

## Guía certamen #1 IND163C - Análisis de Negocios

1. El departamento de marketing desea utilizar un presupuesto de 500000 USD en publicidad en tres medios distintos: youtube, facebook y periódicos, con el fin de aumentar las ventas de sus productos. Para ello, el equipo de inteligencia de negocios analizará datos pasados de presupuestos asignados en los mismos medios. Como modelo inicial y *benchmark* se utilizará modelos lineales para resolver la problemática. Utilizando el conjunto de datos **marketing** del paquete **datarium**. (medido en miles de dólares y miles de unidades vendidas).

Para cargar los datos utilizar:

```
#install.packages("devtools")  
library(devtools)  
devtools::install_github("kassmbara/datarium")  
library(devtools)  
data("marketing", package = "datarium")
```

Alternativamente, cargar .csv anexo.

- (a) Realizar un análisis exploratorio del conjunto de datos **marketing**
  - (b) Realizar un ajuste lineal simple para la variable **sales** para cada uno de los medios posibles. Enuncie los modelos teóricos y supuestos utilizados.
  - (c) ¿Qué modelo de regresión lineal simple es mejor? Justifique e interprete los resultados.
  - (d) Realice un ajuste lineal múltiple para la variable **sales** sin incorporar interacciones. Interprete los resultados
  - (e) Investigar la viabilidad del modelo de regresión múltiple y compare los resultados con el mejor modelo de regresión lineal simple. Obtenga intervalos de confianza para los parámetros de la regresión.
  - (f) ¿Se puede justificar una incorporación de interacción entre las variables en estudio en el modelo de regresión lineal múltiple? Proponer un modelo, analizarlo y compararlo con los modelos anteriores. ¿Qué modelo sería el más adecuado?
  - (g) Bajo el modelo lineal escogido ¿Cuál sería la distribución óptima del presupuesto?
2. Una inmobiliaria desea establecer el precio por metros cuadrados de los terrenos en un sector particular, para ello se analizará las transacciones de venta pasadas para establecer el precio adecuado de las propiedades que desean vender. Utilizando el conjunto de datos **realestate.csv**, sin considerar las variables temporales y de longitud y latitud:
    - (a) Realizar un análisis exploratorio del conjunto de datos **realestate.csv**.
    - (b) Realizar un ajuste lineal simple para la variable **house price of unit area** para cada una de las variables en estudio. Enuncie los modelos teóricos y supuestos utilizados.
    - (c) ¿Qué modelo de regresión lineal simple es mejor? Justifique e interprete los resultados.
    - (d) Realizar un ajuste lineal múltiple para la variable **sales** sin incorporar interacciones. Interprete los resultados.
    - (e) Realizar intervalos de confianza para los coeficientes de regresión y el precio medio esperado.
    - (f) ¿Es el intercepto de la regresión significativo? ¿Es adecuado no incorporar el intercepto en el modelo? Justifique.
    - (g) Predecir el precio por metro cuadrado para una propiedad con 28 años de antigüedad, a 500 metros de una estación de metro y con 4 *minimarkets* cercanos. Realice un intervalo de para esta predicción.

3. Suponga que se tiene un conjunto de datos con 5 predictores:  $X_1$  : Notas promedio,  $X_2$  : Coeficiente intelectual,  $X_3$  : Nivel educacional (1 para universidad, 0 para enseñanza media),  $X_4$  : Interacción entre las notas promedio y el C.I.,  $X_5$  : Interacción entre las notas promedios y el nivel educacional. La variable respuesta es el salario inicial tras graduarse (en miles de dólares). Suponga que se usa un método de mínimos cuadrados para ajustar un modelo de regresión múltiple, obteniendo  $\hat{\beta}_0 = 50$ ,  $\hat{\beta}_1 = 20$ ,  $\hat{\beta}_2 = 0,07$ ,  $\hat{\beta}_3 = 35$ ,  $\hat{\beta}_4 = 0,01$ ,  $\hat{\beta}_5 = -10$ .

(a) ¿Qué afirmación es correcta? Justifique

- I. Para un valor fijo del C.I. y notas promedio, graduados de educación media ganan más, en promedio, que graduados universitarios.
- II. Para un valor fijo del C.I. y notas promedio, graduados universitarios ganan más, en promedio, que graduados de enseñanza media.
- III. Para un valor fijo del C.I. y notas promedio, graduados de enseñanza media ganan más, en promedio, que los graduados universitarios si es que tienen notas promedio suficientemente altas.
- IV. Para un valor fijo del C.I. y notas promedio, graduados universitarios ganan más, en promedio, que graduados de enseñanza media si es que tienen notas promedio suficientemente altas.

(b) Predecir el salario de un graduado universitario con C.I. 110 y notas medio 4.

(c) Verdadero o Falso: Debido a que el coeficiente para la interacción entre notas promedio y C.I. es muy pequeño, existe poca evidencia de interacción entre estas variables. Justifique su respuesta