STAT475/675 TUT06

http://www.sfu.ca/~zza115/teaching.html
http://people.stat.sfu.ca/~joanh/stat475-675web.html

Zhiyang Zhou (zhiyang_zhou@sfu.ca)

2018-02-18

Recall: Hypothesis testing on independence for $I \times J$ contingency table

- $H_0 : X \perp\!\!\!\perp Y$ vs $H_1 : X \not\perp\!\!\!\perp Y$
  - or, equivalently, $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$ for all i, j vs $H_1 : \pi_{ij} \neq \pi_{i+}\pi_{+j}$ for some i, j
- Testing procedures
  - Pearson's chi-squared (χ^2-) test
  - likelihood ratio test (LRT)

Recall: χ^2-test (with purposive or multinomial sampling)

- $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$ for all i, j vs $H_1 : \pi_{ij} \neq \pi_{i+}\pi_{+j}$ for some i, j
- Test statistic:

  $$\chi^2_{\text{obs}} = \sum_{i,j} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}$$

  with $\hat{\mu}_{ij} = n_{i+}n_{+j}/n$
- Caculate p-value $= \Pr_{H_0}(\chi^2 \geq \chi^2_{\text{obs}})$ based on $\chi^2 \sim \chi^2(df)$ approximately under H_0 with $df = (I-1)(J-1)$
- Draw conclusion
  - p-value $< \alpha$: there is a strong evidence against H_0, i.e., there is a signficant association between X and Y
- Remark
  - The χ^2-approximation is good usually when $\mu_{ij} \geq 5$ for all i, j.

Recall: LRT (with purposive or multinomial sampling)

- $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$ for all i, j vs $H_1 : \pi_{ij} \neq \pi_{i+}\pi_{+j}$ for some i, j
- Test statistic:

  $$G^2_{\text{obs}} = 2\sum_{i,j} n_{ij} \ln \frac{n_{ij}}{\hat{\mu}_{ij}}$$

  with $\hat{\mu}_{ij} = n_{i+}n_{+j}/n$
- Caculate p-value $= \Pr_{H_0}(G^2 \geq G^2_{\text{obs}})$ based on $G^2 \sim \chi^2(df)$ approximately under H_0 with $df = (I-1)(J-1)$
- Draw conclusion
  - p-value $< \alpha$: there is a strong evidence against H_0, i.e., there is a signficant association between X and Y
- Remark

- The χ^2-approximation is good usually when $\mu_{ij} \geq 5$ for all i, j.

Multi-way contingency table

- Categorical variables X (of I levels), Y (of J levels), and Z (of K levels)
- Study pairs (X, Y), (X, Z) and (Y, Z) using two-way contingency tables
  - two-way partial table: all the cell counts associated with two of the three variables and with the 3rd variable fixed at a level
  - two-way marginal table: combining the two-way partial tables according to the 3rd variable

$2 \times 2 \times K$ contingency table

- X-Y conditional OR: describe X-Y association conditional on $Z = k$

$$\theta_{XY(k)} = \frac{\pi_{11k}/\pi_{12k}}{\pi_{21k}/\pi_{22k}}$$

estimated by

$$\hat{\theta}_{XY(k)} = \frac{n_{11k}/n_{12k}}{n_{21k}/n_{22k}}$$

- $\theta_{XY(k)} \equiv$ constant for all $k \Rightarrow$ "homogeneous" conditional X-Y association

---

- X-Y marginal OR: describe X-Y association conditional on $Z = k$

$$\theta_{XY} = \frac{\pi_{11+}/\pi_{12+}}{\pi_{21+}/\pi_{22+}}$$

estimated by

$$\hat{\theta}_{XY} = \frac{n_{11+}/n_{12+}}{n_{21+}/n_{22+}}$$

- $X \perp\!\!\!\perp Y | Z \Leftrightarrow \theta_{XY(k)} = 1$ for all $k \Leftrightarrow \frac{\pi_{11k}}{\pi_{1+k}} = \frac{\pi_{21k}}{\pi_{2+k}}$ for all k
- $X \perp\!\!\!\perp Y \Leftrightarrow \theta_{XY} = 1 \Leftrightarrow \frac{\pi_{11+}}{\pi_{1++}} = \frac{\pi_{21+}}{\pi_{2++}}$
- $X \perp\!\!\!\perp Y \not\Leftrightarrow X \perp\!\!\!\perp Y | Z$ in general
  - Simpson's paradox

Demo

The data included in "Arthritis" in package "vcd" is from Koch & Edwards (1988) after a double-blind clinical trial investigating a new treatment for rheumatoid arthritis.

- Check the dependence between the treatment and improvement.
