Independence

Statistically, independence is equivalent the property that all joint Probabilities equal the product of their marginal probabilities,

$$\pi_{ij} = (\pi_{i+}) \cdot (\pi_{+j})$$

Chi-Squared test of Independence

1. Formulation of Hypothesis

$$H_0: \pi_{ij} = (\pi_{i+})(\pi_{+j})$$
$$H_1: \pi_{ij} \neq (\pi_{i+})(\pi_{+j})$$

or

H_0: The two variables of classification are independent (there is no association between two variables of classification)

H_1: The two variables of classification are not independent (there is association between two variables of classification).

Scanned with CamScanner

② level of Significance.

Choose α.

commonly used $\alpha = 0.05, 0.01$

③ test statistic:

$$\chi^2 = \sum \left(\frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \right) \Rightarrow \sum \left(\frac{(O_{ij} - e_{ij})^2}{e_{ij}} \right)$$

~~which, If H_0 is true,~~ has an approximate chi-square distribution with $(I-1)(J-1)$ degree of freedom

where

n_{ij} = frequency counts / observed frequencies

$\hat{\mu}_{ij}$ = estimated expected frequencies

The expected frequency in any cell is equal to the product of the marginal totals common to that cell divided by the total number of observations.

$$\hat{\mu}_{ij} = \frac{(n_{i+})(n_{+j})}{n}$$

④ Calculation:

Compute the expected frequencies for each cell and calculate the value of χ^2 and also the degree of freedom.

⑤ Critical Region / Rejection Region:-

If $\chi^2 \geq \chi^2_{\alpha, [(I-1)(J-1)]}$ then reject

H_o otherwise do not reject H_o.

⑥ Decision:

⇒ We use chi-square test when two nominal variables having two categories are to be compared.

Expected count ≥ 5 in $> 80\%$ of cells.