Universidad del Valle
Facultad de Ingeniería
Escuela de Estadística
Programa Académico de Estadística
Estadística Aplicada II
Cali, febrero de 2018

Tarea 2:El modelo de regresión lineal simple

Kevin Steven García[a], Alejandro Vargas[b]

1. Introducción

El objetivo del estudio es estimar el peso de una persona a partir de las otras variables (Altura, edad y sexo), para ello se proponen varios modelos lineales y se evaluaran con respecto a algunos estadísticos.

Para este trabajo, contamos con una base de datos de 99 personas, de las cuales, 12 eran mujeres y el resto (87) eran hombres. Nos pedian trabajar con una muestra de 24 personas, fijando las 12 mujeres, es decir, tenemos 87 hombres de los cuales debemos seleccionar 12 para completar la muestra. Con la ayuda del software R Core Team (2017) generamos 12 números aleatorios (todos los números tienen la misma probabilidad de salir) entre 1 y 87, donde cada número representa la posición en la base de datos del hombre seleccionado. Ya con nuestra base de datos conformada, procedimos a responder cada uno de los literales dados.

2. Punto 1:Modelo simple

El modelo ajustado para la variable 'Peso' con la variable de predicción 'Altura' fue el siguiente:

$$Peso = -99.0330 + 0.9778 Estatura$$

Para darnos una idea de que tan bueno es nuestro modelo, sin necesidad de evaluar a fondo cada uno de sus coeficientes, generamos una gráfica de dispersión entre las dos variables involucradas donde se observa la recta de regresión ajustada.

[a]Código: 1533173. E-mail: kevin.chica@correounivalle.edu.co
[b]Código: 1525953. E-mail: jose.alejandro.vargas@correounivalle.edu.co

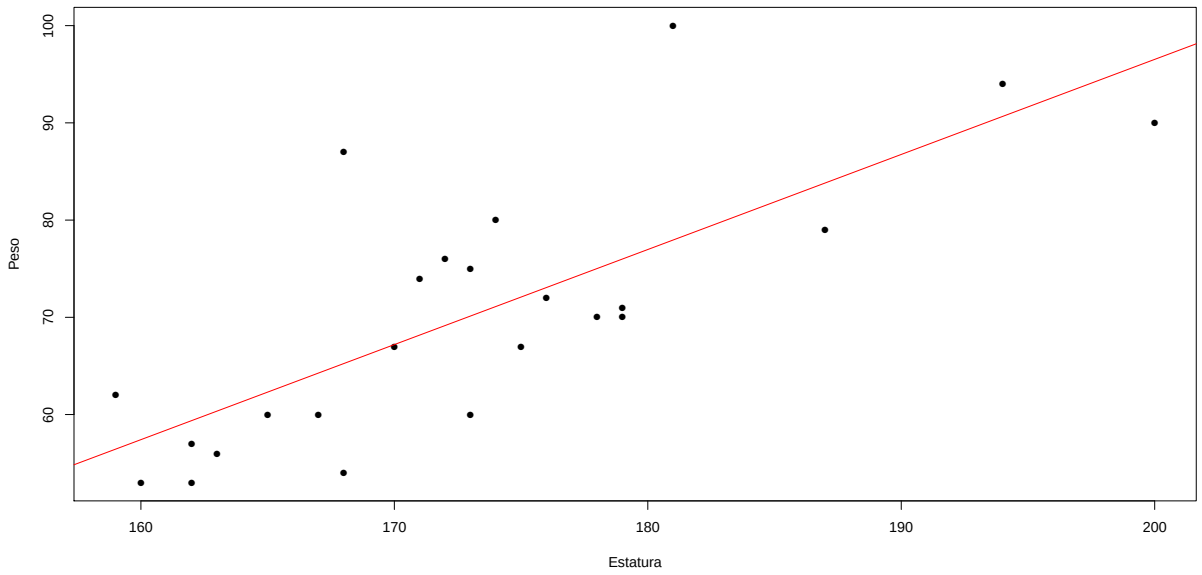

FIGURA 1: Gráfica de dispersión con la recta de regresión ajustada

En este gráfico podemos ver que nuestro modelo no es tan bueno a la hora de predecir el peso dando una estatura, ya que en algunos casos, el peso obtenido va a estar muy por encima o muy por debajo del peso real, es decir, los errores de nuestro modelo son aparentemente altos. Sin embargo, el modelo explica el comportamiento creciente de los puntos, por lo cuál observando solo esta gráfica, podríamos decir que el modelo ajustado, más específicamente la 'Altura' podría explicar aproximadamente entre el 40 y el 60 % de la variabilidad total de la variable 'Peso'.

3. Punto 2:Bondad del modelo e interpretaciones

Con respecto a las interpretaciones del modelo, se obtuvieron los valores p, de cada una de las pruebas de significancia para los coeficientes del modelo (β_0 y β_1), y se obtuvo el R^2.

- $\beta_0 = -99.0330$:
- $\beta_1 = 0.9778$:
- $R^2 = 0.5754$:
- p-valor $\beta_0 = 0.00425$:
- p-valor $\beta_1 = 0.0000174$:

4. Punto 3:Intervalos de confianza para β_0 y β_1

Para generar los intervalos de confianza, utilizamos directamente la función confint en R, los intervalos generados fueron los siguientes:

- $\beta_{0_{95\,\%}} = (-163.4559651; -34.60997)$

- $\beta_{1\,95\,\%} = (0.6064059; 1.34922)$

5. Punto 4:Inclusión de la variable 'Sexo' en el modelo

Para incluir la variable sexo al modelo, generamos una variable indicadora (variable 'Dummy') que básicamente codifica la variable categórica (en nuestro caso es la variable 'Sexo') 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' y teniendo en cuenta la interacción entre las variables 'Altura' y 'Sexo' fue el siguiente:

$$Peso = -92.8094 + 0.9271 Altura + 19.4939 Sexo - 0.0813(Altura \cdot Sexo)$$

6. Punto 5:Interpretación y comparación de modelos

Poner interpretación de los betas

Para comparar los modelos y decidir si uno es mejor que el otro, generamos la siguiente tabla, donde se comparan con respecto al $R^2_{ajustado}$ y al $CME = \sigma^2$

Tabla 1: 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

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

El modelo ajustado incluyendo la variable 'Edad' es el siguiente:

$$Peso = -108.63737 + 0.91223 Altura + 14.77551 Sexo + 0.97851 Edad - 0.05999(Altura \cdot Sexo)$$

Poner interpretacion de β's

Para comparar los tres modelos obtenidos, simplemente adicionamos este ultimo a la tabla comparativa, la cual quedo de la siguiente manera:

Tabla 2: 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

Referencias

R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*https://www.R-project.org/