



# ¿Qué tan caro sería mi seguro de gastos médicos?

Muñiz Salazar R., Castillo Cabello C.

Universidad Autónoma de Nuevo León



## Introducción

Los cotizadores de los seguros de gastos médicos son muy útiles en la actualidad, existen diferentes compañías que te cotizan en línea y posteriormente se ponen en contacto contigo o bien existen instituciones como la CONDUSEF que prestan el servicio de simular una cotización, que compares con distintas compañías y que puedas jugar con el nivel hospitalario, los honorarios y las distintas coberturas.

Vemos importante la implementación de este trabajo en la **difusión de la importancia que puede tener contar con un seguro de gastos médicos** y ser beneficioso para ambas partes tanto para los contratantes que tengan una cifra transparente y que les resulte informativo y bien para alguna compañía o institución, fungir como un cotizador rápido y que le resulte al usuario amigable en el solo en llenar unos pocos datos tener una noción de cuanto le podría costar.

## Objetivos

Como mencionábamos anteriormente algunas compañías o instituciones ya prestan un servicio similar, pero ninguno deja visible su metodología. Se tiene la hipótesis de que hay ciertas características de una persona y su estilo de vida que tienen mayor impacto en lo que se le puede cobrar a una persona y que **dadas estas variables ver si se podría llegar a un monto.**

## Metodología

Para la limpieza de la base de datos se revisaron y normalizaron los datos nulos, no se le retiró ninguna columna. Se le modificaron las columnas, la de fumador como una variable indicadora, la de bmi (índice de masa corporal o IMC) y la de edad se separaron por clases.

En cuanto a estadísticas la base de datos cuenta con 662 mujeres y 676 hombres, en cuanto al IMC la mayoría se encuentra en un nivel medio con 801 personas, le sigue con un nivel bajo 466 y con alto solo 71 personas, notando asimismo el nivel de fumadores siendo un poco más del 20 por ciento (115 mujeres y 159 hombres) de la población total. En un análisis por región podemos encontrar:

Región	Edad	IMC	Cargos	Hijos	Fumador
Noreste	39.26	29.17	13,406.38	1.04	67
Noroeste	39.19	29.19	12,417.57	1.14	58
Sudeste	38.93	33.35	14,735.41	1.04	91
Sudoeste	39.45	30.59	12,346.93	1.14	58

Analizando los gráficos de la derecha se puede concluir:

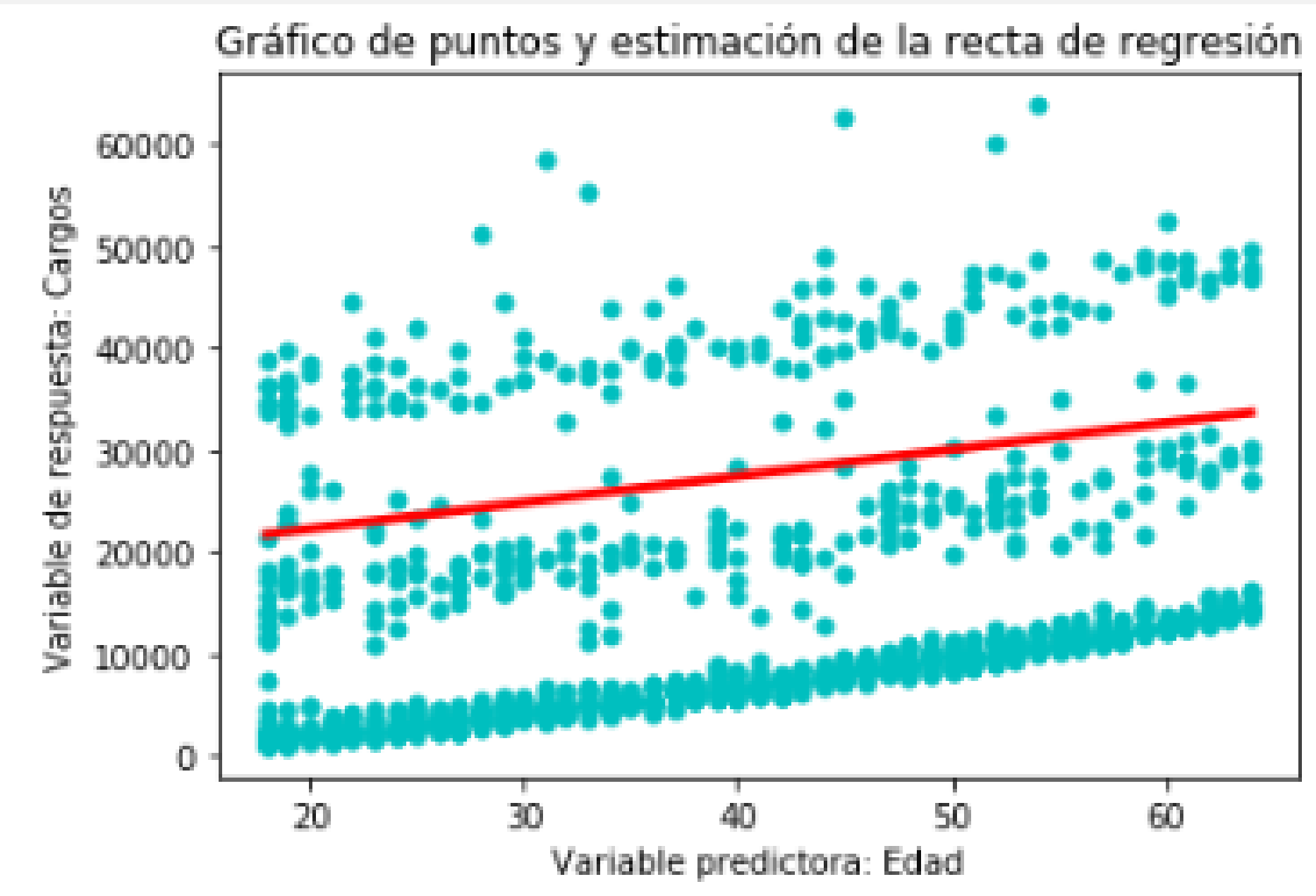
- Ser fumador aumenta los cargos, mas en una edad avanzada. (1)
- Valores atípicos en no fumadores por un alto IMC y en fumadores por un bajo a normal IMC. (2)
- Variables con mayor relacion a los cargos: Fumador, Edad, IMC. (3)

## Resultados

Se implementó un modelo de regresión lineal múltiple (1) con el objetivo de construir un modelo probabilístico que relaicone los cargos con las distintas variables independientes, obteniendo la ecuación (2), con un ajuste del 0.7530. Así mismo se obtuvo la siguiente recta fijando las variables de: IMC = Bajo, Hijos = 0, Sexo = Mujer, Región = Noreste, Fumador = Sí

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + .... + \beta_kx_k + \xi \tag{1}$$

$$Cargos = -13338.43 + 265.28Edad + 338.13IMC + 452.43Hijos + 23791.75Fumador - 317.29Sexo + 241.46Región \tag{2}$$



## Recursos

Las librerías utilizadas principalmente para la visualización de los datos y realizar graficas:

- Matplotlib: para graficas categóricas y datos univariantes.
- Seaborn: para datos bi-variantes, gráficos de dispersión y mapas de calor.
- Sklearn: Para realizar la regresión lineal y obtener sus coeficientes.

Github



Base de datos



## Conclusiones

Existen varios factores que influyen al determinar el costo de un seguro de gastos médicos, como el sexo, la edad, el índice de masa corporal y si eres fumador o no, hasta en que región del país vives.

Después de este análisis se puede concluir que las variables que más tienen impacto en coste son **la edad** y si eres **fumador o no**, no es ningún secreto que estas dos pueden ser indicadoras del estado de salud de una persona.

Respecto a la regresión que se obtuvo, se puede concluir que **se puede ajustar** un modelo de regresión lineal para obtener el coste estimado, aunque no va a tener tanta precisión, en un futuro se probarán otros tipos de modelo para encontrar alguno que tenga un mejor ajuste.