THE CHI SQUARE TEST

An introduction

ANTONY UGONI B.Sc. (Hons).*

BRUCE F. WALKER D.C., M.P.H. †

Abstract: The Chi square test is a statistical test which measures the association between two categorical variables. A working knowledge of tests of this nature are important for the chiropractor and osteopath in order to be able to critically appraise the literature.

Key Indexing Terms: Chi square, chiropractic, osteopathy.

THE CHI SQUARE TEST

The constant collation of data in medical research provides statisticians and researchers with various types of data. The most recognizable of these is data in a quantitative form. For example, straight leg raising (SLR) in subjects able to raise their legs greater than 0 degrees allows us to calculate the average SLR for say two groups and perform a t-test. Unfortunately, not all data is in this quantitative form.

For example, instead of measuring an individuals SLR we may be interested in the patients' subjective improvement (using just "Yes" or "No" responses) after 2 types of treatment. Can we then calculate the average improvement for each group and perform a t-test? Is it possible to calculate the difference between levels of improvement? Is it possible to calculate the ratio of improvement?

The answer to all these questions, of course, is a resounding 'no', and other methods need to be employed. The most common method used to analyze such data is the Chi Squared (χ^2) test of association, and the outline for the simplest scenario is given below in table 1.

* DEPARTMENT OF PUBLIC HEALTH & COMMUNITY MEDICINE.
THE UNIVERSITY OF MELBOURNE. PARKVILLE. VIC, 3052.

† PRIVATE PRACTICE.
33 WANTIRNA RD, RINGWOOD, VIC. 3134. PH 03 879 5555

Table 1

		Category II 1	Category II 2	
Category I	1	a	b	$n_1=a+b$
	2	c	d	$n_2=c+d$
				$n=n_1+n_2$

In words, the elements of the table are,

a = number of individuals who are of type 1 in category I and type 1 in category II

b = number of individuals who are of type 1 in category I and type 2 in category II

c = number of individuals who are of type 2 in category I and type 1 in category II

d = number of individuals who are of type 2 in category I and type 2 in category II

n_1 = the number of individuals who are of type 1 in category 1

n_2 = the number of individuals who are of type 2 in category 1

n = total number of individuals studied

To illustrate this, consider for example two groups of patients with sciatica who undergo 6 weeks of spinal manipulative therapy (SMT) or 6 weeks of intermittent motorized traction (IMT). We wish to know whether there is an association between improvement and the type of treatment received for these sciatica patients.

In our example 190 patients receive IMT and 200 receive SMT. After 6 weeks we ask them whether they have improved. For IMT, 85 reply 'Yes' and 85 reply 'No', and for SMT 45 reply 'Yes' and 155 reply 'No'.

We can display this data in a 2×2 contingency (frequency) table, shown in table 2.

Table 2

	Improved Yes	Improved No	
IMT	95 $_a$	95 $_b$	190
SMT	45 $_c$	155 $_d$	200
	140	250	390

In our example our observations are categorical and not quantitative, so our focus should move from means to proportions. We now display the following table (table 3) to explain.

Table 3

		Category II 1	2
Category I	1	p_1	p_2
	2	q_1	q_2

where

p_1 = the proportion of individuals who are of type 1 in category I and type 1 in category II

p_2 = the proportion of individuals who are of type 1 in category I and type 2 in category II

q_1 = the proportion of individuals who are of type 2 in category I and type 1 in category II

q_2 = the proportion of individuals who are of type 2 in category I and type 2 in category II

Notice that $p_1+p_2=q_1+q_2=1$. Thus p_1 and p_2 can be thought of as the way people who are of type 1 in category 1 are distributed across category 2, and q_1 and q_2 can be thought of as the way people who are of type 2 in category 1 are distributed across category 2.

In an earlier paper (1), it was stated that the statistical hypothesis of interest is always ***nothing happens*** (null hypothesis). This can be extended to this case by testing the hypothesis of $p_1=q_1$, and $p_2=q_2$. That is, the distribution of individuals across category 2 is the same for all types of category 1. In other words, the distribution of individuals across category 2 is independent of category 1.

To test this hypothesis, we need to compare what would be expected if the hypothesis were true, against what has actually been observed.

If we analyse our example above, we observed 140 patients who subjectively improved. This represents 140 out of the total 390 in the trial, or 36%. So, if there is **no** association between treatment and improvement (as hypothesised), then we would expect 36% of each treatment group to improve regardless of management.

Therefore, using our example again,

36% of 190 = 68 on the IMT should improve, and
36% of 200 = 72 on the SMT should improve.

But what about the "no improvement" patients? We observed 250 out of the 390 who did not improve (ie 64%). So, if there is **no** association between treatment and improvement then we would expect 64% of both treatment groups not to improve. That is,

64% of 190 = 122 on the IMT should not improve, and
64% of 200 = 128 on the SMT should not improve.

So our contingency table can be drawn thus (table 4), where the figures in brackets are the expected frequencies.

Table 4

	Improved YES	NO	
IMT	95 (68)	95 (122)	190
SMT	45 (72)	155 (128)	200
	140	250	390

There exists a simple formula to calculate the expected value for any cell in the above table.

Equation 1
Expected value = (Row total)×(Column total)/(Grand total)

For example, the expected number of individuals who receive IMT and improve is,

$190\times140/390 = 68.2 \approx 68$

It should be noted that the expected cell frequencies add up to the same row and column totals as the observed frequencies. It should also be noted that the cell frequencies are calculated under the null hypothesis of no association between treatment and improvement.

Having obtained these expected values, we now need to compare them with what has actually been observed. To do this, we calculate the χ^2 statistic, which is shown below.

Equation 2

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

That is, take each expected value and subtract from the corresponding expected value. Square this result, and divide by the corresponding expected value. Calculate this quantity for each cell in the table, and add together.

The calculations for the example above, are shown below in table 5.

Table 5

Obs	Exp	Obs-Exp	$(Obs-Exp)^2$	$(Obs-Exp)^2/Exp$
95	68	27	729	10.72
95	122	-27	729	5.98
45	72	-27	729	10.13
155	128	27	729	5.70
				32.53

Thus, the value of χ^2 is 32.53.

Inspection of the formula for χ^2 will show that the value of χ^2 will be small when the null hypothesis is true. This is due to the fact that expected values are calculated under the assumption that the null hypothesis is true, and that the term (Observed-Expected) will be small if the observed data lies close to the expected data. Alternatively, if the null hypothesis is false, then the expected values will not be close to the observed values, and the value of χ^2 will be large.

The question to be addressed now is 'How large should χ^2 be to reject the null hypothesis?'

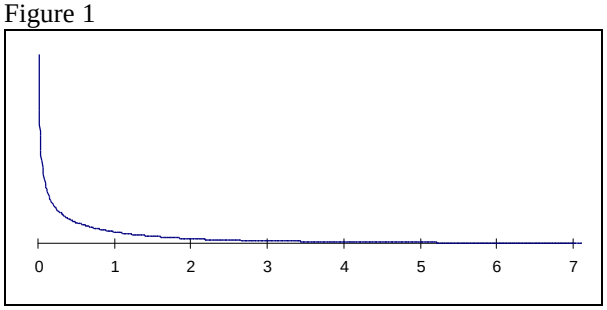
The value of χ^2 comes from a Chi Square distribution. This distribution is defined by 1 parameter, which is known as the degrees of freedom. The degrees of freedom is dependent on the size of the table being studied, and can be calculated using the following simple formula.

Equation 3

Degrees of freedom = (# Rows - 1) × (# Columns - 1)

A Chi Squared distribution with 1 degree of freedom is shown in figure 1.

Figure 1



nb. The range of the horizontal axis is $0 \rightarrow \infty$.

The p-value associated with our test (or any Chi Squared test with a 2×2 table) is the area under the curve and to the right of the calculated value of Chi Squared. The area under the curve and to the right of 6.64 is less than 0.01 (or 1%). Since the calculated value of Chi Squared is 32.53, it is clear that the p-value is less than 0.01 (2). The conclusion is that we reject the null hypothesis. That is, the proportion of improved individuals who received IMT and improved, is different to the proportion of individuals who received SMT and improved.

In many trials involving improvement, more than 2 levels of improvement is used. For example, let us examine a comparison trial between spinal manipulation with the use of hot packs (Trt 1) and spinal manipulation with the use of cold packs (Trt 2) for acute low back pain. For our improvement scale we could use a 5 point categorical scale such that shown in table 6.

Table 6

	None	Mild	Noticeable	Definite	Complete	
Trt 1	39	43	89	126	87	384
Trt 2	12	32	65	98	65	272
	51	75	154	224	152	656

The null hypothesis is that the distribution of improvement is the same for both treatments.

Expected values need to be calculated first, and equation 1 can be applied. The expected value for the *Trt 1/None* cell is 384×51/656=29.85. For the *Trt 1/Mild* cell, 384×75/656=43.90 etc. Once all the expected values are calculated, the value for Chi Square can be computed (table 7).

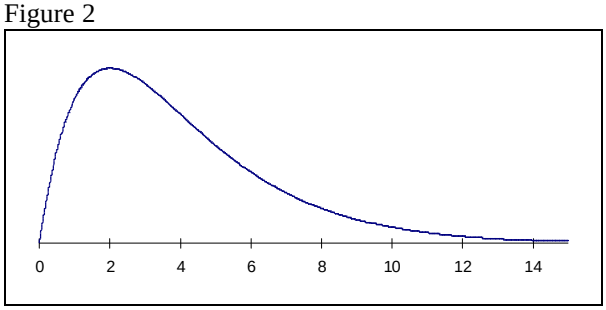
Table 7

Obs	Exp	Obs-Exp	$(Obs-Exp)^2$	$(Obs-Exp)^2/Exp$
39	29.85	9.15	83.72	2.80
43	43.90	-0.90	0.81	0.02
89	90.15	-1.15	1.32	0.02
126	131.12	-5.12	26.21	0.20
87	88.98	-1.98	3.92	0.04
12	21.15	-9.15	83.72	3.96
32	31.10	0.90	0.81	0.03
65	63.85	1.15	1.32	0.02
98	92.88	5.12	26.21	0.28
65	63.02	1.98	3.92	0.06
				7.43

Thus, the value of χ^2 is 7.43.

Using equation 3, the degrees of freedom are (2-1)×(5-1)=4. A Chi Square distribution with 4 degrees of freedom looks like.

Figure 2



The p-value is the area beneath the curve and to the right of 7.43. This turns out to be 0.1148. If we use a significance level of 0.05, then we do not reject the null hypothesis. Therefore there is no difference between the two treatment outcomes. To interpret this further, consider table 8, where the data has been transformed into row percentages.

Table 8

	None	Mild	Noticeable	Definite	Complete
Trt 1	10.2%	11.2%	23.2%	32.8%	22.7%
Trt 2	4.4%	11.8%	23.9%	36.0%	23.9%

Strictly speaking, these distributions differ from each (10.2%≠4.4%, 11.2%≠11.8%,.....,22.7%≠23.9%). However, when we consider the possibility of random error being present in the data, we do not have enough evidence to state that the differences observed are indicative of a true underlying difference.

There are key assumptions which need to be adhered to when using the χ^2 test. They are,

1. Each individual appears in the table once only.
2. The result for each individual is independent of all other individuals.
3. The table of expected values should have 80% of all expected values greater than 5.

CONCLUSION

The chi-square test is a statistical test of association between two categorical variables. It is used very commonly in clinical research and a good understanding of the test is useful for chiropractors and osteopaths to be able to critically appraise the literature.

REFERENCES

1. Ugoni A. On the subject of hypothesis testing. COMSIG Review, 1993; 2(2): 45-8.
2. Neave, HR. Statistics Tables for Mathematicians, Engineers, Economists, and the Behavioural and Management Sciences. Unwin Hyman Ltd, 1988: 42-3