) in each skin type will also be produced. The question of whether there is evidence that the true proportions differ can also be addressed: click on the box labelled Chi-Square… and select at least Chi-Square analysis (also selecting Expected cell counts can be useful) before clicking successively on OK. By now the screen will have looked as shown below.
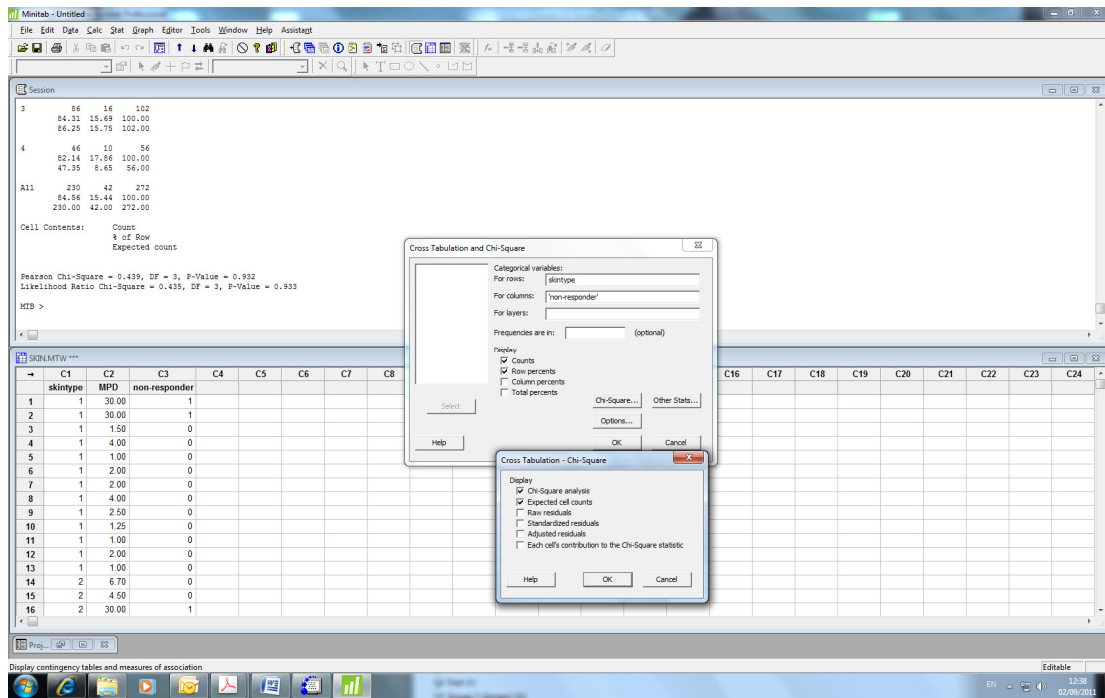
The output obtained is as follows.

Tabulated statistics: skintype, non-responder

```
Rows: skintype    Columns: non-responder

               0        1      All

1             32        5       37
           86.49    13.51   100.00
           31.29     5.71    37.00

2             66       11       77
           85.71    14.29   100.00
           65.11    11.89    77.00

3             86       16      102
           84.31    15.69   100.00
           86.25    15.75   102.00

4             46       10       56
           82.14    17.86   100.00
           47.35     8.65    56.00

All          230       42      272
           84.56    15.44   100.00
          230.00    42.00   272.00

Cell Contents:        Count
                      % of Row
                      Expected count


Pearson Chi-Square = 0.439, DF = 3, P-Value = 0.932
Likelihood Ratio Chi-Square = 0.435, DF = 3, P-Value = 0.933
```

The column headed '1' gives the numbers of non-responding patients for each skin type. This is the first entry in each cell, so for skin type 1 it is 5. The second entry,

13.51 for skin type 1, is the proportion of patients of that skin type who do not respond. These percentages are the values requested in the question. The third entry is the number you would expect, if the null hypothesis that the proportion of non-responders is the same for each skin type, were true.

At the bottom of the output is the χ^2 statistic and the corresponding P-value (the Pearson, rather than likelihood-ratio is the more conventional version). At P=0.933 there is no evidence against the null hypothesis.