27th Feb '2024

Welcome back! Today we'll look at G.O.F and test of Homogeneity examples, to refresh ourselves on the "theory" from before the break.

☑ Let's revisit the Example 2.2.1 (G.O.F test)

⟵ Example 4.5.1 ⟶

$Y \sim Geom(1-\theta)$

H_0 : The assumed model is a good fit.

previously, we found $\tilde{\theta} = 1/2$. [See 19th Jan Notes]

Basic Model : $(X_1, X_2, X_3, X_4) \sim Multinomial(200; p_1, p_2, p_3, p_4)$

$\left[\sum_{i=1}^{4} p_i = 1\right]$

$$L(p_1,p_2,p_3,p_4) = p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4}$$

Then, $\ell(\underset{\sim}{p}) = \sum_{i=1}^{4} x_i \ln(p_i)$; $p_i \in (0,1)$

$\hat{p}_i = x_i/n$, for $i = 1,2,3,4$.

$k = 3$ parameters needed to be estimated.

Hypothesized Model : $p_i = p_i(\theta)$, for $i = 1,2,3,4$.

$$p_1 = 1-\theta \,,\; p_2 = \theta(1-\theta) \,,\; p_3 = \theta^2(1-\theta) \,,\; p_4 = \theta^3$$

$q = 1$ parameter.

$$L_H(\theta) = [1-\theta]^{x_1}[\theta(1-\theta)]^{x_2}[\theta^2(1-\theta)]^{x_3}[\theta^3]^{x_3}$$

Maximize this to get $\tilde{\theta}$, MLE under H_0

$\tilde{\theta} = 1/2 \Rightarrow \tilde{p}_i$; $i = 1,2,3,4$.

$$D = 2\left[\ell(\hat{\underline{p}}) - \ell(\tilde{\underline{p}})\right] = 2\sum_{i=1}^{4} x_i \ln\left(\frac{\hat{p}_i}{\tilde{p}_i}\right) \approx \chi^2_{3-1} \text{ under } H_0$$

$$d_{obs} = 7.048$$

$$p\text{-value} \simeq \mathbb{P}\left[\chi^2_{(2)} \geq 7.048\right] = 0.02948 \text{ using R}$$

$$\in (0.025, 0.05) \text{ using tables}$$

We have evidence against H_0, with p-value of $0.02948 < 0.05$

The data are not consistent with a geometric model.

☆ Exercise: Do G.O.F test with the enhanced fractured plastic model, using (λ, θ).

Test of Homogeneity : Ex 4.4.1 (Weed Legalization)

$\hookrightarrow$ Y_i = # Yes votes in i^{th} province ; $i = 1, 2, 3, 4$

$Y_i \sim \text{Bin}(\overset{n_i}{100}, p_i)$ independent.

Basic Model:

$$f(y_1, y_2, y_3, y_4 ; p_1, p_2, p_3, p_4) = \prod_{i=1}^{4} \binom{100}{y_i} p_i^{y_i} (1-p_i)^{100-y_i}$$

$$L(p_1, p_2, p_3, p_4) = \prod_{i=1}^{4} p_i^{y_i} (1-p_i)^{100-y_i}$$

Algebra, you should practice.

$$\ell(p_1, p_2, p_3, p_4) = \sum_{i=1}^{4} \left[y_i \ln(p_i) + (100-y_i)\ln(1-p_i)\right] ;$$

$$p_i \in (0,1)$$

$$\hat{p}_i = \frac{y_i}{n_i} = \frac{y_i}{100}, \text{ for } i = 1, 2, 3, 4. \quad \boxed{k = 4}$$

(1) $p_1 = p_2 = p_3 = p_4 = \boxed{p}$, unspecified $\quad \boxed{q = 1}$

$$\ell_H(p) = \sum_{i=1}^{4} \left[y_i \ln(p) + (100 - y_i) \ln(1-p)\right]$$

Maximize this to find $\tilde{p}$.

$$\left.\frac{\partial l_H}{\partial p}\right|_{\tilde{p}} = 0 \quad \text{to get } \tilde{p}.$$ Exercise !!

$$\tilde{p} = \frac{\sum_{i=1}^{4} y_i}{\sum_{i=1}^{4} n_i} = \sum_{i=1}^{4} y_i \Big/ 400 = 79/400 = 0.1975.$$

$$D = 2\left[l(\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4) - l(\tilde{p}, \tilde{p}, \tilde{p}, \tilde{p})\right]$$

$$= 2 \sum_{i=1}^{4} \left[y_i \ln\left(\frac{\hat{p}_i}{\tilde{p}}\right) + (100 - y_i) \ln\left(\frac{1-\hat{p}_i}{1-\tilde{p}}\right)\right]$$

$$d_{obs} = 10.76$$

$$D \approx \chi^2_{(3)}$$
$(4-1)$

$$p\text{-value} \approx P(\chi^2_{(3)} \geq 10.76) \approx 0.013$$

There is evidence against H_0, given p-value $= 0.013 < 0.05$

The data do not support the hypothesis that all provinces have the same proportion of yes votes.

Under H_0: $E(y_i) = n\tilde{p} = 100(0.1975) = 19.75$.