

Cheat sheet Práctica 10: Regresión lineal simple y correlación

Juan Antonio Villatoro García – Grupo BioestadísticaR



**UNIVERSIDAD
DE GRANADA**

Todo el material para el conjunto de actividades de este curso ha sido elaborado y es propiedad intelectual del grupo BioestadísticaR formado por:

Antonio Martín Andrés

Juan de Dios Luna del Castillo,

Pedro Femia Marzo,

Miguel Ángel Montero Alonso,

Christian José Acal González,

Pedro María Carmona Sáez,

Juan Manuel Melchor Rodríguez,

José Luis Romero Béjar,

Manuela Expósito Ruíz,

Juan Antonio Villatoro García.

Todos los integrantes del grupo han participado en todas las actividades, en su elección, construcción, correcciones o en su edición final, no obstante, en cada una de ellas, aparecerán uno o más nombres correspondientes a las personas que han tenido la máxima responsabilidad de su elaboración junto al grupo de BioestadísticaR.

Todos los materiales están protegidos por la Licencia Creative Commons CC BY-NC-ND que permite "descargar las obras y compartirlas con otras personas, siempre que se reconozca su autoría, pero no se pueden cambiar de ninguna manera ni se pueden utilizar comercialmente".

Práctica 10: Regresión lineal simple y correlación



UNIVERSIDAD DE GRANADA

Diagramas de dispersión

Diagrama base de R: Función plot

plot(x, y, main, xlab, ylab, pch, col)

x: variable que se representa en el eje X

y: variable que se representa en el eje Y

main: título del gráfico

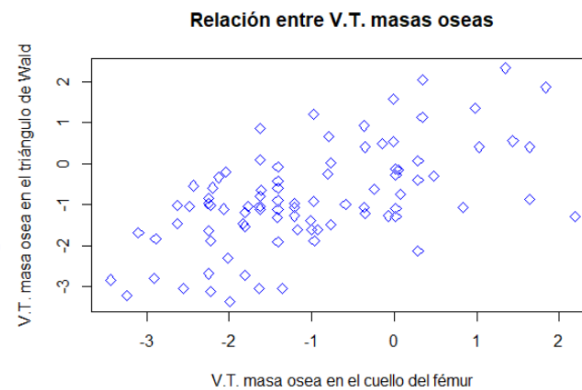
xlab: nombre del eje X

ylab: nombre del eje Y

pch: tipo de puntos (valores del 1-25)

col: color de los puntos

```
> plot(x=osteo$szcue, y=osteo$sztri,
+      main = "Relación entre V.T. masas oseas",
+      xlab = "V.T. masa osea en el cuello del fémur",
+      ylab = "V.T. masa osea en el triángulo de Wald",
+      pch=5, col="blue")
```



Librería ggformula:

Función gg_point: gráfico de dispersión identificando variables cualitativas

gg_point(y ~ x, data, color=~variable, shape,alpha)

x: variable que se representa en el eje X

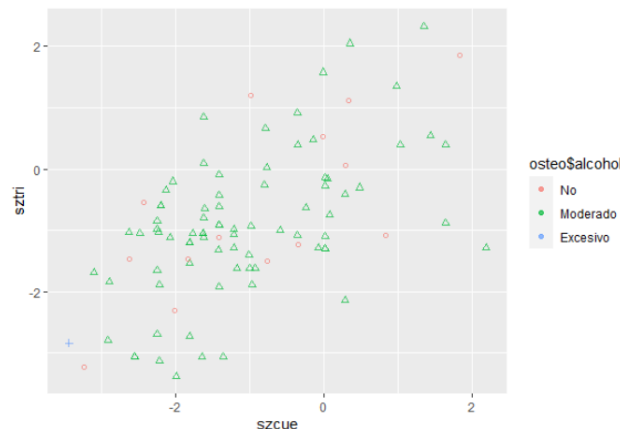
y: variable que se representa en el eje Y

color: variable cualitativa que se distingue por colores

shape: variable cualitativa mediante la cual se quiere señalar en la nube de puntos con una forma atendiendo a una modalidad

apha: intensidad del color de los puntos (de 0 a 1)

```
> library(ggformula)
> gf_point(sztri~szcue,
+          data=osteo,
+          color=~osteo$alcohol,
+          shape=osteo$alcohol,
+          alpha = 0.7)
```



Librería car:

Función scatterplot()

scatterplot(y~x, data, smooth, regLine, boxplots = "", main, xlab, ylab, col, pch)

x: variable que se representa en el eje X

y: variable que se representa en el eje Y

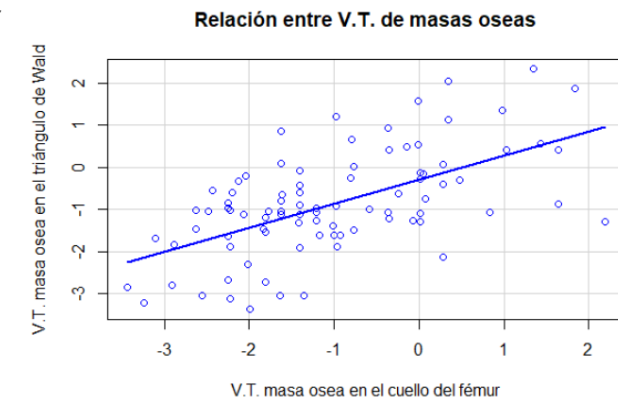
smooth: indica si debe hacer un suavizado. Ponemos TRUE por defecto

regLine: indica si debe representarse la recta de regresión (TRUE) o no (FALSE)

boxplots: indica si hay representar los diagramas de cajas. Poner "" por defecto

main, xlab, ylab, col, pch: igual que en la función plot

```
> scatterplot(sztri ~ szcue, data=osteo,
+             xlab="V.T. masa osea en el cuello del fémur",
+             ylab="V.T. masa osea en el triángulo de Wald",
+             main="Relación entre V.T. de masas oseas",
+             smooth=FALSE, boxplots="", col = "blue")
```



Función scatterplotMatrix(): gráficos matriciales de dispersión

scatterplotMatrix(~v1+v2+v3+..., data, smooth, main)

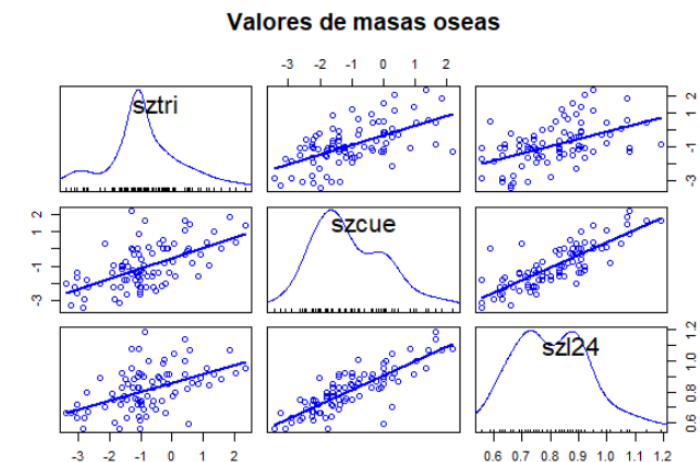
v1,v2,v3: nombres de las variables del dataframe

data: dataframe que contiene los datos

smooth: indica si debe hacer un suavizado (TRUE) o no (FALSE). Ponemos TRUE

main: título del gráfico

```
> scatterplotMatrix(~sztri+szcue+szl24,
+                  data=osteo,
+                  main="Valores de masas oseas",
+                  smooth=FALSE)
```



Práctica 10: Regresión lineal simple y correlación



Regresión lineal simple

Modelo: Función lm

lm(variable dependiente ~ variable independiente)

```
> Modelo.osteo<-lm(osteo$szl24~osteo$sztri)
```

```
> Modelo.osteo
```

```
Call:
lm(formula = osteo$szl24 ~ osteo$sztri)
```

```
Coefficients:
(Intercept)  osteo$sztri
0.85851      0.05714
```

$\widehat{szl24} = 0.859 + 0.057 \cdot sztri$

β_0 β_1

Test de Regresión Lineal Simple:

Función summary(modelo.lm)

summary(modelo.lm)

```
> summary(Modelo.osteo)
```

```
Call:
lm(formula = osteo$szl24 ~ osteo$sztri)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.29880 -0.07359 -0.01182  0.06555  0.38177
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.85851    0.01572   54.608 < 2e-16 ***
osteo$sztri  0.05714    0.01047    5.456 4.09e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.1223 on 92 degrees of freedom
Multiple R-squared:  0.2444,    Adjusted R-squared:  0.2362
F-statistic: 29.76 on 1 and 92 DF,  p-value: 4.093e-07
```

I.C. de los coeficientes: Función confint

Confint(modelo.lm, level = 0,95)

```
> confint(Modelo.osteo)
                2.5 %      97.5 %
(Intercept) 0.82729019 0.88973789
osteo$sztri 0.03633592 0.07793704
```

Análisis descriptivo de los residuos

Significación de la ordenada en el origen

Significación de la pendiente

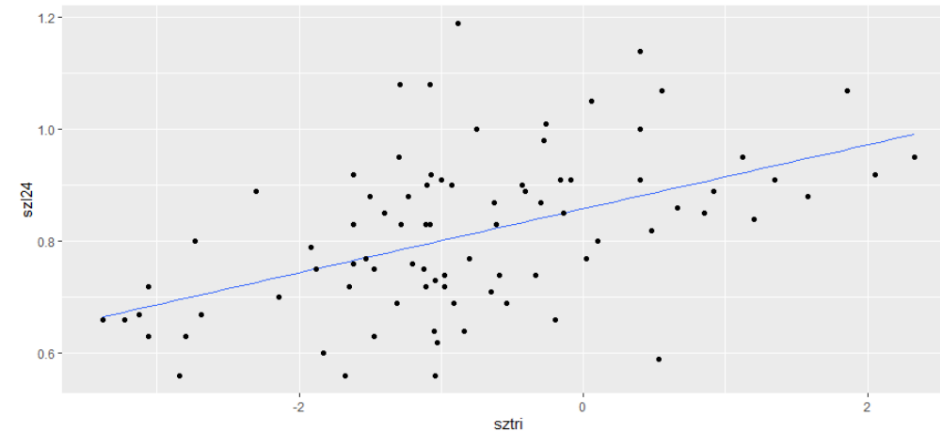
R²: Coeficiente de determinación

Significación de modelo de correlación lineal

Representación de la recta en el diagrama de dispersión

gf_lm(variable dependiente ~ variable independiente, datos = data.frame) %>% gf_point()

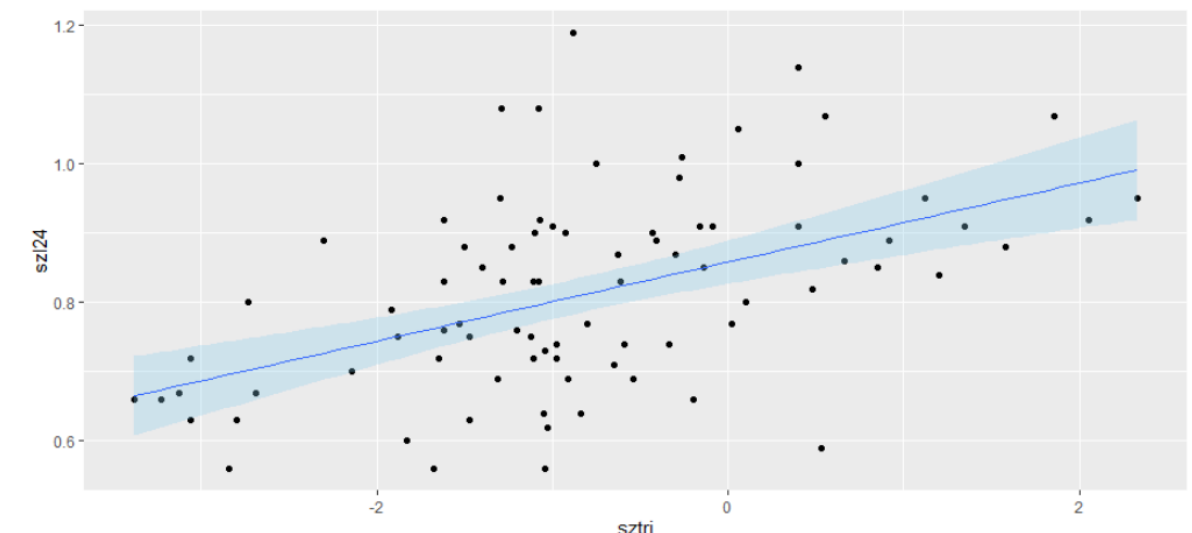
```
> Modelo.osteo<-lm(osteo$szl24~osteo$sztri)
> gf_lm(szl24~sztri, data=osteo)%>% gf_point()
```



Representar la recta y los intervalos de confianza

gf_point(variable dependiente ~ variable independiente, datos = data.frame) %>% gf_lm(interval="confidence", fill="skyblue")

```
> gf_point(szl24~sztri,data=osteo)%>%
+   gf_lm(interval="confidence", fill="skyblue")
```



Práctica 10: Regresión lineal simple y correlación

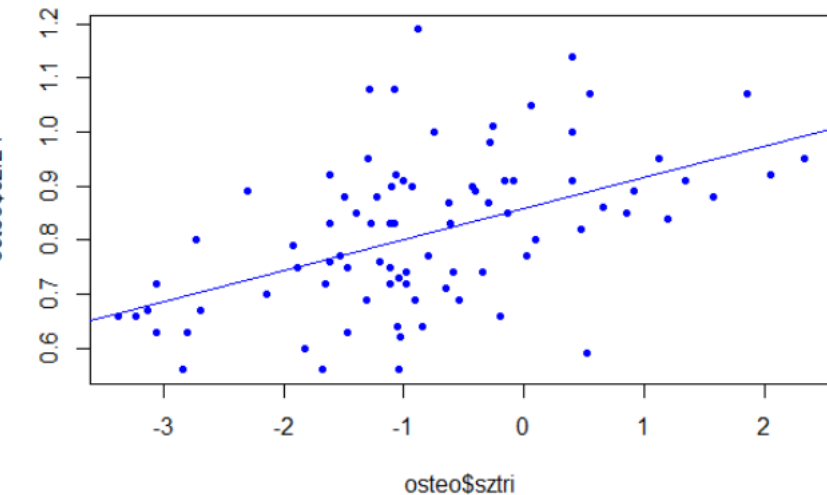


Regresión lineal simple

Otra forma de representar la recta diagrama de dispersión

`abline(modelo.lm)`

```
> plot(x=osteo$sztri,y=osteo$szl24,pch=20,col="blue")
> abline(Modelo.osteo,col="blue")
```



Residuos: Función resid()

`resid(modelo.lm)`

```
> resid(Modelo.osteo)
      1      2      3
0.1800608945 0.0640470505 0.0446206482
      4      5      6
-0.0750925530 -0.2025247608 -0.0556438217
      7      8      9
0.1843383151 0.0349074470 -0.0825202950
     10     11     12
-0.1165198485 0.1540470505 0.0566282400
```

Valores ajustados: Función resid()

`fitted(modelo.lm)`

```
> fitted(Modelo.osteo)
      1      2      3      4      5
0.8899391 0.7659529 0.7853794 0.7950926 0.7625248
      6      7      8      9     10
0.9756438 0.8156617 0.7950926 0.8025203 0.8065198
     11     12     13     14     15
0.7659529 0.8533718 0.8425158 0.8413731 0.8065198
     16     17     18     19     20
```

Correlación lineal

Coeficiente de Pearson

`cor(variable dependiente, variable independiente)`

```
> cor(osteo$sztri,osteo$szl24)
[1] 0.494401
```

Coeficiente de Spearman

`cor(variable dependiente, variable independiente, method="spearman")`

```
> cor(osteo$sztri,osteo$szl24, method = "spearman")
[1] 0.4735113
```

Test de correlación: Función cor.test

Pearson

`cor.test(variable 1, variable 2)`

```
> cor.test(osteo$sztri,osteo$szl24)
```

```
      Pearson's product-moment correlation

data: osteo$sztri and osteo$szl24
t = 5.4555, df = 92, p-value = 4.093e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3242672 0.6335525
sample estimates:
      cor
0.494401
```

Spearman

`cor.test(variable 1, variable 2, method="spearman")`

```
> cor.test(osteo$sztri,osteo$szl24,
+          method = "spearman")
```

```
      Spearman's rank correlation rho

data: osteo$sztri and osteo$szl24
S = 72874, p-value = 1.435e-06
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.4735113
```

Inferencias sobre el modelo

Call:
`lm(formula = osteo$szl24 ~ osteo$sztri)`

Coefficients:
(Intercept) osteo\$sztri $\rightarrow \widehat{szl24} = 0.859 + 0.057 \cdot sztri$

`modelo.lm$coefficients[[1]] +`
`modelo.lm$coefficients[[2]] * valor`

Por ejemplo: valor del szl24 para un sztri de 0.2856

```
> Modelo.osteo$coefficients[[1]]+
+ Modelo.osteo$coefficients[[2]]*0.2856
[1] 0.8748322
```