Lec 3/24

Friday, March 24, 2017 14:56

Analysis of $r \times c$ table:

Statistic: $$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \sim \chi^2_{(r-1)(c-1)}$$

$H_0: \theta_{1j} = \theta_{2j} = \cdots = \theta_{rj} \quad \forall j.$ $\quad H_1: \exists j, i, i' \text{ s.t. } \theta_{ij} \neq \theta_{i'j}$

or $H_0: \theta_{ij} = \theta_{i\cdot} \cdot \theta_{\cdot j} \quad \forall i, j$ $\quad H_1: \exists i, j \text{ s.t. } \theta_{ij} \neq \theta_{i\cdot} \cdot \theta_{\cdot j}$

Test, reject H_0 when $\chi^2 \geq \chi^2_{\alpha, (r-1)(c-1)}$

Ex: Blood types.

	A	B	AB	O	total:
FL	122	117	19	244	502
IA	1781	1351	289	3301	6722
MO	353	269	60	71	1395
total	2256	1737	368	4252	8619

Row totals: $f_{i\cdot}$; column totals: $f_{\cdot j}$; grand total: f

Test $H_0: \theta_{1j} = \theta_{2j} = \theta_{3j} \quad j = 1, 2, 3, 4$

H_1: some not equal

$$e_{ij} = \frac{f_{i\cdot} \cdot f_{\cdot j}}{f}$$

$$e_{11} = \frac{502 \cdot 2256}{8619} = 131.4$$

$$\chi^2 = \frac{(122 - 131.4)^2}{131.4} + \cdots + \frac{(f_{34} - e_{34})^2}{e_{34}} = 5.65$$

with $\alpha = 0.05$, CR is $X^2 \geq \chi^2_{0.05,6} = 12.4$

our $X^2 < 12.4$ so fail to reject H_0.

Ex: 6800 German Men

eye\hair	Brown	Black	fair	red	total
Brown	438	288	115	16	857
Grey/Green	1387	746	446	53	3132
Blue	807	189	1768	47	2811
total	2632	1223	2829	116	6800

(row totals: $f_{i\cdot}$; column totals: $f_{\cdot j}$)

H_0: $\theta_{ij} = \theta_{i\cdot} \cdot \theta_{\cdot j}$ $\forall i,j$ vs H_1: $\theta_{ij} \neq \theta_{i\cdot} \cdot \theta_{\cdot j}$ for at least one pair i,j.

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \qquad e_{11} = \frac{f_{1\cdot} \cdot f_{\cdot 1}}{f} = \frac{2632 \cdot 857}{6800} = 331.71$$

$$= \underbrace{\frac{(438 - 331.71)^2}{331.71} + \cdots}_{12 \text{ terms}} = 1024$$

Since $X^2 = 1024 > \chi^2_{0.05,6} = 12.4$ so reject H_0.

Hair & eye color not independent.

Note: Can't use this method if size is small, use when expected counts ≥ 5
if $e_{ij} < 5$ for some, combine some variables (columns/rows).

Extra credit: theoretical correction for X^2, formula for degrees of freedom in $r \times c$ table

Goodness of fit

Basic problem: Determine whether a particular data set may be looked at as a RS from a particular dist.

Ex: Coin thrown until heads.

# tosses	1	2	3	4	5	6	7	8
freq	130	60	34	12	9	1	3	1

6, 7, or 8, 5 count.

does Data fit a geo $(p = \frac{1}{2})$ dist at level $\alpha = 0.05$?