INTRODUCCIÓN A REGRESIÓN LINEAL

Simple y Múltiple

Introducción

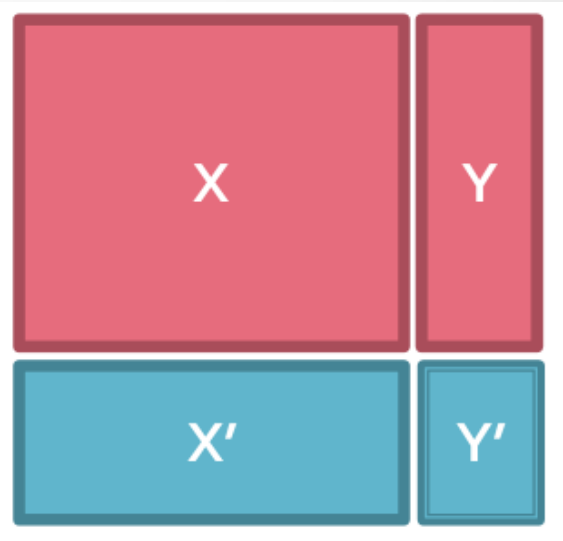
Aprendizaje Supervisado

Predicción: estimar una función **f(x)** de forma que **y = f(x)**

Donde Y puede ser:

- **Número real:** Regresión
- **Categorías**: Clasificación

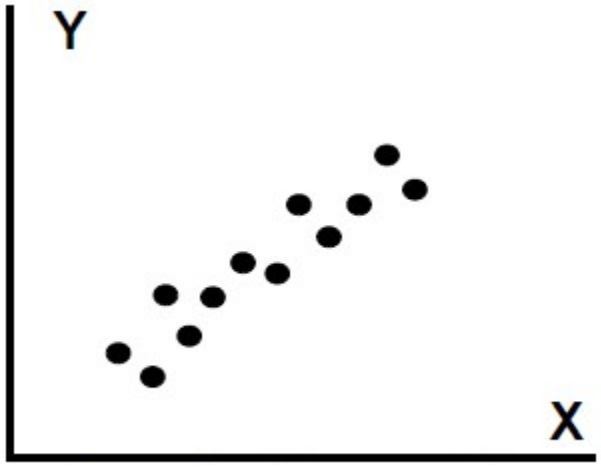
La regresión lineal se ocupa de investigar la relación entre dos o más variables continuas.



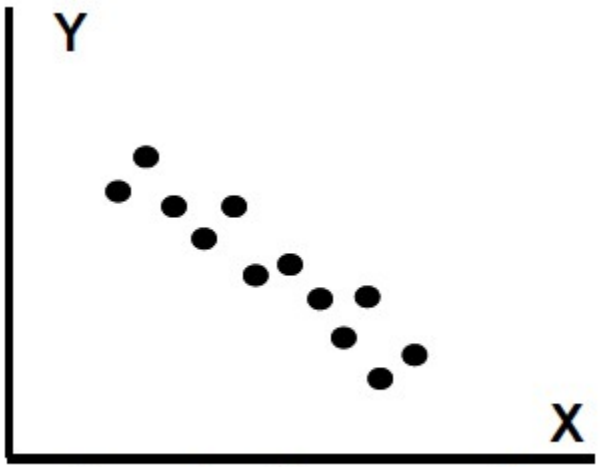
Conjunto de **Training** y **test**

Coeficiente de correlación de Pearson

Buscamos explicar situaciones donde aparece una relación entre X e Y, como las siguientes:



Asociación positiva



Asociación negativa

Definición

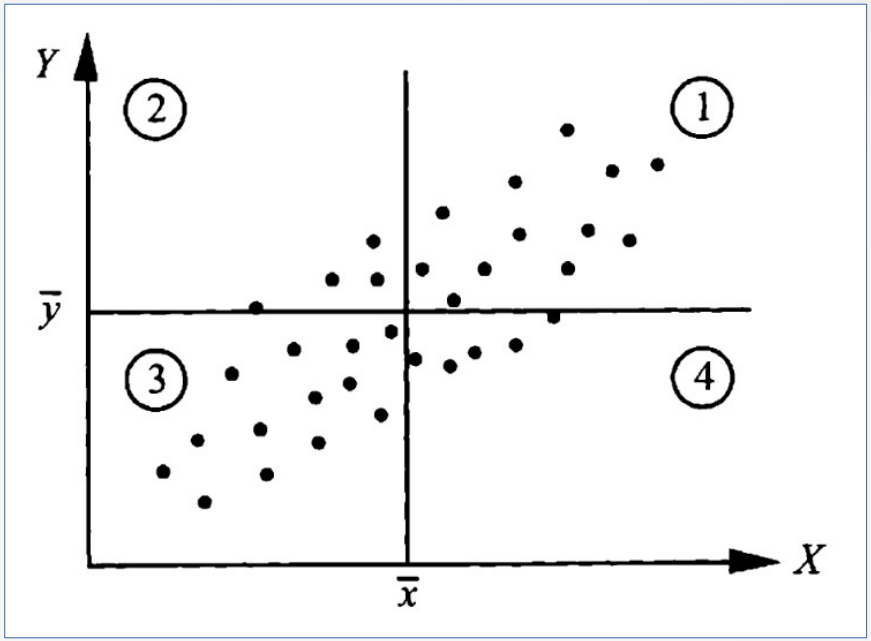
Coeficiente de Correlación Muestral

$$\text{Cor}(Y,X) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{s_y}\right)\left(\frac{x_i-\bar{x}}{s_x}\right)$$

$$\begin{aligned}\text{Cor}(Y,X) &= \frac{\text{Cov}(Y,X)}{s_y s_x} \\ &= \frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{\sqrt{\sum(y_i-\bar{y})^2\sum(x_i-\bar{x})^2}}\end{aligned}$$

El signo de R

El signo depende de **COV(Y,X)**



Cuadrante	$y_i - \bar{y}$	$x_i - \bar{x}$	$(y_i - \bar{y})(x_i - \bar{x})$
1	+	+	+
2	+	−	−
3	−	−	+
4	−	+	−

El problema de la COV es que es sensible a las unidades de las observaciones

Propiedades de r (y de rho ρ)

- **$-1 \leq r \leq 1$**. El valor del coeficiente r está entre 1 y menos 1 porque puede probarse que el denominador es más grande que el numerador
- El **valor absoluto de r**, |r| mide la fuerza de la asociación lineal entre X e Y, a mayor valor absoluto, hay una asociación lineal más fuerte entre X e Y
- El caso particular **r = 0** indica que no hay asociación lineal entre X e Y.
- El caso **r = 1** indica asociación lineal perfecta. O sea que los puntos están ubicados sobre una recta de pendiente (o inclinación) positiva.
- En el caso **r = −1** tenemos a los puntos ubicados sobre una recta de pendiente negativa (o sea, decreciente).

Propiedades de r (y de rho ρ)

- El **signo de r** indica que hay asociación positiva entre las variables (si r > 0); o asociación negativa entre ellas (si r < 0).
- **r = 0,90** indica que los puntos están ubicados muy cerca de una recta creciente.
- **r = 0,80** indica que los puntos están cerca, pero no tanto, de una recta

  creciente.
- **r no depende de las unidades** en que son medidas las variables.
- Los roles de X e Y son simétricos para el cálculo de r.

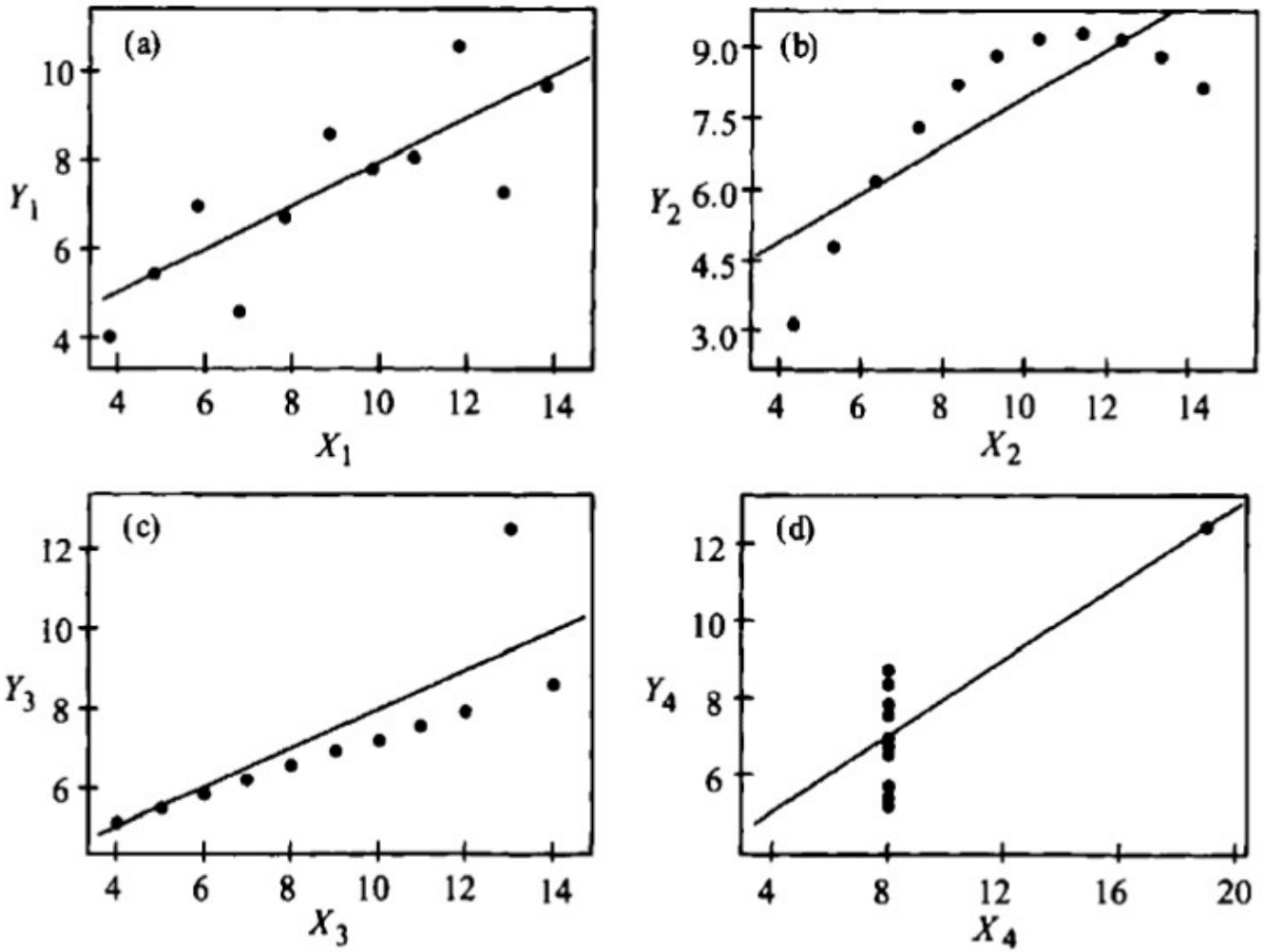
Ojo al piojo: el coeficiente de correlación de Pearson es muy sensible a observaciones atípicas. Hay que hacer siempre un scatter plot de los datos antes de resumirlos con r.

Warning! Con r

- Cabe hacer un comentario respecto de la interpretación del coeficiente de correlación.
  - **Altos grados de asociación lineal** entre X e Y **no son señales de causalidad**, es decir, una relación de causa y efecto entre ambas variables.

Cuarteto de Anscombe

Propiedad	Valor
Media de X	9.0
Varianza X	11.0
Media de Y	7.5
Varianza Y	4.12
Cor(X,Y)	**0.816**



Moraleja: Siempre hay que visualizar los datos

https://es.wikipedia.org/wiki/Cuarteto_de_Anscombe

Regresión Lineal Simple: Introducción

Un modelo de regresión es un modelo que permite describir cómo influye una variable X sobre otra variable Y.

- X : variable explicativa, independiente o covariable.
- Y : variable dependiente o respuesta.

El objetivo es obtener estimaciones razonables de Y para distintos valores de X a partir de una muestra de n pares:

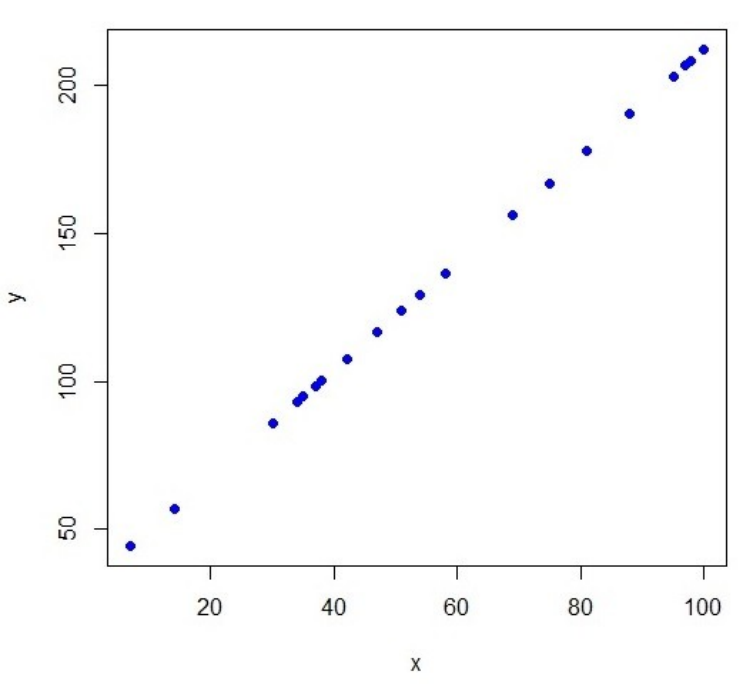
$$(x_1 , y_1), \ldots , (x_n , y_n).$$

Tipos de relación

Determinística: Conocido el valor de **X** , el valor de **Y** queda perfectamente establecido. Es decir, **Y = f(X)**

Ejemplo: La relación existente entre la temperatura en grados centígrados (X) y grados Fahrenheit (Y) es:

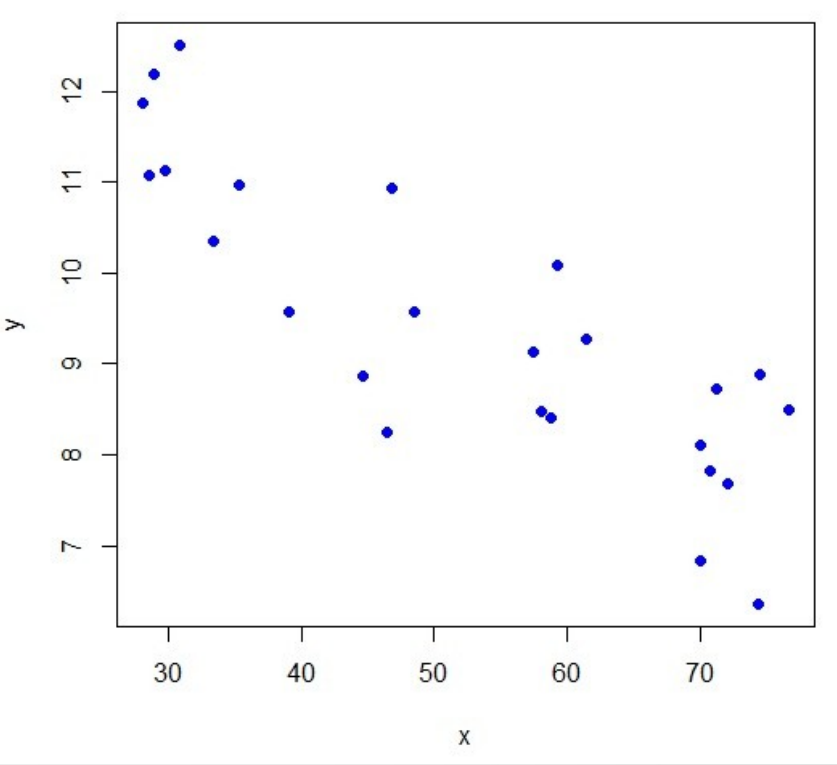
Y = 32 + 1.8X



Tipos de relación

No determinística: Conocido el valor de **X** , el valor de **Y** no queda perfectamente establecido. Son del tipo: **Y = f (X) + ε** donde **ε** es un error desconocido (**variable aleatoria**).

Ejemplo: En una planta a vapor, en 25 meses, se observó el promedio mensual de temperatura atmosférica (en Farenheit) (X) y la cantidad de vapor consumido (en libras) (Y).



Regresión lineal simple

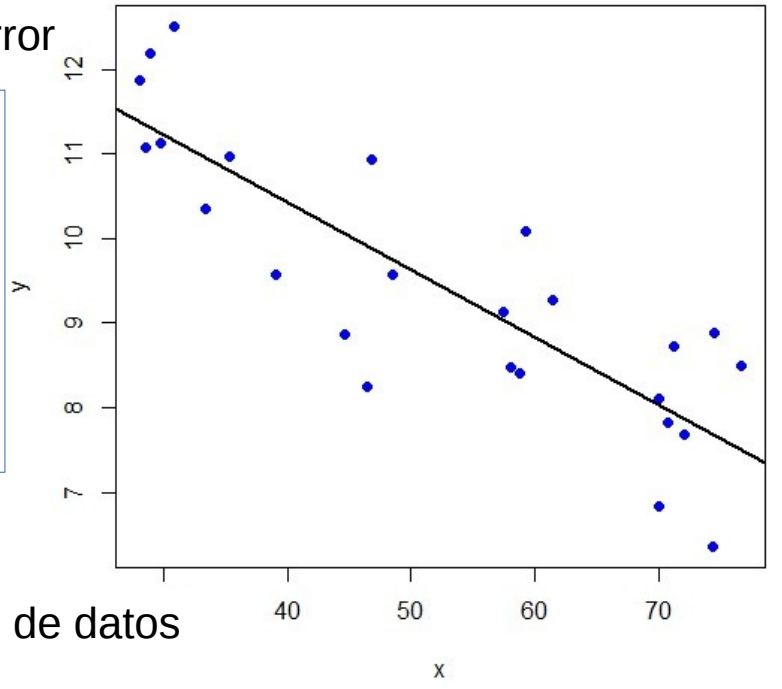
Consiste en describir la relación entre las dos variables mediante una recta.
El caso **Determinístico**, no nos interesa. Con dos puntos ya bastaría para encontrar la recta.

No determinística: El ejemplo de la planta a vapor:

- La función que proponemos para modelar la relación es $Y = \beta_0 + \beta_1 x + \varepsilon$ pero, en este caso, β_0 y β_1 son constantes desconocidas
  - también llamados **parámetros** o coeficientes del modelo de regresión
  - En cuanto a ε es la perturbación aleatoria o error

- Se asume que en el rango de las observaciones estudiadas, la ecuación lineal provee una aceptable aproximación a la relación existente entre **X** e **Y**.

- **Y** es aproximadamente una función lineal de **X** y ε mide las discrepancias en esa aproximación.



Problema: Ajustar la recta que represente al conjunto de datos de la mejor manera

Regresión lineal simple

A través de los **parámetros**:

β_0: ordenada al origen (o *intercept*)

β_1: pendiente (o *slope*)

calculamos

$\hat{y}_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ valores ajustados o predichos

$e_i = y_i - \hat{y}_i$ residuos.

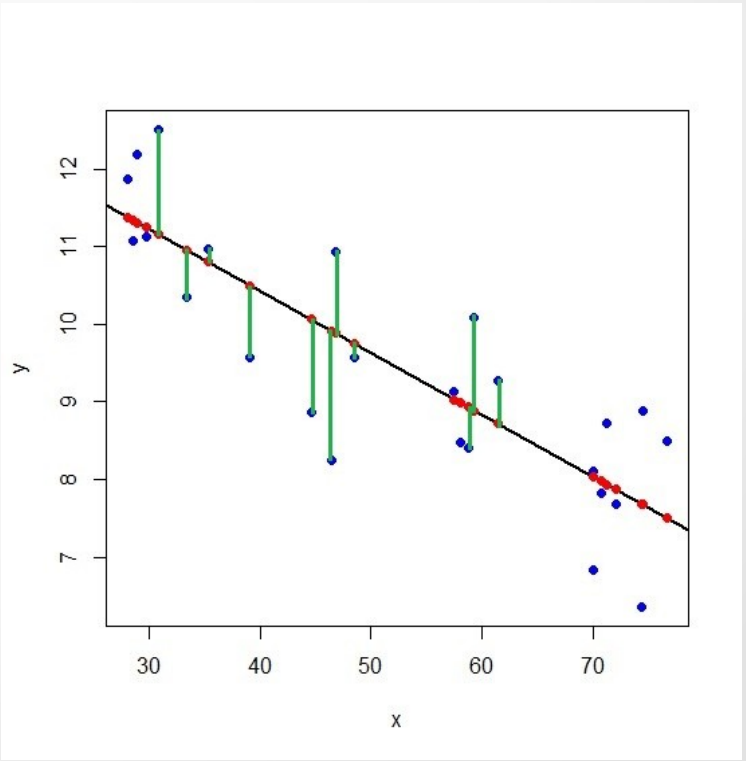
Objetivo: Hallar los mejores coeficientes β_0 y β_1 que representan la relación lineal entre las variables.

Estimación de los parámetros

Los parámetros son estimados utilizando el método de los **mínimos cuadrados**, que minimiza la suma de los cuadrados de las **distancias verticales** desde cada punto a la recta.

Las distancias verticales representan los errores en la variable de respuesta.

Gráficamente, lo que se resuelve es la **minimización** de las distancias entre los **valores observados** y los **valores predichos**



Estimación de los parámetros

Los parámetros estimados de β_0 y β_1 es equivalente a encontrar la recta que mejor ajusta los puntos en el gráfico de dispersión.

Los errores pueden ser escritos como:

$$\varepsilon_i = y_i - \beta_0 - \beta_1 X_i, i = 1,2,\ldots,n$$

Las sumas de cuadrados de las distancias (o SSE Suma de cuadrado del Error) pueden escribirse como:

$$S(\beta_{0,}\beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 X_i)^2$$

Estimación de los parámetros

Los parámetros estimados se denotan $\hat{\beta}_0$ y $\hat{\beta}_1$ estos son los que minimizan la $S(\beta_{0,}\beta_1)$

Para encontrar los parámetros calculamos las derivadas parciales de SSE respecto a β_0 y β_1 .

$$\frac{\partial E}{\partial \beta_0} = -2\sum_{i=1}^{m}(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial E}{\partial \beta_1} = -2\sum_{i=1}^{m}(y_i - \beta_0 - \beta_1 x)x_i = 0 \qquad (1)$$

Luego igualamos las derivadas a cero y resolvemos la ecuación para encontrar los parámetros.

Del sistema de ecuaciones (1) se obtienen las soluciones normales:

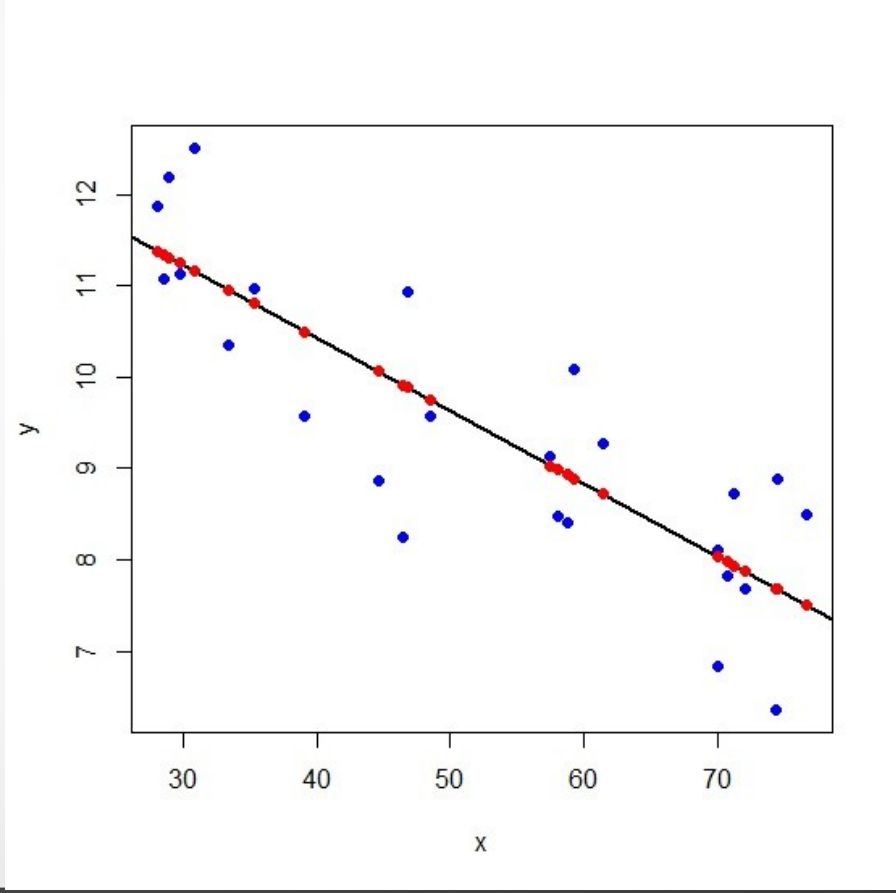
$$\hat{\beta}_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Estimación de los parámetros

Un vez hallada la recta, es decir, hallados $\hat{\beta}_0$ y $\hat{\beta}_1$ tenemos que los valores ajustados en cada punto son:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$



Validación del Modelo: R^2

- Una vez ajustado nuestro modelo lineal debemos evaluar la calidad del modelo.
- Una medida muy común es el **Coeficiente de Determinación R^2**
- Para calcularlo se deben calcular otros errores distintos a los errores cuadráticos SSE.
- Se define a la **suma cuadrática total** (**SST**) como el error predictivo cuando usamos la media de Y para predecir la variable de respuesta Y (es muy similar a la varianza de la variable)

$$SST = \sum_{i}^{n} (y_i - \bar{y})^2$$

- Luego tenemos a la **suma de los cuadrados explicada por el modelo** (**SSM**) que nos indica la variabilidad de los valores predichos por el modelo respecto a la media:

$$SSM = \sum_{i}^{n} (\hat{y}_i - \bar{y})^2$$

Validación del Modelo: R^2

- Se define el coeficiente de determinación para un modelo lineal, como:

$$R^2 = \frac{SSM}{SST} = \frac{\sum_i^n (\hat{y}_i - \bar{y})^2}{\sum_i^n (y_i - \bar{y})^2}$$

- El coeficiente adquiere valores entre 0 y 1.
- Mientras mas cercano a 1 sea, mejor será la calidad el modelo ajustado.
- El valor de R^2 es equivalente a la correlación lineal de Pearson entre y e ŷ al cuadrado.

$$R^2 = cor(y, \hat{y})^2$$

- El coeficiente de determinación da una idea de la proporción de puntos que son explicados por el modelo ajustado.

Supuestos de la RLS

Los supuestos bajo los cuales serán **válidas las inferencias** que haremos más adelante sobre el modelo

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \varepsilon_i$$

son los siguientes:

1. los $\boldsymbol{\varepsilon_i}$ tiene media cero, $\mathbf{E(\varepsilon_i)=0}$.
2. los $\boldsymbol{\varepsilon_i}$ tienen todos la misma varianza desconocida que llamaremos $\boldsymbol{\sigma^2}$ y que es el otro parámetro del modelo, $\mathbf{Var(\varepsilon_i) = \sigma^2}$. A este requisito se lo llama **homoscedasticidad**. O varianza constante.
3. los $\boldsymbol{\varepsilon_i}$ tienen **distribución Normal**.
4. los $\boldsymbol{\varepsilon_i}$ son **independientes** entre sí, y son **no correlacionados** con las X_i.

En resumen: $\varepsilon_i \sim N(0, \sigma^2)$, $1 \leq i \leq n$, independientes entre sí.

Ejemplo USArrests

- Ejemplo, construcción de un modelo lineal. En R podemos usar **lm()** y~x (y = f(x))
- Dataset **USArrests** que tiene información de arrestos ocurridos en Estados Unidos durante 1973 con una observación por estado.

- Variables:

  Murder: arrestos por homicidio (cada 100.000 hab.).

  Assault : arrestos por asalto (cada 100.000 hab.).

  UrbanPop: porcentaje de la población total del estado.

  Rape: arrestos por violación (cada 100.000 hab.).

Ejemplo USArrests

Verificamos las relaciones entre las variables y analizamos si es posible hacer un análisis de regresión lineal

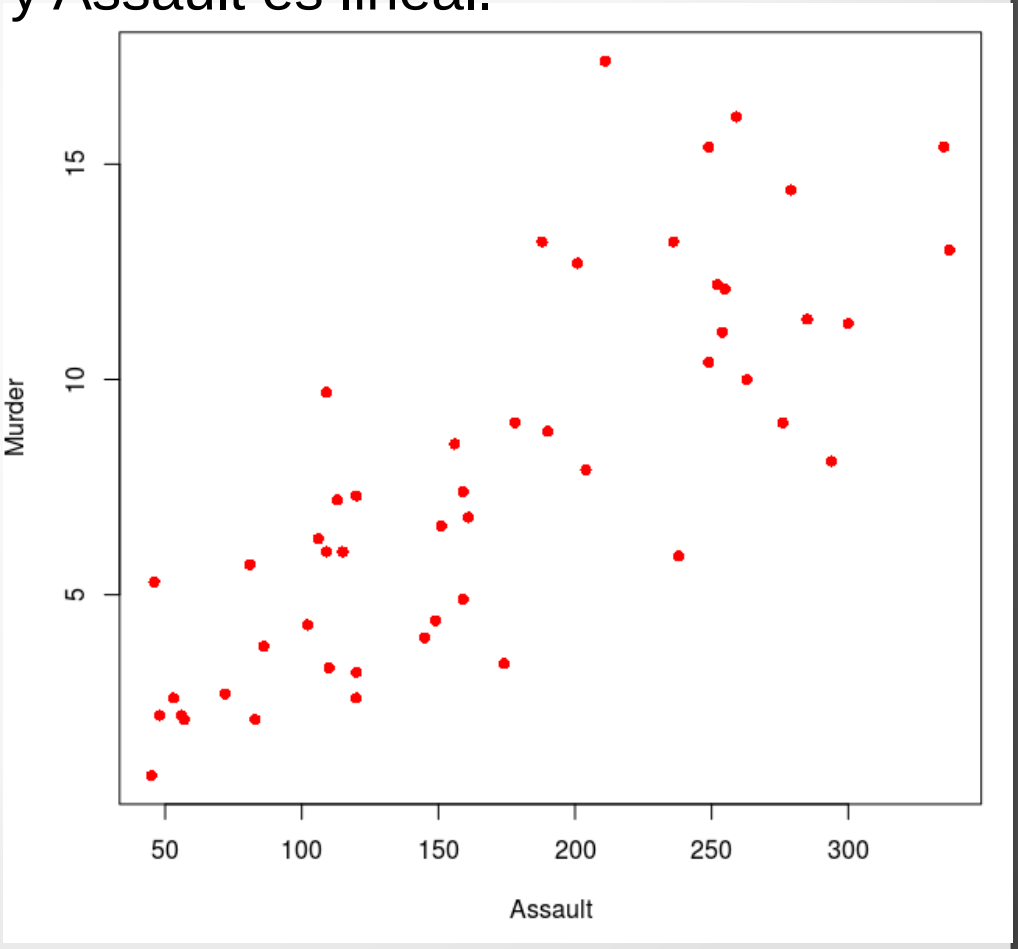
```
> cor(USArrests)
             Murder   Assault   UrbanPop       Rape
Murder   1.00000000 0.8018733 0.06957262 0.5635788
Assault  0.80187331 1.0000000 0.25887170 0.6652412
UrbanPop 0.06957262 0.2588717 1.00000000 0.4113412
Rape     0.56357883 0.6652412 0.41134124 1.0000000
```

- Existe una correlación positiva entre Murder y Assault.

Ejemplo USArrests

Además verificamos a través del **gráfico de dispersión** si la relación positiva entre Murder y Assault es lineal.

- Si bien los puntos no siguen una recta perfecta, se puede observar que conforme se incrementan los asaltos las muertes también se incrementan.



Ejemplo USArrests

- Modelamos los asesinatos en función de los asaltos utilizando una regresión lineal simple:

$$Murder(Assault) = \beta_0 + \beta_1 * Assault$$

```
> ajuste.lineal <- lm(Murder~Assault,USArrests)
> ajuste.lineal
Call:
lm(formula = Murder ~ Assault, data = USArrests)
Coefficients:
(Intercept)      Assault
    0.63168      0.04191
```

$$Murder(Assault) = 0.63168 + 0.04191 * Assault$$

Ejemplo USArrests

- Verificamos la significancia de los coeficientes

  Se debe contrastar la hipótesis nula que tanto β_0 como β_1 son distintos de cero. Es decir:

  $H_0: \beta_0 \neq 0$ $\quad$ $H_0: \beta_1 \neq 0$

  $H_1: \beta_0 = 0$ $\quad$ $H_1: \beta_1 = 0$

- De manera similar lo hace con todos los coeficientes involucrados en el problema de estimación.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.631683   0.854776   0.739    0.464
Assault     0.041909   0.004507   9.298  2.6e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.629 on 48 degrees of freedom
Multiple R-squared:  0.643,   Adjusted R-squared:  0.6356
F-statistic: 86.45 on 1 and 48 DF,  p-value: 2.596e-12
```

Intervalos de Confianza

- Los IC permiten complementar la información que proporcionan los contraste de hipótesis a la hora de expresar el grado de incertidumbre en nuestras estimaciones.
- Obtenemos los correspondientes intervalos de confianza para cada parámetro del modelo con nivel significación al 95%

$$H_0: \beta_0 = \beta_1 = 0$$

```
> confint(ajuste.lineal, level = 0.95)
                    2.5 %      97.5 %
(Intercept) -1.08695906 2.35032438
Assault      0.03284621 0.05097104
```

Interpretamos los intervalos:

- La intersección β_0 mantiene la coherencia que observamos en la prueba t, el IC contiene al cero.
- Con una probabilidad del 95%, el IC de Assault está entre 0.009 y 0.09 .

Regresión Lineal Múltiple

- Supongamos que tenemos n observaciones para una variable dependiente Y, además, tenemos p variables independientes o predictoras X_1, X_2, …, X_p.

Observation Number	Response Y	Predictors X_1	X_2	…	X_p
1	y_1	x_{11}	x_{12}	…	x_{1p}
2	y_2	x_{21}	x_{22}	…	x_{2p}
3	y_3	x_{31}	x_{32}	…	x_{3p}
⋮	⋮	⋮	⋮	⋮	⋮
n	y_n	x_{n1}	x_{n2}	…	x_{np}

- La relación entre Y y X_1, X_2, …, X_p es formulada como un modelo lineal de la forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$

Donde:

$\beta_0, \beta_1, \beta_2 \ldots \beta_p$ son los coeficientes del modelo de regresión y ε es el error o perturbación aleatoria.

Regresión Lineal Múltiple

- Se asume que para un conjunto de valores de $X_1, X_2, \ldots, X_p$ que caen dentro del rango de los datos la ecuación:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$

provee una aceptable aproximación de la verdadera relación entre Y y las X

- Decimos que Y es **aproximadamente una función lineal** de las X y ε mide la discrepancia en esa aproximación.

- ε contiene información **no sistemática** para determinar Y que no es capturada por las X.

Estimación de los parámetros I

- A partir de una muestra queremos estimar los parámetros β_0, β_1, β_2 ... β_p
- De manera similar a la Regresión Lineal Simple vamos a utilizar el método de **ajuste por mínimos cuadrados,** que es minimizar la Suma de los Cuadrados del Error.

- El error se calcula como:

$$\varepsilon_i = Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \ldots - \beta_p X_{ip}, i = 1,2,\ldots,n.$$

- La suma de cuadrados del error:

$$S(\beta_0, \beta_1, \ldots, \beta_p) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \ldots - \beta_p X_{ip})^2$$

Estimación de los parámetros II

- Los mínimos cuadrados estimados $\hat{\beta}_0, \hat{\beta}_1, \ldots \hat{\beta}_p$ que minimizan la $S(\beta_0, \beta_1, \ldots, \beta_p)$ se obtienen por la solución de un sistema de ecuaciones lineales conocidos como **ecuaciones normales.**

- El $\hat{\beta}_0$ estimado es denominado intersección o constante y los $\hat{\beta}_j$ estimados son los coeficientes de la regresión de la variable predictora X_j

- Se asume que el sistema de ecuaciones tiene solución y es única.

NOTA: No vamos a ver la resolución del sistema de ecuaciones normales. Para consultar el modelo matricial se puede leer el apéndice del Cap. III de Chatterjee (pág. 82)

Estimación de los parámetros III

- Utilizando los coeficientes de regresión estimados $\hat{\beta}_0, \hat{\beta}_1, \ldots \hat{\beta}_p$ podemos escribir ahora la ecuación de regresión de mínimos cuadrados ajustados:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_p X_p$$

- Para cada observación podemos calcular los valores ajustados, como:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \ldots + \hat{\beta}_p X_{ip}, i = 1,2,\ldots,n$$

- Los residuales de los mínimos cuadrados ordinarios serían:

$$\varepsilon_i = Y_i - \hat{Y}_i, i = 1,2,\ldots,n$$

Interpretación de los coeficientes

- La interpretación de los coeficientes en regresión lineal múltiple suele prestarse a confusión.
  - La Ecuación de Regresión Lineal Simple representa una recta.
  - Mientras que el sistema de ecuaciones de la regresión múltiple representa:
    - un plano (para 2 predictores)
    - un hiperplano (para más de 2 predictores)

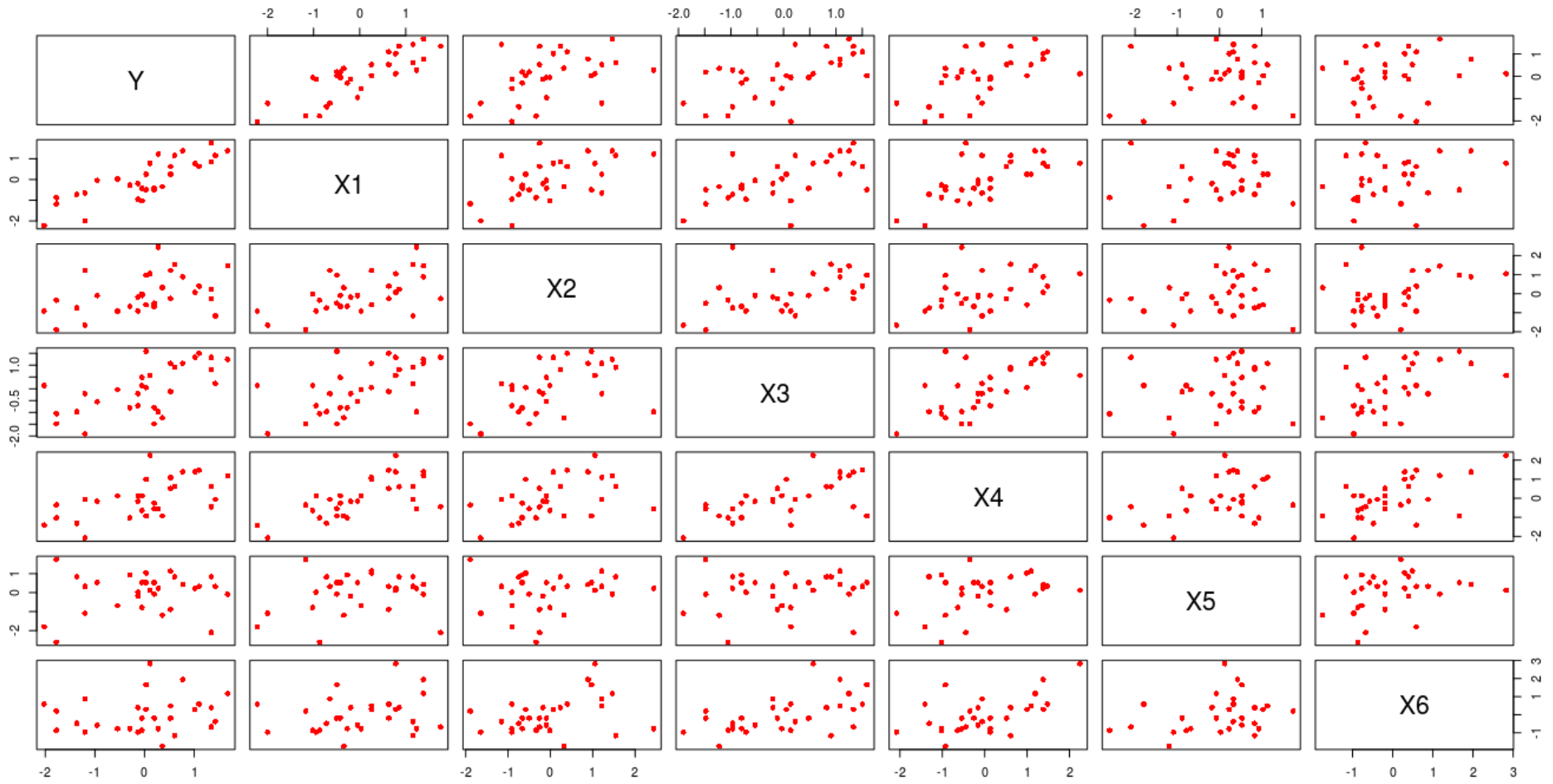
- En regresión lineal múltiple, el coeficiente β_0 es llamado coeficiente constante y es el valor de Y cuando $X_1 = X_2 = \ldots = X_p = 0$
- El coeficiente de regresión β_j, j=1, 2, …, p tiene varias interpretaciones.
  - Este puede ser interpretado como el cambio en Y cuando X_j se modifica en una unidad y todos los demás predictores permanecen constantes.

Ejemplo: Desempeño de Supervisores

Table 3.2 Description of Variables in Supervisor Performance Data

Variable	Description
Y	Overall rating of job being done by supervisor
X_1	Handles employee complaints
X_2	Does not allow special privileges
X_3	Opportunity to learn new things
X_4	Raises based on performance
X_5	Too critical of poor performance
X_6	Rate of advancing to better jobs

Ejemplo: Desempeño de Supervisores



Ejemplo: Desempeño de Supervisores

```
> cor(sup.scaled)
           Y         X1         X2         X3         X4         X5         X6
Y  1.0000000 0.8254176 0.4261169 0.6236782 0.5901390 0.1564392 0.1550863
X1 0.8254176 1.0000000 0.5582882 0.5967358 0.6691975 0.1877143 0.2245796
X2 0.4261169 0.5582882 1.0000000 0.4933310 0.4454779 0.1472331 0.3432934
X3 0.6236782 0.5967358 0.4933310 1.0000000 0.6403144 0.1159652 0.5316198
X4 0.5901390 0.6691975 0.4454779 0.6403144 1.0000000 0.3768830 0.5741862
X5 0.1564392 0.1877143 0.1472331 0.1159652 0.3768830 1.0000000 0.2833432
X6 0.1550863 0.2245796 0.3432934 0.5316198 0.5741862 0.2833432 1.0000000
```

Ejemplo: Desempeño de Supervisores

```
> ajuste.lineal.m=lm(Y~., sup.scaled)
> summary(ajuste.lineal.m)

Call:
lm(formula = Y ~ ., data = sup.scaled)

Residuals:
     Min       1Q   Median       3Q      Max
-0.89889 -0.35781  0.02595  0.45533  0.95288

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.717e-16  1.060e-01   0.000 1.000000
X1           6.707e-01  1.761e-01   3.809 0.000903 ***
X2          -7.343e-02  1.364e-01  -0.538 0.595594
X3           3.089e-01  1.625e-01   1.901 0.069925 .
X4           6.981e-02  1.892e-01   0.369 0.715480
X5           3.120e-02  1.195e-01   0.261 0.796334
X6          -1.835e-01  1.506e-01  -1.218 0.235577
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5806 on 23 degrees of freedom
Multiple R-squared:  0.7326,  Adjusted R-squared:  0.6628
F-statistic:  10.5 on 6 and 23 DF,  p-value: 1.24e-05
```

Evaluación del modelo R^2 ajustado

- El valor de R2 es usado como una medida de resumen para juzgar el ajuste de un modelo lineal a un conjunto de datos.
- Un valor alto de R2 no significa necesariamente ajusta bien a los datos.
- El **R^2-ajustado** permite juzgar la efectividad del ajuste.
- Este se define como:

$$R_a^2 = 1 - \frac{SSE/(n-p-1)}{SST/(n-1)}$$

Verificación de los supuestos

Linealidad (**Datos**): Error de especificación. Lo podemos verificar gráficamente a través de un diagrama o gráfico de dispersión

Independencia (**Residuos**). Este supuesto asume que los residuos no están auto-correlacionados, por lo cual son independientes. Se puede verificar con la prueba de **Durbin Watson.**

Homocedasticidad (**Residuos**). Los residuos en las predicciones son constantes en cada predicción (**varianza constante**). Este supuesto valida que los residuos no aumenta ni disminuye cuando se predicen valores cada vez más altos o mas pequeños. Lo verificamos gráficamente.

Normalidad (**Residuos**) Se asume que los residuos deben **seguir una distribución Norma**l, la ausencia de normalidad supone poca precisión en los intervalos de confianza creados por el modelo. Lo verificamos gráficamente y con una prueba de Shapiro-Wilk.

No-Colinealidad (**Datos**).

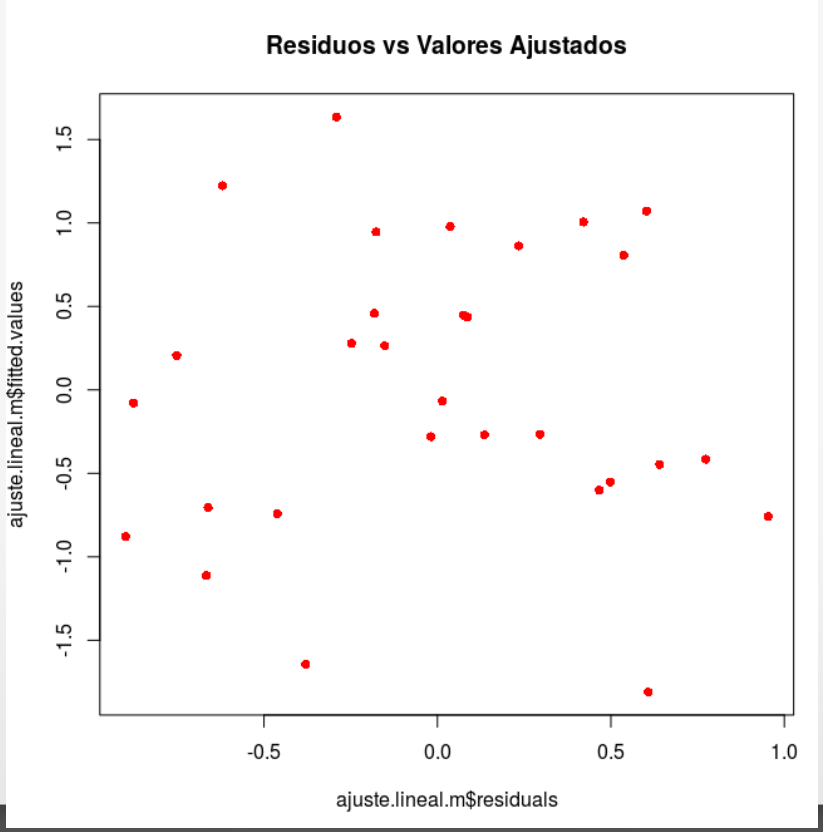
Verificación de los supuestos: Independencia

- **Independencia** (**Residuos**). Este supuesto asume que los residuos no están auto-correlacionados, por lo cual son independientes. Se puede verificar con la prueba de **Durbin Watson.**
- El valor de la prueba tiene que dar entre **1.5 y 2.5** para garantizar la independencia de los residuos.

```
> dwtest(Y~., data=sup.scaled)
        Durbin-Watson test
data:  Y ~ .
DW = 1.7953, p-value = 0.2875
alternative hypothesis: true autocorrelation is greater
than 0
```

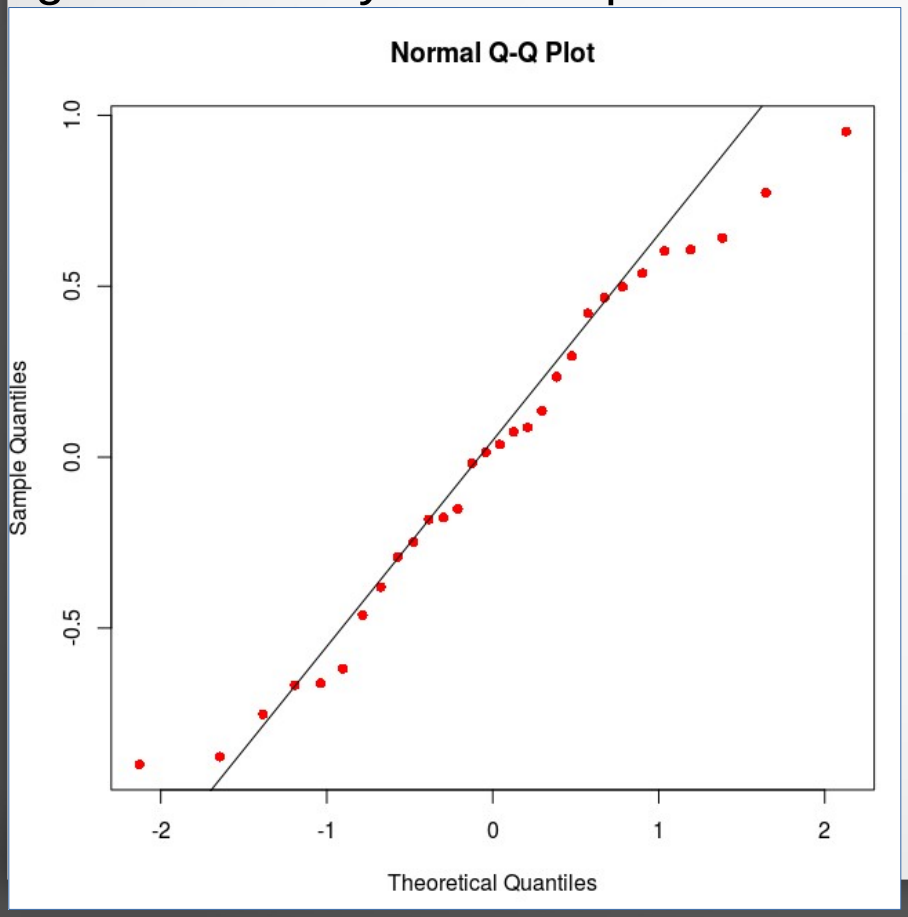
Verificación de los supuestos: Homocedasticidad

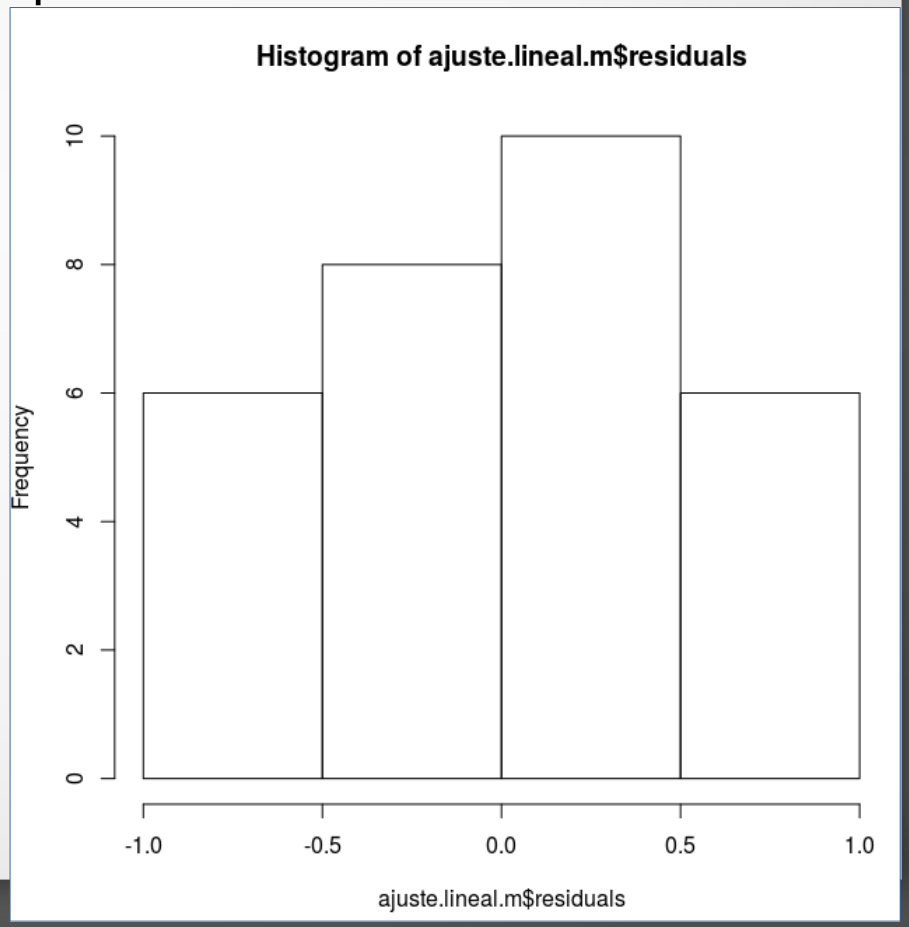
Homocedasticidad (**Residuos**). Los residuos en las predicciones son constantes en cada predicción (**varianza constante**). Este supuesto valida que los residuos no aumentan ni disminuyen cuando se predicen valores más altos o mas pequeños.



Verificación de los supuestos: Normalidad

Normalidad (**Residuos**) Se asume que los residuos deben **seguir una distribución Normal**, la ausencia de normalidad supone poca precisión en los intervalos de confianza creados por el modelo. Lo verificamos gráficamente y con una prueba de Shapiro-Wilk.





Verificación de los supuestos: Normalidad

En las pruebas formales o analíticas tenemos:

- Prueba de Shapiro-Wilk:

La hipótesis a probar es:

H_0: Los errores siguen una distribución normal

H_1: Los errores no siguen una distribución normal

La hipótesis H_0 se rechaza al 5% si p-value < 0.05

```
> shapiro.test(ajuste.lineal.m$residuals)
        Shapiro-Wilk normality test
data:  ajuste.lineal.m$residuals
W = 0.96884, p-value = 0.5081
```

Como el p-value > 0.05 se acepta la hipótesis de normalidad. Los residuos siguen una distribución normal.

REFERENCIAS

Chatterjee, S., & Hadi, A. S. (2015). Regression analysis by example. John Wiley & Sons.

Noste, M. E. S. (2013). Apunte de Regresión Lineal. http://mate.dm.uba.ar/~meszre/apunte_regresion_lineal_szretter.pdf