

Actividad10_Modulo1_Regresion_Lineal

Rodolfo Jesús Cruz Rebollar

2024-08-31

La recta de mejor ajuste

Leer Los datos del archivo csv

```
estatura_peso = read.csv("Estatura_Peso.csv")
```

```
head(estatura_peso)
```

```
##   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
```

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

Obtener La matriz de correlación de Los datos

```
datos_M = subset(estatura_peso, estatura_peso$Sexo == "M")
```

```
datos_H = subset(estatura_peso, estatura_peso$Sexo == "H")
```

```
datos_general = data.frame(datos_H$Estatura,
                           datos_H$Peso, datos_M$Estatura, datos_M$Peso)
```

Obtener matriz de correlación de Los datos separados

```
cor(datos_general)
```

```
##               datos_H.Estatura datos_H.Peso datos_M.Estatura
datos_M.Peso
## datos_H.Estatura    1.0000000000  0.846834792  0.0005540612
0.04724872
## datos_H.Peso        0.8468347920  1.0000000000  0.0035132246
0.02154907
## datos_M.Estatura    0.0005540612  0.003513225  1.0000000000
0.52449621
## datos_M.Peso        0.0472487231  0.021549075  0.5244962115
1.00000000
```

Interpretación de la matriz de correlación

En la matriz de correlación obtenida previamente, es posible observar que existe un grado de correlación mayormente elevado entre la variable de la estatura de los hombres y el peso también de los hombres, siendo el nivel de correlación igual a 0.8468, por lo cual al ser un nivel de correlación positivo, eso indica que la estatura de los hombres es directamente proporcional al peso de los mismos, motivo por el cual, se espera que conforme la estatura de los hombres incrementa, también aumenta su peso. Además de lo anterior, en la matriz de correlación también se puede apreciar que en cuanto a las mujeres, la estatura de las mismas tiene la mayor correlación de todas con el peso de las mismas, siendo el grado de correlación entre ambas variables (estatura de mujeres y peso de mujeres) igual a 0.5244 que además al ser positivo, indica que se espera que entre mayor sea la estatura de las mujeres, su peso también será mayor, aunque dicha relación no es tan fuerte como en el caso de los hombres, por lo cual a partir de la matriz de correlación se concluye que en el caso de los hombres, la relación entre peso y estatura es mucho más fuerte que en el caso de las mujeres, por lo cual se afirma que el sexo de la persona, ya sea masculino, o femenino, tiene una influencia significativa en el peso que se espera que tenga dicha persona en función de su estatura, lo cual demuestra que los hombres son aquellas personas que ganan mucho mayor peso a medida que aumentan de estatura que las mujeres.

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

```
# Definir cantidad de variables para separar los datos principales

# 4 variables: estatura de hombres, peso de hombres, estatura de mujeres,
# peso de mujeres

n = 4

# Definir una matriz de valores nulos NA de n filas por 7 columnas para
# rellenarla con el resumen de las medidas estadísticas para cada una de
las
# las variables del conjunto de datos original

d = matrix(NA, ncol = 7, nrow = n)

# Ciclo for para calcular el resumen de medidas estadísticas por cada
variable del
# dataframe de datos original y por cada resumen calculado, guarde los
valores de los
# estadísticos en cada fila de la matriz d definida anteriormente

for(i in 1:n){

  # Por cada iteración, calcular el resumen de medidas estadísticas de
  cada una de las
```

```

# variables y guardar los valores de los estadísticos en cada fila i de
la matriz d
# anterior

d[i, ] = c(as.numeric(summary(datos_general[, i])), sd(datos_general[
,i]))

}

# Convertir la matriz d ya con las medidas estadísticas para las
diferentes variables
# en un dataframe para una mejor comprensión de los valores de los
estadísticos

medidas_estadisticas = as.data.frame(d)

# Definir nombres de las filas del dataframe creado

row.names(medidas_estadisticas)=c("H-Estatura", "H-Peso", "M-Estatura", "M-
Peso")

# Definir los nombres de las columnas para el dataframe creado con
anterioridad

names(medidas_estadisticas)=c("Mínimo", "Q1", "Mediana", "Media", "Q3", "Máxim
o", "Desv Est")

# Mostrar dataframe con valores de las medidas estadísticas por variable

medidas_estadisticas

##           Mínimo      Q1 Mediana      Media      Q3 Máximo      Desv Est
## H-Estatura    1.48  1.6100    1.650  1.653727  1.7000    1.80  0.06173088
## H-Peso        56.43 68.2575    72.975 72.857682 77.5225    90.49  6.90035408
## M-Estatura    1.44  1.5400    1.570  1.572955  1.6100    1.74  0.05036758
## M-Peso        37.39 49.3550    54.485 55.083409 59.7950    80.87  7.79278074

```

La recta de mejor ajuste

Para encontrar la ecuación de regresión que representa el mejor ajuste para los datos en cuestión, a continuación se realizarán 3 modelos de regresión: uno que explique la relación entre peso y estatura de los hombres, otro que explique la relación entre peso y estatura en mujeres y un tercero explicando la relación entre peso y estatura de forma general entre ambos sexos de personas, tanto mujeres como hombres.

Modelo 1: relación peso y estatura de mujeres

```

# Modelo 1 que representa la relación entre el peso y la estatura en las
mujeres

```

```

modelo.mujeres = lm(Peso ~ Estatura, datos_M)

# Mostrar el modelo que mejor representa el peso en función de la
# estatura en las mujeres

modelo.mujeres

##
## Call:
## lm(formula = Peso ~ Estatura, data = datos_M)
##
## Coefficients:
## (Intercept)      Estatura
##      -72.56         81.15

```

Modelo 2: la relación entre peso y estatura en hombres

Modelo 2 para representar la relación entre el peso y la estatura en los hombres

```

modelo.hombres = lm(Peso ~ Estatura, data = datos_H)

# Mostrar el modelo que mejor representa el peso en función de la
# estatura en hombres

modelo.hombres

##
## Call:
## lm(formula = Peso ~ Estatura, data = datos_H)
##
## Coefficients:
## (Intercept)      Estatura
##      -83.68         94.66

```

Modelo 3: relación del peso con el sexo y la estatura tanto en hombres como en mujeres

Modelo 3 para explicar la relación del peso con el sexo y la estatura de la persona
ya sea hombre o mujer

```

modelo.HM = lm(Peso ~ Estatura + Sexo, estatura_peso)

modelo.HM

##
## Call:
## lm(formula = Peso ~ Estatura + Sexo, data = estatura_peso)
##
## Coefficients:
## (Intercept)      Estatura      SexoM
##      -74.75         89.26        -10.56

```

Ecuaciones de las rectas de mejor ajuste encontradas

*Modelo 1: relación peso-estatura en hombres: $\text{Peso} = 94.66 * \text{Estatura} - 83.68$*

*Modelo 2: relación peso-estatura en mujeres: $\text{Peso} = 81.15 * \text{Estatura} - 72.56$*

*Modelo 3: relación del peso con la estatura y el sexo: $\text{Peso} = 89.26 * \text{Estatura} - 10.56 * \text{Sexo} - 74.75$*

Summary del modelo peso vs estatura para las mujeres

Resumen del modelo para Las mujeres

```
summary(modelo.mujeres)

##
## Call:
## lm(formula = Peso ~ Estatura, data = datos_M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3256  -4.1942   0.4004   4.2724  17.9114
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -72.560     14.041  -5.168 5.34e-07 ***
## Estatura      81.149       8.922   9.096 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.65 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
```

Summary del modelo peso vs estatura para los hombres

Resumen del modelo de regresión peso vs estatura para Los hombres

```
summary(modelo.hombres)

##
## Call:
## lm(formula = Peso ~ Estatura, data = datos_H)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##  -8.3881  -2.6073  -0.0665   2.4421  11.1883
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -83.685       6.663  -12.56 <2e-16 ***
## Estatura      94.660       4.027   23.51 <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.678 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
```

Summary del modelo peso vs estatura y sexo tanto en hombres como en mujeres

Resumen del modelo de regresión del peso en función de la estatura y el sexo

```
summary(modelo.HM)
```

```
##
## Call:
## lm(formula = Peso ~ Estatura + Sexo, data = estatura_peso)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.9505  -3.2491   0.0489   3.2880  17.1243
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -74.7546     7.5555  -9.894  <2e-16 ***
## Estatura      89.2604     4.5635  19.560  <2e-16 ***
## SexoM        -10.5645     0.6317 -16.724  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.381 on 437 degrees of freedom
## Multiple R-squared:  0.7837, Adjusted R-squared:  0.7827
## F-statistic: 791.5 on 2 and 437 DF,  p-value: < 2.2e-16
```

Validación de los diferentes modelos generados

Prueba de Hipótesis para validar la significancia de los modelos de regresión:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

Validación del modelo de regresión 1 (peso vs estatura en mujeres)

En cuanto al primer modelo de regresión lineal realizado para explicar la relación entre el peso y la estatura en las mujeres, es posible observar en el summary de dicho modelo que el coeficiente β_1 del mismo posee un valor de significancia inferior a $2e-16$, lo cual al ser mucho menor que el nivel de significancia de 0.03 para la prueba, se cuenta con suficiente evidencia estadística para rechazar H_0 , motivo por el cual, se afirma que el coeficiente β_1 sí es significativo en el contexto estadístico.

Adicionalmente, también cabe mencionar que en cuanto a la significancia del modelo en general, éste modelo posee un valor p inferior a $2.2e-16$, lo cual al ser menor que 0.03 se tiene evidencia estadística suficiente para rechazar H_0 , lo cual nos lleva a

afirmar que el modelo sí es estadísticamente significativo. Además, en relación al porcentaje de variación explicada por el modelo (R^2), es posible observar en el summary del modelo que el coeficiente R^2 del modelo es igual a 0.2718, lo cual señala que el modelo es capaz de explicar el 27.18% de la variabilidad total presente en los datos de mujeres para estatura y peso, por lo cual, se puede concluir que aunque estadísticamente hablando el modelo sea significativo, éste mismo no es capaz de explicar un porcentaje de variabilidad suficiente de los datos en cuestión, por lo que las predicciones derivadas de dicho modelo no serán suficientemente confiables como para reflejar la relación verdadera entre el peso y la estatura en la población femenina.

Validación del modelo de regresión 2 (peso vs estatura en hombres)

En cuanto al segundo modelo de regresión lineal realizado para explicar la relación entre el peso y la estatura en los hombres, es posible observar en el summary de dicho modelo que el coeficiente β_1 del mismo posee un valor de significancia inferior a $2e-16$, lo cual al ser mucho menor que el nivel de significancia de 0.03 para la prueba, se cuenta con suficiente evidencia estadística para rechazar H_0 , motivo por el cual, se afirma que el coeficiente β_1 sí es significativo en el contexto estadístico. Adicionalmente, también cabe mencionar que en cuanto a la significancia del modelo en general, éste modelo posee un valor p inferior a $2.2e-16$, lo cual al ser menor que 0.03 se tiene evidencia estadística suficiente para rechazar H_0 , lo cual nos lleva a afirmar que el modelo sí es estadísticamente significativo. Además, en relación al porcentaje de variación explicada por el modelo (R^2), es posible observar en el summary del modelo que el coeficiente R^2 del modelo es igual a 0.7158, lo cual señala que el modelo es capaz de explicar el 71.58% de la variabilidad total presente en los datos de hombres para estatura y peso, por lo cual, se puede concluir que además de que estadísticamente hablando el modelo sea significativo, éste mismo también es capaz de explicar un porcentaje de variabilidad suficientemente aceptable de los datos en cuestión, por lo que las predicciones derivadas de dicho modelo serán en su mayoría suficientemente confiables como para reflejar la relación verdadera de la estatura con el peso en la población masculina.

Validación del modelo de regresión 3 (peso vs estatura y sexo)

En cuanto al primer modelo de regresión lineal realizado para explicar la relación del peso con la estatura y el sexo de las personas, es posible observar en el summary de dicho modelo que los coeficientes β_i del mismo poseen un valor de significancia inferior a $2e-16$, lo cual al ser mucho menor que el nivel de significancia de 0.03 para la prueba, se cuenta con suficiente evidencia estadística para rechazar H_0 , motivo por el cual, se afirma que los coeficientes β_i sí son significativos en el contexto estadístico. Adicionalmente, también cabe mencionar que en cuanto a la significancia del modelo en general, éste modelo posee un valor p inferior a $2.2e-16$, lo cual al ser menor que 0.03 se tiene evidencia estadística suficiente para rechazar H_0 , lo cual nos lleva a afirmar que el modelo sí es estadísticamente significativo. Además, en relación al porcentaje de variación explicada por el modelo (R^2), es posible observar en el summary del modelo que el coeficiente R^2 del modelo es igual a 0.7827, lo cual señala

que el modelo es capaz de explicar el 78.27% de la variabilidad total presente en los datos de ambos sexos de personas para estatura y peso, por lo cual, se puede concluir que además de que estadísticamente hablando el modelo sea significativo, éste mismo es capaz de explicar un porcentaje de variabilidad suficiente de los datos en cuestión, incluso mayor que el porcentaje de variabilidad explicado por el segundo modelo (peso vs estatura en hombres), por lo que las predicciones derivadas de dicho modelo tendrá incluso una mayor confiabilidad que las predicciones resultantes del segundo modelo, por lo cual lograrán reflejar de una forma mayormente confiable, la relación verdadera entre del sexo y la estatura con el peso en la población general.

Gráfico de dispersión de los datos y de la recta de mejor ajuste

Gráfico de dispersión y recta de mejor ajuste para las mujeres

*# Gráfico de dispersión de peso contra estatura de Los datos correspondientes a
Las mujeres*

```
plot(datos_M$Estatura, datos_M$Peso, col = "red",  
      main = "Mujeres: Peso vs Estatura", xlab = "Estatura", ylab =  
      "Peso")  
  
abline(modelo.mujeres, col = "darkgreen", lwd = 3)
```

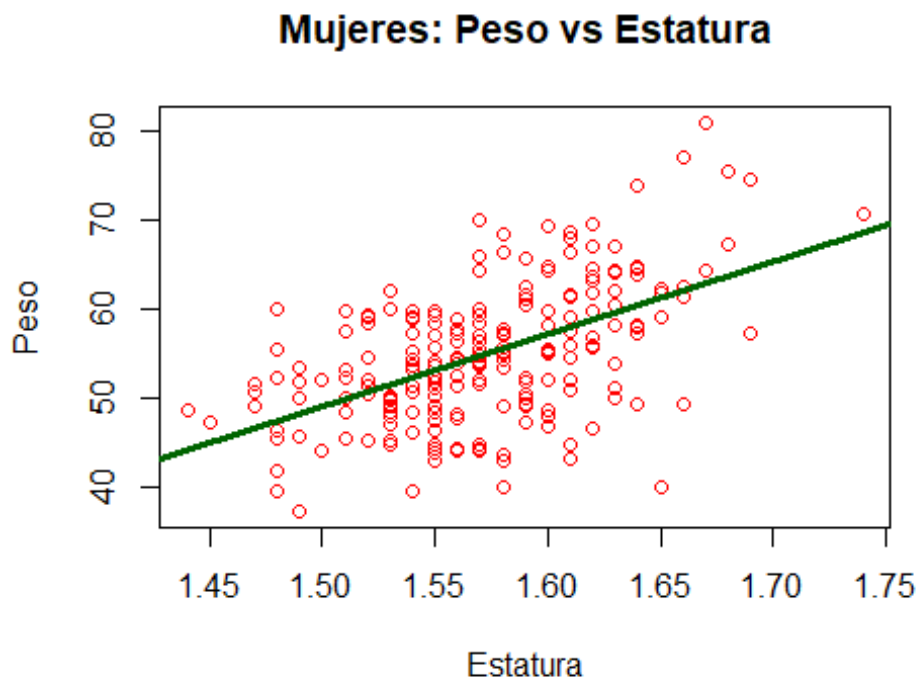


Gráfico de dispersión y recta de mejor ajuste para los hombres

Gráfico de dispersión de peso contra estatura de Los datos correspondientes a


```
# Los hombres

plot(datos_H$Estatura, datos_H$Peso, col = "blue",
      main = "Hombres: Peso vs Estatura", xlab = "Estatura", ylab =
"Peso",
      pch = 19)

abline(modelo.hombres, col = "darkred", lwd = 2.5)
```

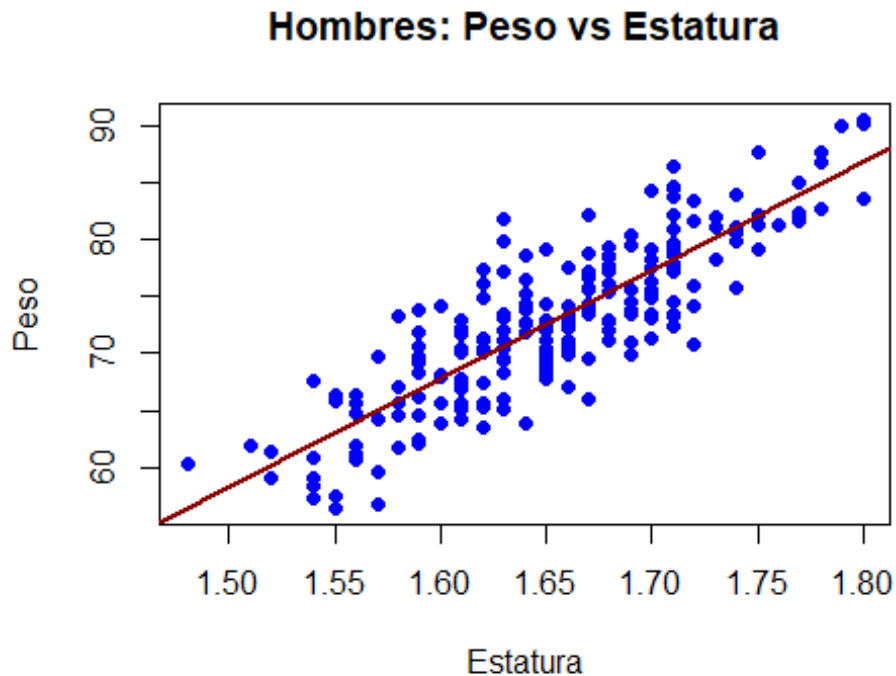


Gráfico de dispersión y recta de mejor ajuste para peso en función de sexo y estatura

Separar los coeficientes del modelo en 3 variables por separado

```
b0 = modelo.HM$coefficients[1] # intercepto del modelo beta0
b1 = modelo.HM$coefficients[2] # coeficiente para la estatura beta1
b2 = modelo.HM$coefficients[3] # coeficiente para el sexo beta2
```

Definir función para predecir el peso en función del sexo y estatura para las mujeres

```
Ym = function(x){b0 + b2 + b1 * x}
```

Definir función para predecir el peso en función del sexo y estatura para las mujeres

```
Yh = function(x){b0 + b1 * x}
```

```

# Colores para graficar los puntos y las rectas correspondientes a
# mujeres y hombres
# rosado: mujeres, azul: hombres

colores = c("pink2", "blue")

# Graficar la estatura de la población general vs su peso con los colores
# indicados para
# distinguir los datos correspondientes a las mujeres y a los hombres

plot(estatura_peso$Estatura, estatura_peso$Peso, col = colores, pch = 19,
      ylab="Peso",
      xlab="Estatura", main="Relación peso vs estatura")

# Datos de estatura desde la estatura mínima hasta la máxima con
# incrementos de 0.01

x = seq(1.40, 1.80, 0.01)

# Graficar recta estatura contra peso predicho por el modelo para las
# mujeres

lines(x, Ym(x), col = "pink2", lwd = 2)

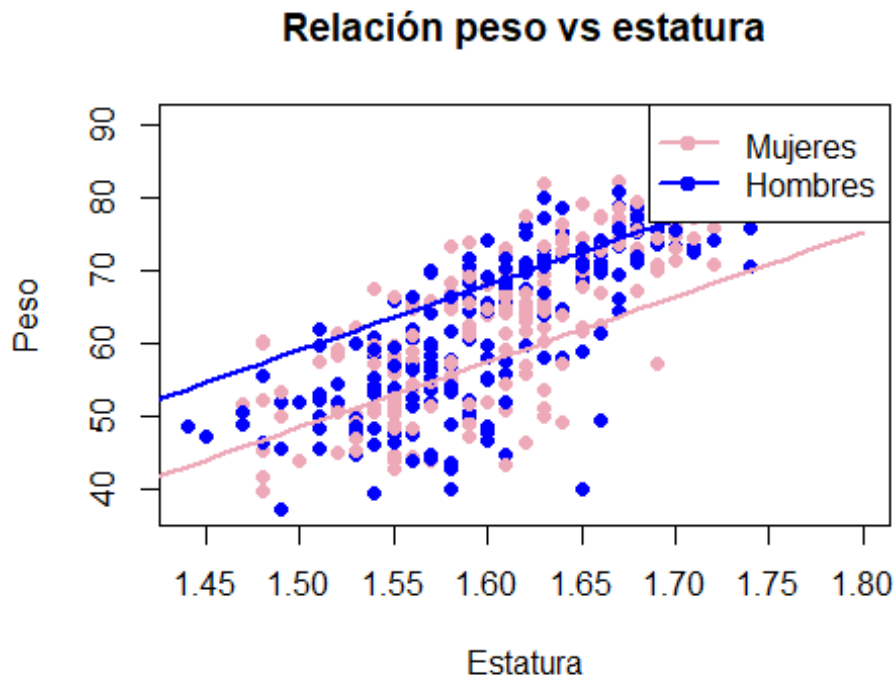
# Graficar recta de estatura contra peso predicho por el modelo para
# hombres

lines(x, Yh(x), col = "blue", lwd = 2)

# Agregar una leyenda al gráfico para distinguir fácilmente entre hombres
# y mujeres

legend("topright", legend = c("Mujeres", "Hombres"), col =
      c("pink2", "blue"), lwd = 2,
      pch = 19)

```



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

En primera instancia, en el primer análisis realizado referente al modelo de peso en función de la estatura exclusivamente en los hombres, cabe mencionar que en el contexto del problema, a medida que se incrementa la estatura en los hombres, también se incrementa de forma gradual su peso, lo cual se evidencia en el gráfico de dispersión de los datos y de la recta del modelo para los datos de los hombres, mientras que de forma similar, aunque con una menor fuerza de correlación, las mujeres también tienen una tendencia a experimentar un aumento en su peso a medida que su estatura también incrementa, aunque en el caso particular de las mujeres, se observa que hay algunos datos que se encuentran más alejados de la línea principal conforme la estatura aumenta, lo cual respalda el hecho de que el grado de correlación entre estatura y peso en el caso de las mujeres sea menor que en el caso de los hombres. Por otro lado, en cuanto al modelo que explica la relación entre el peso en función de la estatura y el sexo de la persona, es posible interpretar que la probabilidad de que una persona tenga un mayor peso se incrementa significativamente cuando la persona es de sexo masculino (hombres), mientras que si es mujer, dicha probabilidad disminuye, lo cual tiene sentido, ya que los hombres tienden a realizar actividades que implican un mayor grado de esfuerzo, mientras que las mujeres no, motivo por el cual, los hombres también tienen la tendencia a tener una dieta más o menos balanceada en función del grado de intensidad de la actividad física que realicen, esto con el propósito de contar con los nutrientes necesarios y suficientes para poder llevarlas a cabo de una forma saludable y para que a su vez, su rendimiento en ellas sea el mejor posible.

Interpreta en el contexto del problema

¿Qué información proporciona β_0 sobre la relación entre la estatura y el peso de hombres y mujeres?

En cuanto a los modelos de regresión lineal realizados, el parámetro del modelo β_0 que corresponde al valor del intercepto del mismo, hace referencia al valor que adopta la variable dependiente del modelo cuando el valor de la variable independiente del modelo es igual a 0, lo cual aplicado a este problema en específico, indica que hipotéticamente hablando, cuando la estatura de una persona es igual a 0, lo cual no posee ningún sentido físico, el peso de dicha persona será igual a -74.75 lo cual tampoco tiene sentido en un contexto físico, ya que tanto el peso como la estatura no pueden ser negativos, o iguales a 0, motivo por el cual, será necesario seguir generando posibles modelos para explicar de una mejor forma los datos en cuestión, principalmente quitando aquellos factores o variables que más allá de resultar ser estadísticamente significativas, no tengan ningún sentido en el contexto del problema, y después de quitar esas variables una por una, volver a generar otro modelo para explicar los datos y ver de qué manera mejoran los aspectos generales del modelo, tales como la significancia de dicho modelo en general, la significancia de los coeficientes correspondientes a cada una de las variables predictoras, entre otros aspectos relevantes del modelo, asegurando que al mismo tiempo, cada uno de los coeficientes del modelo, incluyendo su intercepto β_0 , sí tenga sentido en el contexto del problema que se esté intentando resolver.

¿Cómo interpretas β_1 en la relación entre la estatura y el peso de hombres y mujeres?

Adicionalmente, en cuanto al coeficiente β_1 en el contexto de la relación entre la estatura de las personas y el peso de las mismas, dicho coeficiente del modelo hace referencia al hecho de que por cada unidad que cambie el valor de la estatura de una persona, el peso de la misma cambiará β_1 cantidad de unidades, es decir, que por cada metro que incremente la estatura de una persona, el incremento en su peso será de la cantidad de unidades especificada por dicho coeficiente β_1 , por ejemplo, para el caso particular de las mujeres, por cada metro que aumenten de estatura, su peso experimentará un incremento de 81.15 kilogramos, mientras que por otro lado, en el caso específico de los hombres, por cada metro que aumenten de estatura, su peso experimentará un aumento de 94.66 kilogramos.