16th Feb '2024

# Recap :-> LRTs for Composite Hypothesis.

The remaining sections of Ch-4 refer to the three LRT's for frequency tabled data ("Chi-square tests" in general)

↳ Goodness of fit (G.O.F) tests.

↳ Tests of Homogeneity.

↳ Tests for Independence.

o——— * ———o

## Section 4.5 : Goodness of fit (G.O.F) tests

A G.O.F test is performed with one variable and a single population

| Events | $A_1$ | $A_2$ | ...... | $A_a$ | total |
|---|---|---|---|---|---|
| freq | $x_1$ | $x_2$ | ...... | $x_a$ | $n$ |

$H_0$ : The model is a good fit.

$$D = 2\left[l(\underset{\sim}{\hat{\theta}}) - l(\underset{\sim}{\tilde{\theta}}_0)\right] \approx \chi^2_{(k-q)}$$

Basic Model : $(x_1, x_2, \dots, x_a) \sim \text{Multinomial}(n; p_1, p_2, \dots, p_a)$,

$$\sum_{i=1}^{a} p_i = 1.$$

↳ $k = a - 1$.

$\hat{p}_i = x_i/n$ , for $i = 1(1)a$

$$l(\hat{p}_1, \dots, \hat{p}_a) = \sum_{i=1}^{a} x_i \ln(\hat{p}_i) = \sum_{i=1}^{a} x_i \ln\left(\frac{x_i}{n}\right)$$

we know this; let's use this w/o proof.

Hypothesized Model: Something pertaining to Simplifying $p_1, p_2, \dots, p_a$;

$$0 \le q \le a-2.$$

$\left[\text{If, } q = a-1,\ D \sim \chi^2_{\underbrace{[\overset{k}{(a-1)} - \overset{q}{(a-1)}]}_{=0}} \times \qquad \text{If, } q = a;\ D \sim \chi^2_{\underbrace{[\overset{k}{(a-1)} - \overset{q}{a}]}_{=-1}} \times \right]$

$\tilde{p}_i$'s either stated or solved for as MLE's under $H_0$.

$$l_H(\tilde{p}_1, \dots, \tilde{p}_a) = \sum_{i=1}^{a} x_i \ln(\tilde{p}_i)$$

$$D = 2\left[\sum_{i=1}^{a} x_i \ln\left(\frac{x_i}{n}\right) - \sum_{i=1}^{a} x_i \ln(\tilde{p}_i)\right]$$

$$= 2\left[\sum_{i=1}^{a}\left(x_i \ln\left(\frac{x_i}{n}\right) - x_i \ln(\tilde{p}_i)\right)\right]$$

$$= 2\sum_{i=1}^{a} x_i \left(\ln\left(\frac{x_i}{n}\right) - \ln(\tilde{p}_i)\right)$$

$$\Rightarrow \boxed{D = 2\sum_{i=1}^{a} x_i \ln\left(\frac{x_i}{n\tilde{p}_i}\right)}$$

Note: $n\tilde{p}_i$ is $E(x_i)$ under $H_0$.

## Section 4.4: Test of Homogeneity.

We consider one variable across several populations (indep).
Data in this test is provided in a contingency table.

populations

| | pop 1 | pop 2 | ....... | pop b | Total |
|---|---|---|---|---|---|
| $A_1$ | $x_{11}$ | $x_{12}$ | ⋮ | $x_{1b}$ | $x_{1\cdot}$ |
| $A_2$ | $x_{21}$ | $x_{22}$ | ⋮ | $x_{2b}$ | $x_{2\cdot}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $A_a$ | $x_{a1}$ | $x_{a2}$ | ⋮ | $x_{ab}$ | $x_{a\cdot}$ |
| Total | $n_1$ | $n_2$ | ....... | $n_b$ | $n$ |

Event

Marginal Frequencies.

Sample Sizes

$H_0$ : The population distns are homogeneous

$$D \approx \chi^2_{(K-q)}$$

$$D = 2\left[\sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij} \ln\left(\frac{x_{ij}}{n_j \tilde{p}_{ij}}\right)\right]$$

Basic Model : Joint pmf $(x_{ij})_{i,j}$ : $\prod_{j=1}^{b}$ Multinomial $(n_j, p_{1j}, p_{2j}, \ldots, p_{aj})$

$$\hat{p}_{ij} = \frac{x_{ij}}{n_j}, \text{ for } i = 1(1)a,\ j = 1(1)b$$

$$K = b(a-1)$$

Hypothesized : $P(x_{i1}) = P(x_{i2}) \cdots\cdots = P(x_{ib})$, $\forall\ i = 1(1)a$ for all rows.

$$q \le b - 1$$

$\tilde{p}_{ij}$ either specified or need to be solved for.

In the Binomial Case, as in the notes :

| | pop 1 | pop 2 | ...... | pop b | Total |
|---|---|---|---|---|---|
| $A_1$ | $x_{11}$ | $x_{12}$ | ...... | $x_{1b}$ | $x_{1\cdot}$ |
| $A_2$ | $x_{21}$ | $x_{22}$ | ...... | $x_{2b}$ | $x_{2\cdot}$ |
| Total | $n_1$ | $n_2$ | ...... | $n_b$ | $n$ |

Here, $K = b(2-1) = b$

$\left.q \le b - 1\right\}$ If testing $H_0$ : Subset of populations are Homogeneous.

Intuitively, this is like a G.O.F test.

★ Cannabis legalization Example in Sec 4.4 and news article in Supplementary Material (Ch-4)

Reading Break !!!