Degrees of freedom

χ^2 goodness of fit

$$df = \underbrace{\#\text{ categories} - 1}_{} - \#\text{ parameters}$$

because we know the total sample size, n, which leaves only # categories − 1 freely determined cells.

	category 1	cat. 2	...	cat c	total
counts	y_1	y_2		y_c	n

$$y_c = n - (y_1 + y_2 + \cdots + y_{c-1})$$

Degrees of freedom

χ^2 test for homogeneity

- for each population we have
  \# categories -1
  (same justification as for goodness of fit, i.e. we know n for each pop.).

- also need to estimate parameters,

  → need a p_i for each category:

$$\hat{p}_1 = \frac{y_{\cdot 1}}{n} \qquad \hat{p}_2 = \frac{y_{\cdot 2}}{n} \quad \cdots \quad \hat{p}_c = \frac{y_{\cdot c}}{n}$$

BUT $\hat{p}_c = 1 - (\hat{p}_1 + \hat{p}_2 + \cdots + \hat{p}_{c-1})$

so really only estimate $c-1$ parameters

- put it all together:

  r populations and c categories:

$$\underbrace{r(c-1)}_{r \times \text{ \# freely determined cells in each pop.}} - \overbrace{(c-1)}^{\text{\# parameters}} = (r-1)(c-1)$$

Degrees of freedom

χ^2 test for independence

df = # freely determined cells
— # of parameters estimated

* # freely determined cells

$= r \times c - 1$ (conditional on knowing total sample size).

* # parameters needing to be estimated

- marginal probabilities (row and column)

rows: $\hat{p}_{1\cdot}, \hat{p}_{2\cdot}, \cdots, \hat{p}_{r\cdot}$ BUT $\hat{p}_{r\cdot} = 1 - (\hat{p}_{1\cdot} + \cdots + \hat{p}_{r-1\cdot})$

cols: $\hat{p}_{\cdot 1}, \hat{p}_{\cdot 2}, \cdots \hat{p}_{\cdot c}$ BUT $\hat{p}_{\cdot c} = 1 - (\hat{p}_{\cdot 1} + \cdots + \hat{p}_{\cdot c-1})$

so we only need to estimate

$(r-1) + (c-1)$ probabilities.

$\therefore$ $df = (rc - 1) - (r-1) - (c-1) = (r-1)(c-1)$