

χ^2 Contingency Table - Combining Categories.

Example 1 : A random sample of 62 pupils at a certain school was asked if they used the Milk Bar at lunch times. Their replies were:

f_o	Boys	Girls
most days	1	2
some days	12	16
never	19	12

H_0 : Gender and Milk Bar use are not associated

H_1 : they are associated.

Assume H_0 to be true

$\alpha = 5\%$, 1 tail test

Under H_0 , we generate expected frequencies:

f_e	Boys	Girls
most	1.55	1.45
some	14.45	13.55
never	16	15

now four out of six are ≥ 5 which is less than 80%.

\therefore we combine rows or columns.... we'll combine rows 1+2 here:

f_o	Boys	Girls
most/some	13	18
never	19	12

and we regenerate expected frequencies:

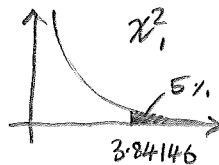
f_e	boys	girls
most/some	16	15
never	16	15

this now has all $f_e > 5$, so we continue with test on this 2×2 table

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 2.325$$

we have $(2-1) \times (2-1) = 1 \times 1 = 1$ degree of freedom

χ^2 has a χ^2_1 distribution



as $2.325 < 3.84146$, we are not in critical region

\therefore we have no evidence to reject H_0

We conclude that this sample suggests no association between gender and use of the milk bar.

Example 2

A random sample of 115 students leaving schools in England in 1979-80 with at least one pass in the GCE A-Level examinations was classified by what they did after leaving school and by their number of A level passes.

f_o	Number A-Level Passes			
	1	2	3	4 or more
Degree course	1	10	29	16
Other course	7	7	4	1
Employment	14	12	11	3

H_0 : no association between post-school destination and number of A-levels

H_1 : there is an association

Assume H_0 to be true

$\alpha = 5\%$, 1 tail test

we generate expected frequencies:

f_e	1	2	3	4+
Degree	10.7	14.1	21.4	9.7
Other	3.6	4.8	7.3	3.3
Emplby	7.7	10.1	15.3	6.9

now 9 out of 12 ($= \frac{3}{4}$) are ≥ 5 which is less than the threshold of 80% being ≥ 5 , so

we combine categories.

we have a choice of how

note: it made more sense to combine columns, rather than rows, due to the context

f_o	1-2	3	4+
Degree	11	29	16
Other	14	4	1
Employ	26	11	3



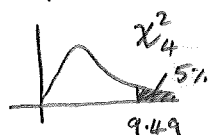
f_e	1-2	3	4+
Degree	24.8	21.4	9.7
Other	8.4	7.3	3.3
Employ	17.7	15.3	6.9

no $f_e < 1$ and $\frac{8}{9} = 89\%$ are ≥ 5



$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 28.48$$

χ^2 has χ^2_4 Distribution



we are in critical region, etc, etc.

f_o	1	2	3+
Degree	1	10	45
Other	7	7	5
Employ	14	12	14



f_e	1	2	3+
Degree	10.7	14.1	31.2
Other	3.6	4.8	10.6
Employ	7.7	10.1	22.3

no $f_e < 1$, but $\frac{7}{9} = 78\%$ are ≥ 5



combine rows/columns again!