



Chi-square and Fisher's exact tests

**Dr. M.M. Mweu,
Level II MBChB Biostatistics,
5 December, 2016**

Chi-square test

- Is a *significance test of association*
- Association means that knowing the value of one variable tells us something about the value of the other variable
- Two variables are said to *be associated* if the *distribution of one variable varies according to the value of the other variable*

Example

Water source	Ethnic group			
	A	B	C	Total
Well	37 (57%)	18 (33%)	24 (50%)	79
Spring	14 (22%)	17 (31%)	14 (29%)	45
River/Stream	12 (19%)	19 (35%)	10 (21%)	41
Total	63 (100%)	54 (100%)	48 (100%)	165

- In the example, ethnic group (predictor) & water source (outcome) are *associated if the distribution of water source varies between the ethnic groups*

Chi-square test

Hypothesis:

H_0 = No association i. e. distribution of water usage is same regardless of ethnic group

H_a = There's an association i. e. distribution of water usage differs across the ethnic groups

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\text{Where } E = \frac{\text{Row total} \times \text{Column total}}{\text{Overall total}}$$

$$\chi^2 =$$

$$\begin{aligned} & \frac{\left(37 - \frac{79 \times 63}{165}\right)^2}{\frac{79 \times 63}{165}} + \frac{\left(18 - \frac{79 \times 54}{165}\right)^2}{\frac{79 \times 54}{165}} + \frac{\left(24 - \frac{79 \times 48}{165}\right)^2}{\frac{79 \times 48}{165}} + \frac{\left(14 - \frac{45 \times 63}{165}\right)^2}{\frac{45 \times 63}{165}} + \\ & \frac{\left(17 - \frac{45 \times 54}{165}\right)^2}{\frac{45 \times 54}{165}} + \frac{\left(14 - \frac{45 \times 48}{165}\right)^2}{\frac{45 \times 48}{165}} + \frac{\left(12 - \frac{41 \times 63}{165}\right)^2}{\frac{41 \times 63}{165}} + \frac{\left(19 - \frac{41 \times 54}{165}\right)^2}{\frac{41 \times 54}{165}} + \frac{\left(10 - \frac{41 \times 48}{165}\right)^2}{\frac{41 \times 48}{165}} = \mathbf{8.46} \end{aligned}$$

Chi-square test

$$df = (r - 1) \times (c - 1)$$

$$df = (3 - 1) \times (3 - 1) = 4$$

$$\chi^2_{0.05} = 9.49$$

$8.46 < 9.49$ so accept H_0 that there's **no association between water usage and ethnicity**

Uses of χ^2 :

- If total sample size $n > 40$ and smallest expected value ≥ 5
- Otherwise use **Fisher's exact test**

Fisher's exact test

- As Chi-square, it's a significance test of association
- Calculates the *exact probability* of observing cell values in a 2 by 2 table given that the *row & margin totals are fixed*

Example

	Disease +	Disease -	
Factor +	4 (a)	1 (b)	5 (a+b)
Factor -	4 (c)	6 (d)	10 (c+d)
	8 (a+c)	7 (b+d)	15 (n)

$$p = \frac{\binom{a+c}{a} \binom{b+d}{b}}{\binom{n}{a+b}} = \frac{\binom{8}{4} \binom{7}{1}}{\binom{15}{5}} = \frac{70 \times 7}{3003} = \mathbf{0.163}$$

Hence **no association** betwn factor and disease