Tutorial 11

STAT3015/4030/7030 Generalised Linear Modelling

The Australian National University

Week 11, 2017

Overview

1 Summary

Two-way contingency table

Sampling Schemes for $R \times C$ tables

- **Poisson distribution** - a fixed amount of time (space, volume, money, etc.) is devoted to collecting a random sample from a single population, and each member of the population falls into one of the $R \times C$ cells. None of the marginal totals are known in advance.

Two-way contingency table

- **Multinomial distribution** - Similar to Poisson sampling but the total sample size is known in advance.
- **Product multinomial distribution** - row totals are fixed. Within each row, the response can fall into one of column $j, (j = 1, .., C)$. That is, each row defines a multinomial population. (Vice versa, can have column totals fixed and each column defines a multinomial population).

Test of independence/homogeneity

- fixed total $n \longrightarrow$ multinomial distribution $\longrightarrow$ test of independence
- fixed column totals $Y_{\bullet j} \longrightarrow$ product multinomial distribution $\longrightarrow$ test of homogeneity within each column
- different philosophies but identical analysis techniques

Two tests

There are two "classical" tests of independence:

To test H_0: no association between Factor 1 and Factor 2

- Likelihood ratio

$$2\sum_{i=1}^{R}\sum_{j=1}^{C} O_{ij} \log\left(\frac{O_{ij}}{E_{ij}}\right)$$

- Pearson Chi-squared

$$\sum_{i=1}^{R}\sum_{j=1}^{C} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Both have an asymptotic χ^2 distribution with $(r-1)(c-1)$ degrees of freedom

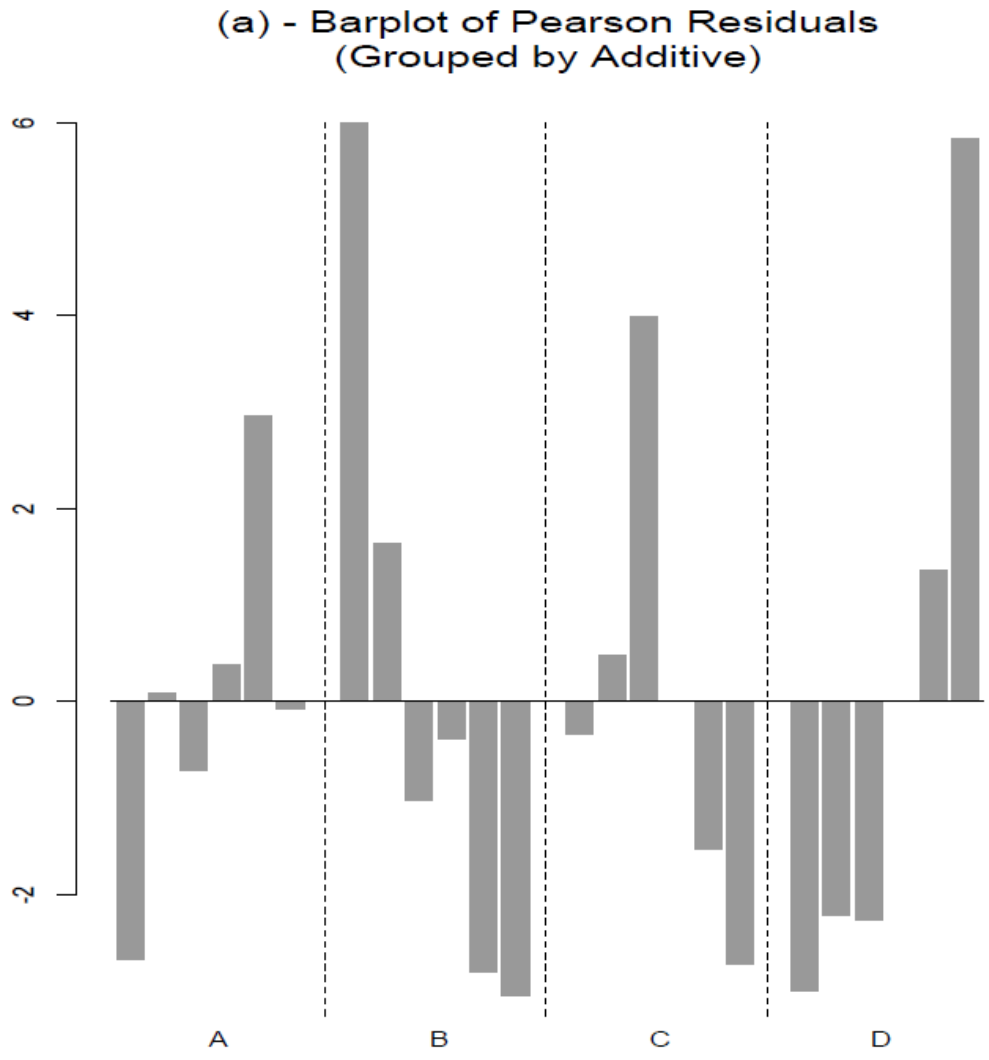
Pearson residuals

Interpretation of Pearson residuals: (Page 78 of brick)

- $(i, j)^{\text{th}}$ cell with a large positive residual $\longrightarrow O_{ij} \gg E_{ij}$, where E_{ij} is the expected value under the independence assumption.
- It indicates that individuals in the j^{th} column are more likely to be in the i^{th} row than individuals in the other columns
- vice versa

```
> rij1
             A          B          C          D
III -2.66666667  6.0000000 -0.3333333 -3.000000
 IV  0.09622504  1.6358258  0.4811252 -2.213176
  V -0.70278193 -1.0151295  3.9824309 -2.264520
 VI  0.37796447 -0.3779645  0.0000000  0.000000
VII  2.96237085 -2.8022427 -1.5212175  1.361089
VIII -0.08219949 -3.0413813 -2.7125833  5.836164
```

(Cheese Tasting Example on Page 81) III stands for "best" and VIII for "worst"



Within each section (i.e. $A, \cdots, D$) from left to right on the horizontal axis: $III, IV, \cdots, VIII$