

MRLS

Diana Lizeth Reyes

2023-12-10

Modelo de Regresión lineal simple

Lectura de matriz de datos

1.- Exportar la matriz penguins.xlsx

1.1.- Instalación de paquetería {r,message=FALSE, warming=FALSE}

```
install.packages("readxl")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

1.2.- Abrir librería

```
library("readxl")
```

1.3.- Exportación de la matriz de datos

```
penguins<-read_excel("penguins.xlsx")
```

2.- Exploración de la matriz

2.1.- Dimensión de la matriz

```
dim(penguins)
```

```
## [1] 344 9
```

2.2.- Nombre de las columnas

```
str(penguins)
```

```
## tibble [344 x 9] (S3: tbl_df/tbl/data.frame)  
## $ ID : chr [1:344] "i1" "i2" "i3" "i4" ...  
## $ especie : chr [1:344] "Adelie" "Adelie" "Adelie" "Adelie" ...  
## $ isla : chr [1:344] "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...  
## $ largo_pico_mm : num [1:344] 39.1 39.5 40.3 37.8 36.7 39.3 38.9 39.2 34.1 42 ...  
## $ grosor_pico_mm : num [1:344] 18.7 17.4 18 18.1 19.3 20.6 17.8 19.6 18.1 20.2 ...  
## $ largo_aleta_mm : num [1:344] 181 186 195 190 193 190 181 195 193 190 ...  
## $ masa_corporal_g: num [1:344] 3750 3800 3250 3700 3450 ...  
## $ genero : chr [1:344] "male" "female" "female" "female" ...  
## $ año : num [1:344] 2007 2007 2007 2007 2007 ...
```

2.3.- Tipo de variables

```
colnames(penguins)
```

```
## [1] "ID" "especie" "isla" "largo_pico_mm"
```

```
## [5] "grosor_pico_mm" "largo_aleta_mm" "masa_corporal_g" "genero"
## [9] "año"
```

2.4.- En busca de datos perdidos

```
anyNA(penguins)
```

```
## [1] FALSE
```

Configuración de la matriz

1.- Visualización de las columnas: *especie, isla, género y año*

1.1.- Especie

```
penguins$especie
```

```
## [1] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [7] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [13] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [19] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [25] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [31] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [37] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [43] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [49] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [55] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [61] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [67] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [73] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [79] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [85] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [91] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [97] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [103] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [109] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [115] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [121] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [127] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [133] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [139] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [145] "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie"
## [151] "Adelie" "Adelie" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [157] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [163] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [169] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [175] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [181] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [187] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [193] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [199] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [205] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [211] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [217] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [223] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [229] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
## [235] "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo" "Gentoo"
```



```
## [185] "female" "male" "female" "male" "female" "male" "female" "male"
## [193] "female" "male" "female" "male" "male" "female" "female" "male"
## [201] "female" "male" "female" "male" "female" "male" "female" "male"
## [209] "female" "male" "female" "male" "female" "male" "female" "male"
## [217] "female" "male" "female" "male" "female" "male" "female" "male"
## [225] "male" "female" "female" "male" "female" "male" "female" "male"
## [233] "female" "male" "female" "male" "female" "male" "female" "male"
## [241] "female" "male" "female" "male" "female" "male" "female" "male"
## [249] "male" "female" "female" "male" "female" "male" "female" "male"
## [257] "female" "male" "female" "male" "female" "male" "female" "male"
## [265] "female" "male" "female" "male" "female" "male" "female" "male"
## [273] "female" "male" "female" "male" "female" "male" "male" "female"
## [281] "male" "female" "female" "male" "female" "male" "female" "male"
## [289] "female" "male" "female" "male" "male" "female" "female" "male"
## [297] "female" "male" "female" "male" "female" "male" "female" "male"
## [305] "female" "male" "female" "male" "female" "male" "male" "female"
## [313] "female" "male" "female" "male" "male" "female" "male" "female"
## [321] "female" "male" "female" "male" "male" "female" "female" "male"
## [329] "female" "male" "female" "male" "female" "male" "male" "female"
## [337] "male" "female" "female" "male" "female" "male" "male" "female"
```

1.4.- Año

```
penguins$año
```

```
## [1] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [16] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [31] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [46] 2007 2007 2007 2007 2007 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [61] 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [76] 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [91] 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2009 2009 2009 2009 2009
## [106] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [121] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [136] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [151] 2009 2009 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [166] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [181] 2007 2007 2007 2007 2007 2007 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [196] 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [211] 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [226] 2008 2008 2008 2008 2008 2008 2008 2009 2009 2009 2009 2009 2009 2009 2009
## [241] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [256] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [271] 2009 2009 2009 2009 2009 2009 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [286] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007
## [301] 2007 2007 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008
## [316] 2008 2008 2008 2008 2008 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
## [331] 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009
```

2.- Convertir las variables categóricas a factores

2.1.- Especie

```
penguins$especie<-factor(penguins$especie,
                          levels=c("Adelie", "Gentoo", "Chinstrap"))
```

2.2.- Isla

```
penguins$isla<-factor(penguins$isla,
                      levels=c("Torgersen", "Biscoe", "Dream"))
```

2.3.- Género

```
penguins$genero<-factor(penguins$genero,
                        levels=c("male", "female"))
```

2.4.- Año

```
penguins$año<-factor(penguins$año,
                     levels=c("2007", "2008", "2009"))
```

3.- Visualizar el tipo de variables que tenemos ahora

```
str(penguins)
```

```
## tibble [344 x 9] (S3: tbl_df/tbl/data.frame)
##  $ ID          : chr [1:344] "i1" "i2" "i3" "i4" ...
##  $ especie     : Factor w/ 3 levels "Adelie","Gentoo",...: 1 1 1 1 1 1 1 1 1 1 ...
##  $ isla        : Factor w/ 3 levels "Torgersen","Biscoe",...: 1 1 1 1 1 1 1 1 1 1 ...
##  $ largo_pico_mm : num [1:344] 39.1 39.5 40.3 37.8 36.7 39.3 38.9 39.2 34.1 42 ...
##  $ grosor_pico_mm : num [1:344] 18.7 17.4 18 18.1 19.3 20.6 17.8 19.6 18.1 20.2 ...
##  $ largo_aleta_mm : num [1:344] 181 186 195 190 193 190 181 195 193 190 ...
##  $ masa_corporal_g: num [1:344] 3750 3800 3250 3700 3450 ...
##  $ genero      : Factor w/ 2 levels "male","female": 1 2 2 2 2 1 2 1 2 1 ...
##  $ año         : Factor w/ 3 levels "2007","2008",...: 1 1 1 1 1 1 1 1 1 1 ...
```

Selección de variables

1.- Se seleccionarán los datos de la especie gentoo y se crea una nueva matriz llamada “gentoo”

1.1.- Se selecciona la variable tomada de la matriz original

```
penguins$especie
```

```
## [1] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [8] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [15] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [22] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [29] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [36] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [43] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [50] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [57] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [64] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [71] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [78] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [85] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [92] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [99] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [106] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [113] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [120] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [127] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [134] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
## [141] Adelie Adelie Adelie Adelie Adelie Adelie Adelie
```

```
## [148] Adelie      Adelie      Adelie      Adelie      Adelie      Gentoo      Gentoo
## [155] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [162] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [169] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [176] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [183] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [190] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [197] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [204] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [211] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [218] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [225] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [232] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [239] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [246] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [253] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [260] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [267] Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo      Gentoo
## [274] Gentoo      Gentoo      Gentoo      Chinstrap  Chinstrap  Chinstrap  Chinstrap
## [281] Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap
## [288] Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap
## [295] Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap
## [302] Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap
## [309] Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap
## [316] Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap
## [323] Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap
## [330] Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap
## [337] Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap  Chinstrap
## [344] Chinstrap
## Levels: Adelie Gentoo Chinstrap
```

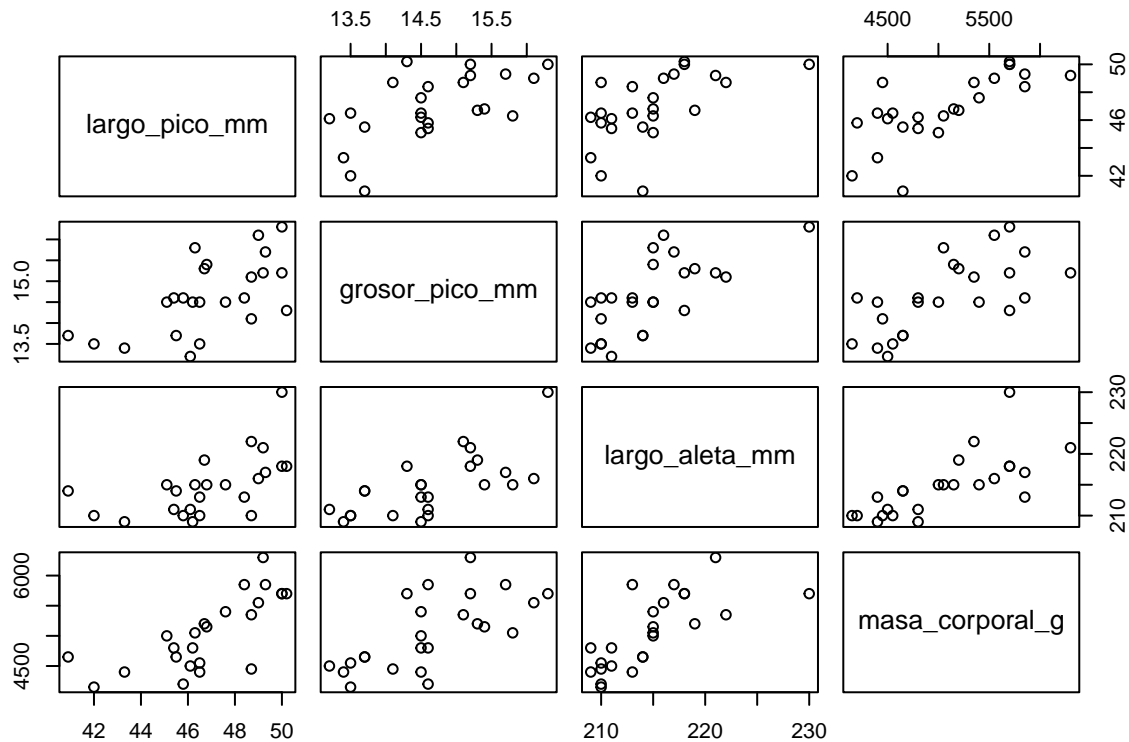
1.2.- Se seleccionan los datos con los que se quieren trabajar y las variables numéricas que se ocuparán : largo del pico, grosor del pico, largo de la aleta y masa corporal.

```
gentoo<-penguins[153:176,c(4,5,6,7)]
```

Gráfico de dispersión

1.- Mediante la función *pairs* se creará un gráfico de dispersión para visualizar nuestra nueva matriz

```
pairs(gentoo)
```



Cálculo de la correlación de Pearson

Mediante la función *cor*, se calcula la correlación de Pearson

```
cor(gentoo)
```

```
##                largo_pico_mm grosor_pico_mm largo_aleta_mm masa_corporal_g
## largo_pico_mm      1.0000000      0.6185638      0.5781154      0.7386365
## grosor_pico_mm      0.6185638      1.0000000      0.6931901      0.6735989
## largo_aleta_mm      0.5781154      0.6931901      1.0000000      0.7205205
## masa_corporal_g      0.7386365      0.6735989      0.7205205      1.0000000
```

Gráfico de dispersión con línea de regresión

1.- Se instala la paquetería *ggplot2* y la librería

1.1.- Instalamos la paquetería “ggplot2”

```
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

1.2.- Abrimos librería

```
library(ggplot2)
```

2.- Se generará un gráfico de dispersión en donde se muestre la línea de regresión.

2.1.1.- El primer objeto se llamará MRL, se ocuparán las variables: *masa corporal* y **largo del pico*. Mediante la función ***ggplot**, creará el gráfico:

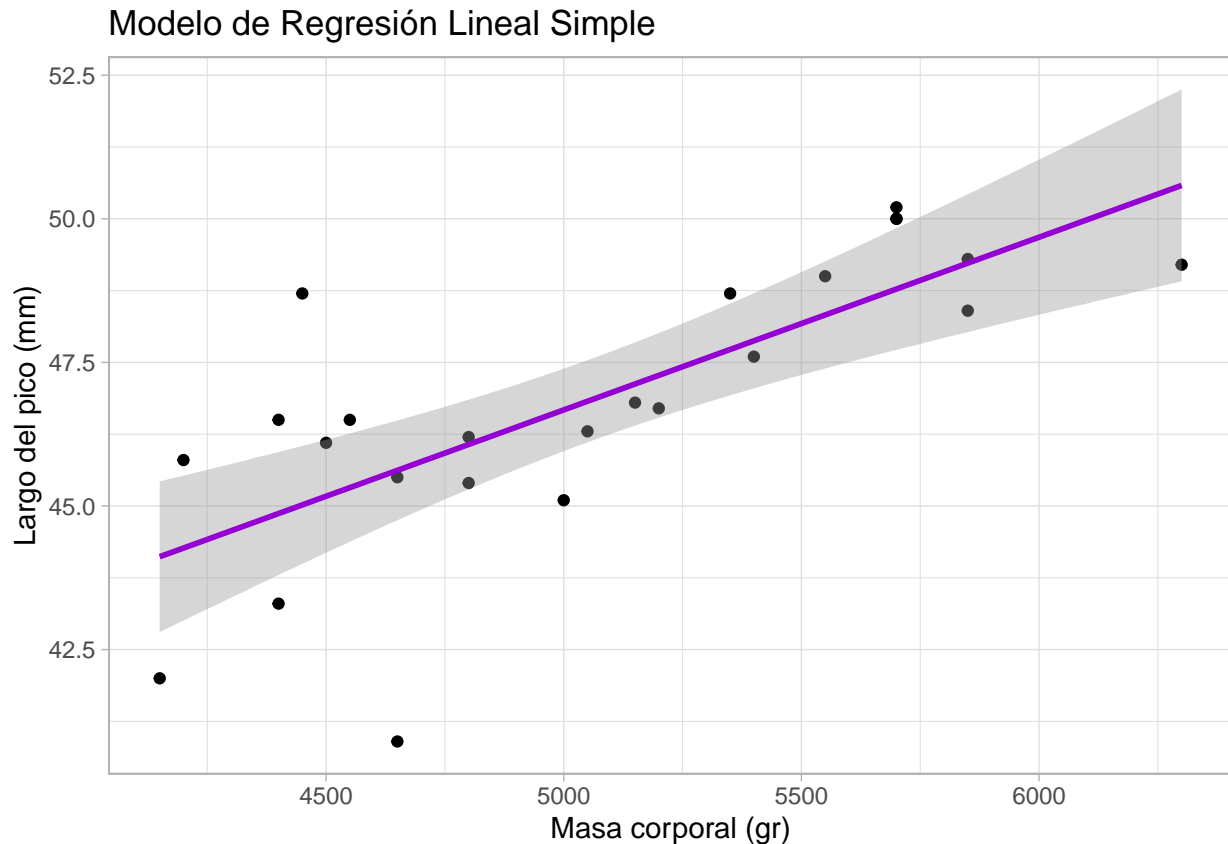
```
MRL<-ggplot(gentoo, aes(x=masa_corporal_g, y=largo_pico_mm))+
  geom_point()+
```



```
geom_smooth(method = "lm", formula=y~x, col="darkviolet")+
ggtitle("Modelo de Regresión Lineal Simple")+
xlab("Masa corporal (gr)")+
ylab("Largo del pico (mm)")+
theme_light()
```

2.1.2.- Se visualiza el objeto para verificar

MRL



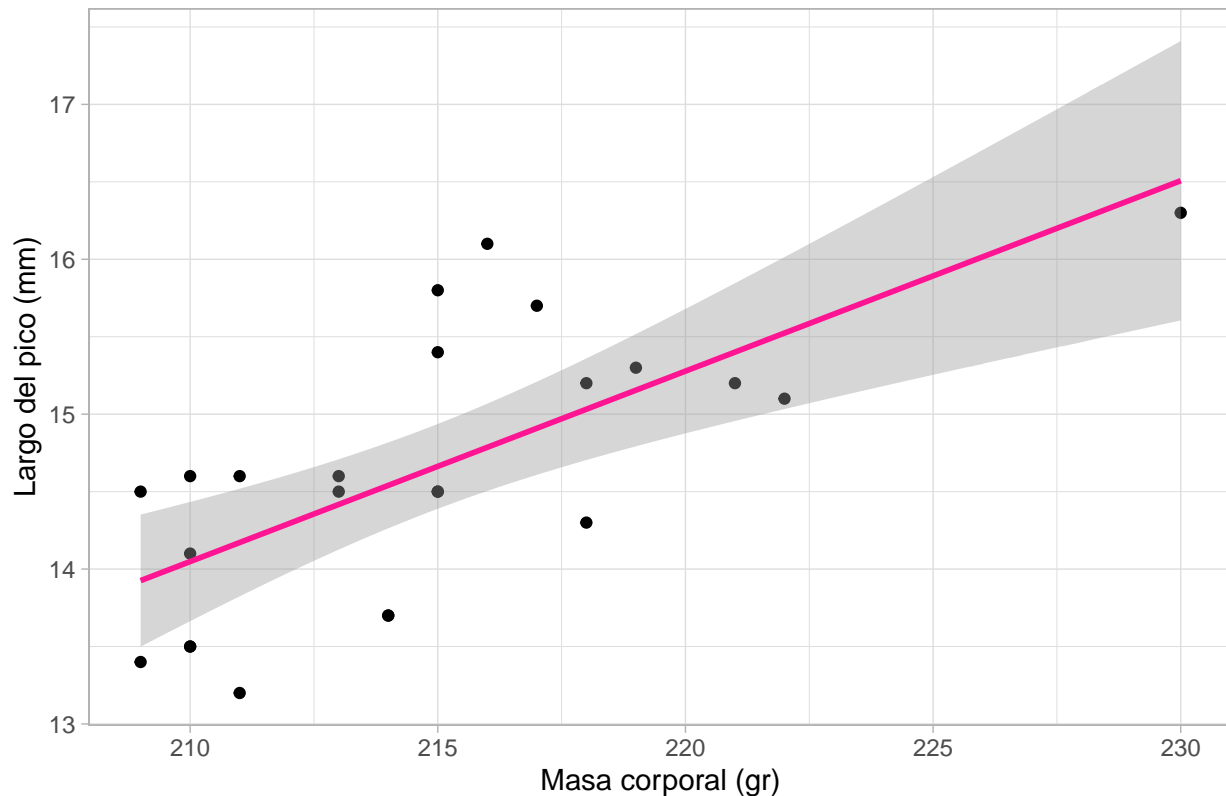
2.1.1.- El segundo objeto se llamará MRL2, se ocuparán las variables: *largo de la aleta* y *grosor del pico*.

```
MRL2<-ggplot(gentoo, aes(x=largo_aleta_mm, y=grosor_pico_mm))+
geom_point()+
geom_smooth(method = "lm", formula=y~x, col="deeppink1")+
ggtitle("Modelo de Regresión Lineal Simple")+
xlab("Masa corporal (gr)")+
ylab("Largo del pico (mm)")+
theme_light()
```

2.1.2.- Visualización el objeto

MRL2

Modelo de Regresión Lineal Simple



Cálculo y representación de la recta por mínimos cuadrados

1.- Con la función `lm` se calcularán datos como: el intercepto, el error, el valor crítico p-valor, entre otros

1.1.1.- Primer modelo, se ocupan las variables *largo del pico* y *masa corporal*

```
regresion<-lm(gentoo$largo_pico_mm~gentoo$masa_corporal_g,  
              data=gentoo)
```

1.1.2.- Se visualiza el objeto

```
summary(regresion)
```

```
##  
## Call:  
## lm(formula = gentoo$largo_pico_mm ~ gentoo$masa_corporal_g, data = gentoo)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -4.7203 -0.7105 -0.0242  1.1910  3.6810   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)    31.640147   2.977858  10.625 3.96e-10 ***  
## gentoo$masa_corporal_g  0.003007  0.000585   5.139 3.76e-05 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 1.69 on 22 degrees of freedom
```

```
## Multiple R-squared:  0.5456, Adjusted R-squared:  0.5249
## F-statistic: 26.41 on 1 and 22 DF,  p-value: 3.761e-05
```

1.2.1- Segundo modelo, ocupando las variables *largo de la aleta* y *grosor del pico*

```
regresion2<-lm(gentoo$largo_aleta_mm~gentoo$grosor_pico_mm,
               data=gentoo)
```

1.2.2.- Visualización el objeto

```
summary(regresion2)
```

```
##
## Call:
## lm(formula = gentoo$largo_aleta_mm ~ gentoo$grosor_pico_mm, data = gentoo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.2542 -2.7111 -0.3458  2.0882  8.7105
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    157.5811     12.7041  12.404  2.1e-11 ***
## gentoo$grosor_pico_mm  3.9085      0.8664   4.511 0.000173 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.643 on 22 degrees of freedom
## Multiple R-squared:  0.4805, Adjusted R-squared:  0.4569
## F-statistic: 20.35 on 1 and 22 DF,  p-value: 0.0001731
```

Coeficiente de Correlación de Pearson (r)

1. Con la función *sqrt* se saca la raíz cuadrada del dato *Multiple R-squared* para sacar el valor del coeficiente de correlación de Pearson

1.1.1.- r del largo del pico y la masa corporal

```
r1<- sqrt(0.5456)
```

1.1.2 Se visualiza el objeto

```
r1
```

```
## [1] 0.7386474
```

1.2.1 r del largo de la aleta y el grosor del pico

```
r2<-sqrt(0.4805)
```

1.2.2 Visualización del objeto

```
r2
```

```
## [1] 0.6931811
```