

Programación y Uso de Librerías en R: Herramientas de Análisis y Visualización de Datos en la Enseñanza y la Investigación Científica

Juan Luis Peñaloza Figueroa
Universidad Complutense de Madrid

Milagros Dones Tacero
Universidad Autónoma de Madrid

Carmen Gladys Vargas Pérez
Universidad Complutense de Madrid

AÑO: 2025

SCRIPT_8: ESTADÍSTICA DESCRIPTIVA BÁSICA

```
## OPERACIONES ESTADÍSTICAS EN R
# Estadísticas simples :
> x<-seq(1:10)
> x
> cumsum(x)
> median(x)
# Generación de 100 datos
> x<-rnorm(100)
> summary(x)
# Generación de dos muestras correladas
> x<-seq(1:10)
> y<-2*x+rnorm(10)
> cor(x,y)
# Creación de permutaciones:
> sample(10)
# Muestreo sin reposición
> simple(1:10, 5, rep = F)
# Muestreo con reposición
> sample(1:10,5,rep=T)
# Simulación de 10 tiradas de un dado equilibrado
> sample(1:6,10,rep=T)
# Cálculo de algunos percentiles de un conjunto de datos
> x<-rnorm(200)
> quantile(x,probs=c(0.1,0.4,0.9))
# Redondeo a k decimales
> round(y,2)
> round(y%*%x,2)
> x<-matrix(c(1,2,3,4),nrow=2)
> x
# Cálculo de la inversa
> y<-solve(x)
> y%*%x

## MEDIDAS DE TENDENCIA CENTRAL EN R
```

```

#Media aritmética
> mean(dforbes2$sales, na.rm=T)
#Media recortada al 5% (+2,5 y -2,5)
> mean(dforbes2$sales, na.rm=T, trim=0.025)
# Cálculo de la mediana
# Mediana de la variable
> median(dforbes2$sales, na.rm = TRUE)
# Cálculo de la moda
> install.packages("modeest")
> library(modeest)
# Valores que aparecen con más frecuencia
> mfv(sexo)
> mfv(dforbes2$category)
## MEDIDAS DE DISPERSIÓN
# Varianza
> var(dforbes2$sales, na.rm = TRUE)
#Desviación típica
> sd(dforbes2$sales, na.rm = TRUE)
# Covarianza
> duration=faithful$eruptions
> waiting=faithful$waiting
> cov(duration,waiting)
> datos <-data.frame(math=c(84,82,81,89,73,94,92,70,88,95),
ciencia=c(85, 82, 72, 77, 75, 89, 95, 84, 77, 94),
historia=c(97, 94, 93, 95, 88, 82, 78, 84, 69, 78))
> cov(datos)
> cov(datos, method="pearson")
> cov(datos, method="kendall")
> cov(datos, method="spearman")
# coeficiente de variación
>cv<-sd(dforbes2$sales,na.rm=T)/mean(dforbes2$sales, na.rm=T)
> cv
# Correlación
x <- c(1, 3, 5, 10)
y <- c(2, 4, 6, 20)
> cor(x,y,method="pearson")
> cor(x,y,method="kendall")
> cor(x,y,method="spearman")
## MEDIDAS DE TENDENCIA NO CENTRAL
# Percentiles o cuantiles de las variables: 'sales' y 'profits'
> library(readr)
> dforbes<-read_csv("E:\\Actividad Docente\\Curso
20222023\\Asignaturas de masters\\Introducción a R\\Dataset-
Libro\\Forbes2000.csv")
> dforbes1<-data.frame(dforbes)
> head(dforbes1,2)
> quantile(dforbes1$sales, c(0.25,0.50,0.75,0.90,0.95))
# Quantile
> quantile(dforbes1$profits,c(0.25,0.50,0.75,0.90,0.95),
na.rm=T)

```

##TABLAS DE FRECUENCIAS O CONTINGENCIA

Datos

```
> edad=c(18,19,NA,18,24,17,22,15,22,25, 27, 20)
> sexo=c(0,1,0,0,1,0,0,1,1,0,0,1)
> estudios=c(1,2,0,1,3,2,3,1,2,3,1
> sexo=factor(sexo, levels=c(0,1),
> labels=c("Hombre", "Mujer"))
```

Tabla de frecuencias absolutas

```
> table(sexo)
> sexo
```

Tabla de frecuencias relativas

```
> prop.table(table(sexo))
> sexo
> prop.table(table(edad))
> edad
> table(edad,useNA="ifany")
> edad
```

Tabla de frecuencias relativas: estudios vs. sexo

```
> prop.table(table(estudios, sexo))
```

Frecuencias relativas por filas (1)

```
> prop.table(table(estudios, sexo), 1)
```

Frecuencias relativas por columnas (2)

```
> prop.table(table(estudios, sexo), 2)
```

Edad agrupada en intervalos

```
> nclass.Sturges(edad)
> seq(15,27,length=nclass.Sturges(edad))
> intervalEdad=cut(edad,breaks=seq(15,27,length =
nclass.Sturges(edad)),include.lowest=TRUE)
> intervalEdad
```

Construcción de los intervalos mediante la función cut()

```
> intervalEdad=cut(edad,breaks=seq(15,27,length=
nclass.Sturges(edad)),include.lowest=TRUE)
> intervalEdad
```

Edades observadas

```
> edad=c(18,19,NA,18,24,17,22,15,22,25, 27, 20)
```

Por último, construimos la tabla de frecuencias de la variable edad.

```
> table(intervalEdad) intervalEdad
```

Representación gráfica de los resultados

```
> pie(table(estudios))
> barplot(table(sexo))
> hist(edad)
> boxplot(edad~sexo)
```

SIMETRÍA, CURTOSIS Y NORMALIDAD

```
> install.packages("psych")
```

```
> library(psych)
```

```
> skew(dforbes1$sales)
```

```
> skew(dforbes1$profits)
```

#Asimetría estandarizada

```
> skew(dforbes2$sales)/sqrt(6/2000)
```

```
> skew(dforbes2$profits)/sqrt(6/2000)
```

Cálculo de la kurtosis

```
> install.packages("psych")
```

```

> library(psych)
> kurtosi(dforbes1$sales)
> kurtosi(dforbes1$profits)
# Kurtosis estandarizada
> kurtosi(dforbes2$profits)/sqrt(6/2000)
> kurtosi(dforbes2$sales)/sqrt(6/2000)
## Tests de normalidad
#Prueba de Shapiro-Wilks
> shapiro.test(dforbes2$sales)
#Prueba de Kolmogorov-Smirnov
> ks.test(dforbes2$sales,"pnorm",mean=mean(dforbes2$sales,
na.rm=T), sd=sd(dforbes2$sales, na.rm=T))
#Prueba de Anderson-Darling
> ad.test(dforbes2$profit)
#Prueba de Cramer-von Mises
> cvm.test(dforbes2$profit)
#Prueba de Lilliefors (Kolmogorov-Smirnov)
> lillie.test(dforbes2$profit)
#Prueba de Pearson chi-square
> pearson.test(dforbes2$profit)
#Prueba de Jarque Bera
> jb.norm.test(MIT[,5])
#Prueba de Frosini
> frosini.norm.test(MIT[,5])
#Prueba de Geary
> geary.norm.test(MIT[,5])
#Prueba de Hegazy-Green
> hegazy1.norm.test(MIT[,5])
#Prueba de Jarque-Bera
> jb.norm.test(MIT[,5], nrepl=2000)
#Prueba de Kurtosis
> kurtosis.norm.test(MIT[,5], nrepl=2000)
# Prueba de Skewness
> skewness.norm.test(MIT[,5], nrepl=2000)

```