

A6-Regresión Poisson

Eliezer Cavazos

2024-10-29

```
data<-warpbreaks
```

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

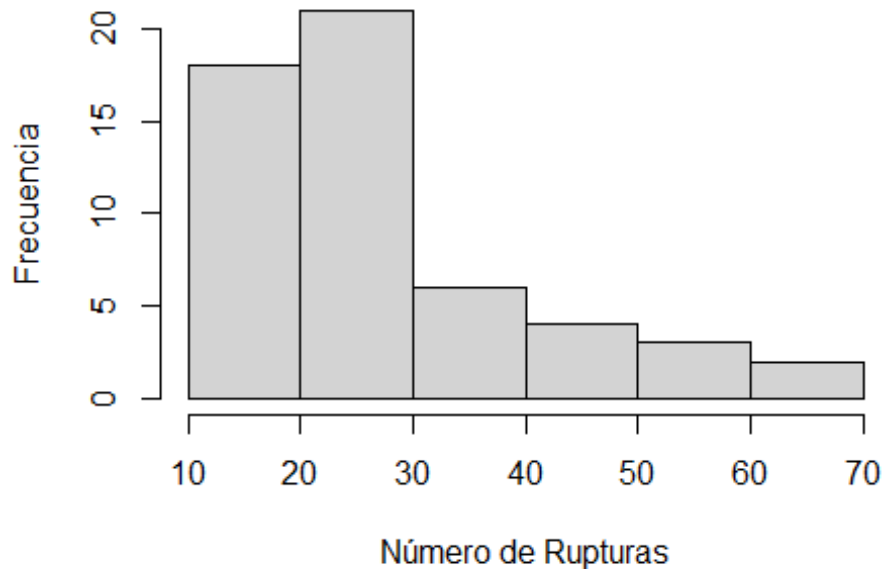
```
##      breaks wool tension
## 1         26    A       L
## 2         30    A       L
## 3         54    A       L
## 4         25    A       L
## 5         70    A       L
## 6         52    A       L
## 7         51    A       L
## 8         26    A       L
## 9         67    A       L
## 10        18    A       M
```

I. Análisis Descriptivo

Histograma del número de rupturas

```
hist(data$breaks, main="Histograma del Número de Rupturas", xlab="Número de Rupturas", ylab="Frecuencia")
```

Histograma del Número de Rupturas



Obtén la media y la varianza de la variable dependiente

```
fMedia = mean(data$breaks)
fVarianza = var(data$breaks)
cat("Media:", fMedia, "\nVarianza:", fVarianza)

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

Interpreta en el contexto de una Regresión Poisson

La regresión de Poisson nos ayuda a analizar tanto los datos de recuento como los datos de tasa al permitirnos determinar qué variables explicativas, que serían nuestros valores X, tienen un efecto en una variable de respuesta dada.

II. Ajusta dos modelos de Regresión Poisson

Ajusta el modelo de regresión Poisson sin interacción

```
oPoissonMod1 <- glm(breaks ~ wool + tension, data, family = poisson(link = "log"))
summary(oPoissonMod1)

##
## 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
```

Ajusta el modelo de regresión Poisson con interacción

```
oPoissonMod2 <- glm(breaks ~ wool * tension, data, family = poisson(link =
"log"))
summary(oPoissonMod2)

##
## 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
```

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.

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 χ^2 para significancia del modelo.

```
oSummary1 <- summary(oPoissonMod1)
oSummary2 <- summary(oPoissonMod2)

oG11 <- oSummary1$null.deviance - oSummary1$df.residual
oG12 <- oSummary2$null.deviance - oSummary2$df.residual
```

AIC: Criterio de Aikaik

```
qchisq(0.05, oG11)
```

```
## [1] 211.9578
```

```
qchisq(0.05, oG12)
```

```
## [1] 213.81
```

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

```
oDR1 <- oSummary1$deviance
oVP1 <- 1 - pchisq(oDR1, oG11)
cat("Estadístico de prueba =", oDR1, "\nValor p =", oVP1)

## Estadístico de prueba = 210.3919
## Valor p = 0.9575667

oDR2 <- oSummary2$deviance
oVP2 <- 1 - pchisq(oDR2, oG12)
cat("\nEstadístico de prueba =", oDR2, "\nValor p =", oVP2)

##
## Estadístico de prueba = 182.3051
## Valor p = 0.999506

oCoeficientes <- data.frame(
  Modelo = c("Sin Interacción", "Con Interacción"),
  Coeficientes = c(coef(oPoissonMod1), coef(oPoissonMod2)),
  Errores = c(summary(oPoissonMod1)$coefficients[,2],
summary(oPoissonMod2)$coefficients[,2])
)
print(oCoeficientes)
```

| ## | Modelo | Coeficientes | Errores |
|-------|-----------------|--------------|------------|
| ## 1 | Sin Interacción | 3.6919631 | 0.04541069 |
| ## 2 | Con Interacción | -0.2059884 | 0.05157117 |
| ## 3 | Sin Interacción | -0.3213204 | 0.06026580 |
| ## 4 | Con Interacción | -0.5184885 | 0.06395944 |
| ## 5 | Sin Interacción | 3.7967368 | 0.04993753 |
| ## 6 | Con Interacción | -0.4566272 | 0.08019202 |
| ## 7 | Sin Interacción | -0.6186830 | 0.08440012 |
| ## 8 | Con Interacción | -0.5957987 | 0.08377723 |
| ## 9 | Sin Interacción | 0.6381768 | 0.12215312 |
| ## 10 | Con Interacción | 0.1883632 | 0.12989529 |

Desviación residual (Prueba de X^2)

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.

```
gl = oSummary1$df.null-oSummary1$df.residual
qchisq(0.05,gl)
```

```
## [1] 0.3518463
```

```
dr = oSummary1$deviance
cat("Estadístico de prueba =",dr, "\n")
```

```
## Estadístico de prueba = 210.3919
```

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

```
## Valor p = 0
```

La prueba de X^2 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:

H_0 : Deviance = 0 H_1 : Deviance > 0 $gl = gl_{\text{desviación residual}} (n-(p+1))$

```
gl = oSummary2$df.null-oSummary1$df.residual
qchisq(0.05,gl)
```

```
## [1] 0.3518463
```

```
dr = oSummary2$deviance
cat("Estadístico de prueba =",dr, "\n")
```

```
## Estadístico de prueba = 182.3051
```

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

```
## Valor p = 0
```

Compara los AIC de cada modelo. Recuerda que un menor AIC indica un mejor modelo.

```
cat("AIC sin interacción:", AIC(oPoissonMod1), "\nAIC con interacción:",  
AIC(oPoissonMod2))
```

```
## AIC sin interacción: 493.056  
## AIC con interacción: 468.9692
```

Compara los coeficientes

Compara los coeficientes de ambos modelos (haz una tabla para que se facilite la comparación)

Compara el error estándar de cada estimador de Bi de ambos modelos (haz una tabla para que se facilite la comparación)

```
oCoeficientes <- data.frame(  
  Modelo = c("Sin Interacción", "Con Interacción"),  
  Coeficientes = c(coef(oPoissonMod1), coef(oPoissonMod2)),  
  Errores = c(summary(oPoissonMod1)$coefficients[,2],  
summary(oPoissonMod2)$coefficients[,2])  
)  
print(oCoeficientes)
```

```
##           Modelo Coeficientes      Errores  
## 1 Sin Interacción   3.6919631 0.04541069  
## 2 Con Interacción  -0.2059884 0.05157117  
## 3 Sin Interacción  -0.3213204 0.06026580  
## 4 Con Interacción  -0.5184885 0.06395944  
## 5 Sin Interacción   3.7967368 0.04993753  
## 6 Con Interacción  -0.4566272 0.08019202  
## 7 Sin Interacción  -0.6186830 0.08440012  
## 8 Con Interacción  -0.5957987 0.08377723  
## 9 Sin Interacción   0.6381768 0.12215312  
## 10 Con Interacción  0.1883632 0.12989529
```

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.

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

dwtest(oPoissonMod1, alternative = "two.sided")

##
## Durbin-Watson test
##
## data:  oPoissonMod1
## DW = 2.0332, p-value = 0.7791
## alternative hypothesis: true autocorrelation is not 0
```

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

```
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:lmtest':
##
##      lrtest

# Para el modelo sin interacción
poisgof(oPoissonMod1)

## $results
## [1] "Goodness-of-fit test for Poisson assumption"
##
## $chisq
## [1] 210.3919
```

```
##
## $df
## [1] 50
##
## $p.value
## [1] 1.44606e-21

# Para el modelo con interacción
poisgof(oPoissonMod2)

## $results
## [1] "Goodness-of-fit test for Poisson assumption"
##
## $chisq
## [1] 182.3051
##
## $df
## [1] 48
##
## $p.value
## [1] 1.582538e-17

library(MASS)
oModelNB <- glm.nb(breaks ~ wool * tension, data, control = glm.control(maxit
= 1000))
summary(oModelNB)

##
## 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

poisgof(oModelNB)

## $results
## [1] "Goodness-of-fit test for Poisson assumption"
##
## $chisq
## [1] 53.50616
##
## $df
## [1] 48
##
## $p.value
## [1] 0.2711637
```

Define si usas defines 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

Con el modelo Binomial Negativa con interaccion ya que no rechaza la Hipótesis nula (H_0): No hay sobredispersión en el modelo, es decir, el modelo se ajusta adecuadamente sin requerir una mayor varianza.