

A4. Regresion Poisson

Francisco Castorena, A00827756

2023-10-17

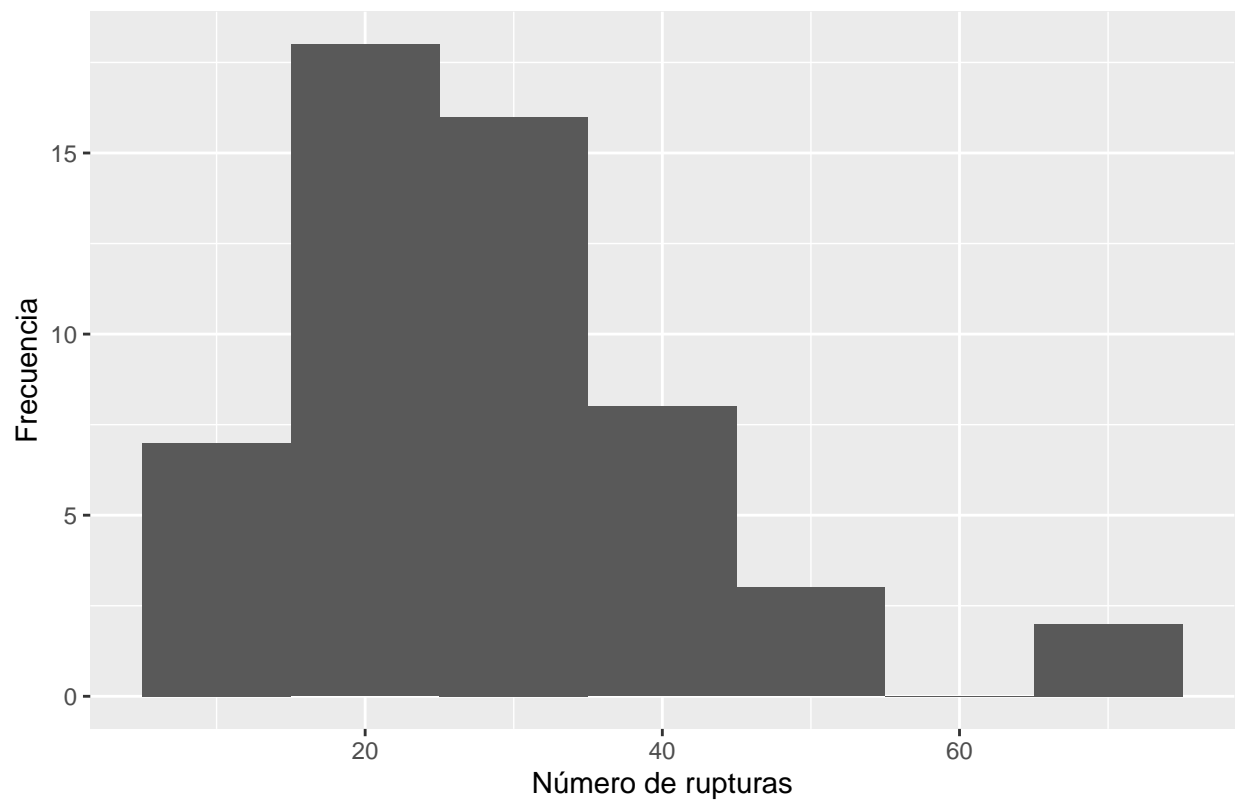
```
data<-warpbreaks  
head(data,10)
```

```
##      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
```

Histograma del número de rupturas

```
library(ggplot2)  
  
# Histograma del número de rupturas  
ggplot(data, aes(x = breaks)) +  
  geom_histogram(bins=sqrt(length(data$breaks))) +  
  labs(title = "Histograma del número de rupturas",  
        x = "Número de rupturas",  
        y = "Frecuencia")
```

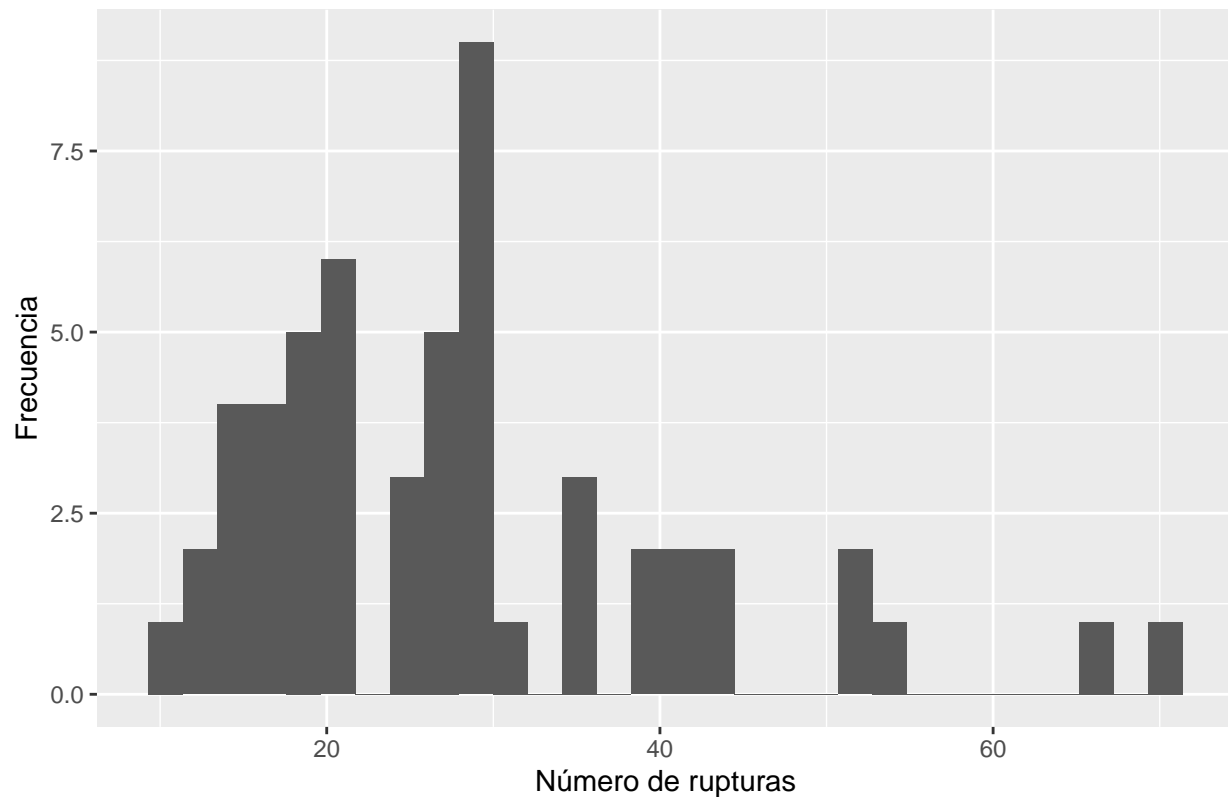
Histograma del número de rupturas



```
# Histograma del número de rupturas
ggplot(data, aes(x = breaks)) +
  geom_histogram() +
  labs(title = "Histograma del número de rupturas",
       x = "Número de rupturas",
       y = "Frecuencia")
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

Histograma del número de rupturas



Media y Varianza

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

print(paste("Media:", media))
```

```
## [1] "Media: 28.1481481481481"
```

```
print(paste("Varianza:", var))
```

```
## [1] "Varianza: 174.204053109713"
```

Ajustando el modelo de regresión Poisson

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

```
##
## Call:
```

```
## glm(formula = breaks ~ wool + tension, family = poisson(link = "log"),
##     data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6871  -1.6503  -0.4269   1.1902   4.2616
##
## 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
```

Interpretando la información obtenida

Podemos aquí ver como las variables predictoras wool y tension influyen en la variable break. Podemos observar que los residuos parecen distribuirse de forma uniforme entre cuartiles.

Tanto las variables “woolB”, “tensionM” y “tensionL” son negativas, lo cual indica una disminución en la tasa de eventos a medida que estas variables aumentan. Por otra parte, podemos ver que el $\Pr(>|z|)$ de “woolB”, “tensionM” y “tensionL” son de alta significancia debido a que tienen tres asteriscos. Tomando esto en cuenta, decimos que el modelo de regresión de Poisson sugiere que las variables “wool” y “tension” tienen un efecto significativo en la tasa de eventos representada por la variable “breaks”.

Se observa que la desviación residual es mayor a los grados de libertad, lo cual indica que el modelo no tiene una dispersión excesiva en sus estimaciones.

A continuación se realizará un modelo cuasi Poisson con las mismas variables predictoras y con la variable “break” para analizar si su ajuste a los datos es mejor que el realizado anteriormente.

Modelo cuasi Poisson

```
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)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
```

```

## -3.6871 -1.6503 -0.4269 1.1902 4.2616
##
## 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

```

Aquí se observa que la varianza no es igual a la media, lo que permite una mayor flexibilidad en la modelado de la varianza. Vemos que los coeficientes del modelo son los mismos que con el modelo Poisson lo cual indica que el modelo es bueno.