

Ejemplo de Regresión.

```
> ## Datos por menú
> ## Cargar conjunto de datos
> ## buscar la carpeta y el archivo
> load("/home/norberto/norberto/Bilbao_2020-2021/Datos_2020/A_P.Rdata")

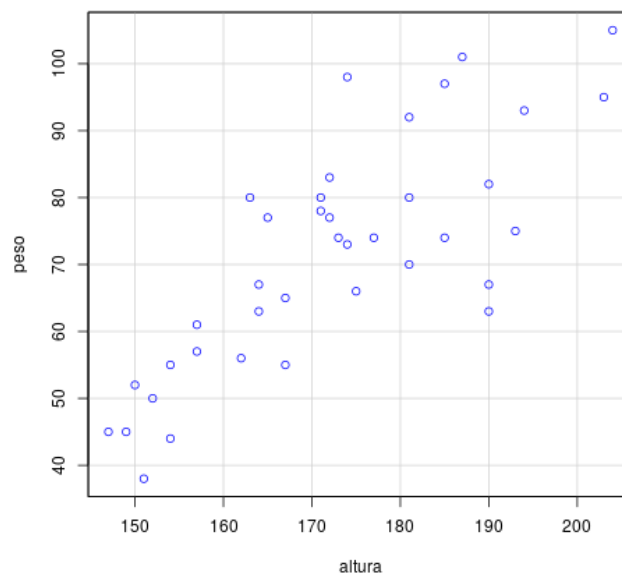
> ## Describir los datos por menú
> ## Resúmenes
> ## Conjunto de datos activos ( describe todo el fichero)
> ## Resúmenes numéricos (sñolo para variables cuantitativas)
> library(abind, pos=16) ## la carga R por defecto
> library(e1071, pos=17) ## la carga R por defecto

> numSummary(A_P[,c("altura", "peso"), drop=FALSE], statistics=c("mean", "sd", "IQR",
"quantiles"), quantiles=c(0,.25,.5,.75,1))
      mean      sd   IQR  0%   25%   50%  75% 100%  n
altura 172.26316 15.34765 21.75 147 162.25 172.0 184 204 38
peso    71.23684 17.13450 22.00 38  58.00  73.5  80 105 38

> numSummary(A_P[,c("altura", "peso")]) ## hace lo mismo que el anterior
      mean      sd   IQR  0%   25%   50%  75% 100%  n
altura 172.26316 15.34765 21.75 147 162.25 172.0 184 204 38
peso    71.23684 17.13450 22.00 38  58.00  73.5  80 105 38

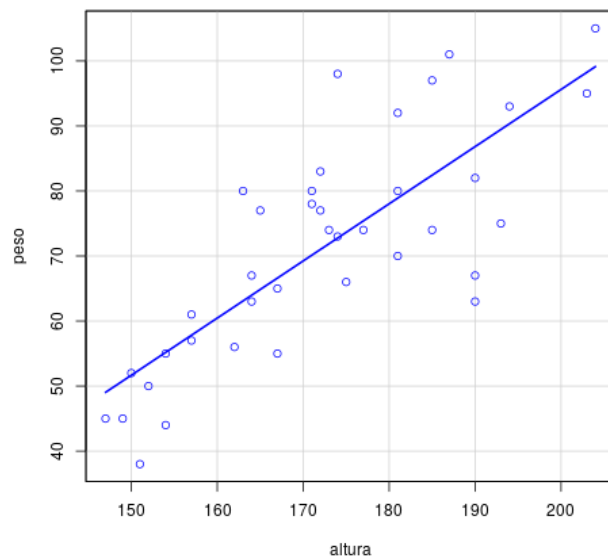
> ## gráficos de dispersión
> ## Gráficas por menú
> ## Diagrama de dispersión
> scatterplot(peso~altura, regLine=FALSE, smooth=FALSE, boxplots=FALSE, data=A_P)

## Para salvar la imagen con formato png.
> png(file = "/home/Sal/graf1.png")
> scatterplot(peso~altura, regLine=FALSE, smooth=FALSE, boxplots=FALSE, data=A_P)
> dev.off()null device      1
```



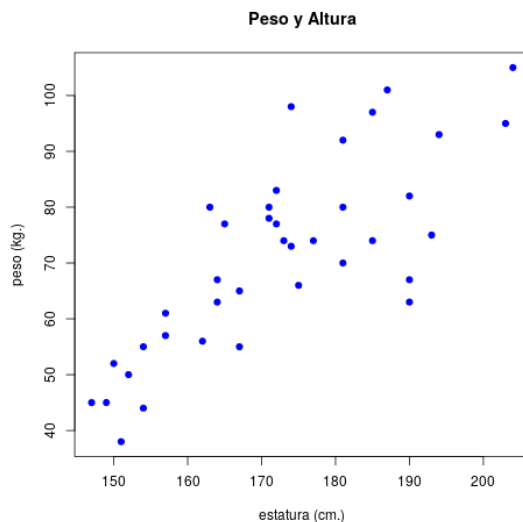
El diagrama de dispersión es muy útil para analizar si existe relación entre dos variables y de que tipo es.

```
> png(file = "/home/Sal/graf3.png")
> scatterplot(peso~altura, regLine=TRUE, smooth=FALSE, boxplots=FALSE, data=A_P)
## opciones: línea de mínimos cuadrados
> dev.off()
null device
1
```



La gráfica sugiere que la nube de puntos está “colocada” alrededor de la recta. Esto indica que tiene sentido predecir el peso en función de la altura con un modelo lineal.

```
> ## gráfico directamente y con título y etiquetas en los ejes
> png(file = "/home/Sal/graf4.png")
> plot(A_P$altura, A_P$peso, main="Peso y Altura", xlab="estatura (cm.)", ylab="peso
(kg.)", col="blue", pch=19)
> dev.off()
null device
```



1

```
> ## Regresión por menú
>   ## Estadísticos
>   ## Ajuste de modelos
>   ## Regresión lineal
> RegModel.1 <- lm(peso~altura, data=A_P)
> summary(RegModel.1) ## resultados de la regresión

> Resumen<- summary(RegModel.1) ## guarda en "Resumen" los resultados de la regresión
```

```

> Resumen ## Muestra lo que hay en Resumen
Call:
lm(formula = peso ~ altura, data = A_P)

Residuals:
    Min       1Q   Median       3Q      Max
-23.8231  -6.1122  -0.5062   5.9644  25.2369

Coefficients:
            Estimate Std. Error t value    Pr(>|t|)
(Intercept) -80.1394    19.8470  -4.038    0.00027 ***
altura       0.8788     0.1148   7.657 0.00000000461 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.71 on 36 degrees of freedom
Multiple R-squared:  0.6195,    Adjusted R-squared:  0.609
F-statistic: 58.62 on 1 and 36 DF,  p-value: 0.000000004607

## Comentarios sobre la salida de la regresión
## beta1= 0.8788
## significa que al aumentar 1 cm la altura el peso predicho aumenta en 0.8788 kg.
## Std. Error representa la desviación típica de los estimadores

## ## t value. Se usa para contrastar la hipótesis nula  $H_0: \beta_1=0$ .
## Si  $\beta_1=0$  equivale a decir que el peso no tiene relación lineal con la altura.
## t value mide lo "separado" que está la estimación de cero la estimación del parámetro
## t value =  $\beta_1 / \text{Std. Error}$ . Por ejemplo:  $0.8788 / 0.1148 = 7.657$ .

##  $\text{Pr}(>|t|)$  es el p-valor del contraste. Hipótesis nula  $H_0: \beta_1=0$ .
## p-valor del contraste: verosimilitud de la muestra bajo la hipótesis nula  $H_0$ 
##  $\text{Pr}(>|t|) = 0.00000000461$ . Significa que si  $\beta_1$  fuera cero, el resultado obtenido
ocurriría menos de cinco veces cada mil millones de pruebas. Por tanto se rechaza que
 $\beta_1=0$ .

## Multiple R-squared: 0.6195, es el coeficiente de determinación. Significa que el
cambio de la estatura explica el 61,95 de la variabilidad del peso.

## contraste de regresión.  $H_0$ : los coeficientes de las variables independientes son cero
## En este ejemplo sólo hay una variable independiente. Si  $H_0$  es cierta entonces la
variable dependiente no tiene relación lineal con las variables predictoras y el modelo
carece de interés
## F-statistic: 58.62 on 1 and 36 DF, p-value: 0.000000004607. Se rechaza  $\beta_1=0$ .

> names(RegModel.1) ## nombres de los objetos contenidos en RegModel
[1] "coefficients" "residuals" "effects" "rank" "fitted.values"
"assign" "qr" "df.residual" "xlevels" "call"
[11] "terms" "model"

> RegModel.1$coefficients ## "coefficients" : valores estimados de los betas
(Intercept) altura
-80.1394391 0.8787502

> RegModel.1$residuals[1:5] ## "residuals": residuales de la regresión
1 2 3 4 5
25.236908 2.115658 -11.188089 7.873158 16.903160

> ## "fitted.values": predicciones del peso para la muestra.

> names(Resumen) ## nombres de los objetos contenidos en summary(RegModel)
[1] "call" "terms" "residuals" "coefficients" "aliases"
"sigma" "df" "r.squared" "adj.r.squared" "fstatistic"
[11] "cov.unscaled"

> Resumen$sigma ## "sigma": estimación de la desviación típica de los residuales
[1] 10.71453

> Resumen$adj.r.squared ## "adj.r.squared" :R2 ajustado por el número de variables
independientes
[1] 0.6089757

```

```
> ## Resumen$fstatistic contiene: valor del estadístico y los grados de libertad
> esta<- Resumen$fstatistic ## copia en "esta" el contenido de Resumen$fstatistic
## pf(58.62, 1, 36) calcula la función de distribución de una F, con 1 y 36 grados de libertad, en el punto 58.62.
> 1-pf(esta[1], esta[2], esta[3]) ## p-valor del contraste
      value
0.000000004607337
```

Cuestiones a resolver:

Interpreta los resultados del modelo de regresión. Comentada en clase

De acuerdo con este modelo, cuál será el peso de dos personas cuyas estaturas sean 180 cm y 90 cm? Analiza si tiene sentido estos resultados. Comentada en clase