

A6-Regresión Poisson

Ricardo Salinas

2024-10-29

```
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

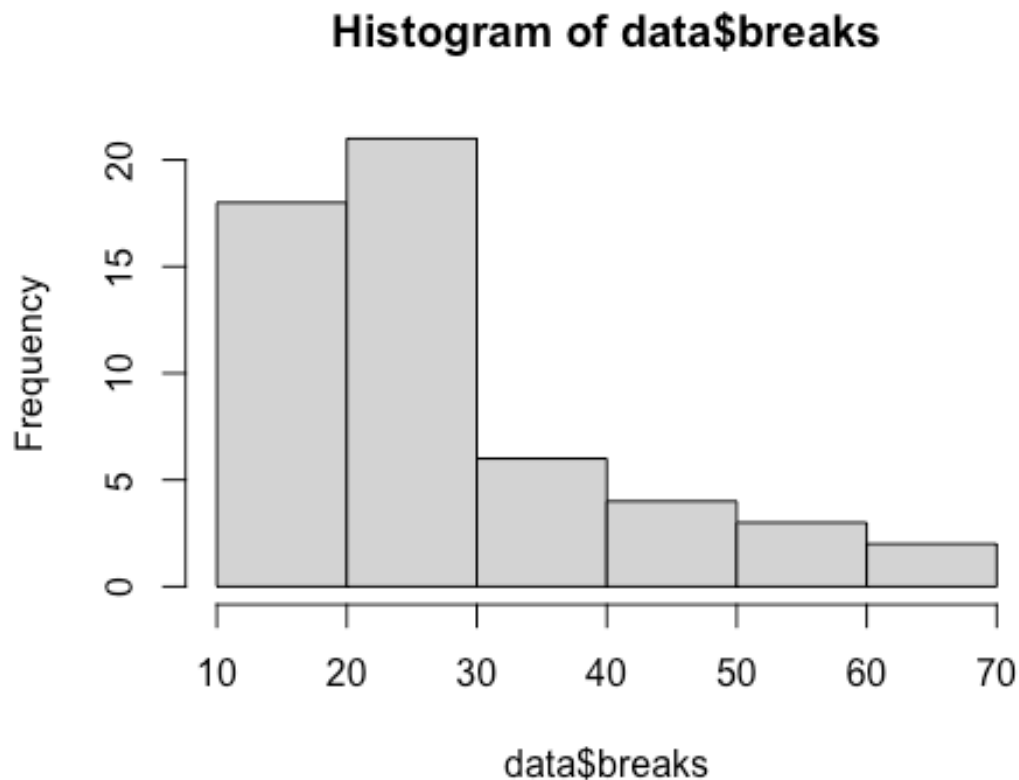
#wool: tipo de lana (A o B)

#tensión: el nivel de tensión (L, M, H)

Análisis Descriptivo

#Histograma del número de rupturas

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



```
breaks = data$breaks
#Obtén la media y la varianza de la variable dependiente
print("Media: ")
## [1] "Media: "
mean(data$breaks)
## [1] 28.14815
print("Varianza: ")
## [1] "Varianza: "
var(data$breaks)
## [1] 174.2041
```

Interpreta en el contexto de una Regresión Poisson

Todos los valores son positivos y consecutivos, lo cual es una característica distintiva de la regresión Poisson, por lo cual se puede considerar como un buen recurso para su análisis.

II. Ajusta dos modelos de Regresión Poisson

#Ajusta el modelo de regresión Poisson sin interacción

```
poisson_models1<-glm(breaks ~ wool + tension, data, family =
quasipoisson(link = "log"))
S1=summary(poisson_models1)
S1

##
## 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
```

#Ajusta el modelo de regresión Poisson con interacción

#Usa Los comandos:

```
poisson_model<-glm(breaks ~ wool * tension, data, family = poisson(link =
"log"))
S=summary(poisson_model)
S

##
## Call:
## glm(formula = breaks ~ wool * tension, family = poisson(link = "log"),
##     data = data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.79674    0.04994  76.030 < 2e-16 ***
## woolB        -0.45663    0.08019  -5.694 1.24e-08 ***
## tensionM     -0.61868    0.08440  -7.330 2.30e-13 ***
## tensionH     -0.59580    0.08378  -7.112 1.15e-12 ***
```

```
## woolB:tensionM  0.63818    0.12215    5.224 1.75e-07 ***
## woolB:tensionH  0.18836    0.12990    1.450    0.147
## ---
## 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: 182.31  on 48  degrees of freedom
## AIC: 468.97
##
## Number of Fisher Scoring iterations: 4

pm = poisson_model
```

Interpreta los coeficientes de las variables Dummy. Escribe el modelo obtenido. Toma en cuenta que R genera variables Dummy para las variables categóricas. Para cada variable genera k-1 variables Dummy en k categorías.

El modelo Poisson sin interacción proporciona un ajuste adecuado cuando no existe sobre-dispersión en los datos, con coeficientes que indican cómo cada variable afecta el número esperado de rupturas de forma independiente.

El modelo quasipoisson sin interacción ajusta mejor los datos al corregir la sobre-dispersión. Este modelo utiliza los mismos coeficientes, pero ajusta los errores estándar, lo que influye en los valores p y la significancia estadística de las variables.

III. Selección del modelo

#Para seleccionar el modelo se toma en cuenta:

#Desviación residual: es la suma del cuadrado de los residuos estandarizados que se obtienen bajo el modelo. Con los grados de libertad se realiza una prueba de para significancia del modelo.

#AIC: Criterio de Akaike

#Comparación entre los coeficientes y los errores estándar de de ambos modelos

#Desviación residual (Prueba de)

#Si el modelo nulo explica a los datos, entonces la desviación nula será pequeña. Lo mismo ocurre con la Desviación residual . Puesto que es de suponer que el modelo contiene variables significativas, lo que importa que es la desviación residual del modelo sea suficientemente pequeño.

#La prueba de mide qué tan lejano está del cero la desviación residual del modelo. Entre más lejos esté del cero, el modelo será un buen modelo, entre más cerca, el modelo será un mal modelo que explicará poco la variabilidad de los datos. Su modelo supone:

```

#H0: Deviance = 0
#H1: Deviance > 0

#gl = gl_desviación residual (n-(p+1))
#Usa los siguientes comandos:
#Valor frontera de la zona de rechazo (S es la variable que denota el summary
del modelo):

gl = S$df.null-S$df.residual
cat("Grados de Libertad: ", gl, "\n")

## Grados de Libertad: 5

r = qchisq(0.05,gl)
cat("Valor de Zona de Rechazo", r, "\n")

## Valor de Zona de Rechazo 1.145476

#Estadístico de prueba y valor p:
dr = S$deviance
cat("Estadístico de prueba =",dr, "\n")

## Estadístico de prueba = 182.3051

vp = 1-pchisq(dr,gl)
cat("Valor p =",vp)

## Valor p = 0

gl1 = S1$df.null-S1$df.residual
cat("Grados de Libertad: ", gl1, "\n")

## Grados de Libertad: 3

r1 = qchisq(0.05,gl1)
cat("Valor de Zona de Rechazo", r1, "\n")

## Valor de Zona de Rechazo 0.3518463

#Estadístico de prueba y valor p:
dr1 = S1$deviance
cat("Estadístico de prueba =",dr1, "\n")

## Estadístico de prueba = 210.3919

vp1 = 1-pchisq(dr1,gl1)
cat("Valor p =",vp1)

## Valor p = 0

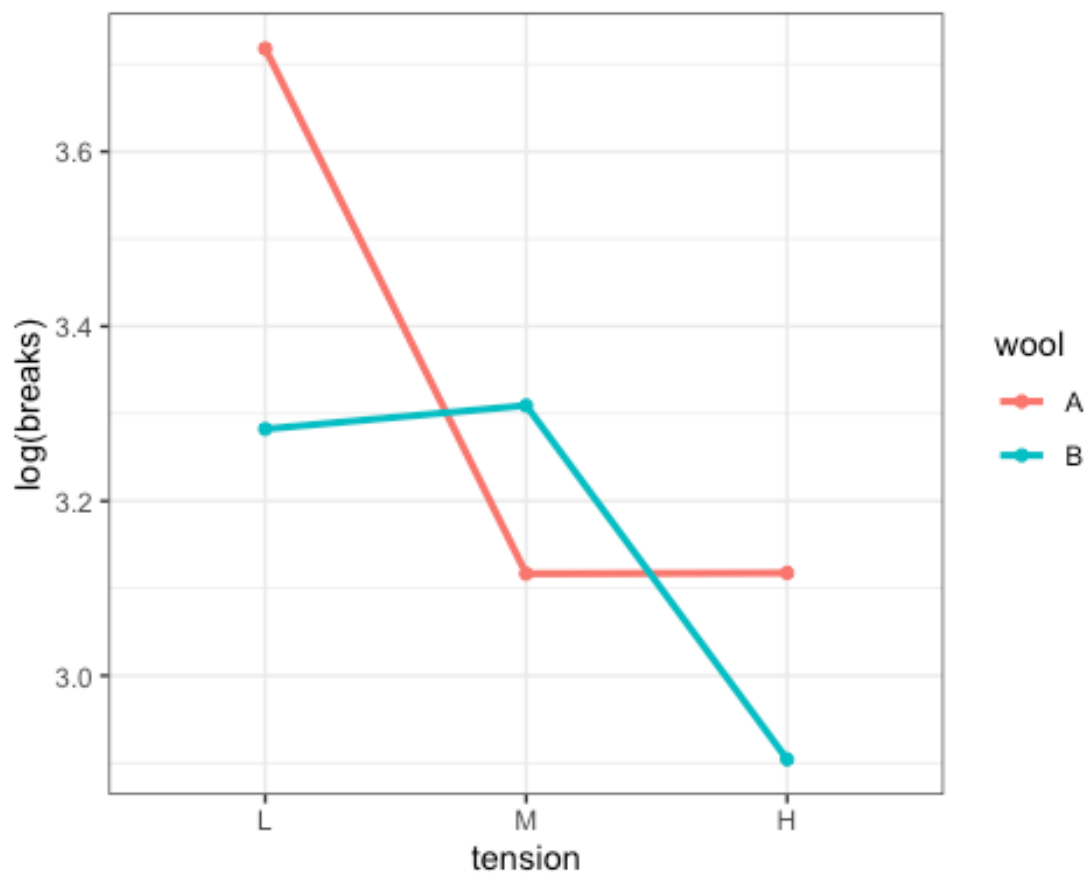
```

Aunque el modelo 1 tiene un mejor valor de AIC, siendo que tiene un valor de 182.31 en comparación a 210.39 para el modelo 2, pero ambos modelos tienen un valor p de 0, lo cual muestra que imposible esta distribución de datos.

#Interpreta los coeficientes de ambos modelos.

#Para interpretar mejor la interacción gráfica con el siguiente código:

```
library(ggplot2)
ggplot(poisson_models1, aes(x = tension, y = log(breaks), group = wool, color = wool)) +
  stat_summary(fun = mean, geom = "point") +
  stat_summary(fun = mean, geom = "line", lwd=1.1) +
  theme_bw() +
  theme(panel.border = element_rect(fill="transparent"))
```



#Define cuál de los dos es un mejor modelo.

IV. Evaluación de los supuestos

#Los supuestos principales que se deben cumplir son:

#Independencia: haz la misma prueba de independencia que usaste en los modelos lineales.

#Sobredispersión de Los residuos. La sobredispersión de Los residuos indicará que el modelo no cumple con el supuesto de que la media es igual a la varianza de Los residuos. Para probarla se usa la prueba posgof, que es una prueba con gl = grados de libertad residual. La desviación estándar se compara con los grados de libertad de la desviación residual, no deben ser muy diferentes. Esto indicará una sobredispersión de Los residuos:

#H0: No hay una sobredispersión del modelo

#H1: Hay una sobredispersión del modelo

#Usa el comando:

```
library(epiDisplay)
```

```
## Loading required package: foreign
```

```
## Loading required package: survival
```

```
## Loading required package: MASS
```

```
## Loading required package: nnet
```

```
##
```

```
## Attaching package: 'epiDisplay'
```

```
## The following object is masked from 'package:ggplot2':
```

```
##
```

```
##      alpha
```

```
poisgof(pm)
```

```
## $results
```

```
## [1] "Goodness-of-fit test for Poisson assumption"
```

```
##
```

```
## $chisq
```

```
## [1] 182.3051
```

```
##
```

```
## $df
```

```
## [1] 48
```

```
##
```

```
## $p.value
```

```
## [1] 1.582538e-17
```

#Si hay un mal modelo, recurre a usar:

#Modelo cuasi Poisson:

```
poisson.model3<-glm(breaks ~ wool + tension, data = data, family =  
quasipoisson(link = "log"))  
summary(poisson.model3)
```

```
##
```

```
## 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

#Modelo Binomial Negativa (intenta imaginar qué es lo que cambia en este
modelo con respecto al Poisson):
bnm = model.nb = glm.nb(breaks ~ wool * tension, data, control =
glm.control(maxit=1000))
summary(bnm)

##
## Call:
## glm.nb(formula = breaks ~ wool * tension, data = data, control =
glm.control(maxit = 1000),
##      init.theta = 12.08216462, link = log)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    3.7967    0.1081  35.116 < 2e-16 ***
## woolB         -0.4566    0.1576  -2.898 0.003753 **
## tensionM      -0.6187    0.1597  -3.873 0.000107 ***
## tensionH      -0.5958    0.1594  -3.738 0.000186 ***
## woolB:tensionM  0.6382    0.2274   2.807 0.005008 **
## woolB:tensionH  0.1884    0.2316   0.813 0.416123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(12.0822) family taken to be 1)
##
##      Null deviance: 86.759  on 53  degrees of freedom
## Residual deviance: 53.506  on 48  degrees of freedom
## AIC: 405.12
##
## Number of Fisher Scoring iterations: 1
##

```



```
##  
##           Theta: 12.08  
##           Std. Err.: 3.30  
##  
## 2 x log-likelihood: -391.125
```

#Define si usas tus modelos con interacción o sin interacción (no hagas Los dos)

#Define el mejor modelo usando las mismas pruebas y criterios que usaste en Los modelos Poisson