

# Análisis exploratorio de datos en R Cheat Sheet by julenx via cheatography.com/168626/cs/35581/

### Tablas y gráficos de frecuencias

table(x)

tabla de frecuencias absolutas

names(table(x))

nombres de las columnas de una tabla unidimensional

z=factor(x, levels=1:7)

table(z)

Si x contiene números que no sean 7, hacemos que el 7 también aparezca en la tabla de frecuencias.

table(x)[4]

Referirnos a tabla de frecuencias por el índice

table(x)["5"]

Referirnos a tabla de frecuencias por el nombre

prop.table(table(x))

Tabla de frecuencias relativas

names(which(table(x)==3))

Elementos cuya frecuencia = 3

names(which(table(x)==max(table(x))))

Moda

## Otros

attach(empleados)

attr(empleados, "variable.labels") = NULL

Eliminar etiquetas de columnas

### Gráfico de puntos

dotchart(salario, xlab='salarios')

stripchart(salario~sexo, method='jitter')

Salario según sexo

# Medidas de tendencia central

as.numeric(names(which(table(edades)-==max(table(edades)))))

Moda

mean(edades)

Media

median(edades)

Mediana

### Medidas de posición

quantile(x, 0.1)

Percentil

quantile(salario)

Todos los cuartiles

quantile(salario, probs=c(0.25,0.5,0.75))

quantile(salario, probs=seq(0.1, 0.9, 0.1))

Todos los deciles

#### summary(x)

R dispone de 9 métodos diferentes para calcular cuantiles, que se pueden especificar dentro de quantile con el parámetro type. En la mayoría de las ocasiones se obtiene el mismo resultado con todos los métodos, pero no siempre. Para saber en detalle las fórmulas que usa quantile para cada valor de type. El método que usa R por defecto si no especificamos otro es type=7.

#### Medidas de dispersión

diff(range(x))

Rango

IQR(x)

Rango intercuatílico

### Medidas de dispersión (cont)

var(x)

Varianza

sd(x)

Desviación típica

cv=sd(consumo)/abs(mean(consumo))

Coeficiente de variación. nos permite, entre otras cosas, comparar dispersiones de variables medidas en diferentes unidades

summary(subset(iris, Species=="setosa",
c("Sepal.Length","Petal.Length")))

#### Tablas bidimensionales de frecuencias

Respuestas=c-

("No","No","Sí","No","Sí","No","No","Sí")

Sexo=c("M","M","M","H","H","H","H","H")

table(Sexo,Respuestas)

Tabla bidimensional de frecuencias. El primer conjunto de valores es las filas y el segundo las columnas.

t(table(Respuestas,Sexo))

Intercambiar filas por columnas

table(Respuestas, Sexo)[1,2]

table(Respuestas,Sexo)["No","M"]

Formas de referirnos a una tabla bidimensional

prop.table(table(Sexo,Respuestas),
margin=1)

Frecuencias relativas marginales por

prop.table(table(Sexo,Respuestas),
margin=2)

Frecuencias relativas marginales por columnas

library(gmodels)

CrossTable(Sexo, Respuestas, prop.chis-q=FALSE)

tabla de frecuencias absolutas y las tres tablas de frecuencias relativas de dos variables



By julenx

cheatography.com/julenx/

Published 20th November, 2022. Last updated 20th November, 2022. Page 1 of 2. Sponsored by Readable.com

Measure your website readability!

https://readable.com



# Análisis exploratorio de datos en R Cheat Sheet by julenx via cheatography.com/168626/cs/35581/

# Tablas bidimensionales de frecuencias (cont)

colSums(table(Sexo,Respuestas))

Sumar columnas

rowSums(prop.table(table(Sexo,Respuestas)))

Sumar filas en tabla de frecuencias relativas

## Diagrama de barras

barplot(table(Sexo\_Ger), col=c("lightb-lue","pink"), main="Title")

Va alternando los colores indicados

barplot(prop.table(table(Respuestas)), main="Diagrama de barras de frecuencias relativas")

barplot(table(x), horiz=TRUE)

Diagrama horizontal

barplot(table(Sexo,Respuestas))

Dibujará una barra dentro de otra (el total) con distinto color según la primera variable

barplot(table(Sexo,Respuestas), beside=TRUE)

Dibuja una barra al lado de otra

legend.text=TRUE

Mostrar leyenda

legend.text=c("A","B")

Cambiar texto leyenda

pie(table(Respuestas), main="Diagrama circular")

Diagrama circular

par(mfrow = c(1, 3))

Matriz de 3 filas donde se mostarán las gráficas

tt <- table(cut(expprev, breaks=5))

tt <- table(cut(expprev, breaks=c(0,40,80,1-50,250,400)))

barplot(tt,xlab="Experiencia previa", main="Categorización en 5 clases")

Uso con variables continuas

#### Tablas multidimensionales de frecuencias

table(Sexo,Respuestas,Pais)

ftable(Sexo,Respuestas,Pais)

Tabla de frecuencias agrupando con tercera variable (la primera opción hace tres tablas, la segunda hace una)

ftable(Sexo,Respuestas,Pais, col.vars=c-("Sexo","Respuestas"))

especificar qué variables queremos que aparezcan como filas o como columnas

table(Sexo,Respuestas,Pais)["H","Sí","Italia"]

table(Sexo,Respuestas,Pais)[ , ,"Italia"]

table(Sexo,Respuestas,Pais)["M", ,"Italia"]

referirnos a una entrada, o a una subtabla

prop.table(table(Sexo,Respuestas,Pais),
margin=3)

Tabla de frecuencias relativas margianales. Con axis = 3, vemos la frecuencia dentro de la tercera variable.

prop.table(table(Sexo,Respuestas,Pais),
margin=c(1,3))

De esta forma se suma según la primera variable dentro de cada país.

prop.table(ftable(Sexo,Respuestas,Pais))

Otra forma de mostrar la tabla de frecuencias relativas

apply(table(Sexo,Respuestas,Pais),
MARGIN=c(1,3), FUN=sum)

# Tablas multidimensionales de frecuencias (cont)

apply(table(Sexo,Respuestas,Pais),
MARGIN=2, FUN=sum)

Aplicación de funciones

#### Histograma

fruta=c(8,11,11,8,9,10,16,6,12,19,13,6,9,1-3,15,9,12,16,8,7,14,11,15,6,14, 14,17,11,-6,9,10,19,12,11,12,6,15,16,16,12,13,12,12,-8,17,13,7,12,14,12)

plot(density(fruta), type="l", xlab="Número de árboles", ylab="Densidad", main="Densidad de la variable \"fruta\"")

representar datos continuos (variables numéricas)

