

Statistics 2

Samples and hypothesis testing

Section 2: Contingency tables

Solutions to Exercise 3C

10. (i) H_0 : 'origin' and 'main purpose of visit' are independent
 H_1 : 'origin' and 'main purpose of visit' are not independent.

(ii) Observed frequencies

	H.R.	G.W.	R.M.	O.A.	TOTAL
A	30	27	16	8	81
B	15	28	13	7	63
other	8	8	9	11	36
TOTAL	53	63	38	26	180

Expected frequencies

	H.R.	G.W.	R.M.	O.A.	TOTAL
A	23.85	28.35	17.1	11.7	81
B	18.55	22.05	13.3	9.1	63
other	10.6	12.6	7.6	5.2	36
TOTAL	53	63	38	26	180

contribution to χ^2

1.5858	0.0643	0.0708	1.1701
0.6794	1.6056	0.0068	0.4846
0.6377	1.6794	0.2579	6.4692

$$\text{so } \chi^2 = 14.712$$

Degrees of freedom $V = (3-1)(4-1) = 6$

For 5% significance level the critical value is 12.59

As $14.712 > 12.59$ the result is significant, so H_0 is rejected, i.e. there is an association between 'origin' and 'main purpose of visit'.

- (iii) The key feature is that people from 'other areas' bring far more old appliances than would be expected if there were no association.

- (iv) If the variables are independent then

$$P(\text{in cell } (i, j)) = P(\text{in row } i) \times P(\text{in row } j)$$

$$= \frac{n_1}{n} \times \frac{n_2}{n}$$
 so expected frequency $= n \left(\frac{n_1 n_2}{n^2} \right) = \frac{n_1 n_2}{n}$