Mathematica Aeterna, Vol. 6, 2016, no. 4, 637 - 639

Application of independence testing in a singing competition

Tingyu Qu

College of Mathematics and Information Science, Hebei University, Baoding 071002, China

Jian Shi

College of Mathematics and Information Science, Hebei University, Baoding 071002, China

Abstract

In this paper, we shall prove that in a Chinese female singing competition, whether a Top 100 contestant could enter the Top 20 group is irrelevant to the region she is from, through the test of independence.

Mathematics Subject Classification: 62A01

Keywords: Test of independence, Contingency table analysis, Data classification

1. Introduction:

Traditionally, we tend to consider that the regions of the contestants would affect the result of the competition significantly. Mostly because of the fact that Sichuan and Hunan contestants are likely to get a higher ranking in this competition, like Chris Li, or Jane Zhang etc.

Definition 1-1(Test of independence, [1]): A test of the independence of the frequency data which are classified by two or more characters.

In this paper, we shall prove that the regions of the contestants would not affect the result of Top 20 contestants list.

2. Main Result:

In this section, we shall prove that the regions of the contestants would not affect the result of Top 20 contestants list, through the test of independence.

The null hypothesis (H_0) states that the regions contestants come from have no effects on the Top20 (Non-Top20) list. While the alternative hypothesis (H_1) states that the regions contestants come from do affect the Top20 (Non-Top20) list. We will

use the alpha level of .05 in this paper.

According to the information we could find from [2] and [3], we divide the Top 100 contestants into 23 groups by regions, including all the regions that contain contestants in China or overseas. The data are recorded in the Table2-1 below.

Region	Hunan	Guangdong	Heilongjiang	Jiangsu	Liaoning	Hubei	Xinjiang	Henan	Fujian	Beijing	Chongqing	Sichuan
Top20	3	2	1	2	1	0	0	0	0	1	0	3
Non-Top20	12	3	2	3	4	4	1	2	2	8	1	9
Region	Tianjin	Shanxi	Yunnan	Shanghai	Hebei	Shandong	Anhui	Zhejiang	Jiangxi	HK/Macao/Taiwan	Overseas	Total
Top20	1	1	1	1	0	0	1	1	0	0	1	20
Non-Top20	8	2	3	2	1	3	0	3	1	1	5	80

Table2-1 Regional data about Top100 contestants

And under the circumstances that the null hypothesis is true, we can obtain the MLE of every parameters ($\hat{p}_{\bullet j}$ and $\hat{p}_{i\bullet}$ represent different proportions that different parameters hold). The results are recorded in the Table2-2 below.

Region	Hunan	Guangdong	Heilongjiang	Jiangsu	Liaoning	Hubei	Xinjiang	Henan	Fujian	Beijing	Chongqing	Sichuan
Top20	3	1	0.6	1	1	0.8	0.2	0.4	0.4	1.8	0.2	2.4
Non-Top20	12	4	2.4	4	4	3.2	0.8	1.6	1.6	7.2	0.8	9.6
$\hat{p}_{\bullet j}$	0.15	0.05	0.03	0.05	0.05	0.04	0.01	0.02	0.02	0.09	0.01	0.12
Region	Tianjin	Shanxi	Yunnan	Shanghai	Hebei	Shandong	Anhui	Zhejiang	Jiangxi	HK/Macao/Taiwan	Overseas	$\hat{p}_{i\bullet}$
Top20	1.8	0.6	0.8	0.6	0.2	0.6	0.2	0.8	0.2	0.2	1.2	0.2
Non-Top20	7.2	2.4	3.2	2.4	0.8	2.4	0.8	3.2	0.8	0.8	4.8	0.8
$\hat{p}_{\bullet j}$	0.09	0.03	0.04	0.03	0.01	0.03	0.01	0.04	0.01	0.01	0.06	

Table2-2 The results of all the $n\hat{p}_{ij}$

Based on Table2-2, the test can be turned into a distribution fitting test. The test statistic would be

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{23}\frac{(n_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}} \qquad (2.1)$$

Thus, based on the data from two tables above, we have

$$\chi^2 = 12.7433 \qquad (2.2)$$

According to the given alpha level (0.05), the critical region is

$$W = \{\chi^2 \geq \chi_{0.95}{}^2(22)\} = \{\chi^2 \geq 33.9244\} \qquad (2.3)$$

Clearly we can see, 12.7433<33.9244. So the null hypothesis would be rejected, leading to the fact that the regions of contestants would not affect the result of Top 20 contestants list.

Acknowledgement:

The author Jian Shi (corresponding author) is supported by Hebei Education Department (No. ZC2016009, QN2016030), Hebei University Funds for Distinguished Young Scientists and Natural Science Foundation of Shandong Province (No. BS2015SF006).

References:

[1] Shisong Mao, Yiming Cheng and Xiaolong Pu, Probability Theory and Mathematical Statistics, Higher Education Press 2011.2
[2] www.baike.baidu.com
[3] www.mgtv.com

Received: August, 2016