

A4-Regresión Poisson

Carlos David Lozano Sanguino - A01275316

2023-10-13

Trabajaremos con el paquete `dataset`, que incluye la base de datos `warpbreaks`, que contiene datos del hilo (yarn) para identificar cuáles variables predictoras afectan la ruptura de urdimbre.

```
data<-warpbreaks  
head(data,10)
```

```
##      breaks wool tension  
## 1         26    A      L  
## 2         30    A      L  
## 3         54    A      L  
## 4         25    A      L  
## 5         70    A      L  
## 6         52    A      L  
## 7         51    A      L  
## 8         26    A      L  
## 9         67    A      L  
## 10        18    A      M
```

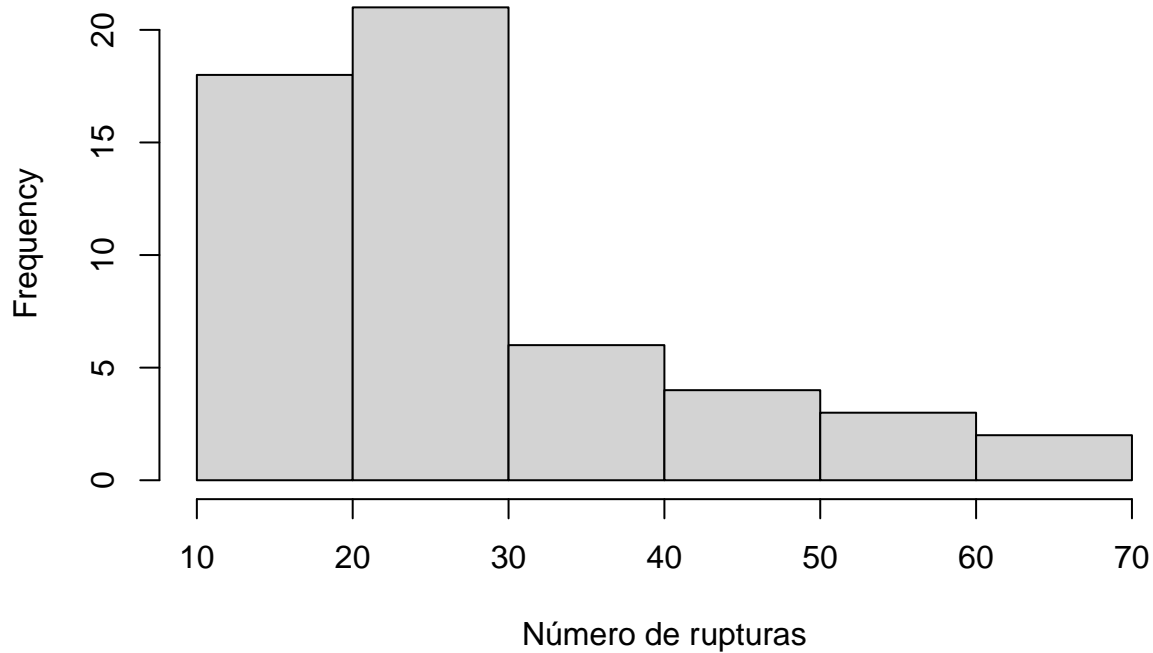
Este conjunto de datos indica cuántas roturas de urdimbre ocurrieron para diferentes tipos de telares por telar, por longitud fija de hilo:

breaks: número de rupturas wool: tipo de lana (A o B) tensión: el nivel de tensión (L, M, H) Obtén:

Histograma del número de rupturas

```
hist(data$breaks, main = "Histograma del número de rupturas", xlab = "Número de rupturas")
```

Histograma del número de rupturas



Obtén la media y la varianza

```
media <- mean(data$breaks)
varianza <- var(data$breaks)
```

```
cat("Media:", media, "\n")
```

```
## Media: 28.14815
```

```
cat("Varianza:", varianza, "\n")
```

```
## Varianza: 174.2041
```

Ajusta el modelo de regresión Poisson.

```
poisson.model <- glm(breaks ~ wool + tension, data = data, family = poisson(link = "log"))
summary(poisson.model)
```

```
##
```

```
## Call:
```

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

```
## data = data)
```

```
##
```

```
## 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
```

Análisis de Resultados: Tras el análisis anterior se obtuvieron los resultados siguientes, para empezar una interpretación de todos los coeficientes: -Intercept (Intercepto): El coeficiente del intercepto es 3.69196 que es la estimación del logaritmo natural del número esperado de rupturas cuando todas las demás variables predictoras son iguales a cero. -woolB: El coeficiente para la variable categórica “woolB” es -0.20599, este coeficiente también está asociado con una disminución en el logaritmo natural del número esperado de rupturas que tiene menos rupturas en comparación con “woolA.” -tensionM: El coeficiente para la variable categórica “tensionM” es -0.32132. Esto indica que, en comparación con la categoría de referencia “tensionL,” un nivel de tensión “tensionM” está asociado con una disminución en el número esperado de rupturas. -tensionH: El coeficiente para la variable categórica “tensionH” es -0.51849 que en comparación “tensionL,” un nivel de tensión “tensionH” está asociado con una disminución aún mayor en el número esperado de rupturas. Teniendo en cuenta los coeficientes y su relación con la cantidad de rupturas, podemos comparar y analizar la significancia estadística de cada coeficiente, en el modelo se aprecia que “woolB,” “tensionM,” y “tensionH” tienen un impacto significativo en el número de rupturas por la cantidad de asteriscos dado el modelo y el valor de cada coeficiente.

También en este modelo se tiene la Null deviance que indica la desviación del modelo cuando solo se incluye el intercepto (gran media) y no se consideran las variables predictoras y Residual deviance que representa la desviación del modelo con la inclusión de las variables predictoras. Una desviación menor con la inclusión de variables sugiere un mejor ajuste del modelo. En este caso, la desviación residual es menor que la desviación nula, lo que indica un ajuste significativamente mejor que un modelo nulo (Residual deviance: 210.39 < Null deviance: 297.37) y como el valor de las desviaciones son mayores que los grados de libertad entonces el modelo toma los errores estándar.

Por último se tiene el AIC que es una medida de la calidad del modelo que tiene en cuenta tanto el ajuste del modelo como su complejidad. En este caso, el AIC es 493.06, y se utiliza para comparar modelos alternativos.

```
poisson.model2<-glm(breaks ~ wool + tension, data = data, family = quasipoisson(link = "log"))
summary(poisson.model2)
```

```
##
## Call:
## glm(formula = breaks ~ wool + tension, family = quasipoisson(link = "log"),
##      data = data)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.69196     0.09374  39.384 < 2e-16 ***
## woolB       -0.20599     0.10646  -1.935 0.058673 .
## tensionM    -0.32132     0.12441  -2.583 0.012775 *
## tensionH    -0.51849     0.13203  -3.927 0.000264 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 4.261537)
```

```
##
##      Null deviance: 297.37  on 53  degrees of freedom
## Residual deviance: 210.39  on 50  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

Analisis con ambos modelos: Al compararlo con el modelo quasipoisson este no tiene AIC por lo que no se puede comparar en ese punto pero si se puede observar que las variables tensionH tiene mas significancia mientras que tensionM tiene una menor y woolB no tiene ninguna, la deviance es la misma que el modelo poisson anterior por lo que se puede concluir que el primer modelo esta mejor ajustado y sugiere que el tipo de lana ("wool") y el nivel de tensión ("tension") tienen un impacto significativo en el número de rupturas. El modelo se ajusta bien a los datos, ya que la desviación residual es menor que la desviación nula, y el AIC es relativamente bajo.