KHÓA BỒI DƯỠNG GIẢNG VIÊN KHU VỰC MIỀN BẮC NĂM 2023

Nguyễn Thị Nhung

TRƯỜNG ĐẠI HỌC THĂNG LONG

Phần III

Bài tập Kiểm định tính độc lập
Kiểm định về phân phối tổng thể

Bài tập 1

Một nhà nghiên cứu cho rằng điểm của các học sinh phụ thuộc vào số lượng thời gian chúng nghe nhạc. Một mẫu ngẫu nhiên gồm 400 học sinh được chọn và được xếp lớp chéo giữa điểm trung bình cuối năm với thời gian nghe nhạc hàng tuần như sau:

Thời gian nghe nhạc	Điểm trung bình					Tổng
	Xuất sắc	Giỏi	Khá	Trung bình	Kém	
< 5h	13	10	11	16	5	55
5h - 10h	20	27	27	19	2	95
11h - 20h	9	27	71	16	32	155
> 20h	8	11	41	24	11	95
Tổng	50	75	150	75	50	400

Ở mức ý nghĩa $\alpha = 5\%$ hãy kiểm tra xem điểm trung bình có phụ thuộc vào thời gian nghe nhạc hay không.

Thực hiện kiểm định tính độc lập trong R

- Để kiểm chứng tính độc lập trong R, ta dùng hàm `chisq.test(A)`, trong đó A là ma trận chỉ bảng gồm các quan sát của hai thuộc tính cần kiểm định tính độc lập.
- Để lập được một ma trận cấp $m \times n$ ta dùng hàm
  `matrix(x, nrow = m, ncol = n, byrow = FALSE, dimnames = NULL)`,
  trong đó
  - `x` là véc tơ chỉ các phần tử của ma trận;
  - `nrow = m` là tham số chỉ số hàng bằng m của ma trận;
  - `ncol = n` là tham số chỉ số cột bằng n của ma trận;
  - `byrow = FALSE (TRUE)` là tham số chỉ việc sắp xếp các phần tử trong véc tơ x theo cột (hàng) trước, mặc định là `FALSE` tức là theo cột trước;
  - `dimnames` là tham số ghi tên cột và hàng của ma trận, mặc định là `NULL`.

Thực hiện kiểm định tính độc lập trong R

- Trong ví dụ về kiểm chứng xem điểm trung bình và thời gian nghe nhạc có phụ thuộc vào nhau hay không, ta thực hiện trong *R* như sau:
- Thiết lập ma trận:

```
> SoHocSinhTheoCot = c(13, 20, 9, 8, 10, 27, 27, 11,
11, 27, 71, 41, 16, 19, 16, 24, 5, 2, 32, 11)
> A = matrix(SoHocSinhTheoCot, nrow = 4)
```

- Hoặc

```
> SoHocSinhTheoDong = c(13, 10, 11, 16, 5,
20, 27, 27, 19, 2, 9, 27, 71, 16, 32, 8, 11, 41, 24, 11)
> A = matrix(SoHocSinhTheoDong, nrow = 4, byrow = T)
```

Thực hiện kiểm định tính độc lập trong R

- Kết quả trong *R* cho ta:

```
> A
     [,1] [,2] [,3] [,4] [,5]
[1,]   13   10   11   16    5
[2,]   20   27   27   19    2
[3,]    9   27   71   16   32
[4,]    8   11   41   24   11
```

Thực hiện kiểm định tính độc lập trong R

- Để thêm tên của các cột và dòng của ma trận trên, ta thực hiện lệnh:

```
> A = matrix(SoHocSinhTheoDong, nrow = 4, byrow = T,
dimnames = list(c("<5h", "5h-10h", "11h-20h", ">20h"),
     c("XuatSac", "Gioi", "Kha", "TrungBinh", "Kem")))
```

- Kết quả trong *R* cho ta:

```
> A
        XuatSac  Gioi  Kha  TrungBinh  Kem
<5h          13    10   11         16    5
5h-10h       20    27   27         19    2
11h-20h       9    27   71         16   32
>20h          8    11   41         24   11
```

Thực hiện kiểm chứng mức phù hợp trong R

Thực hiện kiểm chứng mức phù hợp trong R, ta dùng hàm
`chisq.test(x, p =` p_0`)`,
trong đó

- x là véc tơ chỉ các quan sát;
- p_0 là véc tơ xác suất chỉ qui luật phân phối của tổng thể.

Bài tập 2

Để kiểm định xem quân xúc xắc có cân đối và đồng chất hay không, người ta tiến hành tung con xúc xắc 120 lần và nhận được kết quả như sau:

Số chấm	1	2	3	4	5	6	Tổng
Số lần tung	28	14	26	18	15	19	120

Tại mức ý nghĩa $\alpha = 5\%$ có thể kết luận con xúc xắc là cân đối và đồng chất hay không?

Bài tập 3

Một công ty thương mại dựa vào kinh nghiệm quá khứ đã xác định rằng vào cuối năm thì 80% số hóa đơn đã được thanh toán đầy đủ, 10% khất lại một tháng, 6% khất lại hai tháng và 4% khất lại trên hai tháng. Vào cuối năm nay công ty kiểm tra một mẫu ngẫu nhiên gồm 400 hóa đơn, ta thấy 287 được thanh toán đầy đủ, 49 khất lại một tháng, 30 khất lại hai tháng và 34 khất lại trên hai tháng. Tại mức ý nghĩa $\alpha = 5\%$, những dữ liệu này gợi ý rằng của năm nay có còn giống những năm trước nữa không?

Bài tập 4

Một trong những cách để quyết định ai là tác giả là so sánh tần số xuất hiện của một từ nào đó. Nghiên cứu số lần xuất hiện của từ "có thể" trong một đoạn văn dài xấp xỉ 200 từ người ta ghi lại được như sau:

Số lần xuất hiện	0	1	2	$\geqslant 3$	Tổng
Số đoạn văn	156	63	29	14	262

Tại mức ý nghĩa $\alpha = 5\%$, hãy kiểm tra xem phân phối của từ "có thể" có tuân theo phân phối Poisson với $\lambda = 0.6$ không?

Bài tập 5

File `Covid19_USA.csv` thống kê số ca tử vong do Covid-19 ở các bang của Mỹ trong thời gian 1/1/2020 đến 20/11/2021 theo giới tính và số người tử vong.

a. Tại mức ý nghĩa $\alpha = 5\%$, kiểm định xem có mối liên hệ giữa giới tính (Sex) và độ tuổi (Age.Group) không.

b. Tại mức ý nghĩa $\alpha = 5\%$, kiểm định xem có mối liên hệ giữa độ tuổi (Age.Group) và mức độ tử vong Covid-19 (COVID.19.Deaths_factor) không.

Bài tập 6

File `hostel.csv` cho dữ liệu về giá thuê nhà cũng như điểm đánh giá của khách về giá cả và dịch vụ ở một số thành phố của Nhật. Xét xem dữ liệu về điểm đánh giá của khách du lịch (summary_score) có thể lấy từ một tổng thể có phân phối chuẩn không bằng cách thực hiện các thao tác sau.

a. Vẽ biểu đồ phân phối tần số (histogram), biểu đồ hộp (boxplot), vẽ biểu đồ Q-Q và cho nhận xét.

b. Tính những đại lượng thống kê mô tả: trung bình, trung vị, mode, skewness, kurtosis, khoảng biến thiên, độ trải giữa, độ lệch chuẩn, so sánh các đại lượng một cách phù hợp và cho nhận xét.

c. Tại mức ý nghĩa $\alpha = 5\%$, dùng các kiểm định Jarque-Bera, Shapiro-Wilk để kiểm tra tính chuẩn của điểm đánh giá.