STAT475/675 TUT11

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

Zhiyang Zhou (zhiyang_zhou@sfu.ca)

2018-03-26

The lady testing tea

- 8 randomly ordered cups of tea: 4 prepared by first adding milk and 4 prepared by first adding the tea
- Select the 4 cups prepared by each method
- The lady got all eight cups correct (according to "The Lady Testing Tea" written by David Salsburg)

- Tabulated as a 2×2 contingency table
- Fixed marginals n_{1+}, n_{2+}, n_{+1}, n_{+2} and n_{++} (this happens rarely)
- As long as one element of $\{n_{11}, n_{12}, n_{21}, n_{22}\}$is known, all the others would be determined.

Fisher's exact test

- One-sided test: $H_0 : \mathrm{OR} = 1$ vs $H_1 : \mathrm{OR} > 1$ (the success rate of the speculation is higher than that of a random guess)
- Take n_{ij} as the realization (observed value) of randon variable N_{ij}
- Test statistic: n_{11}, the number of successes
- Under H_0, $N_{11} \sim$ Hypergeometric Distribution, i.e., for $m = 0, \ldots, n_{++}$,

$$\Pr_{H_0}(N_{11} = m) = \frac{\binom{n_{1+}}{m}\binom{n_{2+}}{n_{+1}-m}}{\binom{n_{++}}{n_{+1}}}$$

- p-value $= \Pr_{H_0}(N_{11} \geq n_{11})$
  - $= 1/70$ for $n_{11} = 4$ in the lady tasting tea experiment

Permutation test for independence

- $H_0 : \mathrm{OR} = 1$ vs $H_1 : \mathrm{OR} \neq 1$
- The null distrubition of

$$\chi^2_{\mathrm{obs}} = \sum \frac{(\mathrm{observed} - \mathrm{fitted})^2}{\mathrm{fitted}}$$

  is not well-approximated by χ^2-distrubition when some n_{ij} is not large enough
- Estimate the permutation distrubition by simulation
  - randomly permute the "guess"
  - repeat the permutation for $B \gg 1$ times and obtain $\chi^2_{\mathrm{obs},b}, b = 1, \ldots, B$
  - approximate the null distribution by the permutation distribution
  - p-value $\approx \#\{\chi^2_{\mathrm{obs},b} : \chi^2_{\mathrm{obs},b} \geq \chi^2_{\mathrm{obs}}\}/B$

Demo I

The Lady-Tasting-Tea experiment with different digits.

	Lady said		
Truth	Tea First	Milk First	Row Total
Tea First	10	2	12
Milk First	2	10	12
Column Total	12	12	24

Loglinear-logit connection

- loglinear(XYZ) $\Leftrightarrow$ $Y \sim$ logit(XZ) or multi-logit(XZ)
  - $\ln \mu_{ijk} = \lambda_0 + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$
  - $\ln \frac{\pi_{ijk}}{\pi_{ij'k}} = \beta_{j0} + \beta_{ji}^X + \beta_{jk}^Z + \beta_{jik}^{XZ}$
  - $\beta_{j0} = \lambda_j^Y - \lambda_{j'}^Y$
  - $\beta_{ji}^X = \lambda_{ij}^{XY} - \lambda_{ij'}^{XY}$
  - $\beta_{jk}^Z = \lambda_{jk}^{YZ} - \lambda_{j'k}^{YZ}$
  - $\beta_{jik}^{XZ} = \lambda_{ijk}^{XYZ} - \lambda_{ij'k}^{XYZ}$
- loglinear(XY, YZ, XZ) $\Leftrightarrow$ $Y \sim$ logit(X, Z) or multi-logit(X, Z)
  - $\ln \mu_{ijk} = \lambda_0 + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$
  - $\ln \frac{\pi_{ijk}}{\pi_{ij'k}} = \beta_{j0} + \beta_{ji}^X + \beta_{jk}^Z$
  - $\beta_{j0} = \lambda_j^Y - \lambda_{j'}^Y$
  - $\beta_{ji}^X = \lambda_{ij}^{XY} - \lambda_{ij'}^{XY}$
  - $\beta_{jk}^Z = \lambda_{jk}^{YZ} - \lambda_{j'k}^{YZ}$
- Given $X \perp\!\!\!\perp Z | Y$, loglinear(XY, YZ) $\Leftrightarrow$ $Y \sim$ logit(X, Z) or multi-logit(X, Z)
  - $\ln \mu_{ijk} = \lambda_0 + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}$
  - $\ln \frac{\pi_{ijk}}{\pi_{ij'k}} = \beta_{j0} + \beta_{ji}^X + \beta_{jk}^Z$
  - $\beta_{j0} = \lambda_j^Y - \lambda_{j'}^Y$
  - $\beta_{ji}^X = \lambda_{ij}^{XY} - \lambda_{ij'}^{XY}$
  - $\beta_{jk}^Z = \lambda_{jk}^{YZ} - \lambda_{j'k}^{YZ}$

Demo II

- Data "UCBAdmissions" (included in R default Package "datasets") is on applicants to graduate school at Berkeley for the six largest departments in 1973 classified by admission and sex.
  - Admit: Admitted, Rejected
  - Gender: Male, Female
  - Dept: A, B, C, D, E, F
