

# Ejercitación regresión múltiple

1

Problema 1: Considere el modelo lineal con intercepto y sin variable independiente:

$$Y_i = \mu + e_i, e_i \sim (0, \sigma^2) \quad i = 1 \dots n,$$

- a. Representelo matricialmente
- b. Construya las ecuaciones normales en este caso
- c. Obtenga el estimador de cuadrados mínimos

2

### Problema 2:

Construya un modelo lineal que represente el comportamiento de las muestras de tamaño  $n$  y  $m$  de dos poblaciones normales ( $Y^A, Y^B$ ), con medias respectivas  $\mu_A, \mu_B$ ; independientes y con varianzas iguales.

Represente su modelo matricialmente.

Obtenga los estimadores de cuadrados mínimos de las medias y el estimador de la varianza común

3

### Una situación aplicada

Considere un experimento aleatorio en el que 30 personas son divididas aleatoriamente en dos grupos uno de 16 personas y otro de 14. A los miembros del primer grupo, se les administra una droga llamada “B” y, a los del segundo, otra droga llamada “G”.

Se desea saber cómo las drogas influyen en el tiempo de coagulación.

Planteando el problema con el modelo anterior, la solución del problema me conduce a la comparación de medias de dos poblaciones.

**¿Se podría ver la modelación de otra forma que me permita hacer otro tipo de formulación del problema más cercana a lo que ya vimos en el análisis de regresión?**

4

## Una situación aplicada

Considere un experimento aleatorio en el que 30 personas son divididas aleatoriamente en dos grupos uno de 16 personas y otro de 14. A los miembros del primer grupo, se les administra una droga llamada “B” y, a los del segundo, otra droga llamada “G”.

Se desea saber cómo las drogas influyen en el tiempo (en minutos) de coagulación.

$$tiempo_i = \beta_0 + \beta_1 * droga_i + e_i$$

$$droga = \begin{cases} 1, & \text{si recibió droga B} \\ 0, & \text{si recibió droga G} \end{cases}$$

¿En este caso qué significan los parámetros?

¿Cómo serían los estimadores de cuadrados mínimos en este caso?

5

## Ajuste el modelo de regresión y responda a la pregunta

- (b) A un nivel de significancia del 5% los datos de la muestra evidencian que el tiempo de coagulación es mayor para quienes recibieron la droga G que para quienes recibieron la droga B? (valor 0.3)

$$tiempo_i = \beta_0 + \beta_1 * droga_i + e_i$$

$$droga = \begin{cases} 1, & \text{si recibió droga B} \\ 0, & \text{si recibió droga G} \end{cases}$$

6

# Aplicaciones

7

Una publicación que analiza el mercado y sus tendencias desea saber cuáles son los factores que influyen en el precio de los morrales. Para lograr este objetivo analizan una muestra de morrales en los que evalúan

el precio

su capacidad

lo confortables que pudieran resultar.

La capacidad se midió en pulgadas cúbicas y el confort en una escala de valores de 1 a 5, en la que 5 representa el mayor confort.

Los datos del estudio se encuentran en la base de datos morrales

- (1) Ajuste la recta de cuadrados mínimos, analice los errores de estimación y construya los correspondientes intervalos de confianza
- (2) Suponiendo homocedasticidad, estime la varianza en el modelo
- (3) Valore el efecto conjunto de las variables en el modelo y el efecto individual de cada variable
- (4) Valore el modelo de regresión que construyó
- (5) ¿Sería posible simplificar la ecuación? Valore esta posibilidad y analice cómo quedaría la nueva ecuación y sus posibilidades
- (6) Construya un intervalo de confianza (**resp. intervalo de predicción**) para estimar el valor esperado(**resp. el valor**) del precio de un morral cuando el confort =3 y la capacidad 5000.

**Tasa de error para  
análisis 0.01**

8

**Datos: fumar**

Problema de salud

Estudio de factores que influyen en el hábito de fumar

Variables

Edad: Mediana de la edad

ES: Porcentaje de personas de más de 25 años con educación superior

ingreso: ingreso per cápita

color: % personas raza negra

mujeres: % mujeres

precio: precio promedio (centavos / dólar) de una cajetilla

ventas: per cápita de cajetillas vendidas

Error  
0.01