

▼ Actividad - Estadística básica

- **Nombre:**
- **Matrícula:**

Entregar: Archivo PDF de la actividad, así como el archivo .ipynb en tu repositorio. **Nota:** Recuerda habrá una penalización de **50** puntos si la actividad fue entregada fuera de la fecha límite.

Carga el conjunto de datos `insurance.csv` (se encuentra en el repositorio de la clase) y realiza un análisis estadístico de las variables.

```
# Carga las librerías necesarias.
import pandas as pd
import numpy as np
import random

# Carga el conjunto de datos al ambiente de Google Colab y muestra los primeros
from google.colab import files

uploaded = files.upload()

for fn in uploaded.keys():
    print('User uploaded file "{name}" with length {length} bytes'.format(
        name=fn, length=len(uploaded[fn])))
# 6 renglones.
df = pd.read_csv('insurance.csv')
df1 = pd.read_csv('insurance.csv')
df.head(6)
```

El conjunto de datos contiene información demográfica sobre los asegurados en una compañía de seguros:

- **age**: Edad del asegurado principal
- **sex**: Género del asegurado. female o male
- **bmi**: Índice de masa corporal
- **children**: Número de hijos que estan cubiertos con la poliza.
- **smoke**: ¿El beneficiario fuma? (yes/no)
- **region**: ¿Dónde vive el beneficiario? Estos datos son de Estados Unidos. Regiones disponibles: northeast, southeast, southwest, northwest
- **charges**: Costo del seguro.

```
# Crea una tabla resumen con los estadísticas generales de las variables
df = df.drop('sex', axis=1)
df = df.drop('smoker', axis=1)
df = df.drop('region', axis=1)
# numéricas.
```

```
# ¿Cómo se correlacionan las variables numéricas entre sí?
df.corr()
```

```
# Determina si existe o no una correlación entre el índice de masa corporal
# (bmi) y el costo del seguro.
from scipy import stats
from scipy.stats import pearsonr
```

```
selected = df[['bmi', 'charges']]
r, p = stats.pearsonr(selected['bmi'], selected['charges'])
print(f"Correlación Pearson: r={r}, p-value={p}")
```

```
r, p = stats.spearmanr(selected['bmi'], selected['charges'])
print(f"Correlación Spearman: r={r}, p-value={p}")
```

```
r, p = stats.kendalltau(selected['bmi'], selected['charges'])
print(f"Correlación Kendal: r={r}, p-value={p}")
```

```
print('Correlación Kendal: r=', r, 'p-value=', p)
```

```
print("Se acepta la hipótesis nula, se rechaza la alternativa")
```

```
Correlación Pearson: r=0.1983409688336288, p-value=2.459085535117846e-13  
Correlación Spearman: r=0.11939590358331145, p-value=1.1926059544526874e-05  
Correlación Kendal: r=0.08252397079981415, p-value=6.25690064095591e-06
```

```
# ¿Cuántas personas aseguradas son hombre y cuántas son mujeres?  
df1['sex'].value_counts()
```

```
male      676  
female    662  
Name: sex, dtype: int64
```

```
# ¿Cuántos hombres y mujeres asegurados viven en cada región?  
pd.crosstab(df1['region'], df1['sex'])
```

```
# En promedio, ¿quién paga más de cuota de seguro? ¿Los fumadores o los no  
# fumadores? Muéstralo con los datos.  
df1.groupby(['smoker']).mean()[['charges']]
```


```
# ¿Cuáles son las cuotas mínimas y máximas que las personas pagan dependiendo  
# del género y del número de hijos?  
df.groupby(['sex', 'children']).agg(['min', 'max'])[['charges']]
```

		charges	
		min	max
sex	children		
female	0	1607.51010	63770.42801
	1	2201.09710	58571.07448
	2	2801.25880	47305.30500
	3	4234.92700	46661.44240
	4	4561.18850	36580.28216
	5	4687.79700	19023.26000
male	0	1121.87390	62592.87309
	1	1711.02680	51194.55914
	2	2304.00220	49577.66240
	3	3443.06400	60021.39897
	4	4504.66240	40182.24600

```
# ¿Cuál es el índice de masa corporal promedio para hombre y mujeres dependiendo
# región en la que viven y si son fumadores? ¿Impacta eso en la tarifa del
# seguro?
df.groupby(['region', 'sex', 'smoker']).mean()[['bmi']]
```



bmi



region	sex	smoker	
northeast	female	no	29.777462
		yes	27.261724
	male	no	28.861760
		yes	29.560000
northwest	female	no	29.488704
		yes	28.296897
	male	no	28.930379
		yes	29.983966
southeast	female	no	32.780000
		yes	32.251389
	male	no	34.129552
		yes	33.650000
southwest	female	no	30.050355
		yes	30.128571
	male	no	31.019841
		yes	31.502703

 0 s

completado a las 10:31