First-year Statistics for Psychology Students Through Worked Examples

1. THE CHI-SQUARE TEST

A test of association between categorical variables

by

Charles McCreery, D.Phil

Formerly Lecturer in Experimental Psychology
Magdalen College
Oxford

OXFORD FORUM

Psychological Paper No. 2007-1

Contents

1. Introduction

There are usually three complementary methods for mastering any new intellectual or artistic task; these are, in ascending order of importance:

- reading books about it
- observing how other people do it
- actually doing it oneself

This tutorial focuses on the second of these methods. It is based on handouts that I developed when teaching first-year psychology students at Magdalen College, Oxford. The core of the tutorial is a worked example from an Oxford University Prelims Statistics examination paper. I have therefore placed this section in prime position; however, in teaching the order of events was different, and more nearly corresponded to the three-fold hierarchy of methods given above:

1. The student was invited to read one of the chapters on the Recommended Reading list, given in Section 5. He or she would also have been expected to attend a lecture on the topic in question at the Department of Experimental Psychology.

2. The student would attend a tutorial, in which we would go through the worked example shown here. He or she would take away the handouts shown in Sections 3 and 4, which were designed to give structure to the topic and help them when doing an example on their own.

3. They would be given another previous examination question to take away and do in their own time, which would be handed in later for marking.

I am strongly in favour of detailed worked examples; following one is the next best thing to attempting a question oneself. Even better than either method is doing a statistical test on data which one has collected oneself, and which therefore has some personal significance to one, but that is not usually practicable in a first-year course.

The book which comes closest to giving detailed worked examples is Spiegel (1992, see Recommended Reading). Each chapter has numerous 'solved problems' on the topic in question, which occupy more than half of each chapter. However, the solutions are not as detailed and discursive as the one given here.

I should be pleased to hear from anyone who finds this tutorial helpful, or who spots a typographical or other error which should be corrected. I can be contacted at:

reply@oxfordforumpublisher.com

2. A worked examination question

2.1 Question[1]

Neyzi, Alp and Orhon (1975) investigated the effect of socio-economic class on physical development of Turkish children. Physical development was classified on a scale of 1 (none) to 5 (fully developed) and the socio-economic class of their parents was assessed on a scale of 1 to 4.

The data were as follows:

Socio-economic Class of parents	*Physical development*				
	1	*2*	*3*	*4*	*5*
1	2	14	28	40	18
2	1	21	25	25	9
3	1	12	12	12	2
4	6	17	34	33	6

Plot these data in a meaningful way and report your initial findings.

Stating clearly your hypotheses, carry out an analysis to test for a relationship between physical development and socio-economic class using as many different categories of physical development as possible, and report your conclusions.

Carry out a further analysis comparing those who are fully developed (stage 5) with those who are not (stages 1-4) and report your conclusion.

Provide an explanation for these two conclusions.

For any test where you detect a relationship, report on the nature of that relationship.

(5;5;4;3;3)

[The numbers in the bottom right hand corner of the question are as they appear in the Oxford Prelims exam paper and indicate to the candidate how many marks each part of the question is potentially worth if answered correctly.]

---

[1] The question is taken from the Preliminary Statistics paper for first-year psychology students at Oxford University, Hilary Term, 1999.

2.2 Answer:

2.2.1. How to recognise that this is a chi-square question:

The layout of the data may look superficially like that of a two-way ANOVA, but the data in the cells are raw numbers; to be an ANOVA the entries in the cells would have to be means.

Note also that the measures are both categorical. Class is an ordinal, not an interval, measure. (See Appendix B for some points on this distinction.) Physical development might have been a continuous, interval measure, but here it is not; the five categories are discrete and we do not have any information about how they are arrived at, so we can only safely conclude that they are ordinal.

2.2.2. Method

The first thing we have to do is collapse some cells. A requirement of the chi-square test is that there should be at least 5 observations in each cell. (The wording of the question contains a hint to remember this when it says 'using as many categories of physical development as possible' in paragraph 4.) We therefore amalgamate columns 1 and 2, and columns 4 and 5, to give the following:

	Physical development		
	2	*3*	*4*
Class			
1	16	28	58
2	22	25	34
3	13	12	14
4	23	34	39

2.2.3 Plot

There are two possible ways of plotting this data: as a stacked bar chart of percentages, or as a line chart. The former is sometimes more revealing, but takes much longer. I recommend a line chart in this case. In fact I recommend line charts for all questions in this paper except those which require you to plot a frequency distribution. The *Descriptive Statistics* type questions have in the past always required a bar chart, and occasionally there are other types of questions which require you to plot a frequency distribution, e.g. the goodness-of-fit question from Hilary Term 1999.

The plot (see Appendix A) reveals that in all classes except Class 3 the number of subjects increases as physical development increases. This is presumably a function of how the population was sampled; i.e. there were a disproportionate number of older children in the sample. The reason for Class 3 not showing this pattern may simply be a chance effect of the total numbers in this class being smaller than in any of the others.

The most notable effect revealed by the plot is that fully developed children are apparently over-represented in Class 1. This suggests the hypothesis that children in this class develop faster than those in other social classes, perhaps due to better nutrition, for example, or better living conditions generally.

2.2.4 The calculation

The next step is compute the row and column totals, thus:

	Physical development			
	2	*3*	*4*	*Totals:*
Class				
1	16	28	58	102
2	22	25	34	81
3	13	12	14	39
4	23	34	39	96
Totals:	74	99	145	318

Next we have to compute the expected value for each cell, using the formula:

(Row total / Grand total) * Column total

E.g. for the first cell of Row 1 (Class 1, Physical development 2):

(Row 1 total / Grand total) * Column 1 total
= 102/318 * 74
= 0.32 * 74
= 23.73 = 24 rounded to the nearest whole number.

See **Hoel (1976) p. 253** for a good explanation of why this method gives you the expected values (see 'Recommended Reading' below).

The resulting table in this instance is as follows (expected values are in brackets):

Physical development

Class	2	3	4	**Totals:**
1	16(24)	28(32)	58(46)	102
2	22(19)	25(25)	34(37)	81
3	13(9)	12(12)	14(18)	39
4	23(22)	34(30)	39(44)	96
Totals:	74	99	145	318

Now we have to apply the formula for the chi-square statistic:

$$\chi^2 = \sum [(O-E)^2 / E]$$

where O = the observed frequency in each category

E = the expected frequency in each category

and the summation is made over all categories.

In this case this gives:

$$\chi^2 = (16 - 24)^2/24 + (28 - 32)^2/32 + (58 - 46)^2/46$$
$$+ (22 -19)^2/19 + (25 - 25)^2/25 + (34 - 37)^2/37$$
$$+ (13-9)^2/9 + (12-12)^2/12 + (14-18)^2/18$$
$$+ (23 - 22)^2/22 + (34 - 30)^2/30 + (39 - 44)^2/44$$

$$= 2.67 + 0.50 + 3.13$$
$$+ 0.47 + 0 + 0.24$$
$$+ 1.78 + 0 + 0.89$$
$$+ 0.05 + 0.53 + 0.57$$

$$= 10.83.$$

Degrees of freedom = (No. of rows - l)(No. of columns - 1)
= (4-1)(3-1)
= 6.

For the test to be significant at the 0.05 level, given 6 degrees of freedom, the value for χ^2 has to be at least 12.59 (see Table VII on page 336 of Hoel, 1976, for example). Therefore we cannot reject the null hypothesis of no association between the two variables, parents' socio-economic class and physical development.

Conclusion: any effect in Class 1 as revealed by the plot is swamped by the lack of any effect elsewhere.

2.2.5 The further analysis:

Amalgamating physical development stages 1-4, and recomputing the expected values for each of the new cells by the method described above, gives us the following contingency table:

Class	**Physical development**		
	1-4	**5**	**Totals:**
1	84(91)	18(11)	102
2	72(72)	9(9)	81
3	37(35)	2(4)	39
4	90(85)	6(11)	96
Totals:	283	35	318

Applying the formula:

$$\chi^2 = \sum [(O-E)^2 / E]$$

$$= (84-91)^2/91 + (18-11)^2/11$$
$$+ (72-72)^2/72 + (9-9)^2/9$$
$$+ (37-35)^2/35 + (2-4)^2/4$$
$$+ (90-85)^2/85 + (6-11)^2/11$$
$$= 0.538 + 4.454 + 0 + 0 + 0.114 + 1 + 0.294 + 2.27$$
$$= 8.67$$

D.f. = (No. of rows - l)(No. of columns - 1) = 3 * 1 = 3.

From a table of the χ^2 distribution (e.g. Table VII on p.336 of Hoel, 1976), the critical values for χ^2 with 3 degrees of freedom are 7.81 at the 0.05 level and 11.34 at the 0.01; so in the present case $0.01 < p < 0.05$. I.e. our result is significant at the 1 in 20 level.

Conclusion from the test: there is a relationship between socio-economic status of the parents and physical development of the children, and we can reject the null hypothesis of no relationship.

2.2.6 Explanation for the two differing conclusions:

Reducing the number of cells in the second version of the test reduces the number of degrees of freedom, and hence the size of the χ^2 value required to achieve significance. In other words we have given ourselves fewer opportunities to pick up discrepancies between observed and expected frequencies, and so the chances of such deviations arising fortuitously are correspondingly diminished.

We may also think of the difference between the two tests as follows: by amalgamating the first four categories of physical development we have counteracted the effect mentioned above, namely that the overrepresentation of Class 1 children in the highest category of physical development and the underrepresentation of Class 4 children in that same category was swamped by the lack of any deviation from expected values elsewhere in the table.

2.2.7 The nature of the relationship observed in the second test:

There appears to be a positive association between physical development and parental socio-economic status. This effect is mainly expressed at the upper and lower extremes of the range, i.e. in Classes 1 and 4. The effect is only apparent when physical development has reached its fullest potential.

N.B. (1) To fulfil the instructions for the second test we have had to violate the requirement of >5 observations in each cell, thus illustrating that this is a practical desideratum, rather than an absolute numerical prerequisite. One might comment on the fact that this requirement had been violated in answering the question, pointing out that any such violation is liable to reduce the validity of the test.

N.B. (2) The format of this question - asking you to do one chi-square test, then collapse some cells and do another to compare with the first - is very characteristic of the chi-square questions that have appeared in the Oxford Psychology Prelims Statistics paper of recent years.

3. Summary of some key points about the chi-square test:

- **Suitable for categorical data:**

  I.e. data that can be derived from a merely nominal measure (e.g. gender), though ordinal measures are also suitable.

- **Key concept:**

  'contingency table'

- **Practical desideratum:**

  at least 5 observations per cell

- **Two sorts of application:**

  1. where expected values for cells are known (eg 'goodness-of-fit' tests)

  2. where expected values are not known (and therefore have to be worked out from the observed data, as in the example above)

- **Yatcs's Correction:**

  Subtract 0.5 from each of the '(OE)' terms, before squaring, on the top line of the formula (see below, Section 4: 'Steps').

  N.B. Yates's correction only applies when (a) dealing with a two-by-two table, and (b) when the numbers are small. If in doubt, apply the formula with and without the correction and quote both results, commenting on any difference. (The χ^2 value after the correction should always be smaller.)

4. Summary of steps in a chi-square test

- **Where expected values are not known:**

1. Draw up a contingency table of the observed values.
2. (a) Compute the column totals. (b) Compute the row totals
3. Assume the null hypothesis of no association between the two variables and work out the expected value for each cell under this hypothesis from the row and column totals. (This is done by applying the formula: Row total/grand total * column total.)
4. Compute the value of chi-square from the formula: $\chi^2 = \sum[(O-E)^2 / E]$.
5. Work out the degrees of freedom: d.f. = (No. of rows - l) * (No. of columns - 1).
6. Look up the relevant p-value in a table of the χ^2 distribution (e.g. in Hoel, 1976, p. 336).

- **To test goodness-of-fit against known expected values:**

1. Draw up a table with the following rows:

   (a) The possible values of x (the variable we are interested in)

   (b) The observed frequency for each value (O)

   (c) The expected frequency for each value (E)

   (d) $(O-E)^2$ for each value

   (e) $(O-E)^2 / E$ for each value

2. Apply the chi-square formula ($\chi^2 = \sum[(O-E)^2 / E]$) as for a test where Es were not known.

5. Recommended Reading:

- Hoel, Paul G. (1976). *Elementary Statistics* (4th edition). New York: Wiley. Chapter 10.

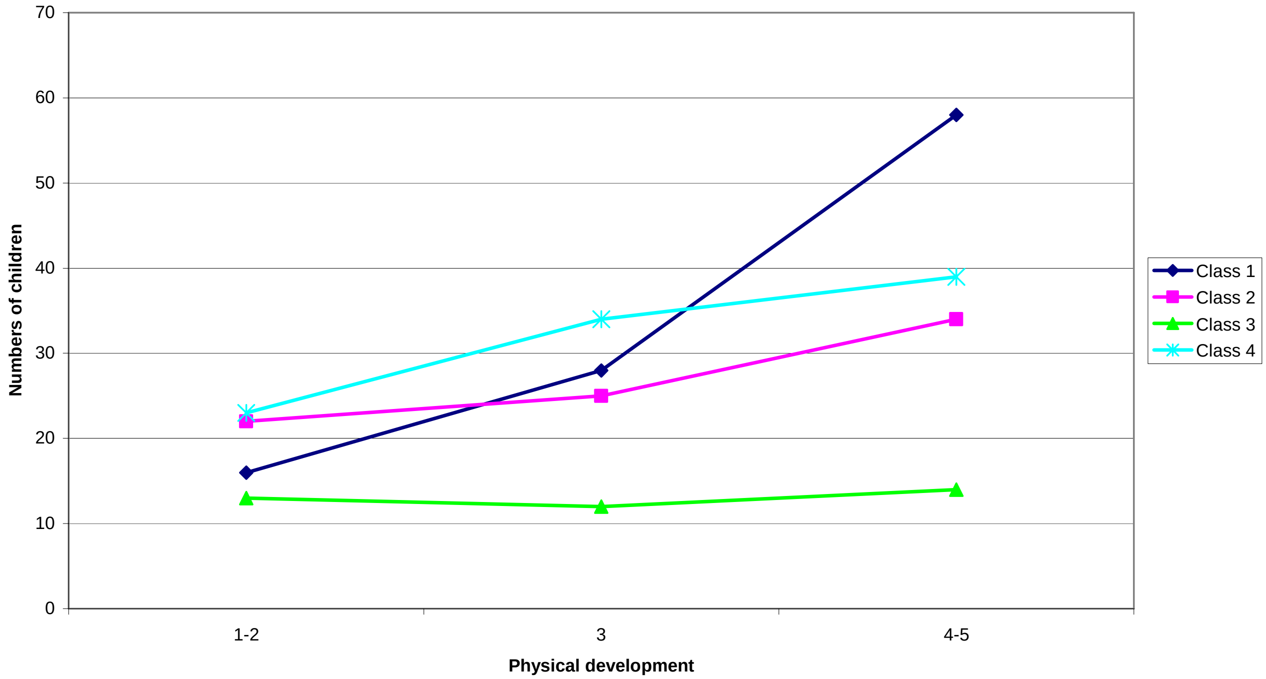
- **OR**: Spiegel, Murray R. (1992). *Schaum's Outline of Theory and Problems of Statistics* (2nd edition). New York: McGraw-Hill. Chapter 12.

- **OR**: Howell, David C. (1997). *Statistical Methods for Psychology* (4th edition). London: Duxbury Press. Chapter 6.

I recommend Hoel (1976) for this topic. It is quite short, but covers all you need to know about the chi-square test for the Oxford Prelims statistics course. Spiegel also covers the ground, but is rather condensed as usual, with a minimum of discursive discussion. Howell contains considerably more than you need to know. Hays[2] is not recommended for this topic; it is too theoretical.

[2] Hays, William L. (1994). *Statistics* (5th edition). Orlando, Florida: Harcourt Brace. This is a book sometimes recommended in connection with the Oxford first-year statistics course.

Appendix A - Plot of Class and Development



Appendix B

How to recognise what type of test to do

Type of measure	Nature of data	Examples	Suitable tests
Nominal	Discontinuous/categorical, having no regard for order	Gender Eye-colour	Non-parametric Chi-square
Ordinal	Discontinuous, but rank ordered	Social class Extraversion	Non-parametric Chi-square Parametric if plenty of ranks and normally distributed data
Interval	Truly quantitative and continuous, so intervals all equal; but zero point arbitrary	Fahrenheit Centigrade	Parametric
Ratio	Truly quantitative and continuous; intervals equal, and zero point not arbitrary, so, for example, a doubling of the measure obtained implies a doubling of the underlying quantity measured	Kelvin Age Weight Height	Parametric

Some other publications from **OXFORD FORUM** are shown below.
For further information about each title please click on the image,
which links to the relevant page on **Amazon UK**.

Alternatively they can be obtained from:

Blackwells Bookshop, Oxford (tel. 01865 792792)

or direct from the distributors:

Book Systems Plus, Linton (tel. 01223 894870)