



TECNOLÓGICO DE MONTERREY
Campus Querétaro

PROFESOR ENCARGADOS:
Pedro Oscar Pérez Murueta

Semana Tec TC1002S

Actividad 6

Daniel Gutiérrez Gómez A01068056

FECHA DE ENTREGA:
22/03/2023

```
[19] import pandas as pd
import numpy as np
from scipy import stats
from scipy.stats import pearsonr
```

Cargar librerías necesarias y primeros 6 renglones

```
insurance = pd.read_csv('./datasets/insurance.csv')
display(insurance.iloc[:6])
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
5	31	female	25.740	0	no	southeast	3756.62160

Crea una tabla resumen con los estadísticas generales de las variables numéricas

```
[21] resumen = pd.DataFrame().assign(age = insurance["age"], bmi = insurance["bmi"], charges = insurance["charges"])
display(resumen)
```

	age	bmi	charges
0	19	27.900	16884.92400
1	18	33.770	1725.55230
2	28	33.000	4449.46200
3	33	22.705	21984.47061
4	32	28.880	3866.85520
...
1333	50	30.970	10600.54830
1334	18	31.920	2205.98080
1335	18	36.850	1629.83350
1336	21	25.800	2007.94500
1337	61	29.070	29141.36030

1338 rows x 3 columns

¿Cómo se correlacionan las variables numéricas entre sí? El coeficiente de correlación da una estimación de la relación entre dos variables. El coeficiente de correlación de Pearson se calcula multiplicando las desviaciones de la media de la variable 1 por las de la variable 2, dividiendo por el producto de las desviaciones estándar

```
[22] print('Correlación Pearson entre Edad e IMC: ', resumen.age.corr(resumen.bmi, method='pearson'))
print('Correlación spearman entre Edad e IMC: ', resumen.age.corr(resumen.bmi, method='spearman'))
print('Correlación kendall entre Edad e IMC: ', resumen.age.corr(resumen.bmi, method='kendall'))

print('Correlación Pearson entre Edad y costo del seguro: ', resumen.age.corr(resumen.charges, method='pearson'))
print('Correlación spearman entre Edad y costo del seguro: ', resumen.age.corr(resumen.charges, method='spearman'))
print('Correlación kendall entre Edad y costo del seguro: ', resumen.age.corr(resumen.charges, method='kendall'))

print('Correlación Pearson entre IMC y costo del seguro: ', resumen.bmi.corr(resumen.charges, method='pearson'))
print('Correlación spearman entre IMC y costo del seguro: ', resumen.bmi.corr(resumen.charges, method='spearman'))
print('Correlación kendall entre IMC y costo del seguro: ', resumen.bmi.corr(resumen.charges, method='kendall'))
```

Correlación Pearson entre Edad e IMC: 0.1092718815485352
Correlación spearman entre Edad e IMC: 0.10773603532023862
Correlación kendall entre Edad e IMC: 0.07327326343408677
Correlación Pearson entre Edad y costo del seguro: 0.29900819333064754
Correlación spearman entre Edad y costo del seguro: 0.5343921337718459
Correlación kendall entre Edad y costo del seguro: 0.47530235429087664
Correlación Pearson entre IMC y costo del seguro: 0.19834096883362895
Correlación spearman entre IMC y costo del seguro: 0.11939590358331145
Correlación kendall entre IMC y costo del seguro: 0.08252397079981415

Determina si existe o no una correlación entre el índice de masa corporal

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

Correlación Pearson: r=0.1983409688336288, p-value=2.459085535116766e-13

Como podemos ver no existe una relación **valor de r=0.198** entre los valores del índice de la masa corporal y del costo del seguro, puesto que el valor de r es cercano a 0, ni -1 ni 1. Aunque correlación no significa causalidad, el hecho de que dos variables no estén correlacionadas no significa que no se afecten entre sí.

¿Cuántas personas aseguradas son hombre y cuántas son mujeres?

```
[24] print("Número de hombres asegurados: ", (insurance["sex"] == "male").sum())
      print("Número de mujeres aseguradas: ", (insurance["sex"] == "female").sum())
```

Número de hombres asegurados: 676
Número de mujeres aseguradas: 662

¿Cuántos hombres y mujeres asegurados viven en cada región

```
[25] display(insurance.groupby([insurance["region"], insurance["sex"]]).agg(By_sex=("sex", "count")))
```

By_sex		
region	sex	
northeast	female	161
	male	163
northwest	female	164
	male	161
southeast	female	175
	male	189
southwest	female	162
	male	163

En promedio, ¿quién paga más de cuota de seguro? ¿Los fumadores o los no fumadores? Muéstralo con los datos.

```
[26] smoker = insurance.groupby([insurance["smoker"]]).agg(MedianSmoker_Charges=("charges", "median"))
      display(smoker)
```

MedianSmoker_Charges	
smoker	
no	7345.40530
yes	34456.34845

Por lo tanto, las personas fumadoras gastan más dinero en seguro que los no fumadores por mucha diferencia

¿Cuáles son las cuotas mínimas y máximas que las personan pagan dependiendo del género y del número de hijos?

```
insurance.groupby([insurance["sex"], insurance["children"]]).agg(MinCharges=("charges", "min"), MaxCharges=("charges", "max"))
```

MinCharges MaxCharges			
sex	children		
female	0	1607.51010	63770.42801
	1	2201.09710	58571.07448
	2	2801.25880	47305.30500
	3	4234.92700	46861.44240
	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
	5	4915.05985	14478.33015

¿Cuál es el índice de masa corporal promedio para hombre y mujeres dependiendo de la región en la que viven y si son fumadores? ¿Impacta eso en la tarifa del seguro?

os

```
insurance.groupby([insurance["sex"], insurance["region"], insurance["smoker"]]).agg(  
    Avg_BMI=("bmi", "mean"), Min_Charges=("charges", "min"),  
    Max_Charges=("charges", "max"), Mean_Charges=("charges", "mean"))
```

Avg_BMIMin_ChargesMax_ChargesMean_Charges

sex	region	smoker	Avg_BMI	Min_Charges	Max_Charges	Mean_Charges
female	northeast	no	29.777462	2196.47320	31620.00106	9640.426984
		yes	27.261724	14283.45940	58571.07448	28032.046398
	northwest	no	29.488704	2117.33885	33471.97189	8786.998679
		yes	28.296897	14711.74380	55135.40209	29670.824946
	southeast	no	32.780000	1607.51010	36580.28216	8440.205552
		yes	32.251389	16577.77950	63770.42801	33034.820716
	southwest	no	30.050355	1727.78500	36910.60803	8234.091260
		yes	30.128571	13844.50600	48824.45000	31687.988430
male	northeast	no	28.861760	1694.79640	32108.66282	8664.042222
		yes	29.560000	12829.45510	48549.17835	30926.252583
	northwest	no	28.930379	1621.34020	30284.64294	8320.689321
		yes	29.983966	15817.98570	60021.39897	30713.181419
	southeast	no	34.129552	1121.87390	27724.28875	7609.003587
		yes	33.650000	17043.34140	62592.87309	36029.839367
	southwest	no	31.019841	1241.56500	27941.28758	7778.905534
		yes	31.502703	15820.69900	52590.82939	32598.862854

Como podemos ver en la tabla agregada anterior, importa mucho la tarifa del seguro si eres fumador... Independientemente de si eres hombre o mujer y también independientemente del genero y de la región en la que vivas el fumar ya te incrementa demasiado la tarifa para pagar el seguro (ver columnas de Mean_charges, Max_charges y Min_charges)