

10. Regresión Lineal

Adrian Pineda Sanchez

2024-08-30

La recta de mejor ajuste (Primera entrega)

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

dataM = subset(data, data$Sexo=="M")
dataH = subset(data, data$Sexo == "H")
data1 = data.frame(dataH$Estatura, dataH$Peso, dataM$Estatura, dataM$Peso)
```

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

Interpreta.

1. Obtener la matriz de correlación de Los datos

```
cor_matrix <- cor(data1)
print(cor_matrix)
```

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

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

Calcular Las estadísticas descriptivas para hombres

```
medidas_hombres <- data.frame(
  Minimo = apply(data1[, c("dataH.Estatura", "dataH.Peso")], 2, min),
  Q1 = apply(data1[, c("dataH.Estatura", "dataH.Peso")], 2, quantile, 0.25),
  Mediana = apply(data1[, c("dataH.Estatura", "dataH.Peso")], 2, median),
  Media = apply(data1[, c("dataH.Estatura", "dataH.Peso")], 2, mean),
  Q3 = apply(data1[, c("dataH.Estatura", "dataH.Peso")], 2, quantile, 0.75),
  Maximo = apply(data1[, c("dataH.Estatura", "dataH.Peso")], 2, max),
  Desv_Est = apply(data1[, c("dataH.Estatura", "dataH.Peso")], 2, sd)
)
```

Calcular Las estadísticas descriptivas para mujeres

```
medidas_mujeres <- data.frame(
  Minimo = apply(data1[, c("dataM.Estatura", "dataM.Peso")], 2, min),
  Q1 = apply(data1[, c("dataM.Estatura", "dataM.Peso")], 2, quantile, 0.25),
  Mediana = apply(data1[, c("dataM.Estatura", "dataM.Peso")], 2, median),
  Media = apply(data1[, c("dataM.Estatura", "dataM.Peso")], 2, mean),
  Q3 = apply(data1[, c("dataM.Estatura", "dataM.Peso")], 2, quantile, 0.75),
)
```

```

Maximo = apply(data1[, c("dataM.Estatura", "dataM.Peso")], 2, max),
Desv_Est = apply(data1[, c("dataM.Estatura", "dataM.Peso")], 2, sd)
)

# Imprimir resultados
print(medidas_hombres)

##           Minimo      Q1 Mediana      Media      Q3 Maximo  Desv_Est
## dataH.Estatura   1.48  1.6100   1.650  1.653727  1.7000   1.80 0.06173088
## dataH.Peso       56.43 68.2575  72.975 72.857682 77.5225  90.49 6.90035408

print(medidas_mujeres)

##           Minimo      Q1 Mediana      Media      Q3 Maximo  Desv_Est
## dataM.Estatura   1.44  1.540   1.570  1.572955  1.610   1.74 0.05036758
## dataM.Peso       37.39 49.355  54.485 55.083409 59.795  80.87 7.79278074

```

3. Encuentra la ecuación de regresión de mejor ajuste:

```

# Regresión para hombres
modelo_hombres <- lm(dataH.Peso ~ dataH.Estatura, data = data1)
modelo_hombres

##
## Call:
## lm(formula = dataH.Peso ~ dataH.Estatura, data = data1)
##
## Coefficients:
## (Intercept)  dataH.Estatura
##      -83.68      94.66

summary(modelo_hombres)

##
## Call:
## lm(formula = dataH.Peso ~ dataH.Estatura, data = data1)
##
## 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 ***
## dataH.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

```

Regresión para mujeres

```
modelo_mujeres <- lm(dataM.Peso ~ dataM.Estatura, data = data1)
modelo_mujeres
```

```
##
## Call:
## lm(formula = dataM.Peso ~ dataM.Estatura, data = data1)
##
## Coefficients:
## (Intercept) dataM.Estatura
## -72.56      81.15
```

```
summary(modelo_mujeres)
```

```
##
## Call:
## lm(formula = dataM.Peso ~ dataM.Estatura, data = data1)
##
## 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 ***
## dataM.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
```

Regresión lineal considerando la estatura y el sexo

```
modelo_ambos <- lm(Peso ~ Estatura + Sexo, data = data)
modelo_ambos
```

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

```
summary(modelo_ambos)
```

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

```
## 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
```

3.1 Realiza la regresión entre las variables involucradas

Gráfico para hombres

```
plot(data1$dataH.Estatura, data1$dataH.Peso, main = "Hombres: Estatura vs
Peso", xlab = "Estatura", ylab = "Peso", pch = 19, col = "blue")
abline(modelo_hombres, col = "red", lwd = 2)
```

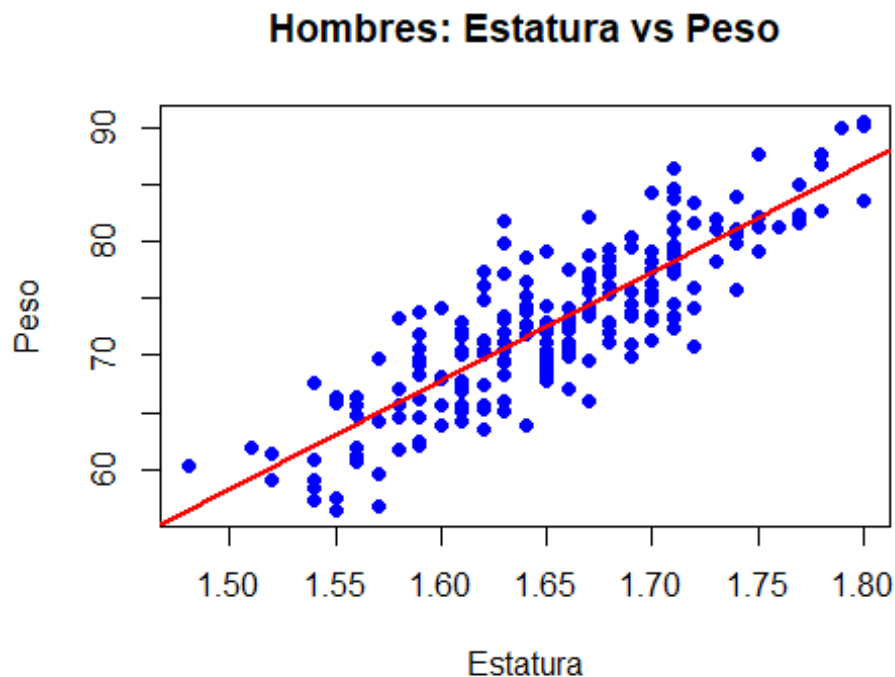
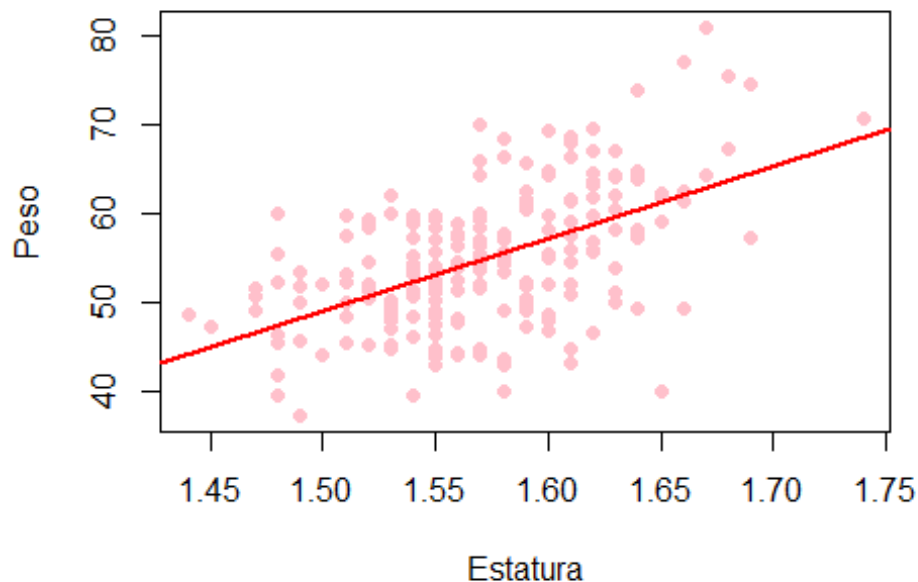


Gráfico para mujeres

```
plot(data1$dataM.Estatura, data1$dataM.Peso, main = "Mujeres: Estatura vs
Peso", xlab = "Estatura", ylab = "Peso", pch = 19, col = "pink")
abline(modelo_mujeres, col = "red", lwd = 2)
```

Mujeres: Estatura vs Peso



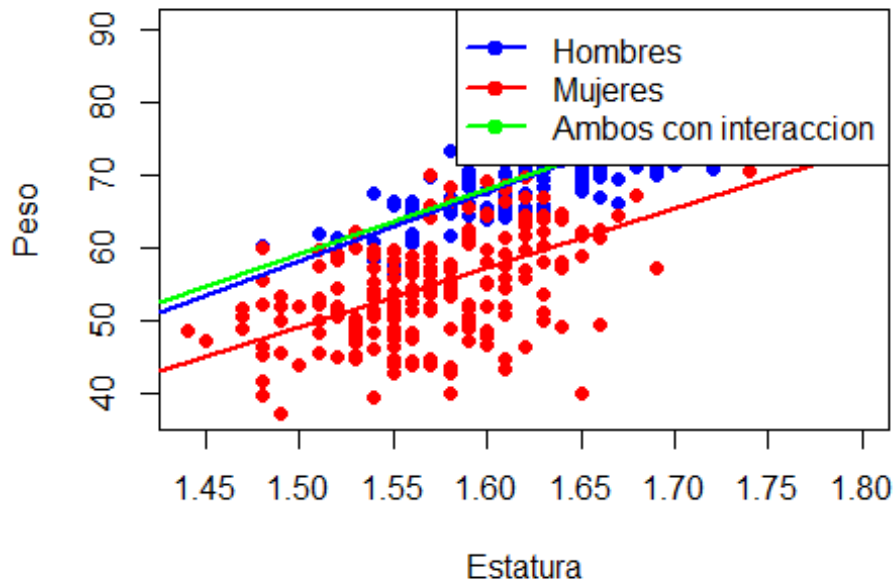
```
# Gráfico de estatura vs peso con distinción por sexo
plot(data$Estatura, data$Peso, col = ifelse(data$Sexo == "H", "blue", "red"),
     pch = 19,
     xlab = "Estatura", ylab = "Peso", main = "Estatura vs Peso con Sexo",
     xlim = c(min(data$Estatura), max(data$Estatura)),
     ylim = c(min(data$Peso), max(data$Peso)))

# Añadir líneas de regresión por sexo
abline(lm(Peso ~ Estatura, data = subset(data, Sexo == "H")), col = "blue",
       lwd = 2)
abline(lm(Peso ~ Estatura, data = subset(data, Sexo == "M")), col = "red",
       lwd = 2)
abline(lm(Peso ~ Estatura + Sexo, data = data), col = "green", lwd = 2)

## Warning in abline(lm(Peso ~ Estatura + Sexo, data = data), col = "green",
## :
## only using the first two of 3 regression coefficients

# Añadir Leyenda
legend("topright", legend = c("Hombres", "Mujeres", "Ambos con interaccion"),
     col = c("blue", "red", "green"), lwd = 2, pch = 19)
```

Estatura vs Peso con Sexo



```
# Coeficientes del modelo para hombres
coef_hombres <- summary(modelo_hombres)$coefficients
print(coef_hombres)

##              Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)  -83.68454   6.663457 -12.55873 1.410453e-27
## dataH.Estatura  94.66024   4.026565  23.50893 1.063532e-61

# Coeficientes del modelo para mujeres
coef_mujeres <- summary(modelo_mujeres)$coefficients
print(coef_mujeres)

##              Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)  -72.56045  14.040773  -5.167839 5.342762e-07
## dataM.Estatura  81.14911   8.921817   9.095581 5.997517e-17

cat("\n")

# Interpretación de  $\beta_0$  y  $\beta_1$  para hombres
cat("Para hombres,  $\beta_0$ :", coef_hombres[1, 1], "y  $\beta_1$ :", coef_hombres[2, 1],
    "\n")

## Para hombres,  $\beta_0$ : -83.68454 y  $\beta_1$ : 94.66024

cat("Interpretación: Para hombres, cuando la estatura es 0 (lo cual es
teórico), el peso promedio sería", coef_hombres[1, 1], "kg. Por cada metro
adicional de estatura, el peso promedio aumenta en", coef_hombres[2, 1],
    "kg.\n")
```

```
## Interpretación: Para hombres, cuando la estatura es 0 (lo cual es
teórico), el peso promedio sería -83.68454 kg. Por cada metro adicional de
estatura, el peso promedio aumenta en 94.66024 kg.
```

```
cat("\n")
```

```
# Interpretación de  $\beta_0$  y  $\beta_1$  para mujeres
```

```
cat("Para mujeres,  $\beta_0$ :", coef_mujeres[1, 1], "y  $\beta_1$ :", coef_mujeres[2, 1],
"\n")
```

```
## Para mujeres,  $\beta_0$ : -72.56045 y  $\beta_1$ : 81.14911
```

```
cat("Interpretación: Para mujeres, cuando la estatura es 0 (lo cual es
teórico), el peso promedio sería", coef_mujeres[1, 1], "kg. Por cada metro
adicional de estatura, el peso promedio aumenta en", coef_mujeres[2, 1],
"kg.\n")
```

```
## Interpretación: Para mujeres, cuando la estatura es 0 (lo cual es
teórico), el peso promedio sería -72.56045 kg. Por cada metro adicional de
estatura, el peso promedio aumenta en 81.14911 kg.
```

3.2 Verifica el modelo:

- $H_0: \beta_1 = 0$
- $H_1: \beta_1 \neq 0$

Hombres

```
# Resumen del modelo para hombres
```

```
summary(modelo_hombres)
```

```
##
```

```
## Call:
```

```
## lm(formula = dataH.Peso ~ dataH.Estatura, data = data1)
```

```
##
```

```
## 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 ***
```

```
## dataH.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
```

3.2.1 Verifica la significancia del modelo con un alfa de 0.03.

Valor p : $< 2.2e-16$ Conclusión: Dado que el valor p es mucho menor que 0.03, rechazamos la hipótesis nula y concluimos que el modelo es altamente significativo. Esto significa que la estatura tiene un efecto significativo sobre el peso en hombres.

3.2.2 Verifica la significancia de β_i con un alfa de 0.03.

β_0 (Intercepto) Valor p para B_0 : $< 2e-16$ Conclusión: El valor p para es mucho menor que 0.03, por lo que rechazamos la hipótesis nula. Esto significa que el intercepto es significativamente diferente de cero. En términos prácticos, aunque representa el peso teórico cuando la estatura es cero (lo cual no es realista), su significancia indica que es un componente importante del modelo.

B_1 (Pendiente) Valor p para B_1 $< 2e-16$ Conclusión: Similar a B_0 , el valor p también es mucho menor que 0.03, lo que nos permite rechazar la hipótesis nula. Esto confirma que existe una relación significativa entre la estatura y el peso en hombres, y por cada metro adicional en la estatura, el peso promedio aumenta en aproximadamente 94.66 kg.

3.2.3 Verifica el porcentaje de variación explicada por el modelo

R^2 : 0.7171 Interpretación: El modelo explica aproximadamente el 71.71% de la variabilidad en el peso de los hombres a partir de la estatura. Esto indica un ajuste bastante bueno, sugiriendo que la estatura es un buen predictor del peso en este grupo.

Mujeres

Resumen del modelo para mujeres

```
summary(modelo_mujeres)
```

```
##
## Call:
## lm(formula = dataM.Peso ~ dataM.Estatura, data = data1)
##
## 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 ***
## dataM.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
```


3.2.1 Verifica la significancia del modelo con un alfa de 0.03.

Valor p: $< 2.2e-16$ Conclusión: Dado que el valor p es mucho menor que 0.03, rechazamos la hipótesis nula y concluimos que el modelo es significativo. Esto significa que la estatura tiene un efecto significativo sobre el peso en mujeres.

3.2.2 Verifica la significancia de β_i con un alfa de 0.03.

β_0 (Intercepto) Valor p para β_0 : $5.34e-07$ Conclusión: El valor p para β_0 es menor que 0.03, por lo que rechazamos la hipótesis nula. Esto significa que el intercepto es significativamente diferente de cero. Como mencionamos antes, aunque β_0 representa el peso teórico cuando la estatura es cero, lo cual no es realista, es un componente importante del modelo.

β_1 (Pendiente) Valor p para β_1 : $< 2e-16$ Conclusión: El valor p para β_1 también es mucho menor que 0.03, lo que nos permite rechazar la hipótesis nula. Esto confirma que existe una relación significativa entre la estatura y el peso en mujeres, y por cada metro adicional en la estatura, el peso promedio aumenta en aproximadamente 81.15 kg.

3.2.3 Verifica el porcentaje de variación explicada por el modelo

El modelo explica aproximadamente el 27.51% de la variabilidad en el peso de las mujeres a partir de la estatura. Esto indica que, aunque la estatura es un predictor significativo del peso, no explica tanto la variabilidad en el peso de las mujeres como lo hizo en el caso de los hombres.

Ambos

Resumen del modelo para ambos sexos

```
summary(modelo_ambos)
```

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

3.2.1 Verifica la significancia del modelo con un alfa de 0.03.

Valor p: $< 2.2e-16$ Conclusión: Dado que el valor p es mucho menor que 0.03, rechazamos la hipótesis nula y concluimos que el modelo es significativo. Esto significa que tanto la estatura como el sexo tienen un efecto significativo sobre el peso.

3.2.2 Verifica la significancia de β_i con un alfa de 0.03.

β_0 (Intercepto) Valor p para β_0 : $< 2e-16$ Conclusión: El valor p para β_0 es mucho menor que 0.03, lo que nos permite concluir que el intercepto es significativamente diferente de cero. Este coeficiente representa el peso promedio cuando la estatura es cero y el sexo está codificado como la categoría de referencia (en este caso, hombres).

β_1 (Estatura) Valor p para β_1 : $< 2e-16$ Conclusión: El valor p para β_1 es también mucho menor que 0.03, lo que indica que existe una relación significativa entre la estatura y el peso cuando se controla por el sexo. Por cada metro adicional de estatura, el peso promedio aumenta en aproximadamente 89.26 kg.

B_2 (SexoM) Valor p para B_2 : $< 2e-16$ Conclusión: El valor p para B_2 es mucho menor que 0.03, lo que indica que el sexo tiene un efecto significativo en el peso. En este modelo, el coeficiente asociado a SexoM (-10.56) indica que, manteniendo la estatura constante, las mujeres (codificadas como "M") pesan en promedio 10.56 kg menos que los hombres.

3.2.3 Verifica el porcentaje de variación explicada por el modelo

R^2 : 0.7827

Interpretación: El modelo explica aproximadamente el 78.27% de la variabilidad en el peso considerando la estatura y el sexo. Este es un ajuste fuerte, lo que sugiere que la estatura y el sexo juntos son buenos predictores del peso.

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

```
# Gráfico para hombres
plot(data1$dataH.Estatura, data1$dataH.Peso, main = "Hombres: Estatura vs
Peso", xlab = "Estatura", ylab = "Peso", pch = 19, col = "blue")
abline(modelo_hombres, col = "red", lwd = 2)
```

Hombres: Estatura vs Peso

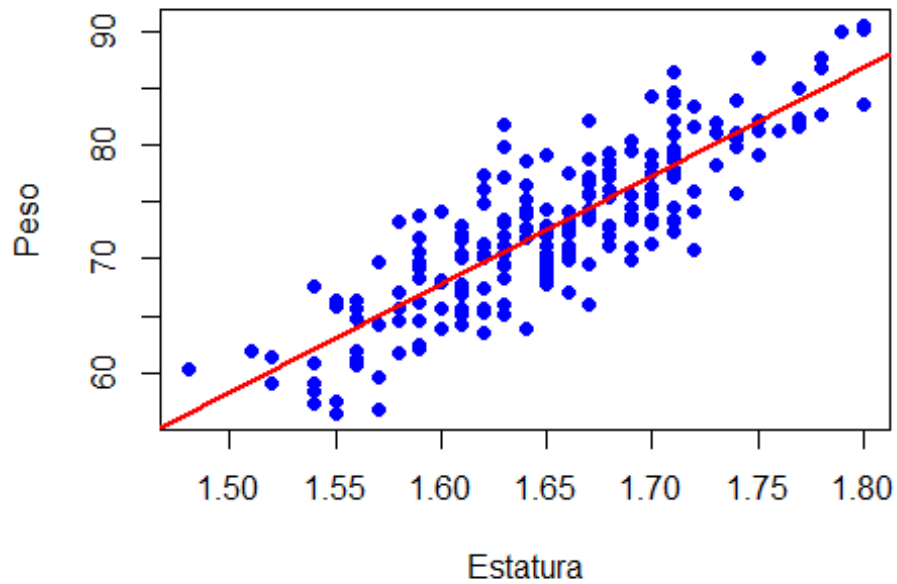
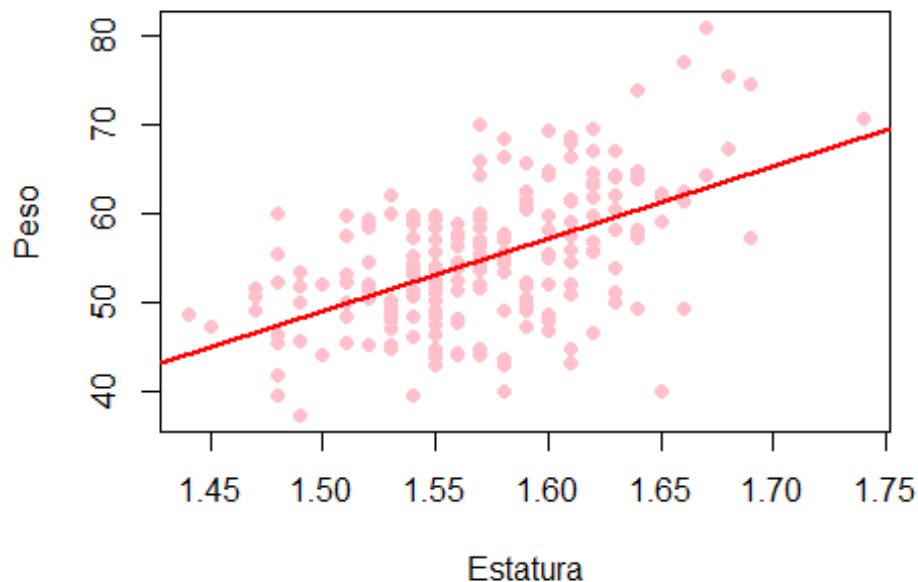


Gráfico para mujeres

```
plot(data1$dataM.Estatura, data1$dataM.Peso, main = "Mujeres: Estatura vs  
Peso", xlab = "Estatura", ylab = "Peso", pch = 19, col = "pink")  
abline(modelo_mujeres, col = "red", lwd = 2)
```

Mujeres: Estatura vs Peso

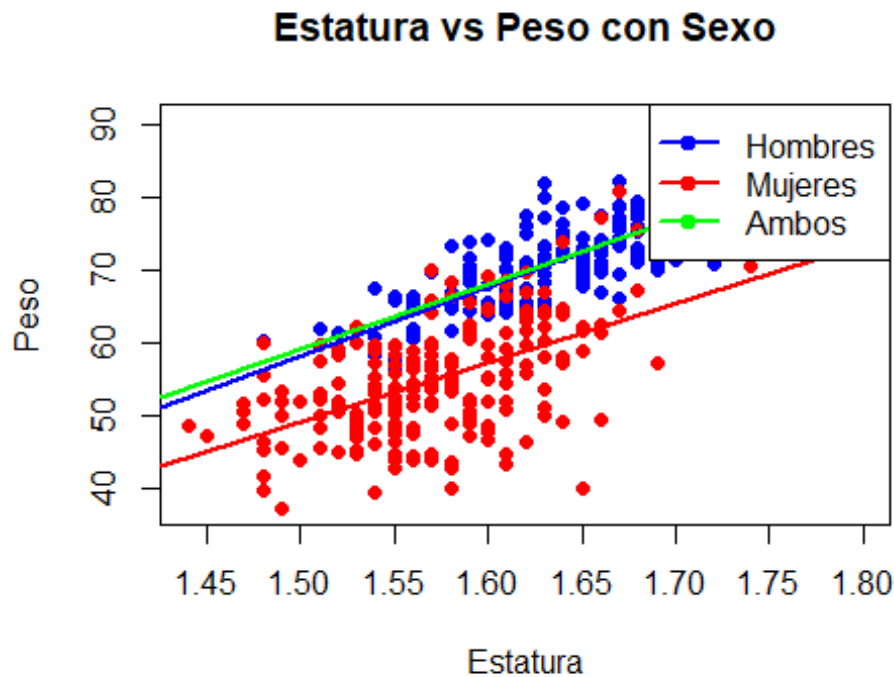


```
# Gráfico de estatura vs peso con distinción por sexo
plot(data$Estatura, data$Peso, col = ifelse(data$Sexo == "H", "blue", "red"),
     pch = 19,
     xlab = "Estatura", ylab = "Peso", main = "Estatura vs Peso con Sexo",
     xlim = c(min(data$Estatura), max(data$Estatura)),
     ylim = c(min(data$Peso), max(data$Peso)))

# Añadir líneas de regresión por sexo
abline(lm(Peso ~ Estatura, data = subset(data, Sexo == "H")), col = "blue",
       lwd = 2)
abline(lm(Peso ~ Estatura, data = subset(data, Sexo == "M")), col = "red",
       lwd = 2)
abline(lm(Peso ~ Estatura + Sexo, data = data), col = "green", lwd = 2)

## Warning in abline(lm(Peso ~ Estatura + Sexo, data = data), col = "green",
## :
## only using the first two of 3 regression coefficients

# Añadir Leyenda
legend("topright", legend = c("Hombres", "Mujeres", "Ambos"), col = c("blue",
"red", "green"), lwd = 2, pch = 19)
```



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

Relación entre Estatura y Peso por Sexo:

*Hombres (Línea Azul): Existe una tendencia positiva entre la estatura y el peso en los hombres. Es decir, a medida que la estatura aumenta, el peso también tiende a aumentar. La línea de regresión azul indica la relación específica para los hombres, mostrando una pendiente que sugiere un incremento considerable en el peso con cada unidad adicional de estatura.

*Mujeres (Línea Roja): De manera similar, hay una relación positiva entre la estatura y el peso en las mujeres, aunque la pendiente es menos pronunciada en comparación con los hombres. Esto significa que, aunque el peso también aumenta con la estatura en mujeres, lo hace a un ritmo menor que en los hombres.

*Modelo Combinado (Línea Verde): La línea verde representa el modelo que considera tanto la estatura como el sexo como predictores del peso. Esta línea refleja una tendencia intermedia, mostrando cómo la estatura influye en el peso cuando se consideran ambos sexos en conjunto. Dado que la línea está más cerca de la línea azul que de la roja, esto sugiere que el efecto de la estatura en el peso es más similar a la tendencia observada en hombres que en mujeres, pero también ajusta por el impacto del sexo.

6. Interpreta en el contexto del problema:

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

B_0 (el intercepto) representa el peso estimado cuando la estatura es cero. Aunque este valor no tiene un significado práctico directo (porque la estatura nunca es cero), es crucial para definir la línea de regresión que describe cómo el peso cambia con la estatura. En términos de interpretación, B_0 ayuda a situar la recta de mejor ajuste en el gráfico.

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

B_1 (la pendiente) indica cuánto se espera que cambie el peso por cada metro adicional de estatura. En este análisis:

Hombres: B_1 es mayor, lo que significa que el peso de los hombres aumenta de manera más pronunciada con la estatura.

Mujeres: B_1 es menor, lo que indica que aunque el peso de las mujeres también aumenta con la estatura, lo hace a un ritmo más lento comparado con los hombres.