

Actividad 4 - Regresión Poisson

A00831314- Sofia Reyes

2023-10-18

Datos

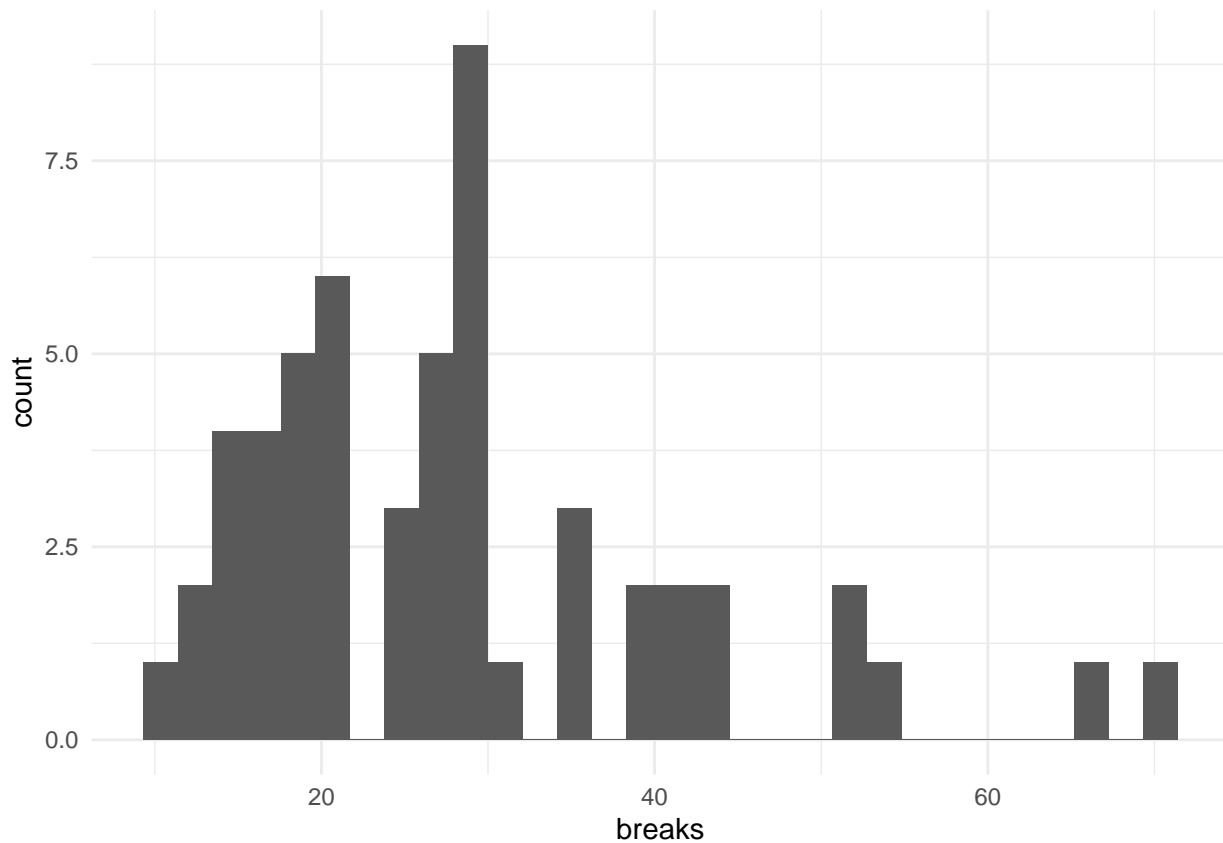
```
data <- warpbreaks  
head(data,10)
```

```
##      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)  
ggplot(data, aes(x=breaks))+geom_histogram()+theme_minimal()
```

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



Media y Varianza

```
cat("Media: ", mean(data$breaks), "\n",
    "Varianza: ", var(data$breaks))
```

```
## Media: 28.14815
## Varianza: 174.2041
```

Ajuste del modelo

```
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
```

Interpretación

En este modelo, se calcula el número de rupturas según variables Dummy del tipo de lana y el nivel de la tensión. Como tal, existe una variable Dummy para cuando el tipo de lana es B, lo que significa que si esta variable es 0, el tipo de lana que considera el modelo es la A. En cambio, en el caso del nivel de tensión, existen 2 variables para Medium y High; entonces cuando ambas tengan valor 0, se está modelando una tensión baja. Los p-valores de todas las variables muestran que están altamente significantes en el modelo.

Por un lado, la desviación residual es mayor que los grados de libertad, lo que permite que la dispersión de las predicciones no sea muy alta. No obstante la diferencia con la desviación nula, es considerable, lo que significa que el modelo estima mejor cuando las variables no son nulas. Por ende, no puede predecir correctamente para el tipo A de lana y la tensión Baja. Esto permite llegar a la consideración que el modelo no es bueno, al no poder estimar el comportamiento de los datos para todas las categorías de las variables.

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
```

Al tener los mismos valores de la desviación nula y residual que el primer modelo, se puede comprobar que el modelo es válido.