14.1 Goodness-of-fit Test When Category probability are completely specified.

1. Theorm:

Provided that $np_i \geq 5$ for $i \in (1, k)$, $\sum N_i = n$

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} = \sum_{\text{All cells}} \frac{(\text{observed} - \text{expected})^2}{\text{Expected}}$$

has approx. chi-squared distribution $k-1$ df.

$np_{i0} \geq 5$

2. Test.

Provided : $np_{i0} \geq 5$ for $\forall i$

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

H_a: at least one p_i does not equal p_{i0}.

Test statistic :

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

P-value: Area under χ^2_{k-1}, to the right of χ^2.

14.3. Two-way Contingency Table.

1. Test for homogeneity.

a) n_{ij} = number of individuals in the ith sample who fall into category j.

$$n_{i.} = \sum_{i=1}^{I} n_{ij}$$

= total number of individuals in j.

b) H_0: $p_{1j} = p_{2j} = p_{3j} = \cdots = p_{Ij}$.

H_a: it is not homogeneous with respect to the categories.

Test Statistic:

$$\chi^2 = \sum_{\text{All Cell}} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

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

$$e_{ij} = n_{i} \cdot \frac{n_{\cdot j}}{n} = \frac{(\text{ith row total})\ (\text{jth row total})}{n}.$$

Assumption: $e_{ij} \geq 5$