On comparison between two square tables using index of marginal inhomogeneity

AASC2018

Kouji Tahata (Tokyo University of Science, JAPAN)

1. INTRODUCTION

Consider the data in Table 1 taken from [1]. These data come from a randomized, double-blind clinical trial comparing an active hypnotic drug with placebo in patients with insomnia. There are many methods available to analyze these data. For example, [1] proposed the log-linear model to characterize differential change among treatments, and [2] analyzed these data by using various marginal models.

Table 1: Time to falling asleep, by treatment and occasion.

		Time to falling asleep			
		Follow-up			
Treatment	Initial	< 20	20-30	30-60	>60
Active	<20	7	4	1	0
	20-30	11	5	2	2
	30-60	13	23	3	1
	>60	9	17	13	8
Placebo	<20	7	4	2	1
	20-30	14	5	1	0
	30-60	6	9	18	2
	>60	4	11	14	22

Table 2 shows sample marginal cumulative distributions for Table 1. We can see that (1) from the initial to follow-up occasions, the time to falling asleep tends to shift downward for both treatments, and (2) the degree of shift seems greater for the active drug. Therefore we are interested (1) in measuring the degree of departure from equality of two marginal distributions between the initial and follow-up occasions for each treatment, and (2) in testing the equality of degree of departure from marginal homogeneity between two treatments.

Table 2: Sample marginal cumulative distributions.

		Time to falling asleep			
Treatment	Occasion	<20	<30	<60	<(>60)
Active	Initial	0.101	0.269	0.605	1.000
	Follow-up	0.336	0.748	0.908	1.000
Placebo	Initial	0.117	0.283	0.575	1.000
	Follow-up	0.258	0.500	0.792	1.000

2. MEASURE

Consider an $R \times R$ square contingency table with ordinal categories. Let p_{ij} denote the (i,j)th cell probability, and let X and Y denote the row and column variables, respectively. Let F_i^X and F_i^Y denote the cumulative marginal probabilities of X and Y, respectively. The marginal homogeneity (MH) model ([3]) is defined as

$$F_i^X = F_i^Y \quad (i = 1, \ldots, R-1).$$

Submeasure I: Let $\Delta_1 = \sum_{i=1}^{R-1}(F_i^X + F_i^Y)$, and

$$F^*_{1(i)} = \frac{F_i^X}{\Delta_1},\ F^*_{2(i)} = \frac{F_i^Y}{\Delta_1},\ Q^*_{1(i)} = \frac{1}{2}(F^*_{1(i)} + F^*_{2(i)}).$$

Assume that $F_1^X + F_1^Y \neq 0$. Consider the submeasure defined by

$$\Omega_{M1} = \left[\frac{2+\sqrt{2}}{2}\sum_{i=1}^{R-1}\sum_{k=1}^{2}\left(F^*_{k(i)} - Q^*_{1(i)}\right)^2\right]^{1/2}.$$

Submeasure II: Let $S_i^X = 1 - F_i^X$, $S_i^Y = 1 - F_i^Y$, $\Delta_2 = \sum_{i=1}^{R-1}(S_i^X + S_i^Y)$ and let

$$S^*_{1(i)} = \frac{S_i^X}{\Delta_2},\ S^*_{2(i)} = \frac{S_i^Y}{\Delta_2},\ Q^*_{2(i)} = \frac{1}{2}(S^*_{1(i)} + S^*_{2(i)}).$$

Assuming that $S_{R-1}^X + S_{R-1}^Y \neq 0$, we shall define the submeasure Ω_{M2}, which represents the degree of departure from MH, by Ω_{M1} with $\{F^*_{1(i)}\}$, $\{F^*_{2(i)}\}$, and $\{Q^*_{1(i)}\}$ replaced by $\{S^*_{1(i)}\}$, $\{S^*_{2(i)}\}$, and $\{Q^*_{2(i)}\}$, respectively.

Measure for MH: Assume that $F_1^X + F_1^Y \neq 0$ and $S_{R-1}^X + S_{R-1}^Y \neq 0$. Consider a measure defined by

$$\Omega_M = \frac{\Omega_{M1} + \Omega_{M2}}{2}.$$

Properties of measure: (i) $0 \leq \Omega_M \leq 1$, (ii) $\Omega_M = 0$ if and only if there is a structure of MH, and (iii) $\Omega_M = 1$ if and only if the degree of departure from MH is the largest, in the sense that $F_i^X = 0$ (then $S_i^X = 1$) and $F_i^Y = 1$ (then $S_i^Y = 0$), or $F_i^X = 1$ (then $S_i^X = 0$) and $F_i^Y = 0$ (then $S_i^Y = 1$), for arbitrary cut point i $(i = 1, 2, \ldots, R-1)$.

3. TEST

Let n_{ij} denote the observed frequency in the (i,j)th cell. The sample version of Ω_M, i.e., $\widehat{\Omega}_M$, is given by Ω_M with $\{p_{ij}\}$ replaced by $\{\widehat{p}_{ij}\}$, where $\widehat{p}_{ij} = n_{ij}/n$ and $n = \sum\sum n_{ij}$. Assuming that a multinomial distribution applies to the $R \times R$ table. Using the delta method, we obtain the following result.

$\sqrt{n}(\widehat{\Omega}_M - \Omega_M)$ has asymptotically a normal distribution with mean zero and variance $\sigma^2[\widehat{\Omega}_M]$.

For tables A and B (with sample sizes n_A and n_B), denote Ω_M by $\Omega_M^{(A)}$ and $\Omega_M^{(B)}$, respectively. Then an estimate of the difference between $\Omega_M^{(A)}$ and $\Omega_M^{(B)}$ is given by the sample difference $\widehat{\Omega}_M^{(A)} - \widehat{\Omega}_M^{(B)}$. When n_A and n_B are large, this difference has approximately a normal distribution with standard error

$$\left(\frac{\widehat{\sigma}^2[\widehat{\Omega}_M^{(A)}]}{n_A} + \frac{\widehat{\sigma}^2[\widehat{\Omega}_M^{(B)}]}{n_B}\right)^{1/2},$$

where $\widehat{\sigma}^2[\widehat{\Omega}_M^{(A)}]$ and $\widehat{\sigma}^2[\widehat{\Omega}_M^{(B)}]$ are the estimated variances. For Table 1, the value of test statistic is 3.20. This is the significant at 5% level. So, we can infer that the active drug is more effective than the placebo.

REFERENCES

[1] Francom, S. F., Chuang-Stein, C. and Landis, J. R. (1989). A log-linear model for ordinal data to characterize differential change among treatments. *Statist. Med.* **8**, 571 – 582.

[2] Agresti, A. (2010). *Analysis of ordinal categorical data*, 2nd edition. Wiley, Hoboken, New Jersey.

[3] Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika* **42**, 412–416.