Exercises: Contingency tables[a]

Lecturer: Georgios Karagiannis georgios.karagiannis@durham.ac.uk

[a]Author: Georgios P. Karagiannis.

Exercise 1

Consider a $I \times J \times K$ contingency table, with classification variables X, Y, Z. Prove that if

1. $X$and Y are conditionally independent on Z; and
2. $X$and Z are conditionally independent on Y

then:

Y and Z are jointly independent from X

Hint: Write down the probability forms involved, and try to derive the result by using simple probability calculus.

Solution 1

I have

$$\pi_{ijk} = \frac{\pi_{i+k}\pi_{+jk}}{\pi_{++k}} \tag{1}$$

$$\pi_{ijk} = \frac{\pi_{ij+}\pi_{+jk}}{\pi_{+j+}} \tag{2}$$

By dividing those two, we get

$$\pi_{i+k}\pi_{+j+} = \pi_{ij+}\pi_{++k}$$

By summing with respect to j we get

$$\pi_{i+k} = \pi_{i++}\pi_{++k} \tag{3}$$

By substituting (3) in (1), we get

$$\pi_{ijk} = \pi_{i++}\pi_{+jk}$$

Hence Y and Z are jointly independent from X

The next exercise is from Homework 1

Exercise 2

Consider a $I \times J \times K$ contingency table, with classification variables X, Y, Z. Prove that if Y and Z are jointly independent from X,

then

1. $X$and Y are conditionally independent on Z; and
2. $X$and Z are conditionally independent on Y

Hint: Write down the probability forms involved, and try to derive the result by using simple probability calculus.

Solution 2

The next exercise is from Homework 1

Exercise 3

[1] The 1988 General Social Survey compiled by the National Opinion Research Center asked: "Do you support or oppose the following measures to deal with AIDS? (1) Have the government pay all of the health care costs of AIDS patients; (2) Develop a government information program to promote safe sex practices, such as the use of condoms. Table 1 summarizes opinions about health care costs (H) and the information program (I), classified also by the respondent's gender (G).

Gender (G)	Information Opinion (I)	Health Opinion (H)	
		Support	Oppose
Male	Support	76	160
	Oppose	6	25
Female	Support	114	181
	Oppose	11	48

Table 1: Source: 1988 General Social Survey, National Opinion Research Center.

1. Compute the marginal GH-table

[1]R-script is available to double check.

2. For the GH-table, compute the MLE of the marginal odds ratio, the confidence intervals. Interpret the result. (sig. level 5%)

3. Perform a hypothesis test, in order to test if the Information Opinion and the Health Opinion are independent at each level of the Gender. (sig. level 5%)

4. Compute the conditional IH odds ratio at each level of the Gender. Interpret the result.

Solution 3

Exercise from PJC' notes

Exercise 4

Let X and Y be discrete random variables with possible values $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ respectively and such that $p_{ij} > 0$ for all i and j. Here, p_{ij} denotes

$$\Pr(X = x_i \cap Y = y_j).$$

Let $p_{i|j}$ denote p_{ij}/p_j where p_j denotes $\sum_i p_{ij}$ so that $p_j = \Pr(Y = y_j)$ and $p_{i|j} = \Pr(X = x_i | Y = y_j)$. Also let p_i denote $\sum_j p_{ij}$.

Then the following statements are equivalent:

1. X and Y are independent;

2. $p_{i|j}$ does not depend on j for all i;

3. there exist $g_1, \ldots, g_m$ and $h_1, \ldots, h_n$ such that $p_{ij} = g_i h_j$ for all i and j;

4. the "odds"
$$\frac{p_{i|j}}{p_{i'|j}}$$
for values of X given $Y = y_j$ do not depend on j for all i and i';

5. the "odds ratios"
$$\left(\frac{p_{i|j}}{p_{i'|j}}\right) \Big/ \left(\frac{p_{i|j'}}{p_{i'|j'}}\right) = 1$$
for all i, i', j and j'.

Solution 4

- (2) trivially implies (4).
- From (3), $p_j = \sum_i (g_i h_j) = h_j \sum_i g_i$ and so $p_{i|j} = g_i / \sum_i' g_{i'}$ which does not depend on j.

  Thus (3) implies (2).
- From (4)

$$\frac{p_{i|j}}{p_{i'|j}} = k_{ii'}$$

  for some matrix k and therefore $p_{ij} = k_{i1} p_{1j}$ and taking $g_i = k_{i1}$ and $h_j = p_{1j}$, we have (3).

  Thus (4) implies (3).
- From (1), $p_{ij} = p_i p_j$ and this is (3) if we take $g_i = p_i$ and $h_j = p_j$.

  Thus (1) implies (3).
- From (2), there exist k_i such that $p_{i|j} = k_i$ for all i and j. This implies that $p_{ij} = k_i p_j$ which implies that $p_i = k_i \sum_j p_j = k_i$ and so $p_{ij} = p_i p_j$ for all i and j.

  Thus (2) implies (1).
- (4) and (5) are trivially equivalent.

The first three bullet points show that (2), (3) and (4) are equivalent. The next two bullets show that they are also equivalent to (1) and the final bullet point that they are equivalent to (5).

The next exercise is from Problem Class 1

Exercise 5

[2] The 674 subjects classified in Table 2 were the defendants in indictments involving cases with multiple murders in Florida between 1976 and 1987. The variables in Table 2 are

Y: death penalty verdict, with categories ($j = 1$: Yes, 2: No)

X: race of Defendant, with categories ($i = 1$: White, 2: Black)

Z: race of Victim, with categories ($k = 1$: White, 2: Black)

[2] R-script is available to double check.

		Death Penalty (Y)	
Victim's Race (Z)	Defendant's Race (X)	Yes	No
White	White	53	414
	Black	11	37
Black	White	0	16
	Black	4	139

Table 2: Death Penalty Verdict by Defendant's Race and Victims' Race

1. Compute the XY-marginal table, and the marginal counts
2. Based on the XY-marginal table,
   (a) Compute the conditional proportion of the White defendants received a death penalty, and that of the Black defendants received a death penalty. What do you observe?
   (b) Compute the MLE of the odd ratio and the asymptotic confidence interval at 5% sig. level. What is the association between Death Penalty and the Defendant's Race?
   (c) Perform a Goodness of fit test to test if the Death Penalty and the Defendant's Race are independent at 5% sig. level.
3. Test the Hypothesis that the Death Penalty and the Defendant's Race are independent across the Victim's Race levels at sig. level 5%.
4. Based on the XY-partial table,
   (a) Compute the MLE of the conditional XY odds ratios at each level of Victim's Race (Z),
   (b) Test the hypothesis that the Death Penalty and Defendant's Race are independent when the Victim is White, and against the alternative hypothesis that it is more likely for the Death penalty to be imposed for a Black defendant when the Victim is White. (Sig. level 5%)
   (c) Test the hypothesis that the Death Penalty and Defendant's Race are independent when the Victim is Black, and against the general alternative hypothesis that they are dependent when the Victim in Black. (Sig. level 5%)
5. Test the hypothesis that the Death Penalty and the Victim's Race are independent across the Defendant's Race levels (Sig. level 5%).
6. Compute the ZY conditional odds ratios for each level of the Defendant's Race, and discuss what they imply.
7.
   (a) Compute the marginal YZ-table.

(b) Compute the marginal YZ-odds ratios.

(c) Test the hypothesis that the Death Penalty and the Victims Race are independent against the alternative that it is more likely for a Death Penalty to be imposed when the Victim is White, than what it is when the Victim is Black. (Sig. level 5%)

8.

(a) Compute the marginal XZ-table.

(b) Compute the marginal XZ-odds ratio.

(c) Test the hypothesis that the Defendant's Race and the Victims rate are independent, against the alternative that it is more likely for a Victim to be White when the defendant is White, than what it is when the Defendant is Black. (Sig. level 5%)

9. Any comments?

HINT: Regarding the computation of Odds ratio, if the matrix has ZERO cells we use a correction by adding $+0.5$ at each cell. Precisely,

$$\theta = \frac{(n_{11} + 0.5)(n_{22} + 0.5)}{(n_{21} + 0.5)(n_{12} + 0.5)}$$

if any of n_{11}, n_{12}, n_{21}, n_{22} is zero. This is just a treatment, so that we do not get in infinite numbers

E.g., in this exercise, as the $(1, 1, 2)$th cell is $n_{112} = 0$, the above remedy is applied for the computation of $\theta^{XY}_{(2)}$, $\theta^{ZY}_{(1)}$, etc...

Solution 5