Lec 3/20

Monday, March 20, 2017 14:56

Tests for diffs in k proportions

$X_i \sim Bin(n_i, \theta_i) \qquad i = 1, 2, \dots, k$

When all n_i large

$$Z_i \sim \frac{X_i - n_i\theta_i}{\sqrt{n_i\theta_i(1-\theta_i)}} \overset{\text{approx}}{\sim} N(0,1)$$

so $$X^2 = \sum_{i=1}^{k} z_i^2 \overset{\text{approx}}{\sim} \chi^2_k$$

$H_0: \theta_1 = \theta_2 = \dots = \theta_k = \theta_0$ vs H_1: at least one $\theta_i \neq \theta_0$

Under H_0, $$X^2 = \sum_{i=1}^{k} \frac{(x_i - n_i\theta_0)^2}{n_i\theta_0(1-\theta_0)}$$

$\boxed{\text{This is not LRT}}$

CR: $X^2 \geq \chi^2_{\alpha,k}$

P-value: $\mathbb{P}(\chi^2_k \geq \text{actual value of } X^2)$.

Q: What if we don't know or want to specify θ_0?

Instead, want to test if they are the same unknown value.

$H_0: \theta_1 = \theta_2 = \dots = \theta_k \qquad H_1: \exists_{i,j}$ s.t. $\theta_i \neq \theta_j$

Idea: use $$\hat{\theta} = \frac{X_1 + \dots + X_n}{n_1 + \dots + n_1}$$ to estimate shared θ_0. (pooped estimate)

Under H_0, test statistic $$X^2 = \sum_{i=1}^{k} \frac{(X_i - n_i\hat{\theta})^2}{n_i\hat{\theta}(1-\hat{\theta})} \sim \chi^2_{\underbrace{k-1}_{\text{use } \hat{\theta} \text{ to estimate } \theta_0.}}$$

(Theorem 8.11)

approx is $X_i \sim N(n_i\theta_i, n_i\theta_i(1-\theta_i))$

There is another way to write test statistic.

	Success	Failure
sample 1	X_1	$n_1 - X_1$
sample 2	X_2	$n_2 - X_2$
$\vdots$	$\vdots$	$\vdots$
sample k	X_k	$n_k - X_k$
Under H_0, expected val:	$n_i \theta_0$	$n_i (1-\theta_0)$
Approx EV: in 2nd case	$n_i \hat{\theta}$	$n_i (1-\hat{\theta})$

Let f_{ij} = observed value in row i, column j.

e_{ij} = expected count in row i, column j. $i = 1, 2, \dots k$, $j = 1, 2$.

$(e_{i1} = n_i \theta_0,\ e_{i2} = n_i(1-\theta_0)$ or $e_{i1} = n_i \hat{\theta},\ e_{i2}\ n_i(1-\hat{\theta}))$.

So table is

f_{11}	f_{12}	e_{11}	e_{12}
f_{21}	f_{22}	$\vdots$	
$\vdots$			
f_{k1}	f_{k2}		

Statistic

$$X^2 = \sum_{i=1}^{k} \frac{(X_i - n_i \theta_0)}{X_i \theta_0 (1-\theta_0)}$$

exercise. show this 13.8 $\longrightarrow$

$$= \sum_{i=1}^{k} \sum_{j=1}^{2} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

This can be extended to more than 2 columns.

Ex (wicking) Assume 3 shifts per day. Machines shut down between shifts & cleaned.

Determine whether proportion of boxes w/ wicking is equal for all shifts.

Data:

	# wicking	# no wicking
Shift 1	9	41
" 2	11	39
" 3	14	36

$n_i = 50 \ \forall i.$

$\alpha = 0.05.$

Test $H_0: \theta_1 = \theta_2 = \cdots = \theta_k$ $\quad H_1: \theta_i \neq \theta_j$ for some i, j.

$$\hat{\theta} = \frac{9+11+14}{150} = \frac{34}{150} = \frac{17}{75} = 0.2267.$$

$e_{i1} = 50\hat{\theta} = 11.335. \quad e_{i2} = 50(1-\hat{\theta}) = 38.665 \quad \forall i.$

$$\chi^2 = \frac{(9-11.335)^2}{11.335} + \frac{(41-38.665)^2}{38.665}$$
$$+ \frac{(11-11.335)^2}{11.335} + \frac{(39-38.665)^2}{38.665}$$
$$+ \frac{(14 - '')^2}{''} + \frac{(36 - '')^2}{''} = 1.445.$$

Under H_0, $\chi^2 \sim \chi^2_2$. and $1.445 < 5.991 = \chi^2_{0.05, 2}$, fail to reject H_0.

Q: if wanted to test whether $\theta_i = 0.2 \ \forall i$,

test: $H_0: \theta_1 = \cdots \theta_k = 0.2$. Test stat: replace e_{i1}, e_{i2} w/ 10 and 40.

value $\chi^2 = 2.25$. $\quad 2.25 < 7.815 = \chi^2_{0.05, 3}$, so fail to reject H_0.