

---

Se presenta una base de datos que recopila información de diferentes clientes de un Ecommerce, a groso modo su tarea consiste en explorar dicha base de datos para ajustar un modelo de regresión adecuado donde la variable respuesta es la cantidad anual gastada por cliente.

Cuadro 1: Vista previa de la base de datos

Email	Avatar	Cantidad gastada al año por cliente
mstephenson@fernandez.com	Violet	587.9511
hduke@hotmail.com	DarkGreen	392.2049
pallen@yahoo.com	Bisque	487.5475
riverarebecca@gmail.com	SaddleBrown	581.8523
mstephens@davidson-herman.com	MediumAquaMarine	599.4061

Su misión como analista es realizar las siguientes tareas usando el software estadístico *R*.

1. Realice la lectura de la base de datos, seleccione únicamente las variables numéricas.
2. Elabore un gráfico de dispersión de las variables para encontrar aquella que presente una mejor relación lineal con respecto a la variable respuesta.
3. Escriba la ecuación del modelo de regresión, junto con sus supuestos. Ajuste un modelo de regresión lineal simple y añada la recta de regresión a la gráfica generada anteriormente. **Nota:** seleccione aleatoriamente el 80 % de los datos para ajustar el modelo.
4. Realice la prueba de significancia para la pendiente, luego realice la prueba de significancia de la regresión usando análisis de varianza. ¿Ambos enfoques permiten llegar a la misma conclusión? ¿Qué relación existe entre una prueba y la otra?
5. De una interpretación de los parámetros  $\beta_0$  y  $\beta_1$  del modelo, claro está, si es posible hacerlo.
6. Calcule el  $R^2$  usando el coeficiente de correlación y usando sumas de cuadrados, compare estos entre sí y compárelos con las salidas de R. Realice una interpretación de este.
7. Use el modelo para predecir las cantidad anual total gastada por cliente en el 20 % de los datos que no usó para ajustar el modelo. Calcule los respectivos intervalos de confianza y de predicción. ¿Cuáles intervalos son más anchos? ¿Por qué cree usted que esto sucede?