



Tecnológico de Monterrey

Campus Estado de México

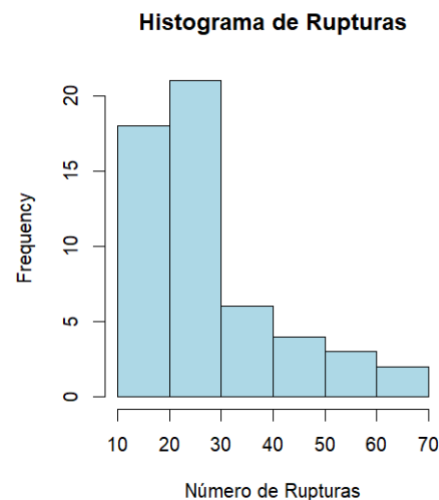
Inteligencia artificial avanzada para la ciencia de datos II (Gpo 101)

Actividad 4.Regresión Poisson

Juan Pablo Castañeda Serrano

A01752030

1. Histograma del número de rupturas



```
data(warpbreaks)
mydata <- warpbreaks
hist(mydata$breaks, main="Histograma de Rupturas", xlab="Número de Rupturas", col="lightblue", border="black")
```

2. Obtén la media y la varianza

```
mean(mydata$breaks)
var(mydata$breaks)
```

```
> mean(mydata$breaks)
[1] 28.14815
> var(mydata$breaks)
[1] 174.2041
```

3. Ajusta el modelo de regresión Poisson

Call:

```
glm(formula = breaks ~ wool + tension, family = poisson(link = "log"),
     data = mydata)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.69196	0.04541	81.302	< 2e-16 ***
woolB	-0.20599	0.05157	-3.994	6.49e-05 ***
tensionM	-0.32132	0.06027	-5.332	9.73e-08 ***
tensionH	-0.51849	0.06396	-8.107	5.21e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 297.37 on 53 degrees of freedom
 Residual deviance: 210.39 on 50 degrees of freedom
 AIC: 493.06

Number of Fisher Scoring iterations: 4

4. Coeficientes e interpretación

(Intercepto): El intercepto es el logaritmo del número esperado de rupturas cuando todas las demás variables son 0. En este contexto, se refiere a cuando se utiliza la lana tipo "A" y la tensión es "L" (baja). El valor estimado es de 3.69196.

woolB: El coeficiente para woolB es -0.20599. Esto indica que, manteniendo todo lo demás constante, el logaritmo del número esperado de rupturas disminuye en 0.20599 al usar la lana tipo "B" en comparación con la lana tipo "A".

tensionM: El coeficiente es -0.32132. Esto significa que, manteniendo todo lo demás constante, el logaritmo del número esperado de rupturas disminuye en 0.32132 cuando se utiliza tensión media en comparación con tensión baja.

tensionH: El coeficiente es -0.51849. Esto indica que el logaritmo del número esperado de rupturas disminuye en 0.51849 cuando se utiliza tensión alta en comparación con tensión baja.

Todos los coeficientes son significativos (todos tienen p-values muy bajos), lo que indica que la lana y la tensión tienen un efecto significativo en el número de rupturas.

5. Desviación y Grados de Libertad

Desviación Nula: La desviación nula es 297.37 con 53 grados de libertad. Esta es la desviación cuando solo tenemos el intercepto en el modelo.

Desviación Residual: La desviación residual es 210.39 con 50 grados de libertad. Esta es la desviación después de incluir las variables predictoras (lana y tensión) en el modelo.

El hecho de que la desviación residual (210.39) sea menor que la desviación nula (297.37) indica que el modelo con las variables predictoras es un mejor ajuste que el modelo solo con el intercepto. Sin embargo, debemos notar que la desviación residual es menor que sus grados de libertad ($210.39 < 250$), lo que podría ser indicativo de sobredispersión.

6. Conclusión

Basándonos en la salida, podemos concluir que la lana y la tensión son predictores significativos del número de rupturas. Específicamente, la lana tipo B, la tensión media y la tensión alta parecen reducir el número de rupturas en comparación con la lana tipo A y la tensión baja.

Sin embargo, dado que la desviación residual es menor que los grados de libertad, es posible que el modelo Poisson no sea el más adecuado para estos datos y podría haber sobredispersión. Es posible que un modelo quasi-Poisson pueda ser una mejor elección, como se sugirió anteriormente.