

Aldo Daniel Villaseñor Fierro

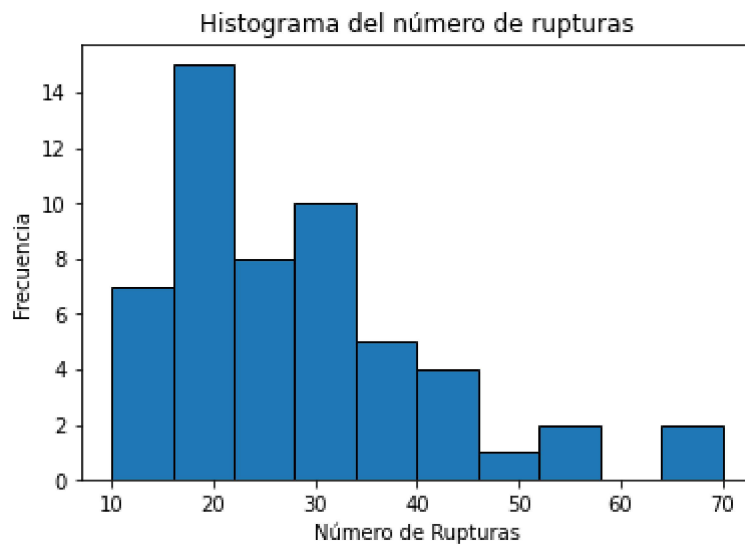
A01637907

```
In [ ]: import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
import scipy.stats as stats
import pylab
```

```
In [ ]: # Cargar los datos de warpbreaks
data = pd.read_csv('warpbreaks.csv', sep=' ')

# Paso 1: Histograma del número de rupturas

plt.hist(data['breaks'], bins=10, edgecolor='black')
plt.xlabel('Número de Rupturas')
plt.ylabel('Frecuencia')
plt.title('Histograma del número de rupturas')
plt.show()
```



```
In [ ]: #creacion de variables dummy para wool y tension
dummy_wool = pd.get_dummies(data['wool'],drop_first=True)
dummy_tension = pd.get_dummies(data['tension'],drop_first=True)
data=data.drop(['wool','tension'],axis=1)
data=pd.concat([data,dummy_wool,dummy_tension],axis=1)
```

```
In [ ]: # Paso 2: Obtener la media y la varianza
mean_breaks = data['breaks'].mean()
variance_breaks = data['breaks'].var()

print(f"Media del número de rupturas: {mean_breaks}")
print(f"Varianza del número de rupturas: {variance_breaks}")

# Paso 3: Ajustar el modelo de regresión Poisson
poisson_model = sm.GLM(data['breaks'], sm.add_constant(data[['B', 'L', 'M']]), prepend=F
```

```
# Paso 4: Resumen del modelo
print(poisson_model.summary())
```

Media del número de rupturas: 28.14814814814815

Varianza del número de rupturas: 174.20405310971347

Generalized Linear Model Regression Results

```
=====
Dep. Variable:          breaks    No. Observations:          54
Model:                  GLM      Df Residuals:              50
Model Family:          Poisson   Df Model:                  3
Link Function:          Log      Scale:                    1.0000
Method:                 IRLS     Log-Likelihood:         -242.53
Date:                   Tue, 17 Oct 2023    Deviance:              210.39
Time:                   18:26:28    Pearson chi2:          213.
No. Iterations:         4          Pseudo R-squ. (CS):      0.8003
Covariance Type:        nonrobust
=====
```

	coef	std err	z	P> z	[0.025	0.975]
B	-0.2060	0.052	-3.994	0.000	-0.307	-0.105
L	0.5185	0.064	8.107	0.000	0.393	0.644
M	0.1972	0.068	2.885	0.004	0.063	0.331
const	3.1735	0.056	57.002	0.000	3.064	3.283

```
=====
```

```
In [ ]: print(f'Desviación Nula',round(poisson_model.null_deviance,2))
```

Desviación Nula 297.37

```
In [ ]: stats.chi2.pdf(poisson_model.null_deviance-poisson_model.deviance , 3)
```

Out[]: 4.820403668362979e-19

Se puede utilizar el p-value de χ^2 donde

$\chi^2 = \text{Desviación Residual} - \text{Desviación Nula}$ para determinar si el modelo es útil para predecir la variable objetivo. Si el p-value es menor a 0.05, entonces el modelo es útil para predecir la variable objetivo. En este caso, el p-value es 0.0, por lo que el modelo es útil para predecir el número de rupturas.