An example motivated discourse of the chi-squared test of association (2 by 2)

Paul White,

Applied Statistics Group,

Faculty of Environment and Technology,

Univeristy of the West of England, Brsitol,

Bristol BS16 1QY, UK

paul.white@uwe.ac.uk

Paul Redford,

Department of Health and Social Sciences

Faculty of Health and Applied Sciences,

University of the West of England, Brsitol,

Bristol BS16 1QY, UK

paul2.redford@uwe.ac.uk

James Macdonald,

Department of Health and Social Sciences

Faculty of Health and Applied Sciences,

University of the West of England, Brsitol,

Bristol BS16 1QY, UK

james.macdonald@uwe.ac.uk

Abstract— An example application of the chi-square test of association is given through a set of questions and answers. The emphasis is on methodological aspects and what can be legitimately inferred, if anything. It is strongly highlighted that statistical conclusions can be drawn from a study, but these statistical conclusions might not translate into research or scientific conclusions.

Keywords— chi-squared, test of association

I. Introduction

The chi-square test of association is a long-established procedure used to statistically examine whether there is an association between two categorical variables based on a sample of *N* independently sampled observations. Other texts (e.g. [1, 2]) give a good simple mathematical description of the underpinning mathematics, statistical approximations, and subtleties. The focus of this short note is to give a worked example, to discuss emerging issues, and to reflect on what might limit the ability to generalize findings. The motivating example is an example taken from a newspaper report (described below). The example will be deconstructed using a series of questions.

II. A Motivating Example

A major UK newspaper reported on a study investigating the relationship between incidence of breast cancer and breast injury. The women in the study were patients attending a North Lancashire breast screening unit and were in the 50 to 64 age group. Of the 67 breast cancer cases 35 reported a breast injury in the previous 5 years. Of the 134 women who did not have breast cancer 16 reported having an injury to their breasts in the five-year period. The data referred to are summarised in Table 1.

To get a better understanding of this table we will consider a series of questions which require minimal calculation.

Question 1. What proportion of women with breast cancer reported a breast injury in the previous 5 years?

Answer
The fraction of women with breast cancer who reported a breast injury is 35/67.

The proportion of women with breast cancer who reported a breast injury is 0.522.

Table 1

	Breast Injury	
	No	Yes
Breast Cancer	32	35
No Breast Cancer	108	16

Question 2. What percentage of women with breast cancer reported breast injury in the previous 5 years?

Answer
The percentage of women with breast cancer who reported a breast injury is 52.2%.

Question 3. What percentage of women without breast cancer reported breast injury in the past 5 years?

Answer
The fraction of women without breast cancer who reported a breast injury is 16/124.

The proportion of women without breast cancer who reported a breast injury is 0.129.

The percentage of women with breast cancer who reported a breast injury is 12.9%.

Question 4 Would it be better to consider the percentage of women with breast injury who subsequently were found to have breast cancer? (i.e. for the given layout of data would it be better to consider column percentages rather than row percentages?).

Answer
The quick answer to this question is "Yes!". The suggestion, or motivation, behind the research is that breast injury is causally related to an increased likelihood of breast cancer (i.e. injury is being considered as a prognostic factor or risk factor for breast cancer). Of course, the data from this observational study cannot be used to prove a causal link, and we can only see if the data is consistent with this position. However, we do acknowledge that there might be a temporal relationship in as much as we are considering injuries which pre-date any diagnosis. For these reasons it perhaps makes better conceptual sense to consider the percentages conditional on breast injury status (as given below).

Question 5. In Table 1, what percentage of women with a reported breast injury in the previous five years were diagnosed as having breast cancer?

Answer
The percentage is 68.6%

Question 6. In Table 1, what percentage of women who did not report a breast injury were found to have breast cancer?

Answer
The percentage is 22.9%.

III. Hypotheses

We will now consider the research question, scientific hypotheses

Question 7. What is the research question being investigated?

Answer
The research question can be phrased many ways including "Is breast injury a risk factor for breast cancer?", or "Does breast injury increase the likelihood of breast cancer?"

Question 8. What would be the scientific hypotheses?

Answer
Example scientific statements would be along the following lines:

S_0: Breast injury is not a risk factor for breast cancer
S_1: Breast injury is a risk factor for breast cancer

S_0: Incidence of breast cancer is independent of breast injury status
S_1: Incidence of breast cancer is affected by incidence of breast injury

S_0: Breast injury does not affect the likelihood of breast cancer
S_1: Breast injury increases the likelihood of breast cancer

S_0: Breast injury and breast cancer are independent events
S_1: If "breast injury" then more likely to have cancer.

Note that some of these scientific statements imply a direction (e.g. an increased risk); however this implied direction does not mean we would conduct a one-sided hypothesis test.

Question 9. What would be the statistical hypotheses for this situation?

Answer
Let's start with a technical explanation.

Let π_1 denote the population proportion of women who have had a diagnosis of breast cancer after having a breast injury in the previous five years and let π_2 denote the population proportion of women who have had a diagnosis of breast cancer who have not had a breast injury (in the previous five years).

The statistical hypotheses would then be

$$\boldsymbol{H_0} : \boldsymbol{\pi_1} = \boldsymbol{\pi_2}$$
$$\boldsymbol{H_1} : \boldsymbol{\pi_1} \neq \boldsymbol{\pi_2}$$

Of course, it may be preferable for some to avoid heavy mathematical notation. If wanting to avoid the use of mathematical symbols, then the hypotheses could be written:

H_0 : There is no association between breast injury and breast cancer
H_1: There is an association between breast injury and breast cancer

or alternatively as

H_0 : Breast cancer and breast injury are independent
H_1: : Breast cancer and breast injury have some dependency

IV. SPSS

Table 2a and Table 2b provides SPSS output from a chi-square test of the above hypotheses. The output has been edited to remove superfluous material. Also note that the rows and columns in Table 1 have been transposed to have (hopefully) an easy table to comprehend.

Chi-square test of association, 2 by 2

Table 2a

			Cancer Status		
			Breast cancer	No breast cancer	Total
Injury Status	No Breast Injury	Count	32	108	140
		Percentage	22.9%	77.1%	100.0%
	Breast Injury	Count	35	16	51
		Percentage	68.6%	31.4%	100.0%
Total		Count	67	124	191
		Percentage	35.1%	64.9%	100.0%

Table 2b Chi-Square Tests

	Value	df	Sig. (2-sided)
Pearson Chi-Square	34.388[a]	1	.000

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 17.89.

Note that the column headed "Sig" in Table 2b is the column for the *p*-value. (See [3] for a commentary on *p*- values.)

Question 10. What statistical conclusion can you draw from your analysis?

The statistical conclusion would be to reject the null hypothesis ($p < .001$).

Question 11. Write a sentence, effectively summarising the results of your analysis.

Analysis of the data using Pearson's chi-square test of association indicates that there is a statistically significant association between occurrences of self-reported breast injury and breast cancer in the sample ($\chi^2 = 34.39$; $df = 1$, $p < .001$).

Note that this is a statistical conclusion for the sample; it is not a scientific conclusion. It is not a research conclusion.

V. Some Reflections

[1] Is this study an experiment? How would you design an experiment to investigate the relationship between breast injury and breast cancer?

No, this is not an experiment. The study does not have random selection nor does it have random allocation and there is no manipulation of any factor or stimulus. This is an observational study (aka a correlational study).

An experimental design would be to take a group of women and randomly allocate some of the women to Group A and others to Group B. The women in Group A would deliberately be subjected to a breast injury and followed-up over five-years. The women in the other group would have their breasts protected for five years. The outcome of interest would be the development of breast cancer in the five-year period.

[2] Would your design be ethical?

Clearly this proposal would not be deemed ethical (i.e. deliberate harm to participants with the possibility of increasing the likelihood of a difficult to treat disease).

[3] What is the target population? Do we have a random sample from the target population? Do we have a representative sample?

The target population is women patients of the North Lancashire breast screening unit in the 50 to 64 age group. A random sample of this screening unit was not taken; instead the sample is a convenience sample of women attending the unit who themselves may have gone for a screening for a whole host of reasons. A representative sample and a random sample are not the same thing; we would need to have further background information on the women patients and those in the target population to make an assessment of representativeness. However, it is likely that this set of women is not representative of women aged 50 to 64 in North Lancashire (nor from any other region).

[4] Is the sample size sufficiently big enough to generalise?

No; it is difficult to see how a sample of $n = 191$ would generalise to the entire population of women in North Lancashire and even harder to generalise to other populations. The research suffers from poor external validity, and poor ecological validity.

[5] What sources of bias do we have in this study?

Selection bias i.e. the women are self-selecting and not representative of the population of interest.

Recall bias i.e. ability to recall past injuries and this ability might be group dependent. The recall bias in itself is sufficient to cast doubt on the internal validity of conclusions. There is no definition of a "breast" injury.

[6] In light of the above, do you think we can draw any scientific conclusions from this study?

No. There are severe questions over both the internal and external validity of data. At best these data present some prima facie evidence to support an application for further investigation using a better methodological approach. Newspaper are often criticised for poor "statistical" journalism and often add to the problems by being "sensational" in their reporting.

It is always perfectly fine to draw statistical conclusions from an analysis but it does not follow that strong scientific conclusions can be drawn. In general, the design of the study dictates the quality of the argument in drawing scientific conclusions.

VI. Summary

The well-established chi-square test of association for the 2 by 2 contingency table has been applied. This note has deliberately avoided being a mathematical exposition. Instead the focus has been on the research question, scientific hypotheses, statistical hypotheses and limitations of inference (internal, external, and ecological validity).

It is evident that it is very difficult to undertake good quality empirical research; there are always lots of potential limitations to overcome. Of course, research per se, is not a one-shot activity; research is a global shared activity concerned with reproducibility and replication of findings, and any single study is simply, at best, provides one piece of evidence which is to be weighted and be synthesised with the totality of evidence on the subject.

Acknowledgments

The authors gratefully acknowledge the support of the Learning and Teaching Initiative in the Faculty of Health and Applied Sciences, University of the West of England, Bristol in supporting the wider Qualitative and Quantitative Methods teaching programme.

Series Bibliography

White P, Redford PC, and Macdonald J (2019) A primer on validity and design terminology in comparative designs, *Qualitative and Quantitative Research Methods Project*, University of the West of England, 1 –4

White P, Redford PC, and Macdonald J (2019) An example motivated discourse of the chi-squared test of association (2 by 2), *Qualitative and Quantitative Research Methods Project*, University of the West of England, 1 –4

White P, Redford PC, and Macdonald J (2019) An example motivated discourse of the independent samples t-test and the Welch test, *Qualitative and Quantitative Research Methods Project,* University of the West of England, 1 –6

White P, Redford PC, and Macdonald J (2019) An example motivated discourse of the paired samples t-test, *Qualitative and Quantitative Research Methods Project,* University of the West of England, 1 –5

White P, Redford PC, and Macdonald J (2019) A primer on statistical hypotheses and statistical errors, *Qualitative and Quantitative Research Methods Project,* University of the West of England, 1 –4

White P, Redford PC, and Macdonald J (2019) A reverse look at p-values, *Qualitative and Quantitative Research Methods Project*, University of the West of England, 1 –5.

White P, Redford PC, and Macdonald J (2019) That assumption of normality, *Qualitative and Quantitative Research Methods Project*, University of the West of England, 1 –11

White P, Redford PC, and Macdonald J (2019) Cohen's *d* for two independent samples, *Qualitative and Quantitative Research Methods Project*, University of the West of England, 1 –4

References

[1] Pandis, N (2016) The chi-square test, *American Journal of Orthodontics and Dentofacial Orthopedics*, Vol 150, No 5, 898-899

[2] Kim HY (2017) Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test, *Restorative Dentistry and Endodontics*, Vol 42, No 2, 152–155.

[3] White P, Redford PC, and Macdonald J (2019) A reverse look at p-values, *Qualitative and Quantitative Research Methods Projectt*, Univeristy of the West of England, 1 –5.