**KIỂM ĐỊNH KHI BÌNH PHƯƠNG**

**Các bài toán liên quan:**

1. Kiểm định cùng phân phối xác xuất?
2. Kiểm định thỏa phân phối xs?
3. Kiểm định tính độc lập của dữ liệu.

Kiểm định với Khi bình phương:

Giả thuyết kiểm định có thể là 1 câu, mệnh đề, biểu thức.

Kiểm định bằng Khi bình phương:

Trong đó:

O: Giá trị mẫu (dữ liệu mẫu)

E: Giá trị theo lý thuyết (E = n\*Pi)

K: số khoảng/giá trị khác nhau của mẫu (K>=5)

H0: Dữ liệu thực == lý thuyết

H1: Dữ liệu thực != lý thuyết

H0: Bị bác bỏ khi:

Trong đó: r: số tham số cần ước lượng

p-value = P(

Trong đó =

r= 2: phân phối chuẩn

r=0: phân phối đều

r=1: phân phối poission

**KIỂM ĐỊNH DỮ LIỆU CÓ PHÙ HỢP VỚI 1 PPXS?**

Ví dụ: Các kỹ sư môi trường thường sử dụng thông tin về số lượng các loài tảo khác nhau và số lượng cụm tế bào ứng mỗi loài để đo tình hình “sức khỏe” của hồ. Những hồ này thấy một vài loài nhưng mỗi loài lại có nhiều cụm tế bào. Trong một cuộc điều tra như vậy, một mẫu hồ được phân tích dưới kính hiển vi để xác định số lượng cụm tế bào trên mỗi loài tảo.

Những dữ liệu được tóm tắt ở đây cho 150 loài được kiểm tra dưới kính hiển vi.



Trong đó:

yi: là số cụm tế bào

ni: số loài tảo

Kiểm định với mức ý nghĩa là 0.05, mẫu dữ liệu các cụm tế bào có phân phối Poisson không?

(Số lượng cụm tế bào trong mỗi loài tảo có phân phối Poisson ?)

1. Lượng giá giá trị trung bình λ của mẫu:
2. Xác suất Poisson cho các giá trị y= 0,1,..,7

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | >=7 |
|  | .0369 | 0.1269 | 0.2056 | 0.2220 | 0.1798 | 0.1165 | 0.0629 | 0.0471 |
| Ej |  |  |  |  |  |  |  |  |

df= K-r-1 = 8-1-1=6

kstar<-qchisq(1-alpha,df)

X<-c(0,1,2,3,4,5,6,7)

> Oj<-c(6,23,29,31,27,13,8,13)

> ld<- sum(X\*Oj)/sum(Oj)

> Pi<- dpois(X,ld)

> Pi[8] <- 1- ppois(6,ld)

> n<- sum(Oj)

> n

[1] 150

> ld

[1] 3.24

> Ej<-n\*Pi

> Q2<- sum((Oj-Ej)^2/Ej)

> alpha<-0.05

> K<-length(Oj)

> r<-1

> df<-K-r-1

> df

[1] 6

> kstar<-qchisq(1-alpha,df)

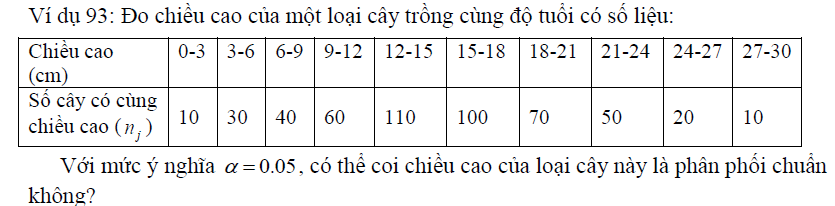
> Q2 >= kstar

[1] FALSE

> # X co phan phoi Poisson

Kết luận: Dữ liệu kiểm định phù hợp với phân phối Possion.

Ví dụ 2: (giáo trình – Ví dụ 93)



Giải: Gọi X biến ngẫu nhiên thể hiện chiều cao của cây

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 1.5 | 4.5 | 7.5 | 10.5 | 13.5 | 16.5 | 19.5 | 22.5 | 25.5 | 28.5 |
| *ni* | 10 | 30 | 40 | 60 | 110 | 100 | 70 | 50 | 20 | 10 |

* Phát biểu giả thuyết:

H0: X có phân phối chuẩn )

H1: X không có phân phối chuẩn.

*Chuẩn hóa dữ liệu để đưa về kiểm tra xem X có phân phối chuẩn N(0,1) ?*

1. Chuẩn hóa dữ liệu ( Z-score)
2. Chuẩn hóa các cận:



thành

> cdn<- (cd-Xbar)/S

> ctn<- (ct-Xbar)/S

3.

> # Neu X1 co phan phoi chuan N(0,1) khi do xs de Xi thuoc (li,li+1)

#

#

> # P(Xi in (li,li+1)) = phi(li+1)-phi(li)

> Pi<- pnorm(ctn)-pnorm(cdn)

> # So cay theo ly thuyet

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 1.5 | 4.5 | 7.5 | 10.5 | 13.5 | 16.5 | 19.5 | 22.5 | 25.5 | 28.5 |
| *Oj* | 10 | 30 | 40 | 60 | 110 | 100 | 70 | 50 | 20 | 10 |
| *Pi* | p1 | p2 | .. |  |  |  |  |  |  | pn |
| *Ej* | p1\*n | p2\*n |  |  |  |  |  |  |  | pn\*n |

Ví dụ:

|  |  |  |  |
| --- | --- | --- | --- |
| n=12 | A | B | C |
| P | p1=1/3 | p2=1/2 | p3=1/6 |
| E | p1\*n=4 | p2\*n=6 | p3\*n=2 |

Xác suất X=1.5 ~ X nằm trong khoảng (0,3)

P(0<=X <=3) = P(X<=3) – P(X<=0)

Giá trị 3 chuan hoa l2=(3-Xbar)/S;

0 chuan hoa l1=(0-Xbar)/S

P(0<=X<=3) ~ P(l1<=X1<=l2)==

> Ej<- Pi\*n

> Oj<- c(10,30,40,60,110,100,70,50,20,10)

> Q2<- sum((Oj-Ej)^2/Ej)

> K<-length(Oj)

> r<-2

> df<-K-r-1

> alpha<-0.05

> kstar<-qchisq(1-alpha,df)

> Q2 >= kstar

[1] FALSE

> #H0: Chap nhan- X co phan phoi chuan

Giải đầy đủ:

> cd<-c(0,3,6,9,12,15,18,21,24,27)

> ct<-c(3,6,9,12,15,18,21,24,27,30)

> X<-(cd+ct)/2

> X

[1] 1.5 4.5 7.5 10.5 13.5 16.5 19.5 22.5 25.5 28.5

> ni<-c(10,30,40,60,110,100,70,50,20,10)

> Xbar<-sum(X\*ni)/sum(ni)

> n<-sum(ni)

> S2<- sum((X-Xbar)^2\*ni)/(n-1)

> S<-sqrt(S2)

> # Du lieu chuan hoa

> X1<- (X-Xbar)/S

> # Chuan hoa cac can

> cdn<- (cd-Xbar)/S

> ctn<- (ct-Xbar)/S

> # Neu X1 co phan phoi chuan N(0,1) khi do xs de Xi thuoc (li,li+1)

> # P(Xi in (li,li+1)) = phi(li+1)-phi(li)

> Pi<- pnorm(ctn)-pnorm(cdn)

> # So cay theo ly thuyet

> Ej<- Pi\*n

> Oj<- c(10,30,40,60,110,100,70,50,20,10)

> Q2<- sum((Oj-Ej)^2/Ej)

> K<-length(Oj)

> r<-2

> df<-K-r-1

> alpha<-0.05

> kstar<-qchisq(1-alpha,df)

> Q2 >= kstar

[1] FALSE

> #H0: Chap nhan- X co phan phoi chuan

Bài tập: 10.37 - 10.39 p.541

**An Introdution to Statistical Methods & Data Analysis, R. Lyman Ott, Michael Longnecker**

HỆ SỐ TƯƠNG QUAN TUYẾN TÍNH

1. Pearson: cor(); *cor.test(x,y)*
2. Spearman*: cor.test(x,y,method=”spearman”)*
3. Kendall*: cor.test(x,y,method=”kendall”)*

HỒI QUI

Tìm phương trình đường thẳng thể hiện mối quan hệ giữa 2 biến ngẫu nhiên có mối quan hệ tuyến tính.

* Trong Toán dùng phương pháp bình phương bé nhất để xác định các hệ số của pt.
* Trong ngôn ngữ R có hàm lm()