0.1 The simulation result of the three statistic

0.1.1 The three statistic

Consider a two-way contingency table, $N = (n_{ij}, i = 1, 2, \ldots, Ij = 1, 2, \ldots, J)$, that cross classifies n individuals/units according to I row categories and J ordered (ascending) column categories.Let the observed frequency in (i,j)th cell be n_{ij},the column total frequency be $n_{+j} = \sum_i n_{ij}$,the row total frequency be $n_{i+} = \sum_j n_{ij}$. the matrix of proportions is denoted by $P = n^{(-1)}N$,the marginal relative of ith row and jth column of N are $p_{i+} = n_{i+}/n, p_{+j} = n_{+j}/n$.

For row i, consider the cumulative frequencies $Rz_{ij} = \sum_{l=1}^{j} n_{il}, l = 1, 2, \ldots, J$,and the relative cumulative proportions $Rd_j = \sum_{l=1}^{j} n_{+l}$, for column j,consider the cumulative frequencies $Cz_{ij} = \sum_{k=1}^{i} n_{kj}, k = 1, 2, \ldots, I$,and the relative cumulative proportions $Cd_i = \sum_{k=1}^{i} n_{k+}$,consider the i and j in the meaning time ,the cumulative frequencies $RCz_{ij} = \sum_{k=1}^{i}\sum_{l=1}^{j} n_{ij}$

1. $F(y|x) - F(y)$ (Nair 1978 JASA)

$$T1 = \sum_{j=1}^{J-1} \omega_j^y \Big[\sum_{i=1}^{I} n_{i+} (\frac{Rz_{ij}}{n_{i+}} - Cd_i)^2 \Big]$$

2. $F(x|y) - F(x)$

$$T2 = \sum_{i=1}^{I-1} \omega_i^x \Big[\sum_{j=1}^{J} n_{i+} (\frac{Cz_{ij}}{n_{+j}} - Rd_j)^2 \Big]$$

3. $F(x, y) - F(x)F(y)$

$$T3 = \sum_{i=1}^{I-1}\sum_{j=1}^{J-1} \omega_{ij}^{xy} n (\frac{RCz_{ij}}{n} - Rd_j Cd_j)^2$$

(Nair 1978 JASA)show the $\omega_j^x = Rd_j(1 - Rd_j)$,and The T1 has good power When $\omega_j^x = 1/J$,so we let $w_i^y = 1/I, w_{ij}^y = I \times J$ and we have the T1,T2,T3 as follow:

1. $F(y|x) - F(y)$ (Nair 1978 JASA)

$$T1 = \frac{1}{J} \sum_{j=1}^{J-1} \Big[\sum_{i=1}^{I} n_{i+} (\frac{Rz_{ij}}{n_{i+}} - Cd_i)^2 \Big]$$

2. $F(x|y) - F(x)$

$$T2 = \frac{1}{I} \sum_{i=1}^{I-1} \Big[\sum_{j=1}^{J} n_{i+} (\frac{Cz_{ij}}{n_{+j}} - Rd_j)^2 \Big]$$

3. $F(x, y) - F(x)F(y)$

$$T3 = \frac{1}{IJ} \sum_{i=1}^{I-1}\sum_{j=1}^{J-1} n (\frac{RCz_{ij}}{n} - Rd_j Cd_j)^2$$

0.1.2 The differnce from Li's paper

The association between $(Y|Z)$ and $(X|Z)$(Li 2010 JASA)

$$\text{X,Y is indenpendent} \rightarrow (Y|Z),\ (X|Z) \text{ indenpendent,but not vice versa}$$

0.1.3 Generate ordinal data(具体产生算法正确与否有待讨论)

Deriving from the (Li 2010 JASA),but don't generate the covariate Z.

The specifics of our four generating scenarios are as follows: we first generated X with five categories using th proportion odds model

$$P(X \le i) = [1 + exp(-(\alpha_i^X))]^{-1}$$

with $\alpha^X = (\alpha_1^X, \alpha_2^X, \alpha_3^X, \alpha_4^X) = (-1, 0, 1, 2)$.The Y was generated with four levels using the proportional odds model

$$P(Y \le j) = [1 + exp(-(\alpha_i^Y + \eta_1 I_{\{X=1\}} + \eta_2 I_{\{X=2\}} + \cdots + \eta_5 I_{\{X=5\}}))]^{-1}$$

with $\alpha^Y = (\alpha_1^Y, \alpha_2^Y, \alpha_3^Y) = (-1, 0, 1)$,and $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_5)$ specified as

1. $\boldsymbol{\eta} = (0, 0, 0, 0, 0)$(the null)

2. $\boldsymbol{\eta} = (-0.4, -0.2, 0, 0, 2, 0, 4)$(linear effect)

3. $\boldsymbol{\eta} = (-0.30, -0.18, 0.20, 0.22, 0.24)$(monotonic nonlinear effect)

4. $\boldsymbol{\eta} = (-0.2, 0, 0.2, 0. - 0.2)$(nonmonotonic effect)

0.1.4 The process of simulation(具体算法正确与否有待讨论)

My process(Different from the (Li 2010 JSAS)'s process)

- Simulate the *null distribution*.

  Generate Nrep=1000000 datasets from the null hypothesis ，each consisting of N=500 subjects. computer Nrep three statistics.

- Computer the *Type I error*:

  Generate Nreptest=100000 datasets from the null hypothesis ，each consisting of N subjects. computer Nreptest three statistics.

- Computer the *power*

  Generate Nreptest datasets from the alternative hypothesis，each consisting of N subjects. computer Nreptest three statistics.

Li's process

- Generate a data sets D1,consisting N=500 subjects, computing the statistic
- Relying on D1 ,generate Nemp=1000 data sets as the null distribution, computing a p-value.
- repeating step1-step2 Nrepl=1000. gain Nrepl=1000 p-value,computing the Type I error and power

0.1.5 The result

Nrep=1000000,Nreptest=100000,N=500,Time=26400.44s=7.3h

Table 1. The result of the three statistics :Type I error and power

	Simulation scenarios			
Analysis method	Null	Linear	Nonlinear	Nonmonotonic
T1	0.04948	0.75114	0.57597	0.28759
T2	0.04999	0.78369	0.47814	0.11663
T3	0.04964	0.89773	0.65047	0.17715
X linear	0.04906	0.91231	0.59056	0.07404
X categorcial	0.04974	0.76194	0.58867	0.29465
Spline	0.04926	0.85319	0.63418	0.38027