Chi-squared Tests:

1. Goodness-of-fit Tests:

Consider a variable which can be broken down into k categories from a population**.** If we take a random sample of size n with the different k categories and then for each category we obtain the observed frequencies denoted by $n_1, \ldots, n_k$ and the expected frequencies denoted by $e_1, \ldots, e_k$. Our aim is to compare how the expected frequencies under the hypothesis match or fit the observed frequencies.

Category	**1**	**2**	**…**	**k**	**Total**
Observed	n_1	n_2	…	n_k	n
Expected	$e_1 = np_{10}$	$e_2 = np_{20}$	…	$e_k = np_{k0}$	n

I. Hypotheses:

$\boldsymbol{H_0}: p_1 = p_{10}, p_2 = p_{20}, \ldots, p_k = p_{k0}$

$\boldsymbol{H_a}: at\ least\ one\ p_i\ does\ not\ equal\ p_{i0}$

II. Test Statistic:

$$\chi^2 = \sum_{i=1}^{k} \frac{(\textbf{observed} - \textbf{expected})^2}{\textbf{expected}}$$

$$= \sum_{i=1}^{k} \frac{(\boldsymbol{n_i} - \boldsymbol{e_i})^2}{\boldsymbol{e_i}} = \sum_{i=1}^{k} \frac{(\boldsymbol{n_i} - \boldsymbol{np_{i0}})^2}{\boldsymbol{np_{i0}}}$$

III. Rejection Region:

Reject H_0 if:

$$\chi^2 > \chi^2_{\alpha,k-1}$$

2. Testing of Homogeneity:

(I populations + J categories)

		Category				
		1	**2**	…	**J**	**Total**
population	**1**	n_{11}	n_{12}	…	n_{1J}	n_1
	2	n_{21}	n_{22}	…	n_{2J}	n_2
	⋮	⋮	⋮	⋱	⋮	⋮
	I	n_{I1}	n_{I2}	…	n_{Ij}	n_I
	Total	$n_{.1}$	$n_{.2}$	…	$n_{.J}$	n

I. Hypotheses:

$\boldsymbol{H_0}: p_{1j} = p_{2j} = \cdots = p_{Ij}, j = 1,2,\dots,J. \equiv (The\ I\ populations\ are\ homogeneous)$

$\boldsymbol{H_a}: The\ I\ populations\ are\ not\ homogeneo$

II. Test Statistic:

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(\text{observed} - \text{estimated expected})^2}{\text{estimated expected}}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(n_{ij} - \hat{e}_{ij}\right)^2}{\hat{e}_{ij}}$$

$$\boldsymbol{Where}, \qquad \hat{e}_{ij} = \frac{n_i\ n_{.j}}{n}$$

III. Rejection Region:

Reject H_0 if:

$$\chi^2 > \chi^2_{\alpha,(I-1)(J-1)}$$

3. Testing for Independence

		Category j				
		1	**2**	**…**	**J**	**Total**
Category i	**1**	$\boldsymbol{n_{11}}$	$\boldsymbol{n_{12}}$	…	$\boldsymbol{n_{1J}}$	$\boldsymbol{n_{1.}}$
	2	$\boldsymbol{n_{21}}$	$\boldsymbol{n_{22}}$	…	$\boldsymbol{n_{2J}}$	$\boldsymbol{n_{2.}}$
	⋮	⋮	⋮	⋱	⋮	⋮
	I	$\boldsymbol{n_{I1}}$	$\boldsymbol{n_{I2}}$	…	$\boldsymbol{n_{Ij}}$	$\boldsymbol{n_{I.}}$
	Total	$\boldsymbol{n_{.1}}$	$\boldsymbol{n_{.2}}$	…	$\boldsymbol{n_{.J}}$	$\boldsymbol{n}$

I. Hypotheses:

$\boldsymbol{H_0}: p_{ij} = p_{i.}\,.\,\,p_{.j} \equiv (The\ two\ variables\ are\ independent)$

$\boldsymbol{H_a}: The\ two\ variables\ are\ not\ independent\ (related)$

II. Test Statistic:

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(\textbf{observed} - \textbf{estimated expected})^2}{\textbf{estimated expected}}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(\mathbf{n_{ij}} - \mathbf{\hat{e}_{ij}})^2}{\mathbf{\hat{e}_{ij}}}$$

$$\boldsymbol{Where}, \qquad \boldsymbol{\hat{e}_{ij}} = \frac{\boldsymbol{n_{i.}\ n_{.j}}}{\boldsymbol{n}}$$

III. Rejection Region:

Reject H_0 if:

$$\chi^2 > \chi^2_{\alpha,(I-1)(J-1)}$$