

Aplicaciones del análisis multivariante con R

Estadística Multivariante - Universidad de Granada

Laura Gómez Garrido

Miguel Lentisco Ballesteros

Antonio Martín Ruiz

Daniel Pozo Escalona

Francisco Javier Sáez Maldonado

22 de enero de 2020

- 1 Introducción: R
- 2 R en el análisis multivariante
 - Distribución Normal Multivariante. Ejemplos
- 3 Para ampliar

1 Introducción: R

2 R en el análisis multivariante

- Distribución Normal Multivariante. Ejemplos

3 Para ampliar

¿Qué es R?

R es un entorno y lenguaje de programación enfocados a la computación estadística y de gráficos. Surge como una reimplementación libre del lenguaje y entorno S. Proporciona una amplia variedad de funcionalidades estadísticas y gráficas y es altamente extensible.

R está disponible como software libre bajo los términos de la GNU General Public License de la Free Software Foundation en forma de código fuente. Puede ser compilado y ejecutado en una gran cantidad de plataformas UNIX, Windows y MacOS.

Entornos de desarrollo para R

Principales librerías

Algunas aplicaciones de R

1 Introducción: R

2 R en el análisis multivariante

- Distribución Normal Multivariante. Ejemplos

3 Para ampliar

Nuestro dataset

Característica del data set	Multivariante	Nº de Instancias	178
Características de los atributos	Enteros, Reales	Nº de Atributos	13
Área	Física	Donado	01/07/1991

Fuente: Machine Learning Repository

Propietarios Originales:

Forina, M. et al, PARVUS -

*An Extendible Package for Data Exploration, Classification
and Correlation.*

*Institute of Pharmaceutical and Food Analysis and Technologies,
Via Brigata Salerno, 16147 Genoa, Italy.*



```
wine <- read.table("http://archive.ics.uci.edu/ml/mach
sapply(wine[2:14], mean)
"Media muestral del conjunto completo"
```

```
In [41]: wine
```

Adults Income <= 14

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
0	1.4621	1.71	2.63	15.5	121	0.81	1.88	8.29	2.29	5.64	1.64	1.82	1985	
1	1.638	1.78	2.14	11.2	100	0.43	2.78	3.26	1.56	4.35	1.01	1.41	1981	
2	1.161	2.38	2.87	18.1	101	2.88	3.24	3.31	2.41	5.68	1.63	3.17	1983	
3	1.447	1.90	2.33	14.8	113	3.40	3.30	3.24	2.18	7.80	0.84	1.98	1984	

Dibujado de datos seleccionados. Scatterplots

```
selection1 <- wine[wine$V1 == "1",]  
selection3 <- wine[wine$V1 == "3",]  
mean1 <- sapply(selection1[2:14], mean)  
mean2 <- sapply(selection3[2:14], mean)  
chemical <- c(2,3,4,5,6,7,8,9,10,11,12,13,14)  
plot(chemical, mean1, col = "red")  
points(chemical, mean2, col="blue", pch="*")  
legend(2,1000, legend=c("Medias en Fabrica 1", "Medias en Fabrica 2"),  
      pch=c("o", "*"))
```

Podemos realizar scatterplots de la siguiente manera

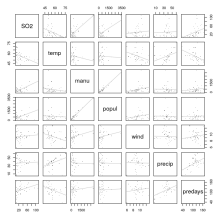
```
plot(popul ~ manu, data = USairpollution)
```

O, dibujar todas las parejas

posibles a la vez, y añadir las rectas de regresión

```
pairs(USairpollution,  
      panel = function(x, y, ...) {  
        points(x, y)      })
```





Implementación Teórica de una DNM

```
DNM <- setRefClass("DNM",  
                  fields = list(p = "numeric",  
                               media = "matrix",  
                               cov = "matrix"))
```

1 Introducción: R

2 R en el análisis multivariante

- Distribución Normal Multivariante. Ejemplos

3 Para ampliar

Computing Machinery and Intelligence Alan Turing (1950)

Artificial Intelligence: A Modern Approach Stuart J. Russell y Peter Norvig

Concrete Problems in AI Safety Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, Dan Mané

The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation Miles Brundage, Shahar Avin et al.