Chi-Squared - Summary of Important Ideas

(3 pages; 23/3/20)

See separate note for further details.

(A) Contingency Tables

(1)

O_i

	voting				
	Con	Lab	Lib	oth	
Male	45	37	12	6	100
Female	24	40	12	9	85
	69	77	24	15	185

Alternative expressions for E_1:

$\frac{100}{185} \times 69$

$\frac{69}{185} \times 100$

$185 \times \frac{69}{185} \times \frac{100}{185}$

(2) Reasons for $(O_i - E_i)^2$ term:

(i) So that positive and negative $O_i - E_i$ don't cancel out.

(ii) So that larger differences are given a bigger weighting (after division by E_i).

(3) When there are just two rows or two columns, record the $(O_i - E_i)^2$ - as a check, as this value will be the same for the two cells in the particular column (row).

(4) The χ^2 distribution has been found to be a good model for the value of $X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$, when there is no association between the factors.

(5) Derivation of $(m-1)(n-1)$ degrees of freedom:

Deduct 1 from mn (the number of cells) for each constraint.

The deductions are as follows:

grand total fixed: -1

row totals fixed $-(m-1)$ [last one covered by grand total]

column totals fixed $-(n-1)$ [last one covered by grand total]

This gives $\nu = mn - 1 - (m-1) - (n-1)$

$= mn - m - n + 1 = (m-1)(n-1)$

(6) A very low value of X^2 could be regarded as suspicious (ie the data might have been rigged). In theory, this could be tested for using the left-hand critical values from the table.

(7) Association does not imply "cause & effect".

(8) Once H_1 has been accepted, values of $\frac{(O_i-E_i)^2}{E_i}$ for particular cells may suggest a theory as to the nature of an association between factors.

(B) Goodness of Fit Test

(1) In general, the null and alternative hypotheses would be:

H_0: the data are drawn from the model population

H_1: this is not the case

(2) A deduction of 1 has to be made, when calculating the degree of freedom, if a parameter has to be estimated from the given data.

(3) If a model has been based on data, then fresh data are needed when carrying out the Goodness of Fit test.