PHCM9795: Foundations of Biostatistics

Timothy Dobbins

Term 2, 2023

Table of contents

# Preface

library(huxtable)  
library(magrittr)

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

1 + 1

[1] 2

# 1. Introduction

This is a book created from markdown and executable code.

Here is a method for including mathematical notation in tables:

**?(caption)**

# 2. Hypothesis testing for categorical data

## Learning objectives

By the end of this module you will be able to:

* Use and interpret the appropriate test for testing associations between categorical data;
* Conduct and interpret an appropriate test for independent proportions;
* Conduct and interpret a test for paired proportions;

## Readings

Kirkwood and Sterne (2001); Chapter 17. [[UNSW Library Link]](http://er1.library.unsw.edu.au/er/cgi-bin/eraccess.cgi?url=https://ebookcentral.proquest.com/lib/unsw/detail.action?docID=624728)

Bland (2015); Chapter 13. [[UNSW Library Link]](http://er1.library.unsw.edu.au/er/cgi-bin/eraccess.cgi?url=https://ebookcentral.proquest.com/lib/unsw/detail.action?docID=5891730)

Acock (2010); Section 7.6.

## 2.1 Introduction

In Module 6, we estimated the 95% confidence intervals of proportions and measures of association for categorical data and conducted a significance test comparing a sample proportion to a known value.

When both the outcome variable and the exposure variable are categorical, a chi-squared test can be used as a formal statistical test to assess whether the exposure and outcome are related. The P-value obtained from a chi-squared test gives the probability of obtaining the observed association (or more extreme) if there is in fact no association between the exposure and outcome.

In this Module, we also include tests for a difference in proportion for paired data.

### 2.1.1 Worked Example

We are using the randomised controlled trial as given in Worked Example 6.4 on the nauseating side effect of a drug.

The research question is whether the active drug resulted in a different rate of nausea than the placebo drug. This is equivalent to testing whether there is an association between nausea and type of drug received (active or placebo). Thus, we will test the null hypothesis that the experience of nausea and the treatment are not related to one another. The null hypothesis is:

* H0: The proportion with nausea in the active drug group is the same as the proportion with nausea in the placebo drug group.

The alternative hypothesis can be stated as:

* Ha: The proportion with nausea in the active drug group is different to the proportion with nausea in the placebo drug group.

## 2.2 Chi-squared test for independent proportions

A chi-squared test is used to test the null hypothesis that of no association between two categorical variables. First a contingency table is drawn up and then we estimate the counts of each cell (i.e. a, b, c and d) that would be expected if the null hypothesis was true. The row and column totals are used to calculate expected counts in each cell of the contingency table as follows:

Expected count = (Row count × Column count) / Total count

Stata will do this for us, as described in the Stata Notes section in this Module.

A chi-squared value is then calculated to compare the expected counts (E) in each cell with the observed (actual) cell counts (O). The calculation is as follows:

With [Number of rows 1] [Number of columns 1] degrees of freedom.

As for many statistics, the deviations between the observed and expected values are squared to prevent the negative and positive values balancing one another out.

If the expected counts are close to the observed counts, the chi-squared statistic will be close to zero, and the P-value will be close to 1. The larger the difference between the observed and expected counts, the larger the chi-squared statistic becomes (and the smaller the P-value). A large chi-squared statistic provides more evidence of an association between the exposure and outcome.

A chi-squared test to examine whether two independent proportions are associated can be obtained in Stata using the *Tables for Epidemiologists*: cc and cs commands. The test can also be conducted using the tab2 command, with the advantage that the tab2 command shows the expected counts. We will show why this is important shortly.

The Stata output for Worked Example 7.1 for the study on nausea is shown in Output 7.1a. It is always important to request the percentages using the options in ‘Cells’ to show the proportion of patients who have the outcome in each exposure group, i.e. the row percents.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (a) Observed counts   |  | Nausea | No nausea | Total | | --- | --- | --- | --- | | Active | 35 (70%) | 15 (30%) | 50 (100%) | | Placebo | 46 (92%) | 4 (8%) | 50 (100%) | | Total | 81 (81%) | 19 (19%) | 100 (100%) | | (b) Expected counts   |  | Nausea | No nausea | Total | | --- | --- | --- | --- | | Active | 40.5 | 9.5 | 50 | | Placebo | 40.5 | 9.5 | 50 | | Total | 81 | 19 | 100 | |

Table 2.1: Nausea side-effect data

We can see from the row percentages that 8% of patients in the placebo group experienced nausea compared to 30% of patients in the active group. If no association existed, we would expect to find approximately the same percent of patients with nausea in each group. The ‘Expected’ counts are higher for the groups with ‘No nausea’ because ‘No nausea’ is more prevalent in the sample than ‘Nausea’.

While the tab2 command will perform the chi-square test, the measure of effect is best obtained using the cs or cc commands, as discussed in Module 6. Using the relative risk obtained from Module 6, a conclusion from the above test can be written as:

The proportion with nausea in those who received the active drug is 30%, compared to 8% in those who received the placebo drug. Nausea was more frequent in those who received the active drug (Relative Risk = 3.75, 95% CI: 1.34 to 10.51). There is strong evidence that the proportion with nausea differs between the two groups ( = 7.86 with 1 df, P=0.005).

### 2.2.1 Assumptions for using a Pearson’s chi-squared test

The assumptions that must be met when using Pearson’s chi-squared test are that:

* each observation must be independent;
* each participant is represented in the table once only;
* at least 80% of the expected cell counts should exceed a value of five;
* all expected cell counts should exceed a value of one.

The first two assumptions are dictated by the study design. The last two assumptions relate to the numbers in the cells and can be explored when running the test. There should not be too many cells with low expected counts. Sometimes, the only way to avoid small cell counts is to recruit a much larger sample size. If small cell counts cannot be avoided, Fisher’s exact test can be used instead. More information on Fisher’s exact test can be found in Chapter 13 of An Introduction to Medical Statistics, Bland (2015).

### 2.2.2 Interpreting chi-squared tests

For the data being considered from Worked Example 7.1 all cells have an expected count greater than 5 and that the minimum cell count is 9.5. Therefore, it is appropriate to use the Pearson’s Chi-Squared test. If one or more cells have an expected cell count less than 5, then the Fisher’s exact test should be used. The P value associated with the Pearson’s chi-squared test is 0.005 indicating that we can reject the null hypothesis at a 5% level of significance. Thus, we can conclude that there is strong evidence that more patients who were randomised to receive the active treatment experienced nausea than patients randomised to the control (placebo) group.

## 2.3 Chi-squared tests for tables larger than 2-by-2

Chi-squared tests can also be used for tables larger than a 2-by-2 dimension. When a contingency table larger than 2-by-2 is used, say a 4-by-2 table if there were 4 exposure groups, the Pearson’s chi-squared can still be used.

### 2.3.1 Worked Example

The Stata file Example\_7.2.dta contains information about the severity of allergic reaction, coded as absent, slight, moderate or severe. We can test the hypothesis that the severity of allergy is not different between males and females. To do this we can use a two-way tabulation in Stata to obtain Stata Output 7.2 which shows the percent of females and males who fall into each severity group for allergy. The table shows that the percentage of males is higher in each of the categories of severity (slight, moderate, severe) than the percentage of females. [Command: tab allergy\_severity sex, col]

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (a) Observed counts   | Sex | Non-allergenic | Slight allergy | Moderate allergy | Severe allergy | Total | | --- | --- | --- | --- | --- | --- | | Female | 150 (62.0%) | 50 (20.7%) | 27 (11.2%) | 15 (6.2%) | 242 (100%) | | Male | 137 (53.1%) | 70 (27.1%) | 32 (12.4%) | 19 (7.4%) | 258 (100%) | | Total | 287 (57.4%) | 120 (24.0%) | 59 (11.8%) | 34 (6.8%) | 500 (100.0%) | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (b) Expected counts   | Sex | Non-allergenic | Slight allergy | Moderate allergy | Severe allergy | Total | | --- | --- | --- | --- | --- | --- | | Female | 138.9 | 58.1 | 28.6 | 16.5 | 242.0 | | Male | 148.1 | 61.9 | 30.4 | 17.5 | 258.0 | | Total | 287.0 | 120.0 | 59.0 | 34.0 | 500.0 | |

Table 2.2: Allergy data

The Pearson chi-squared statistic is calculated as 4.31, with 3 degrees of freedom, providing a P-value of 0.23. Therefore, there is little evidence of an association between gender and the severity of allergy.

## 2.4 McNemar’s test for categorical paired data

If a binary categorical outcome is measured in a paired study design, McNemar’s statistic is used. This statistic is a form of chi-square applied to a paired situation. A Pearson’s chi-squared test cannot be used because the measurements are not independent. However, McNemar’s test can be used to assess whether there is a significant change in proportions between two time points or between two conditions, or whether there is a significant difference in proportions between matched cases and controls.

For McNemar’s test, the data are displayed as shown in Table 7.1. Cells ‘a’ and ‘d’ called concordant cells because the response was the same at both baseline and follow-up or between matched cases and controls. Cells ‘b’ and ‘c’ are called discordant cells because the responses between the pairs were different. For a follow-up study, the participants in cell ‘c’ had a positive response at baseline and a negative response at follow-up. Conversely, the participants in cell ‘b’ had a negative response at baseline and a positive response at follow-up.

For other types of paired data such as twins or matched cases and controls, the data are similarly displayed with the responses of one of the pairs in the columns and the responses for the other of the pairs in the rows. For paired data, the grand total ‘N’ is always the number of pairs and not the total number of participants.

**Table** **:** Table layout for testing matched proportions

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Negative at follow-up** | **Positive at follow-up** | **Total** |
| Negative at baseline | a | b | a + b |
| Positive at baseline | c | d | c + d |
| Total | a + c | b + d | N |

**?(caption)**

### 2.4.1 Worked Example 7.3

Two drugs labelled A and B have been administered to patients in random order so that each patient acts as their own control. The dataset Example\_7.3.dta is available on Moodle. The null hypothesis is as follows:

* H0: The proportion of patients who do better on drug A is the same as the proportion of patients who do better on drug B

Using the tabulate2 command, observed count and cell percentages were obtained in Table 7.2. From the “Total” row in the table, we can see that the number of patients who respond to drug A is 41 (68.3%) and from the “Total” column the number who respond to drug B is less at 35 (58.3%), that is there is a difference of 10.0%.

#### Stata Output 7.3: cell percentages for paired data

. tabulate druga drugb, cell  
  
+-----------------+  
| Key |  
|-----------------|  
| frequency |  
| cell percentage |  
+-----------------+  
  
 Response | Response to Drug B  
 to Drug A | No Yes | Total  
-----------+----------------------+----------  
 No | 5 14 | 19   
 | 8.33 23.33 | 31.67   
-----------+----------------------+----------  
 Yes | 20 21 | 41   
 | 33.33 35.00 | 68.33   
-----------+----------------------+----------  
 Total | 25 35 | 60   
 | 41.67 58.33 | 100.00

The difference in the paired proportions is calculated using the simple equation:

Here,

The cell counts show that 20 patients responded to Drug A but not to drug B, and 14 patients responded to Drug B but not to drug A. McNemar’s statistic is computed from these two discordant pairs (labelled as ‘b’ and ‘c’) as follows:

with 1 degree of freedom.

In computing this statistic, the counts in the concordant cells (‘a’ and ‘d’) are not used and only the information from the discordant cells ‘b’ and ‘c’ is of interest.

The mcc command in Stata is used to perform the McNemar’s test. Its output is shown in Stata Output 7.4. The number of patients who have a response to each drug is labelled as Exposed in the output.

#### Stata Output 7.4: McNemar’s test

. mcc druga drugb  
  
 | Controls |  
Cases | Exposed Unexposed | Total  
-----------------+------------------------+-----------  
 Exposed | 21 20 | 41  
 Unexposed | 14 5 | 19  
-----------------+------------------------+-----------  
 Total | 35 25 | 60  
  
McNemar's chi2(1) = 1.06 Prob > chi2 = 0.3035  
Exact McNemar significance probability = 0.3915  
  
Proportion with factor  
 Cases .6833333  
 Controls .5833333 [95% Conf. Interval]  
 --------- --------------------  
 difference .1 -.1054528 .3054528  
 ratio 1.171429 .8663498 1.583939  
 rel. diff. .24 -.1585239 .6385239  
  
 odds ratio 1.428571 .6862537 3.057277 (exact)

Two versions of the McNemar’s test are given in the output. The exact version is generally recommended. The P value for the exact McNemar’s test is 0.39, providing no evidence against the null hyopthesis.

In this study of 60 participants, where each participant received both drugs, 41 (68%) responded to Drug A and 35 (58%) responded to Drug B. The difference in the proportions responding is estimated as 10.0% (95% CI -10.5% to 30.5%). There is no evidence that the response differed between the two drugs (exact McNemar’s P=0.39).

## 2.5 Summary

In Module 6, we estimated proportions and measures of association for categorical data and conducted a one-sample test of proportions. In this module, we conduct significance tests for two or more independent proportions using the chi-squared test. The chi-squared test can also be used to conduct a significance test when there are more than two categories in both variables. The McNemar’s test is used when we have paired data.

# References

Acock, Alan C. 2010. *A Gentle Introduction to Stata*. 3rd ed. College Station, Tex: Stata Press.

Bland, Martin. 2015. *An Introduction to Medical Statistics*. 4th ed. Oxford, New York: Oxford University Press.

Kirkwood, Betty, and Jonathan Sterne. 2001. *Essentials of Medical Statistics*. 2nd ed. Malden, Mass: Wiley-Blackwell.