

# El modelo de regresión lineal simple

Kevin García - Alejandro Vargas

9 de marzo de 2018

# Selección de la base de datos

La base de datos 'empleados1' cuenta con información sobre la edad, la estatura y el peso de 99 personas, 12 de ellas mujeres. De esas 99 personas debíamos seleccionar una muestra de 24, fijando las 12 mujeres, por lo tanto, nos quedaron 87 hombres de los cuales debíamos seleccionar 12. La selección se hizo con un sample de R, el cuál me arroja 12 números aleatorios entre 1 y 87, esos 12 números generados fueron nuestros hombres seleccionados.

## Punto 1:Modelo lineal simple

El modelo lineal ajustado para la variable peso con la variable predictora 'estatura' fue:

$$Peso = -99,0330 + 0,9778Estatura$$

# Punto 1:Modelo lineal simple

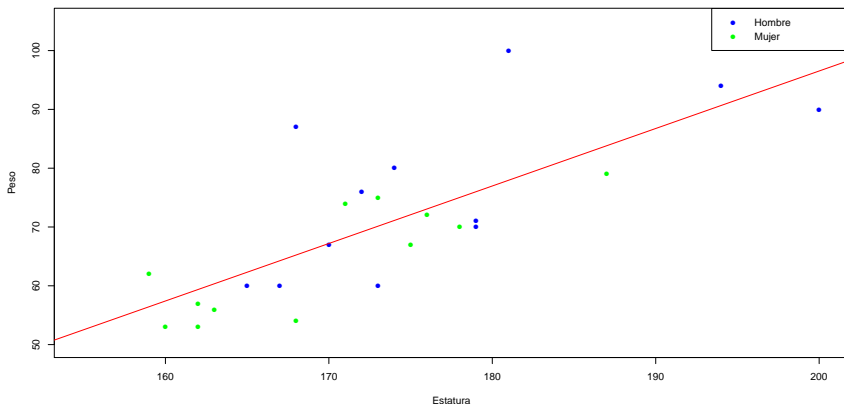


Figura: Gráfica de dispersión con la recta ajustada

## Punto 2: Bondad del modelo e interpretaciones

- $R^2 = 0,5754$ : El 57.54 % de la variabilidad total de la variable Y: 'Peso' es explicada por la variable X: 'Estatura'
- $\beta_0 = -99,0330$ : Este valor del intercepto es para un mejor ajuste del modelo teniendo en cuenta que no tenemos alturas negativas nos ayuda a ajustar el peso en función de la altura
- $\beta_1 = 0,9778$ : Cuando la variable 'Estatura', aumenta en una unidad (1 centímetro), se espera que el 'Peso' de la persona aumente en 0.9778 kg.
- p-valor  $\beta_0 = 0,00425$ : Como mi p-valor es menor que 0.05 rechazo mi hipótesis nula y digo que  $\beta_0$  si es significativo para el modelo
- p-valor  $\beta_1 = 0,0000174$ : Como mi p-valor es menor que 0.05 rechazo mi hipótesis nula y digo que  $\beta_1$  si es significativo para el modelo

## Punto 3: Intervalos de confianza para $\beta_0$ y $\beta_1$

- $\beta_0$ : (-163.456 ; -34.610) ; El verdadero valor de  $\beta_0$  está entre -163.456 y -34.610 con una confianza del 95 %
- $\beta_1$ : (0.6063928 ; 1.3492072) ; El verdadero valor de  $\beta_1$  está entre 0.6064 y 1.3492 con una confianza del 95 %

## Punto 4: Inclusión de la variable 'Sexo' al modelo

Para incluir la variable sexo al modelo, recodificamos la variable en términos binarios, la cual tomaba el valor 0 cuando es mujer y 1 cuando es hombre, además, es claro que la variable 'Altura', también depende del sexo de la persona (normalmente la media de la estatura de los hombres es mayor a la media de las mujeres), por lo cuál se tuvo en cuenta este cambio en la altura dependiendo del sexo, en pocas palabras, se tuvo en cuenta la interacción entre estas dos variables ('Altura' y 'Sexo'), el modelo ajustado incluyendo la variable 'Sexo' fue el siguiente:

$$\text{Peso} = -92,8094 + 0,9271\text{Altura} + 19,4939\text{Sexo} - 0,0813(\text{Altura} \cdot \text{Sexo})$$

## Punto 4: Inclusión de la variable 'Sexo' al modelo

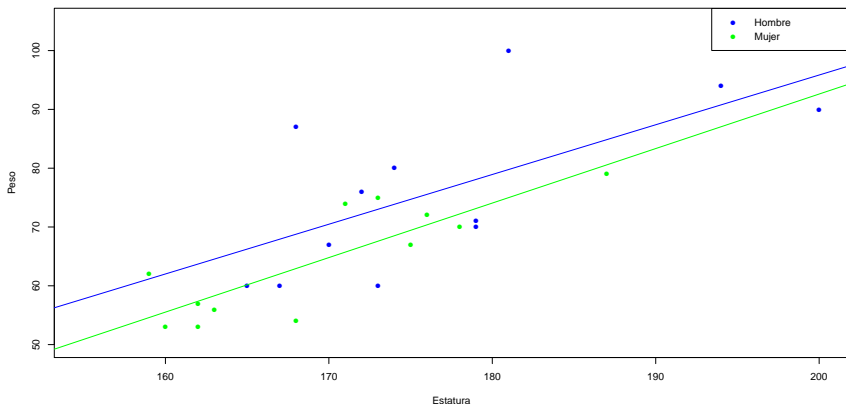


Figura: Gráfica de dispersión con la recta ajustada por sexo



## Punto 5: Comparación de modelos

Cuadro: Tabla comparativa entre los modelos ajustados

	Peso-Altura	Peso-Altura, Sexo
$R^2_{ajustado}$	0.5561	0.5574
$CME = \sigma^2$	77.33809	77.10571

## Punto 6: Inclusión de la variable 'Edad' en el modelo

El modelo ajustado incluyendo la variable edad, es el siguiente:

$$\begin{aligned} \text{Peso} = & -108,63737 + 0,91223\text{Altura} + 14,77551\text{Sexo} + 0,97851\text{Edad} \\ & - 0,05999(\text{Altura} \cdot \text{Sexo}) \end{aligned}$$

## Punto 6: Comparación de todos los modelos

Cuadro: Tabla comparativa entre los modelos ajustados

	Peso-Altura	Peso-Altura,Sexo	Peso-Altura,Sexo,Edad
$R^2_{ajustado}$	0.5561	0.5574	0.5572
$CME = \sigma^2$	77.33809	77.10571	77.14896