



Ejercicios Regresión Lineal

Probabilidad y Estadística - Laboratorio (02-2024)

1. Utilizando los datos *mtcars* disponibles en **R**, ajuste un modelo de regresión lineal para la variable que indica el consumo de combustibles *mpg* (millas recorridas por galón), utilizando distintas variables explicativas según las siguientes preguntas.
 - a) Utilice como variable explicativa el peso del vehículo, *wt*. Interprete el coeficiente que acompaña a tal variable y determine si es significativamente distinto de 0, utilizando $\alpha = 0.05$.
 - b) Además del peso del vehículo, incluya en el modelo la variable que indica la potencia del motor, *hp*. ¿Es significativo el coeficiente que acompaña a esta nueva variable?
 - c) Ahora, incluya en el modelo anterior la variable que indica el número de cilindros, *cyl*. ¿Qué dice ahora de la significancia de los parámetros? ¿Es significativo el coeficiente de la variable *hp*?

Para el detalle de cada variable puede consultar mediante el comando `?mtcars`.

2. Cree un dataset ficticio de tamaño $n = 500$. Para ello, siga los siguientes pasos.
 - a) Simule x_1 de una distribución $Gamma(2, 2)$. Luego simule una variable x_2 desde una distribución $Normal(0, 1)$. Finalmente defina una variable $x_3 = 0.7x_1 + 0.3x_2$.
 - b) Genere un error ε desde una distribución $Normal(0, \sigma^2)$, con $\sigma = 5$.
 - c) Cree la variable respuesta mediante $y = \beta_0 + \beta_1 x_3 + \varepsilon$. Utilice $\beta_0 = 3$ y $\beta_1 = 10$.
 - d) Por último, ajuste 3 modelos de regresión para y . El primero incluyendo solo la variable x_1 . El segundo incluyendo solo la variable x_2 . Y el tercero incluyendo las variables x_1 y x_3 . ¿Qué ocurre con la significancia del coeficiente β para x_1 en este último?
3. Cuando se ajusta un modelo de regresión lineal utilizando una variable explicativa del tipo categórica, es decir, que induce grupos, **R** buscará un grupo de referencia al que le corresponderá el intercepto y luego los demás grupos serán desviaciones de aquel de referencia.

Por ejemplo, se quiere explicar el peso de abalones según el centro de cultivo en el que están. Nuestra variable categórica son los centros de Caldera, Coquimbo, Puerto Montt y Chiloé. Luego, si **R** interpreta a Caldera como referencia, entonces el modelo está dado por

$$Y_i \sim Normal(\beta_0 + \beta_1 Chiloé + \beta_2 Coquimbo + \beta_3 PuertoMontt, \sigma^2).$$

Luego, el peso medio de un abalón en Caldera es de β_0 unidades, mientras que de Chiloé es de $\beta_0 + \beta_1$. Para Coquimbo, el peso medio es de $\beta_0 + \beta_2$ y para Puerto Montt es de $\beta_0 + \beta_3$. Si $\beta_1 < 0$, entonces los abalones de Chiloé son más livianos que los de Caldera. Si $\beta_1 > 0$, éstos

son más pesados. Mientras que si $\beta_1 = 0$, no hay diferencias en los pesos de estas ciudades. Las interpretaciones para los demás centros de cultivo son análogas.

Con lo mencionado anteriormente, responda.

- a) ¿Cómo podría determinar si los abalones de Coquimbo son más o menos pesados que los de Puerto Montt y cual sería la magnitud de tal diferencia?
- b) Utilizando la base de datos *Abalon.xlsx* ajuste el modelo de regresión lineal para el peso total de los abalones (*pesot*) utilizando la variable categórica *centro* como variable explicativa. Interprete los coeficientes de regresión.
- c) ¿Existe diferencia significativa entre los centros de Caldera y Puerto Montt? Utilice un nivel $\alpha = 0.05$.