

Regresion Poisson

Felipe Yépez

2022-11-07

Base de 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
```

El conjunto de datos indica cuántas roturas de urdimbre ocurrieron para diferentes tipos de telares por telas por longitud fija de hilo.

La variable breaks es el número de rupturas, wool es el tipo de lana (A, B) y tensión es el nivel de tensión(L, M, H).

Análisis Base de datos

```
summary(data)
```

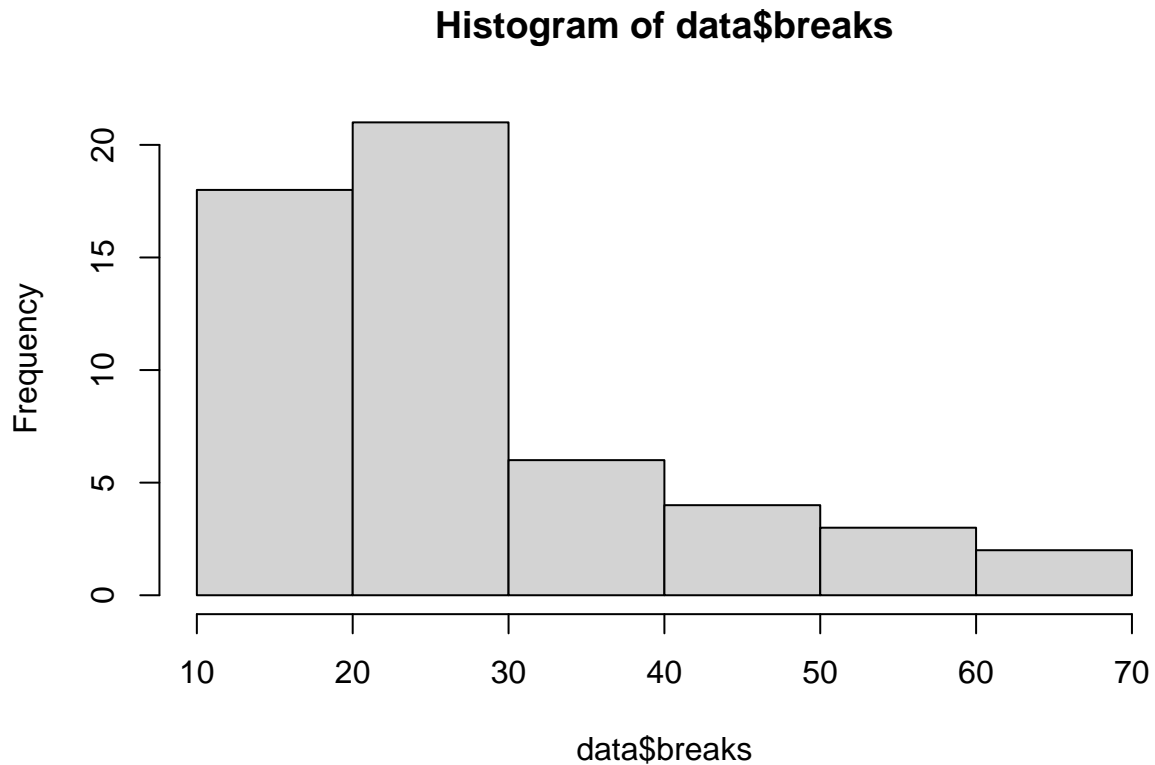
```
##      breaks      wool  tension  
## Min.   :10.00    A:27    L:18  
## 1st Qu.:18.25    B:27    M:18  
## Median :26.00                H:18  
## Mean   :28.15  
## 3rd Qu.:34.00  
## Max.   :70.00
```

Se puede observar que se cuenta con las 3 variables descritas anteriormente en un total de 54 registros.

La variable breaks es numérica con un mínimo de 10 y un máximo de 70 con media de 28. Wool tiene 2 tipos de lana, A o B representadas por letras. El nivel de tensión es igualmente representado por letras L, M y H. Wool y Tension tienen el mismo número de ocurrencias para cada uno de sus posibles valores.

Histograma de rupturas

```
hist(data$breaks)
```



No se puede observar distribución normal en el conjunto de datos de rupturas, pues hay mayor número de rupturas más pequeñas entre el rango 10 a 30.

Media del número de rupturas

```
mean(data$breaks)
```

```
## [1] 28.14815
```

Varianza del número de rupturas

```
var(data$breaks)
```

```
## [1] 174.2041
```

Como se puede observar la varianza y la media del número de rupturas es diferente. Esto indica que se tendría sobredispersión en el conjunto de datos.

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
```

Se utilizará el valor de significancia de 0.05 como en la academia.

Con este nivel de significancia se obtiene que cada una de las variables es significativa para el modelo ya que el p value de cada una es menor a 0.05.

Se obtiene una desviación residual de 210.39 en 50 grados de libertad. Al tener desviación residual mayor, los errores estándar son incorrectos y no son tomados en cuenta por el modelo.

La desviación excesiva nula muestra qué tan bien se predice la variable de respuesta mediante un modelo que incluya solo el intercepto (gran media). El valor de la desviación nula es de 210.39 con 53 grados de libertad, 3 más que la desviación residual. La diferencia entre ambas desviaciones es de 86.98

La diferencia es suficiente como para indicar que el modelo tiene mal ajuste, razón por la cuál se recurre al modelo cuasi Poisson. Se recurre a este modelo cuando sigma cuadrada y la media no son iguales y el modelo sobre estima o subestima a dispersión de los datos.

Regresión 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
```

Para determinar que el modelo es bueno se verifica que los coeficientes entre ambos modelos son iguales.

```
coef(poisson.model)
```

```
## (Intercept)      woolB      tensionM      tensionH
##  3.6919631  -0.2059884  -0.3213204  -0.5184885
```

```
coef(poisson.model2)
```

```
## (Intercept)      woolB      tensionM      tensionH
##  3.6919631  -0.2059884  -0.3213204  -0.5184885
```

El modelo resulta tener los mismos coeficientes por lo que se verifica que es bueno.