



Modelos de Regresión y Series de Tiempo (MRST)

2025 - 02

Clase 7 – Introducción a los modelos de regresión lineal múltiple (MRLM)

Docente: Natalia Jaramillo Quiceno

Escuela de Ingenierías

natalia.jaramilloq@upb.edu.co

Regresión lineal múltiple

Introducción



- El objetivo de un modelo de regresión múltiple es construir un modelo que relacione una variable respuesta con más de un predictor.

- Ejemplo 1:** El precio de una casa podría depender de varias variables predictoras como:

x_1 : Tamaño [m²]

x_2 : Antigüedad [años]

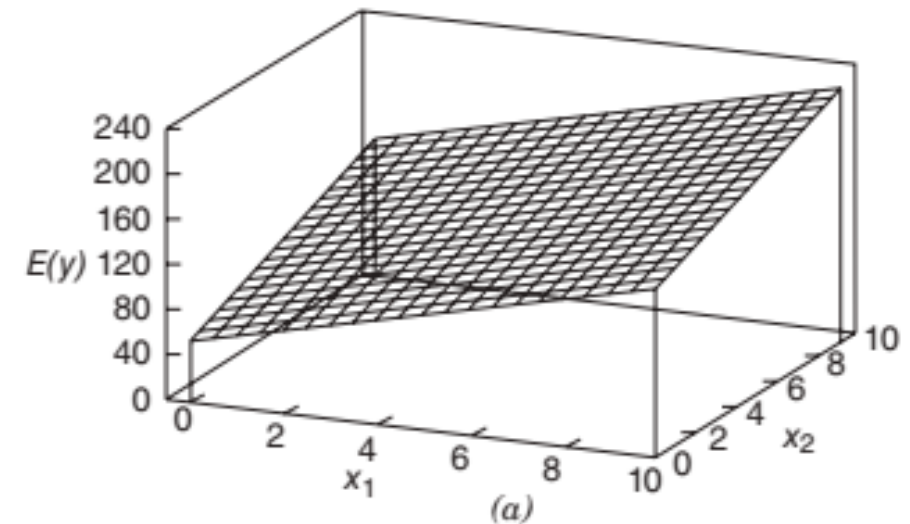
x_3 : Barrio

x_4 : Si es nueva o usada

Entre otras...(x_5, x_6, \dots, x_k)

Ejemplo 2

$$E(y) = 50 + 10x_1 + 7x_2$$



Un modelo con 2 variables predictoras describe un plano en el espacio tridimensional

Regresión lineal múltiple

Ejemplo



- Suponga que el valor del arriendo de un local depende del tamaño [m²] y de la distancia [km] a la universidad.
- Un posible **modelo** que describe esta relación podría ser:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Función lineal de los
parámetros desconocidos
 β_0, β_1 y β_2

- Este modelo de regresión lineal múltiple describe un plano en el espacio tridimensional:
 - β_0 es el intercepto del plano de regresión (cuando $x_1=x_2=0$). **Para que este parámetro tenga una interpretación correcta, el conjunto de datos debe incluir los puntos $x_1=x_2=0$.**
 - β_1 indica el cambio esperado en y (valor del arriendo) por cada cambio unitario en x_1 , mientras que x_2 permanece constante
 - β_2 indica el cambio esperado en y (valor del arriendo) por cada cambio unitario en x_2 , mientras que x_1 permanece constante

Regresión lineal múltiple

Modelo general



- En general, la variable respuesta (y) se puede relacionar con k variables independientes (regresoras, predictoras) así:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

- Es llamado modelo de regresión lineal múltiple (MRLM) y describe un hiperplano en el espacio $k+1$ dimensional
- Los parámetros β_j representan el cambio esperado en la variable respuesta por cada unidad de cambio en x_j , mientras que TODAS las demás permanecen constantes
- Otros modelos más complejos se pueden **transformar** en un MRLM

Regresión lineal múltiple

Otros modelos en MRLM



- Modelos polinómicos:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k + \varepsilon$$

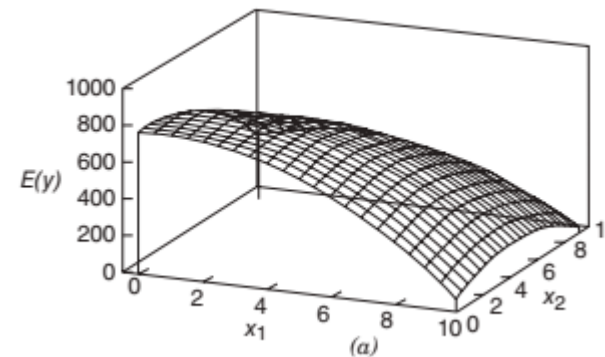
- Modelos que incluyen interacción entre variables:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1,2} x_1 x_2 + \varepsilon$$

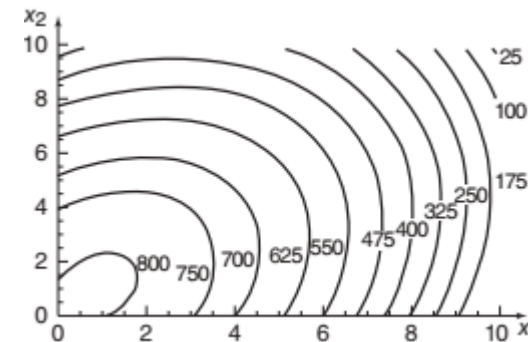
OJO: Si en los modelos hay interacción entre variables, implica que el cambio esperado en y depende no sólo de la variable que aumenta o disminuye sino de otras variables (con las que interactúa), por lo que debe tenerse MUCHO cuidado con la interpretación de los β_j

Ejemplo

$$E(y) = 800 + 10x_1 + 7x_2 - 8.5x_1^2 - 5x_2^2 + 4x_1x_2$$



Gráfica tridimensional del modelo



Curvas de nivel

Cada curva corresponde a un valor de y constante

Regresión lineal múltiple

Ejemplo en RStudio



¿La satisfacción de un paciente con la atención recibida en un hospital puede estimarse a partir de su edad, la severidad de su enfermedad y sus niveles de ansiedad?

Un ingeniero industrial desea evaluar cómo la satisfacción de los pacientes, respecto a la atención recibida en un centro hospitalario, se encuentra asociado a la edad del paciente (x_1), la severidad de su enfermedad (x_2) y su nivel de ansiedad (x_3). En total, se obtuvieron 46 observaciones para las variables:

- **Variable Respuesta (y):** índice de satisfacción del paciente (**satisfacción**)
- **Potencial predictor (x_1):** edad del paciente en años (**edad**)
- **Potencial predictor (x_2):** índice de severidad de la enfermedad padecida por el paciente (**severidad**)
- **Potencial predictor (x_3):** índice de ansiedad del paciente (**ansiedad**)

Base de datos
07 – hospital.txt

Regresión lineal múltiple

Ejemplo en RStudio

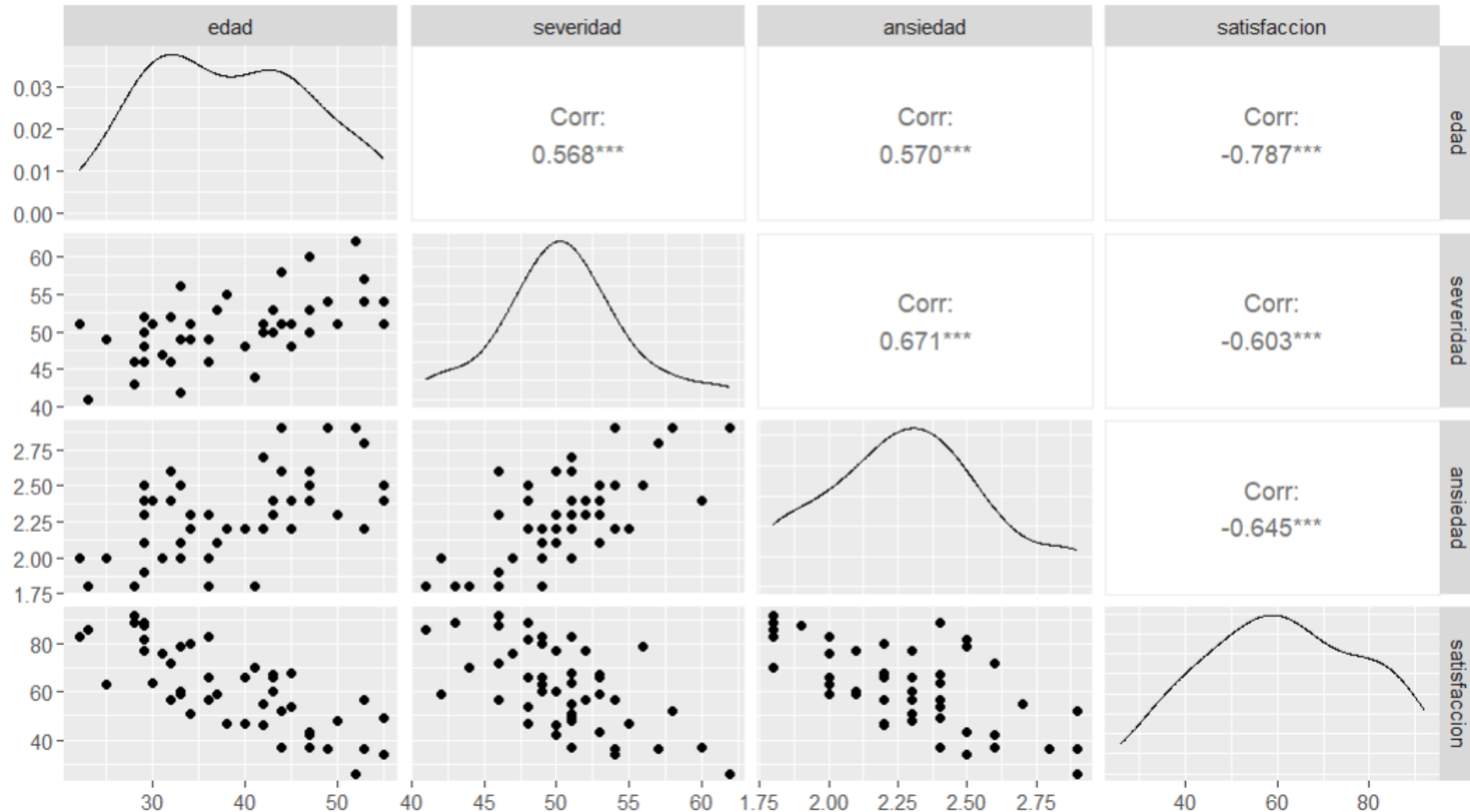


Paso 1 – Graficar los datos

- En el caso de los modelos de regresión múltiple se debe revisar la relación entre cada variable independiente con respecto a la variable dependiente y las posibles relaciones entre ellas.
- Una forma de investigar las relaciones entre todas las variables es por medio de una matriz de gráficos de dispersión (disponible en R)
 - En el gráfico que vamos a generar, para cada diagrama de dispersión, la variable en el eje vertical aparece a la derecha del gráfico y la variable en el eje horizontal aparece en la parte superior del gráfico.
 - Veamos en Rstudio con la base de datos **07 – hospital.txt**:

Regresión lineal múltiple

Ejemplo en RStudio



Regresión lineal múltiple

Seguimiento – Clase 7



A partir del muestreo de 113 hospitales en Estados Unidos, se evaluó la incidencia de diferentes factores en la probabilidad de que un paciente adquiriera una infección mientras está hospitalizado. Las variables analizadas fueron:

- **Variable Respuesta (y):** riesgo de infección en porcentaje (**riesgo**)
- **Potencial predictor (x_1):** tiempo de hospitalización promedio de los pacientes (**tiempo**)
- **Potencial predictor (x_2):** edad promedio de los pacientes (**edad**)
- **Potencial predictor (x_3):** índice de rayos X realizados (**tasarayosx**)

Base de datos
07 – riesgo.txt

A partir de los datos obtenidos, los cuales se encuentran en la base de datos **07 – riesgo.txt**, realice un diagrama de dispersión múltiple para las variables de interés y responda **¿qué variables parecen presentar una relación lineal más fuerte?**

Regresión lineal múltiple

Avance - Taller en Rstudio



Actividad:

Buscar un conjunto de datos que incluya una variable respuesta (y) y más de 2 variables regresoras (x). Las variables regresoras pueden ser de tipo cuantitativo o cualitativo.

Entregables:

- Archivo en formato Word que incluya la descripción del conjunto de datos (contexto, nombre y unidades de cada variable) así como la fuente dónde esta se encuentra reportada (página web, tesis, artículo científico, paquete de R, etc).
- Archivo en formato .txt o .csv con el conjunto de datos.

Modalidad de entrega: Asignación en Teams.

Plazo de entrega: hasta el martes 2 de septiembre a las 11:59 pm.



MUCHAS GRACIAS

Natalia Jaramillo Quiceno

e-mail: natalia.jaramilloq@upb.edu.co