

Práctica 8

Cálculo de Ji cuadrada en R

Luis Eduardo Galindo Amaya (1274895)

| | |
|------------|-----------------------|
| Asignatura | Estadística Avanzada |
| Docente | Olivia Mendoza Duarte |
| Fecha | 27-10-2022 |

Cálculo de Ji cuadrada en R

Luis Eduardo Galindo Amaya (1274895)

27-10-2022

Información del dataset¹

This is one of the best known datasets in statistics and machine learning. Fisher's paper is a classic in the field and is frequently used for tutorial and teaching purposes. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are not linearly separable from each other.

Predicted attribute: class of iris plant.

Attribute information:

- 1. sepal length in cm
- 2. sepal width in cm
- 3. petal length in cm
- 4. petal width in cm
- 5. class:
 - Iris Setosa
 - Iris Versicolour
 - Iris Virginica

¹<https://archive-beta.ics.uci.edu/ml/datasets/iris>

Practica

- Comprender el código de la última diapositiva del Parcial II
- Tomar como base ese código para calcular el estadístico Ji cuadrada con otra base de datos
- Hacer un reporte que contenga el código adaptado y corregido si es necesario y capturas de pantalla de los resultados

Capturas

```

<tos/ESA-2022/practica 8
+ ...
Practica 8.org
Practica 8.pdf
Practica 8.tex
ji.R

archivo <- read.csv("bezdekIris.csv")
clase <- archivo[, 5]
cantidad_datos <- length(clase)
esperado <- c(1)
conteo_clases <- 1
for (i in 2:cantidad_datos) {
  if (clase[i] != clase[i - 1]) {
    conteo_clases <- conteo_clases + 1
    esperado <- c(esperado, 1)
  }
}

por_muestra <- 0.5
cant_muestra <- trunc(cantidad_datos * por_muestra)
indice_muestra <- sample(1:cantidad_datos, cant_muestra)
clase_muestra <- clase[indice_muestra]
esperado <- esperado * trunc(cant_muestra / conteo_clases)
observado <- (1:conteo_clases) * 0

for (i in 1:cant_muestra) {
  clase_contada <- clase_muestra[i]
  observado[clase_contada] <- observado[clase_contada] + 1
}

tabla <- data.frame(esperado, observado)
tabla
chisq.test(tabla)

X-squared = 0.50819, df = 2, p-value = 0.7756

archivo <- read.csv("bezdekIris.csv")
clase <- archivo[, 5]
cantidad_datos <- length(clase)
esperado <- c(1)
conteo_clases <- 1
for (i in 2:cantidad_datos) {
  if (clase[i] != clase[i - 1]) {
    conteo_clases <- conteo_clases + 1
    esperado <- c(esperado, 1)
  }
}
por_muestra <- 0.5
cant_muestra <- trunc(cantidad_datos * por_muestra)
indice_muestra <- sample(1:cantidad_datos, cant_muestra)
clase_muestra <- clase[indice_muestra]
esperado <- esperado * trunc(cant_muestra / conteo_clases)
observado <- (1:conteo_clases) * 0
for (i in 1:cant_muestra) {
  clase_contada <- clase_muestra[i]
  observado[clase_contada] <- observado[clase_contada] + 1
}
tabla <- data.frame(esperado, observado)
tabla
esperado observado
1      25      28
2      25      24
3      25      23
> chisq.test(tabla)

Pearson's Chi-squared test

data: tabla
X-squared = 0.27355, df = 2, p-value = 0.8722
>

```

Figura 1: p-value = 0.8722

```

<tos/ESA-2022/practica 8
+ ...
Practica 8.org
Practica 8.pdf
Practica 8.tex
ji.R

archivo <- read.csv("bezdekIris.csv")
clase <- archivo[, 5]
cantidad_datos <- length(clase)

esperado <- c(1)
conteo_clases <- 1
for (i in 2:cantidad_datos) {
  if (clase[i] != clase[i - 1]) {
    conteo_clases <- conteo_clases + 1
    esperado <- c(esperado, 1)
  }
}

por_muestra <- 0.5
cant_muestra <- trunc(cantidad_datos * por_muestra)

indice_muestra <- sample(1:cantidad_datos, cant_muestra)
clase_muestra <- clase[indice_muestra]
esperado <- esperado * trunc(cant_muestra / conteo_clases)
observado <- (1:conteo_clases) * 0

for (i in 1:cant_muestra) {
  clase_contada <- clase_muestra[i]
  observado[clase_contada] <- observado[clase_contada] + 1
}

tabla <- data.frame(esperado, observado)
tabla
chisq.test(tabla)

> clase_muestra <- clase[indice_muestra]
> esperado <- esperado * trunc(cant_muestra / conteo_clases)
> observado <- (1:conteo_clases) * 0
> for (i in 1:cant_muestra) {
+   clase_contada <- clase_muestra[i]
+   observado[clase_contada] <- observado[clase_contada] + 1
+ }
> tabla <- data.frame(esperado, observado)
> tabla
  esperado observado
1       25        28
2       25        26
3       25        21
> chisq.test(tabla)

Pearson's Chi-squared test

data: tabla
X-squared = 0.53725, df = 2, p-value = 0.7644
> |

```

Figura 2: p-value = 0.7644

```

<tos/ESA-2022/practica 8
+ ...
Practica 8.org
Practica 8.pdf
Practica 8.tex
ji.R

archivo <- read.csv("bezdekIris.csv")
clase <- archivo[, 5]
cantidad_datos <- length(clase)

esperado <- c(1)
conteo_clases <- 1
for (i in 2:cantidad_datos) {
  if (clase[i] != clase[i - 1]) {
    conteo_clases <- conteo_clases + 1
    esperado <- c(esperado, 1)
  }
}

por_muestra <- 0.5
cant_muestra <- trunc(cantidad_datos * por_muestra)

indice_muestra <- sample(1:cantidad_datos, cant_muestra)
clase_muestra <- clase[indice_muestra]
esperado <- esperado * trunc(cant_muestra / conteo_clases)
observado <- (1:conteo_clases) * 0

for (i in 1:cant_muestra) {
  clase_contada <- clase_muestra[i]
  observado[clase_contada] <- observado[clase_contada] + 1
}

tabla <- data.frame(esperado, observado)
tabla
chisq.test(tabla)

> conteo_clases <- 1
> for (i in 2:cantidad_datos) {
+   if (clase[i] != clase[i - 1]) {
+     conteo_clases <- conteo_clases + 1
+     esperado <- c(esperado, 1)
+   }
+ }
> por_muestra <- 0.5
> cant_muestra <- trunc(cantidad_datos * por_muestra)
> indice_muestra <- sample(1:cantidad_datos, cant_muestra)
> clase_muestra <- clase[indice_muestra]
> observado <- (1:conteo_clases) * 0
> for (i in 1:cant_muestra) {
+   clase_contada <- clase_muestra[i]
+   observado[clase_contada] <- observado[clase_contada] + 1
+ }
> tabla <- data.frame(esperado, observado)
> tabla
  esperado observado
1       25        26
2       25        22
3       25        27
> chisq.test(tabla)

Pearson's Chi-squared test

data: tabla
X-squared = 0.28802, df = 2, p-value = 0.8659
> |

```

Figura 3: p-value = 0.8659

Código

```
1 |
2 | archivo <- read.csv("bezdeklris.csv")
3 |
4 | clase <- archivo[, 5]
5 |
6 | cantidad_datos <- length(clase)
7 |
8 | esperado <- c(1)
9 | conteo_clases <- 1
10 | for (i in 2:cantidad_datos) {
11 |   if (clase[i] != clase[i - 1]) {
12 |     conteo_clases <- conteo_clases + 1
13 |     esperado <- c(esperado, 1)
14 |   }
15 | }
16 |
17 |
18 | por_muestra <- 0.5
19 | cant_muestra <- trunc(cantidad_datos * por_muestra)
20 |
21 | indice_muestra <- sample(1:cantidad_datos, cant_muestra)
22 | clase_muestra <- clase[indice_muestra]
23 | esperado <- esperado * trunc(cant_muestra / conteo_clases)
24 | observado <- (1:conteo_clases) * 0
25 |
26 | for (i in 1:cant_muestra) {
27 |   clase_contada <- clase_muestra[i]
28 |   observado[clase_contada] <- observado[clase_contada] + 1
29 | }
30 |
31 |
32 | tabla <- data.frame(esperado, observado)
33 | tabla
34 |
35 | chisq.test(tabla)
```