proba_reglin_fer

Fernanda Pérez

2024-08-30

La recta de mejor ajuste (Primera entrega)

Obtén la matriz de correlación de los datos que se te proporcionan. Interpreta.

```
M = read.csv("D:/Downloads/Estatura-peso_HyM.csv")
head(M)
```

```
##   Estatura  Peso Sexo
## 1     1.61 72.21    H
## 2     1.61 65.71    H
## 3     1.70 75.08    H
## 4     1.65 68.55    H
## 5     1.72 70.77    H
## 6     1.63 77.18    H
```

```
datos = data.frame(Estatura = M$Estatura, Peso = M$Peso)

correlacion = cor(datos)
print(correlacion)
```

```
##           Estatura      Peso
## Estatura 1.0000000 0.8032449
## Peso     0.8032449 1.0000000
```

Se observa que hay una correlación alta y positiva de 0.8032.

```
MM = subset(M,M$Sexo=="M")
MH = subset(M,M$Sexo=="H")
M1=data.frame(MH$Estatura,MH$Peso,MM$Estatura,MM$Peso)

n=4
d=matrix(NA,ncol=7,nrow=n)
for(i in 1:n){
  d[i,]<-c(as.numeric(summary(M1[,i])),sd(M1[ ,i]))
}
m=as.data.frame(d)

row.names(m)=c("H-Estatura","H-Peso","M-Estatura","M-Peso")
names(m)=c("Minimo","Q1","Mediana","Media","Q3","Máximo","Desv Est")
```

Obtén medidas (media, desviación estándar, etc) que te ayuden a analizar los datos.

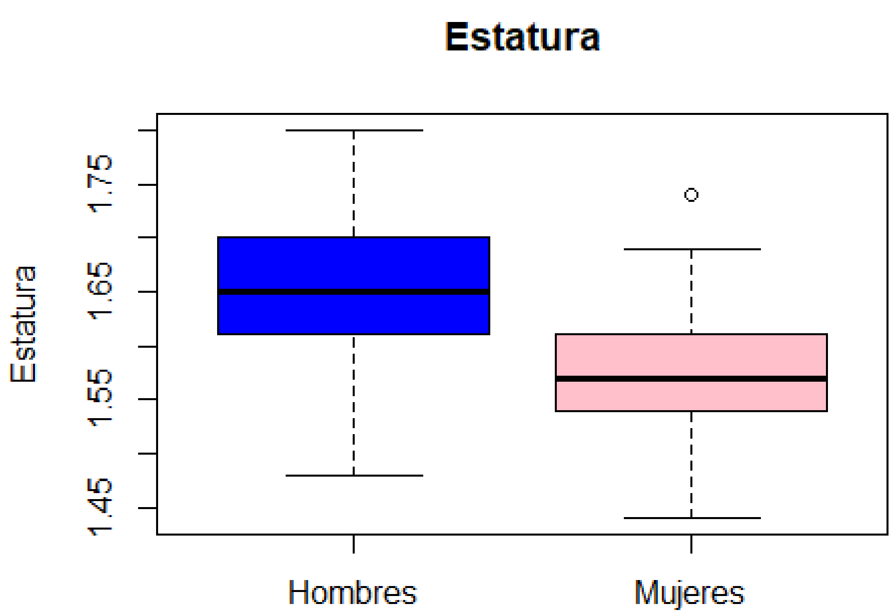
```
resumen = summary(datos)
print(resumen)
```

```
##     Estatura          Peso      
##  Min.   :1.440   Min.   :37.39  
##  1st Qu.:1.560   1st Qu.:54.49  
##  Median :1.610   Median :64.53  
##  Mean   :1.613   Mean   :63.97  
##  3rd Qu.:1.660   3rd Qu.:73.22  
##  Max.   :1.800   Max.   :90.49
```

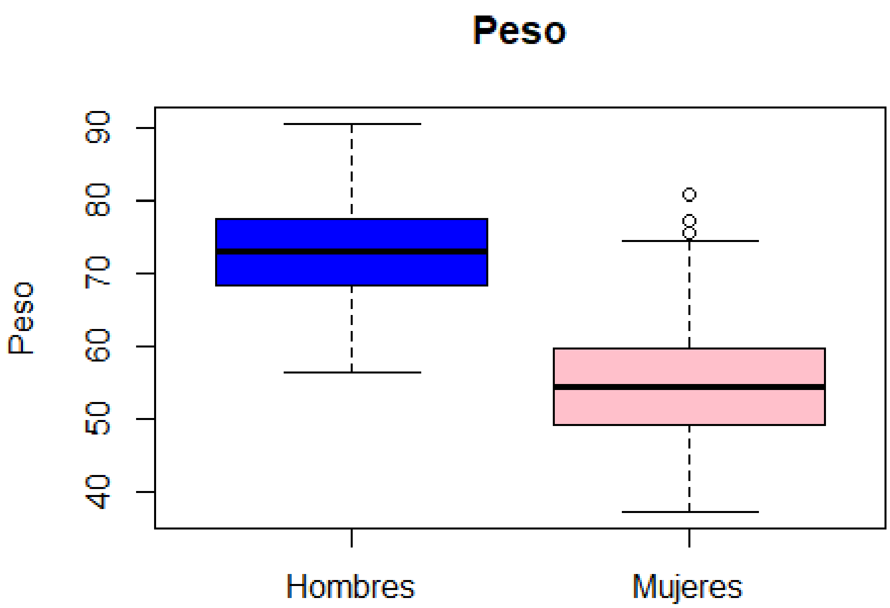
```
desviacion_estandar = apply(datos, 2, sd)
print(desviacion_estandar)
```

```
##    Estatura        Peso 
##  0.06929171 11.54161456
```

```
boxplot(M$Estatura~M$Sexo, ylab="Estatura", xlab="",
col=c("blue","pink"), names=c("Hombres", "Mujeres"), main="Estatura")
```




```
boxplot(M$Peso~M$Sexo, ylab="Peso",xlab="", names=c("Hombres",
"Mujeres"), col=c("blue","pink"), main="Peso")
```



Los hombres parecen tener un peso más alto que el de la smujeres, y el peso de los hombres se encuentra con una distribución más concentrada y el de las mujeres con más variabilidad.

#3. Encuentra la ecuación de regresión de mejor ajuste:

3.1 Realiza la regresión entre las variables involucradas

```
modelo <- lm(Peso ~ Estatura, data = datos)

summary(modelo)

##
## Call:
## lm(formula = Peso ~ Estatura, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.8653  -3.7654   0.6706   5.0142  15.6006
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -151.883      7.655  -19.84   <2e-16 ***
## Estatura     133.793      4.741   28.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.883 on 438 degrees of freedom
```

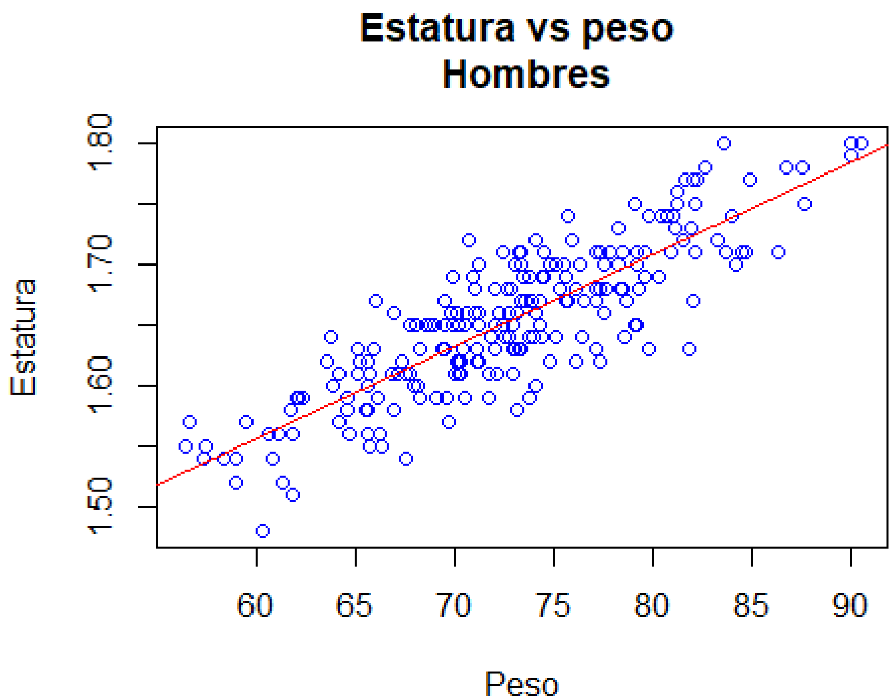
```
## Multiple R-squared:  0.6452, Adjusted R-squared:  0.6444
## F-statistic: 796.5 on 1 and 438 DF,  p-value: < 2.2e-16
```

Hombres

```r
Modelo1H = lm(Estatura~Peso, MH)
Modelo1H

##
## Call:
## lm(formula = Estatura ~ Peso, data = MH)
##
## Coefficients:
## (Intercept)         Peso
##    1.101770     0.007576
```

Hipótesis: * $H_0: \beta_1 = 0$ * $H_1: \beta_1 \neq 0$

```r
summary(Modelo1H)

##
## Call:
## lm(formula = Estatura ~ Peso, data = MH)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.091473 -0.020942  0.001445  0.024020  0.082089
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1017704  0.0235832   46.72   <2e-16 ***
## Peso        0.0075758  0.0003223   23.51   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03291 on 218 degrees of freedom
## Multiple R-squared:  0.7171, Adjusted R-squared:  0.7158
## F-statistic: 552.7 on 1 and 218 DF,  p-value: < 2.2e-16

plot(MH$Peso, MH$Estatura, col="blue", main="Estatura vs peso \n
Hombres", ylab="Estatura", xlab="Peso")
abline(Modelo1H, col="red", lwd=1.5)
```



Mujeres

Verifica el modelo:

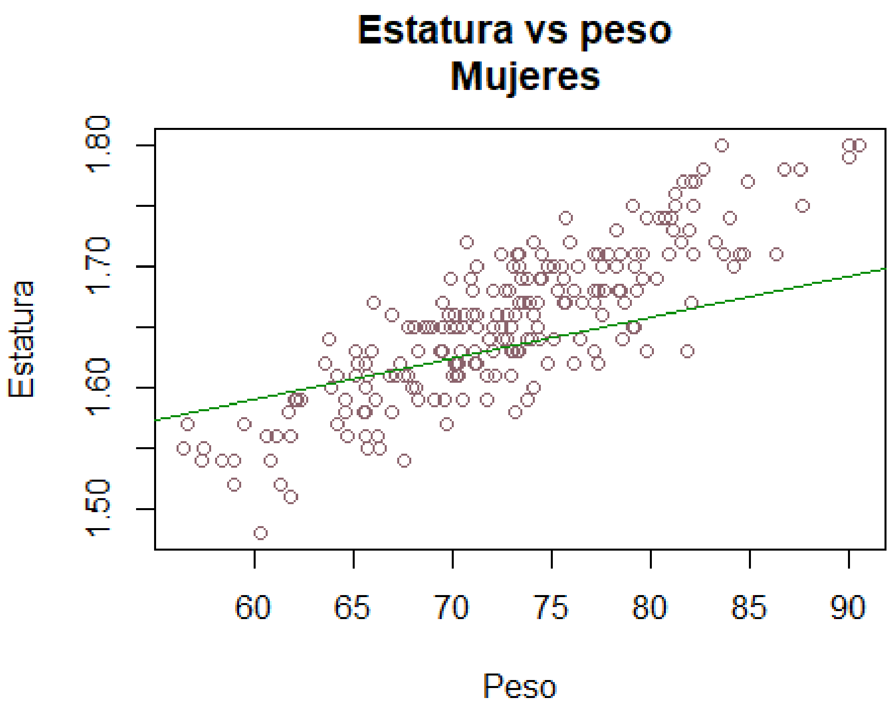
Verifica la significancia del modelo con un alfa de 0.03.

Mujeres

```
Modelo1M = lm(Estatura~Peso, MM)
Modelo1M

##
## Call:
## lm(formula = Estatura ~ Peso, data = MM)
##
## Coefficients:
## (Intercept)         Peso
##     1.38622      0.00339

summary(Modelo1M)

##
## Call:
## lm(formula = Estatura ~ Peso, data = MM)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.11162 -0.02611 -0.00174  0.02806  0.12814
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3862212  0.0207336  66.859   <2e-16 ***
## Peso        0.0033900  0.0003727   9.096   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04298 on 218 degrees of freedom
## Multiple R-squared:  0.2751, Adjusted R-squared:  0.2718
## F-statistic: 82.73 on 1 and 218 DF,  p-value: < 2.2e-16
```

```r
plot(MH$Peso, MH$Estatura, col="pink4", main="Estatura vs peso \n
Mujeres", ylab="Estatura", xlab="Peso")
abline(Modelo1M, col="green4", lwd=1.5)
```



Realiza la regresión entre las variables involucradas

Un modelo

```r
Modelo2 = lm(Estatura~ Peso+Sexo,M)
Modelo2
```

```
##
## Call:
## lm(formula = Estatura ~ Peso + Sexo, data = M)
##
## Coefficients:
```

```
## (Intercept)         Peso        SexoM
##     1.27271      0.00523      0.01218
```

```r
modelo = lm(Peso ~ Estatura, data = datos)
summary(modelo)
```

```
##
## Call:
## lm(formula = Peso ~ Estatura, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.8653  -3.7654   0.6706   5.0142  15.6006
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -151.883      7.655  -19.84   <2e-16 ***
## Estatura     133.793      4.741   28.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.883 on 438 degrees of freedom
## Multiple R-squared:  0.6452, Adjusted R-squared:  0.6444
## F-statistic: 796.5 on 1 and 438 DF,  p-value: < 2.2e-16
```

```r
f_stat = summary(modelo)$fstatistic
f_p_value = pf(f_stat[1], f_stat[2], f_stat[3], lower.tail = FALSE)
f_p_value < 0.03
```

```
## value
##  TRUE
```

Verifica la significancia de $\hat{\beta}$i con un alfa de 0.03.

```r
beta1_p_value = summary(modelo)$coefficients[2, 4]
beta1_p_value < 0.03
```

```
## [1] TRUE
```

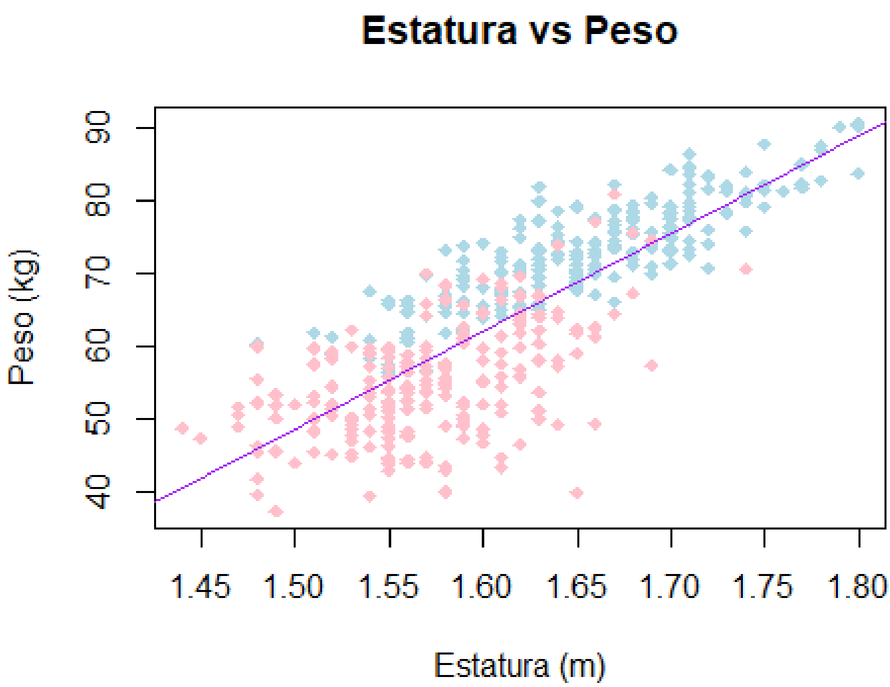
Verifica el porcentaje de variación explicada por el modelo

```r
r_squared = summary(modelo)$r.squared
r_squared
```

```
## [1] 0.6452023
```

Dibuja el diagrama de dispersión de los datos y la recta de mejor ajuste.

```r
colors <- ifelse(M$Sexo == "H", "lightblue", "pink")
plot(datos$Estatura, datos$Peso, main="Estatura vs Peso",
    ylab="Peso (kg)", xlab="Estatura (m)", col=colors, pch=18)
abline(modelo, col="purple", lwd=1.5)
```



Interpreta en el contexto del problema cada uno de los análisis que hiciste.

Vemos que el modelo es significativo, o sea que si existe una relación lineal entre la estura y el peso. El R^2 al ser de 0.645 nos dice aproximadamente un 64.5% de los datos si ajustan es decir que el 64.5% de la variabilidad del peso se explica por la variabilidad en la estatura. Y tambien por los resultados que obtuvimos en el analisis ahora sabemos que el aumento de peso asociado con un aumento de estatura es distinto entre los géneros.

Interpreta en el contexto del problema:

¿Qué información proporciona $\hat{\beta}0$ sobre la relación entre la estatura y el peso de hombres y mujeres?

Los resultados obtenidos no nos hacen sentido ni tienen un significado practico en la vida real de este contexto, sin embargo es uno de los componentes necesarios para la ecuación de regresión.

¿Cómo interpretas $\hat{\beta}1$ en la relación entre la estatura y el peso de hombres y mujeres?

En hombres lo que sucede es que por cada kilo que aumenta de peso su estatura incrementara en 0.007576 metros o sea menos de un cm. Es un cambio pequeño pero es una relación fuerte entre peso y estatura en los hombres.

Y en mujeres lo que sucede es que por cada kilo que aumentan de peso su estatura incrementará en 0.00339 metros. Es un aumento menor que en hombres por lo cual podemos pensr que el peso no está tan directamente relacionado con la estatura.