

Reporte de Investigación

Oscar Ferrer

2024-06-25

Introducción

En el presente trabajo se realizó una recopilación de datos fisiológicos cuantitativos con la intención de conocer y de clasificar a los alumnos que presentan características sobre salientes con relación de sus compañeros, esto para conocer a aquellos posibles candidatos ante una posible competición. Dicha recopilación se realizó utilizando pruebas, toma de medidas y haciendo uso de herramientas virtuales a manera de innovación que lo fue R estudio. Y fue gracias a esta herramienta que se encontro una relación positiva entre la altura y que tan alto saltan los alumnos.

Metodología

Comenzamos realizando la instalación de los paquetes y librerías que serán necesarios para la correcta realización de los comandos así como la lectura de lo datos.

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

```
library(readxl)
```

Posterior a ello le asignaremos un nombre a la matriz de la que se estan analizando los datos para poder identificarla.

```
BD<-read.csv("Archivo 1.csv")
```

Exploración de la Matriz

Se realiza la exploración de la matriz, en la que se definen algunos datos de ella como pueden ser la dimensión, el nombre de las columnas, el tipo de variable que se ocupa y los datos que se hayan perdido.

Dimensión de la matriz

Número de observaciones y variables.

```
dim(BD)
```

```
## [1] 25 6
```

Nombre de las variables

Nombre de cada variable.

```
colnames(BD)
```

```
## [1] "ID" "altura_cm" "peso_kg" "X25m" "salto_cm" "genero"
```

Tipo de variable

Tipo de variable entre cuanti o cualitativa.

```
str(BD)
```

```
## 'data.frame': 25 obs. of 6 variables:
## $ ID : chr "i1" "i2" "i3" "i4" ...
## $ altura_cm: int 163 155 151 167 172 163 162 172 168 156 ...
## $ peso_kg : int 55 50 50 50 53 50 60 69 60 45 ...
## $ X25m : num 4.37 6.14 6.45 4.42 4.42 5.79 4.62 4.72 4.47 4.37 ...
## $ salto_cm : int 260 224 223 255 257 227 242 250 242 233 ...
## $ genero : chr "m" "f" "f" "m" ...
```

Datos perdidos

Nos indica si existen datos que se hayn perdido.

```
anyNA(BD)
```

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

Preprocesamiento de la Matriz

Convertir las variables categóicas a factores

Se realizó la conversión de dos variables categóicas a factores, con la intención de que dejará de ser variables cualitativas y se transformarán en variables cuantitativas.

```
BD$genero<-factor(BD$genero,
                  levels=c("m", "f"))
```

Estadísticas Descriptivas

Se realiza la descripción de las estadísticas de la matriz en la que se nos indican: el valor mínimo y el valor máximo, la mediana y la media, así como el primer y el tercer cuartil.

```
summary(BD)
```

```
##      ID          altura_cm      peso_kg      X25m
## Length:25      Min.   :143.0    Min.   :34.00    Min.   :4.010
## Class :character 1st Qu.:156.0    1st Qu.:47.00    1st Qu.:4.420
## Mode  :character Median :163.0    Median :50.00    Median :4.620
##              Mean  :161.6    Mean  :53.16    Mean  :4.926
##              3rd Qu.:168.0    3rd Qu.:60.00    3rd Qu.:5.340
##              Max.   :175.0    Max.   :86.00    Max.   :6.450
##      salto_cm      genero
## Min.   :201.0    m:14
## 1st Qu.:224.0    f:11
## Median :232.0
## Mean   :234.8
## 3rd Qu.:250.0
## Max.   :270.0
```

Resultados

Gráficos Descriptivos

Comenzamos a realizar la creación de los graficos que describiran la información que estamos trabajando, pero de manera visual.

Se debe realizar nuevamente la instalación de paquetes y librerías.

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

```
library(ggplot2)
```

Creacion del vector de color.

Con el comando que se muestra se realiza la aplicación de los colores que aparecieran en nuestro gráfico.

```
color=c("mediumpurple", "plum2")
```

Creación del gráfico (Boxplot)

Posterior a la aplicación del color del gráfico se comienza a generar el comando para dicho gráfico.

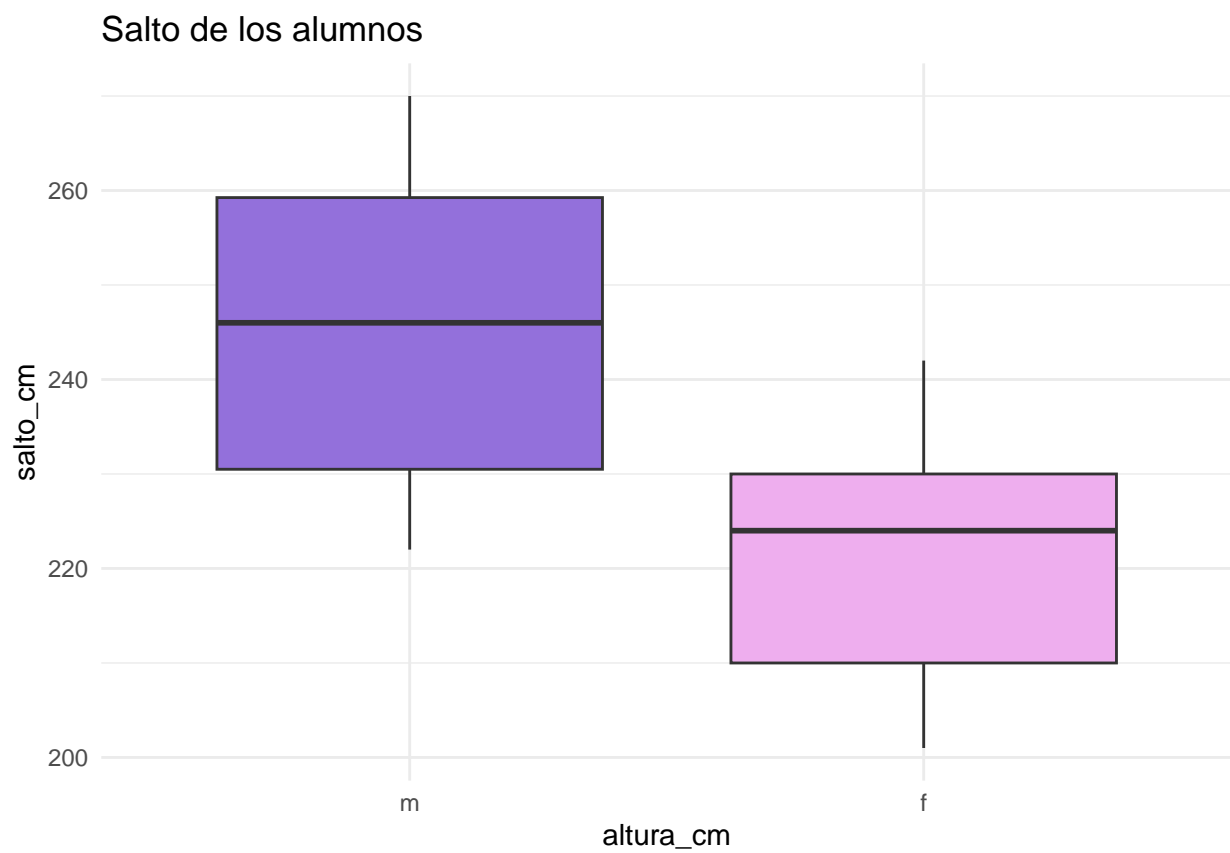
```
BDG<-ggplot(BD, aes(x=genero, y=salto_cm))+  
  geom_boxplot(fill=color)+  
  ggtitle("Salto de los alumnos")+  
  xlab("altura_cm")+  
  ylab("salto_cm")+  
  theme_minimal()
```

```
BDG1<-ggplot(BD, aes(x=genero, y=altura_cm))+  
  geom_boxplot(fill=color)+  
  ggtitle("Altura de los alumnos")+  
  xlab("altura_cm")+  
  ylab("salto_cm")+  
  theme_minimal()
```

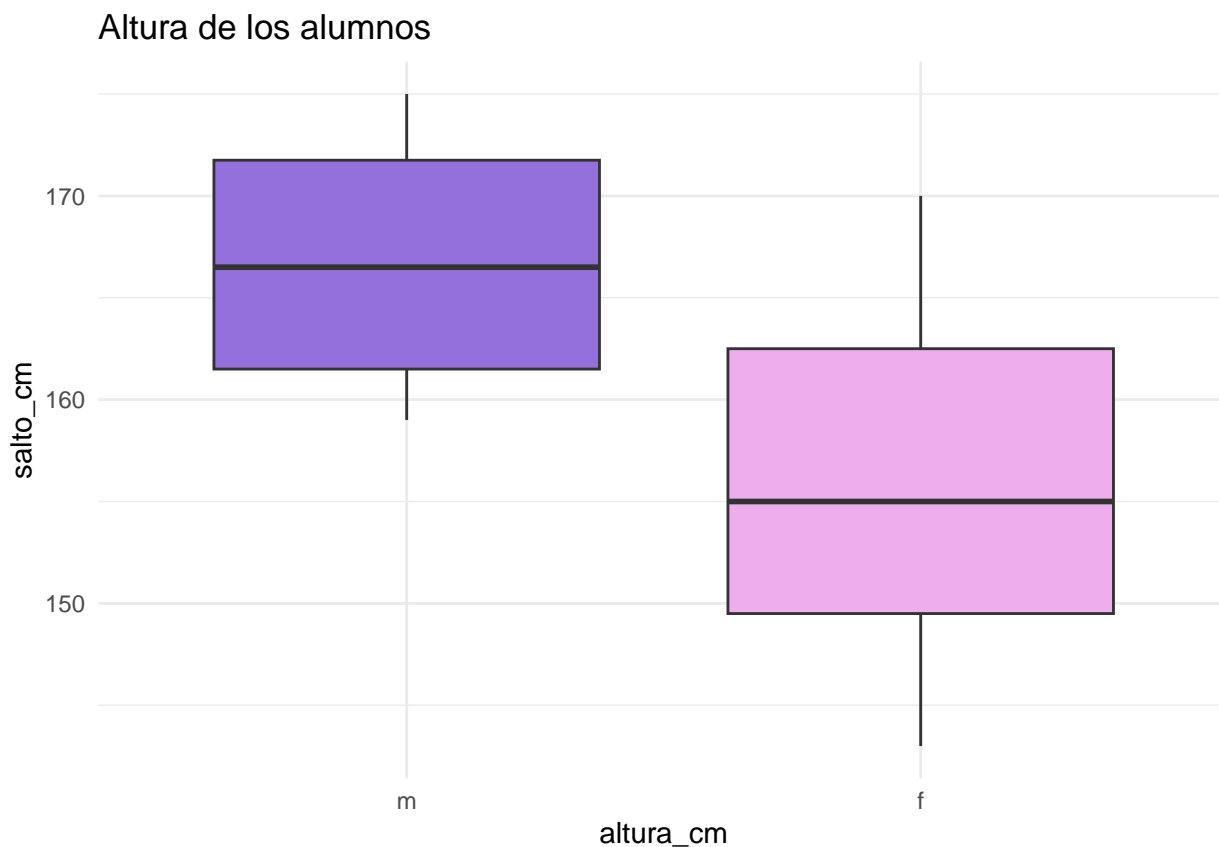
Visualización del gráfico

Una vez realizado el comando se debe ejecutar la asignación que se le dió a cada gráfico que se quiera para poderlo visualizar.

```
BDG
```



BDG1



Analisis de correlación

Se genera un análisis de la correlación para identificar aquellas variables que tienen correlación entre sí.

```
colnames(BD)
```

```
## [1] "ID"          "altura_cm" "peso_kg"   "X25m"      "salto_cm"  "genero"
```

Selección de columnas “altura_cm”, “peso_kg”, “25m”, “salto_cm”

Instalación de paquetería y librería.

```
install.packages("dplyr")
```

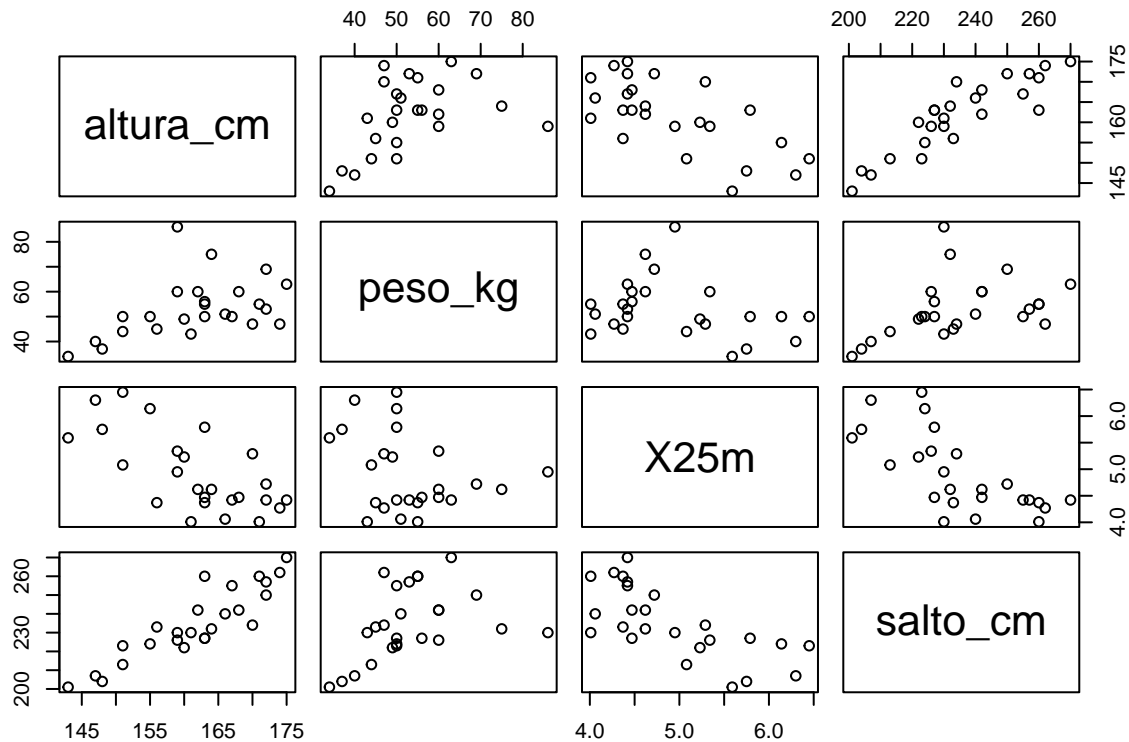
```
library(dplyr)
```

Seguido de eso se realizó la creación de una matriz nueva, en la que se eliminaron variables que no son relevantes.

```
BD_cuanti<-BD %>%
  select(c("altura_cm", "peso_kg", "X25m", "salto_cm"))
```

Visualización

```
plot(BD_cuanti)
```



Modelo de Regresión lineal simple

Calculo de la correlacion de Pearson

Ahora se calcula la correlación que existe entre las variables.

```
cor(BD_cuanti)
```

```
##          altura_cm  peso_kg      X25m  salto_cm
## altura_cm  1.0000000  0.4597339 -0.6820548  0.8928534
## peso_kg    0.4597339  1.0000000 -0.2909901  0.4101489
## X25m      -0.6820548 -0.2909901  1.0000000 -0.7025473
## salto_cm   0.8928534  0.4101489 -0.7025473  1.0000000
```

Gráfico de dispersion con linea de regresion

Paquetería y librería

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

```
library(ggplot2)
```

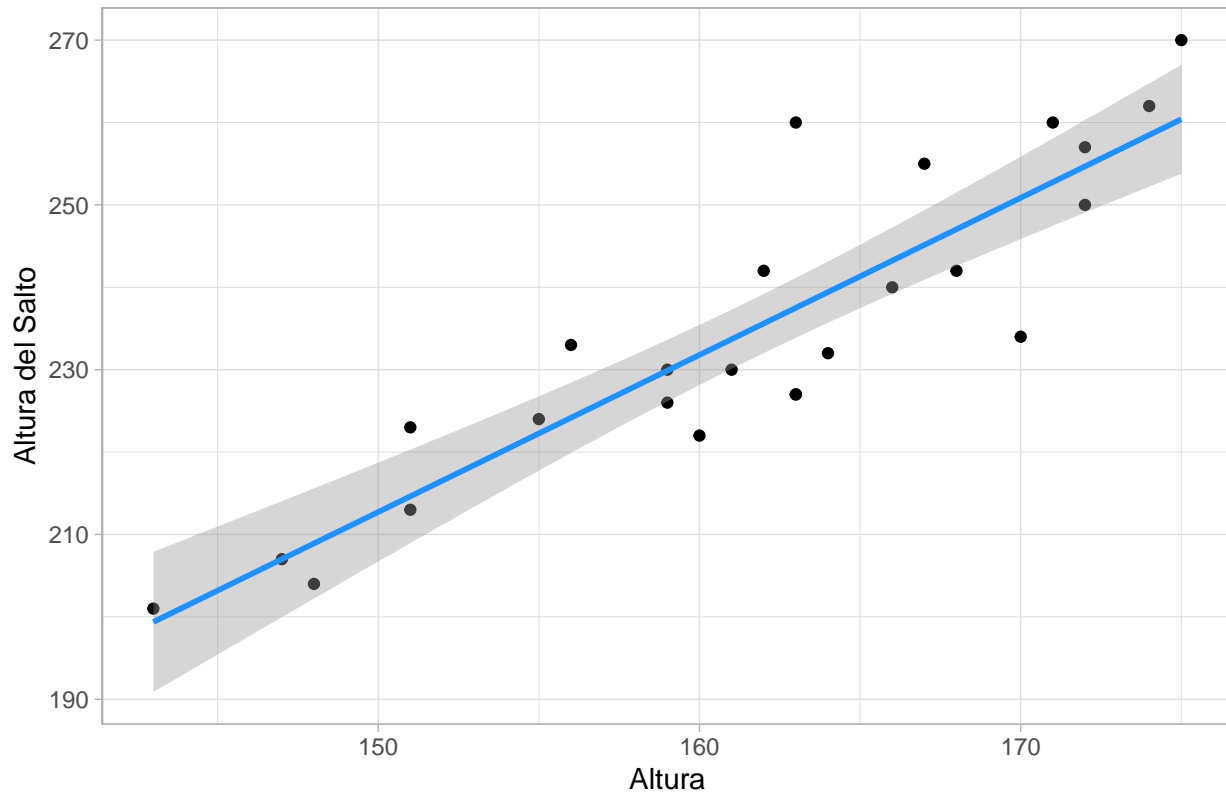
Creación del gráfico de dispersión.

```
MRL<-ggplot(BD_cuanti, aes(x=altura_cm, y=salto_cm))+
  geom_point()+
  geom_smooth(method = "lm", formula=y~x, col="dodgerblue1")+
  ggtitle("Modelo de Regresión Lineal Simple")+
  xlab("Altura")+
  ylab("Altura del Salto")+
  theme_light()
```

Visualizacion del objeto

MRL

Modelo de Regresión Lineal Simple



Cálculo y representación de la recta por mínimos cuadrados

```
regresion<-lm(BD_cuanti$altura_cm~BD_cuanti$salto_cm,  
              data=BD_cuanti)
```

```
summary(regresion)
```

```
##  
## Call:  
## lm(formula = BD_cuanti$altura_cm ~ BD_cuanti$salto_cm, data = BD_cuanti)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -9.1197 -2.5937 -0.5763  3.4063  8.7512   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)   63.41035    10.35877   6.121 3.04e-06 ***  
## BD_cuanti$salto_cm 0.41811     0.04397   9.508 1.96e-09 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 4.061 on 23 degrees of freedom
```

```
## Multiple R-squared:  0.7972, Adjusted R-squared:  0.7884
## F-statistic: 90.41 on 1 and 23 DF,  p-value: 1.963e-09
```

Ecuación de regresión lineal simple

$y = \text{Beta0} + \text{Beta1} x + \text{error}$

y= al valor de la variable dependiente que se pretende hayar

x= al valor de la variable independiete (de esta ya conocemos su valor)

Beta0= al valor de la intercepto entre los ejes y y x de la recta de rgresión

Beta1= al valor de la inclinación de la pendiente de la recta de regresión

Error= al valor del residuo también conocido como error

Coefficients:

(Intercept) 63.41035(Beta0)

archivo_cuanti\$salto_cm 0.41811(Beta1)

Error= 4.061 (error)

Conclusiones

En conclusión se descubrio que la variable independiente de la altura esta relacionada de manera directa con la calidad y altura del salto vertical que realizan los alumnos de secundaria. Además de recalcar que se encontro poca o nula relación entre las demás variables.