28th Feb'2024

Continue with the Test of Homogeneity

(2) $p_1 = p_2 = p_3 = p_\omega$; $p_4 = 1 - 3p$ (p unknown)

$$\boxed{\begin{array}{l} p_1 = p_2 = p_3 = p, \quad p_4 \\ p \text{ and } p_4 \text{ unknown} \end{array}}$$

$\{q = 2\}$

$$l_H(p_\omega, p_4) = \sum_{i=1}^{3}\left[y_i \ln(p_\omega) + (100 - y_i)\ln(1 - p_\omega)\right]$$
$$+ y_4 \ln(p_4) + (100 - y_4)\ln(1 - p_4)$$

$$\left.\frac{\partial l}{\partial p_\omega}\right|_{\tilde{p}_\omega, \tilde{p}_4} = \sum_{i=1}^{3} y_i / \tilde{p}_\omega - \frac{\sum_{i=1}^{3}(100 - y_i)}{1 - \tilde{p}_\omega} = 0$$

$$\Rightarrow \tilde{p}_\omega = \frac{\sum_{i=1}^{3} y_i}{300} = \frac{69}{300} = 0.23$$

$$\left.\frac{\partial l}{\partial p_4}\right|_{\tilde{p}_\omega, \tilde{p}_4} = \frac{y_4}{\tilde{p}_4} - \frac{100 - y_4}{1 - \tilde{p}_4} = 0$$

$$\Rightarrow \tilde{p}_4 = \frac{y_4}{100} = \frac{10}{100} = 0.10$$

Under H_0, $D \approx \chi^2_{(4-2)}$

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

where,

$$l(\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4) = \sum_{i=1}^{4} \left[y_i \ln(\hat{p}_i) + (n_i - y_i) \ln(1-\hat{p}_i) \right]$$

we know, $\hat{p}_i = y_i / n$

and, $$l(\tilde{p}_w, \tilde{p}_w, \tilde{p}_w, \tilde{p}_4) = \sum_{i=1}^{3} \left[y_i \ln(\tilde{p}_w) + (100 - y_i) \ln(1-\tilde{p}_w) \right]$$
$$+ y_4 \ln(\tilde{p}_4) + (100 - y_4) \ln(1-\tilde{p}_4)$$

we know, $\tilde{p}_w$ and $\tilde{p}_4$ value as well.

So, plug in data and estimates $\Rightarrow d_{obs} = 1.814$

d_{obs} from H_0 #1 : 10.76

$$p\text{-value} = \mathbb{P}\left[\chi^2_{(2)} \geq 1.814\right] = 0.404.$$

The data support the hypothesis that the Western province votes are equal but distinct from Ontarios. No evidence against H_0, given p-value $= 0.404 > 0.05$,

for, $i = 1, 2, 3$: $E(y_i) = n_i \tilde{p}_w = 23$

$E(y_4) = n_i \tilde{p}_4 = 10$

Section 4.6 – 4.8: Tests of Independence

We consider two variables for one population.

Variable B events

	B_1	B_2	......	B_b	Total	
Variable A events A_1	x_{11}	x_{12}	......	x_{1b}	r_1	"row 1"
A_2	x_{21}	x_{22}	......	x_{2b}	r_2	"row 2"
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_a	x_{a1}	x_{a2}	----	x_{ab}	r_a	"
Total	c_1	c_2	......	c_b	n	

"Columns"

H_0: Variables A and B are independent

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

$$\frac{x_{ij}}{n\tilde{p}_{ij}} = \frac{\hat{p}_{ij}}{\tilde{p}_{ij}}$$

$n\tilde{p}_{ij} = e_{ij}$ "Expected values under H_0"

Basic Model:

$$(X_{ij})_{i,j} \sim \text{Multinomial}\left(n, p_{ij} \text{ for } i = 1,2,\cdots a,\ j = 1,2,\cdots,b\right)$$

$$(X_{11}, X_{12}, \cdots, X_{1b}, X_{21}, X_{22}, \cdots, X_{2b}, \cdots)$$

$$\hat{p}_{ij} = \frac{x_{ij}}{n} \text{ for all } i, j$$

$$k = ab - 1$$