

The χ^2 Test¹

Given the following:

- s : Number of samples.
- r : Number of variables.
- x_{ij} : Relative frequency of variable j in sample i .
- a_{ij} : Count of variable j in sample i .
- n_i : Size of sample i .
- N : Total size, $N = \sum_i n_i$.
- p_j^* : Estimated expected of variable j , $p_j^* = \frac{1}{N} \sum_i x_{ij}$.
- v : Degrees of freedom, $v = (s-1)(r-1)$.

Sample	Relative frequencies		
	In general	$r = 2$	Totals
1	$x_{11}, x_{12}, \dots, x_{1r}$	$x_{11}, n_1 - x_{11}$	n_1
2	$x_{21}, x_{22}, \dots, x_{2r}$	$x_{21}, n_2 - x_{21}$	n_2
\vdots	\vdots	\vdots	\vdots
s	$x_{s1}, x_{s2}, \dots, x_{sr}$	$x_{s1}, n_s - x_{s1}$	n_s
Totals		$X_1, N - X_1$	N

Calculate Q , such that

$$Q = \sum_{i=1}^s \sum_{j=1}^r \frac{[x_{ij} - n_i p_j^*]^2}{n_i p_j^*}$$

Homogenitetstestet. Om samma uppsättning p_j förekommer i alla serier anses dessa vara *homogena*. Förkasta hypotesen om homogenitet om

$$Q > \chi_{\alpha, v}^2.$$

In the present case $r = 2$, so $x_{i2} = 1 - x_{i1}$ and $p_2^* = 1 - p_1^*$:

$$Q = \sum_{i=1}^s \frac{[x_{i1} - n_i p_1^*]^2}{n_i p_1^*} + \frac{[1 - x_{i1} - n_i(1 - p_1^*)]^2}{n_i(1 - p_1^*)}.$$

If we prefer using absolutes instead of relative frequencies, we get $a_{i1} = n_i \cdot x_{i1}$ and $a_{i2} = n_i(1 - x_{i1}) = n_i - a_{i1}$:

$$\begin{aligned} Q &= \sum_{i=1}^s \frac{[a_{i1} - n_i p_1^*]^2}{n_i p_1^*} + \frac{[n_i - a_{i1} - n_i(1 - p_1^*)]^2}{n_i(1 - p_1^*)} \\ &= \sum_{i=1}^s \frac{[a_{i1} - n_i p_1^*]^2}{n_i p_1^*} + \frac{[a_{i1} - n_i p_1^*]^2}{n_i(1 - p_1^*)} \\ &= \sum_{i=1}^s \frac{(1 - p_1^*)[a_{i1} - n_i p_1^*]^2 + p_1^*[a_{i1} - n_i p_1^*]^2}{n_i p_1^*(1 - p_1^*)} = \sum_{i=1}^s \frac{[a_{i1} - n_i p_1^*]^2}{n_i p_1^*(1 - p_1^*)}. \end{aligned}$$

Finally, we calculate the probability corresponding to Q , p :

$$\begin{aligned} p &= 1 - F_{\chi^2}(x) = 1 - \frac{1}{2^{v/2} \Gamma(\frac{v}{2})} \int_0^x e^{-\frac{t}{2}} t^{\frac{v}{2}-1} dt \\ &= \{t = 2u\} = 1 - \frac{1}{2^{v/2} \Gamma(\frac{v}{2})} \int_0^{\frac{x}{2}} e^{-u} u^{\frac{v}{2}-1} 2^{\frac{v}{2}-1} 2 du \\ &= 1 - \frac{1}{\Gamma(\frac{v}{2})} \int_0^{\frac{x}{2}} e^{-u} u^{\frac{v}{2}-1} du = 1 - P\left(\frac{v}{2}, \frac{x}{2}\right), \end{aligned}$$

where $P(a, x)$ is the incomplete gamma function.

¹ Daniel Zwillinger and Stephen Koska. *CRC Standard Probability and Statistics Tables and Formulae*, pp 233–234. Chapman & Hall/CRC, 2000.