

Regresión Poisson

Federico Medina

2023-10-10

Instrucciones

Obtén:

- Histograma del número de rupturas
- Obtén la media y la varianza
- Ajusta el modelo de regresión Poisson
- Interpreta la información obtenida. Toma en cuenta que R genera variables Dummy para las variables categóricas. Para cada variable genera k-1 variables Dummy en k categorías.
- La desviación residual debe ser mayor que los grados de libertad para asegurarse que no exista una dispersión excesiva. Una diferencia menor, significará que aunque las estimaciones son correctas, los errores estándar son incorrectos y el modelo no los toma en cuenta.
- La desviación excesiva nula muestra que tan bien se predice la variable de respuesta mediante un modelo que incluye solo el intercepto (gran media) mientras que el residual con la inclusión de variables. Una diferencia en los valores significa un mal ajuste.
- Si hay un mal modelo, recurre a usar un modelo cuasi Poisson, si los coeficientes son los mismos, el modelo es bueno

Cargamos los datos

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)

```
data<-warpbreaks
```

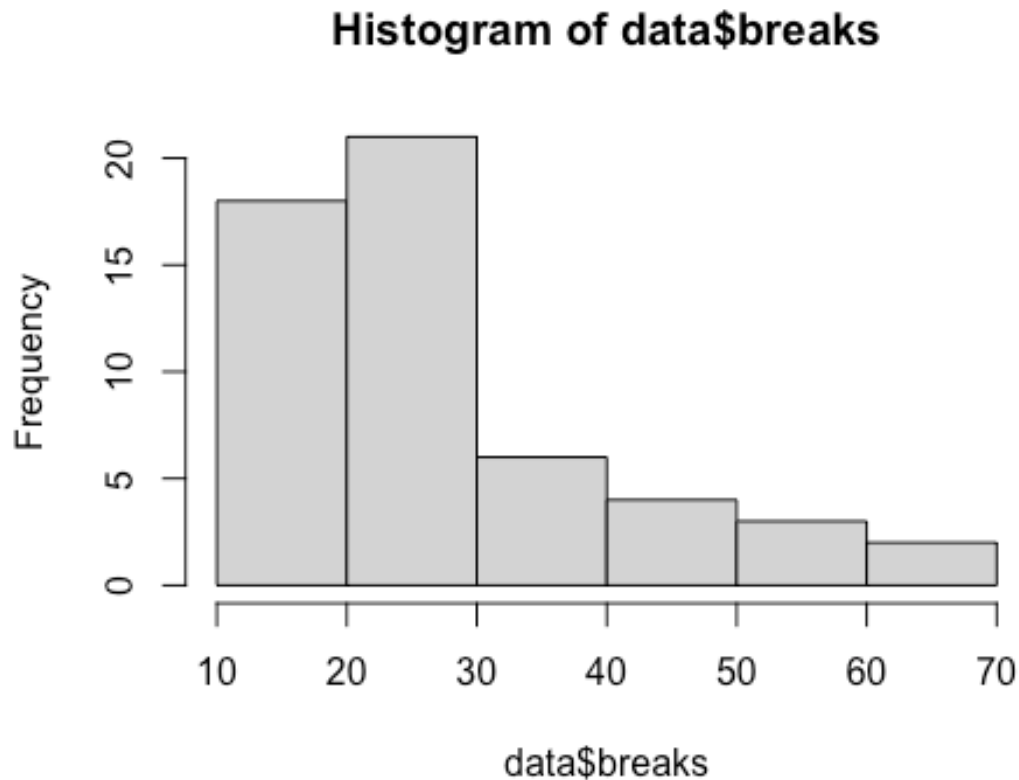
```
head(data,10)
```

```
##      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
```

Crear los histogramas

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



Media y Varianza de Breaks

```
cat('Media de Breaks: ', mean(data$breaks))
```

```
## Media de Breaks: 28.14815
```

```
cat("\n")
```

```
cat('Varianza de Breaks: ', var(data$breaks))
```

```
## Varianza de Breaks: 174.2041
```

Ajusta 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)
##
## 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
```

Observando la desviación nula, vemos que el modelo es bueno ya que se compara el modelo con otro que no incluye los predictores y se obtiene un valor alto, lo cual dice que se ajusta bien. Por otro lado, la desviación residual nos dice que hay mucha dispersión de la varianza ya que es un valor mayor a los grados de libertad, por lo que vemos que no es ideal esto. Igualmente se complementa al observar la media y la varianza ya que son valores distintos, por lo que no sigue una distribución de 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)
##
## 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
```

Observando los resultados con la cuasi Poisson, obtenemos valores distintos de p, por lo que obtenemos nuevos insights sobre qué variables son significativas. En este caso, para un nivel de significancia de 0.05, los valores significantes son de Intercept, tensionM y tensionH, por lo que se volverá a modelar quitando la variable woolB.

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

##
## Call:
## glm(formula = breaks ~ tension, family = quasipoisson(link = "log"),
##      data = data)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.59426    0.08382  42.881  < 2e-16 ***
## tensionM     -0.32132    0.12928  -2.485  0.016260 *
## tensionH     -0.51849    0.13721  -3.779  0.000414 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 4.601927)
##
##      Null deviance: 297.37  on 53  degrees of freedom
## Residual deviance: 226.43  on 51  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

Como podemos ver, el modelo mantiene un mismo valor de las desviaciones lo cual no es ideal por lo explicado anteriormente. Sin embargo, vemos que los estimadores se mantienen iguales en los tres modelos, lo cual significa que hay cierta congruencia entre los modelos. Esto nos indica que a pesar de no ser un modelo ideal, es el mejor que podemos obtener para estos valores.