

TALLER 1

El archivo seguro.csv contiene datos referidos a pólizas de seguros adquiridas por 200 personas, de las cuales se registran los siguientes atributos:

- age: Edad de la persona
- sex: Sexo de la persona
- bmi: Índice de masa corporal
- children: Número de hijos menores de 12 años
- smoker: Indicador de hábito de fumar (0: No, 1: Sí)
- region: Región de residencia de la persona
- charges: Monto mensual que paga, en dólares

Copie cada una de las preguntas y desarróllelas en un formato Rmd, de modo que genere un archivo HTML.

1. Ejecute las siguientes tareas de preprocesamiento de datos:
 - a. Retire la columna region
 - b. Renombre las columnas como edad, sexo, imc, nhijos, fuma, gastos
 - c. Recodifique las categorías de la variable sexo como masculino y femenino, fuma: sí y no.
2. Obtenga la matriz de correlaciones de las variables cuantitativas y presente un gráfico que permita visualizarla
3. Construya un intervalo del 97.5% de confianza para la correlación entre la edad y el imc
4. Pruebe si la correlación entre el imc y el número de hijos es significativamente distinta de cero. En caso lo sea, interprete su valor.
5. Se desea estudiar la **influencia lineal de la edad sobre los gastos**. Presente el modelo de regresión lineal simple estimado.
6. Verifique si se cumple la siguiente igualdad:

$$\hat{\beta}_1 = \frac{S_y}{S_x} R$$

Donde S_y y S_x son las desviaciones estándar de la variable respuesta y predictora, respectivamente, y R es el coeficiente de correlación entre estas variables

7. Utilizando las fórmulas de error estándar, construya un intervalo del 92% de confianza para el intercepto y uno de 98% para la pendiente
8. Calcule las sumas de cuadrados para el análisis de varianza
9. Presente el análisis de varianza, redacte sus conclusiones con un nivel de significancia del 5%
10. ¿A cuánto asciende el valor estimado de σ^2 ? ¿En qué unidades se encuentra establecida esta estimación?
11. Se desea estudiar la **influencia lineal de las variables sobre los gastos**. Presente el modelo de regresión lineal múltiple estimado.
12. Interprete los coeficientes de regresión estimados
13. ¿Qué sucedió con el valor del coeficiente de regresión de la edad cuando se trabajó con un modelo de regresión lineal múltiple en vez de uno simple?
14. Presente el análisis de varianza, redacte sus conclusiones con un nivel de significancia del 5%
15. Plantee y desarrolle una prueba de hipótesis para una de las variables predictoras cuantitativas
16. Plantee y desarrolle una prueba de hipótesis para una de las variables predictoras cualitativas

17. Obtenga e interprete el coeficiente de determinación, así como el de determinación ajustado.
18. Plantee una situación donde el punto evaluado para una predicción NO corresponda a una extrapolación
19. Plantee una situación donde el punto evaluado para una predicción SÍ corresponda a una extrapolación
20. Obtenga los residuales y presente un gráfico (distinto al mostrado en clase)