

Act6-Regresion Poisson

Andrés Villarreal González

2024-11-03

Act 6 Regresión Poisson

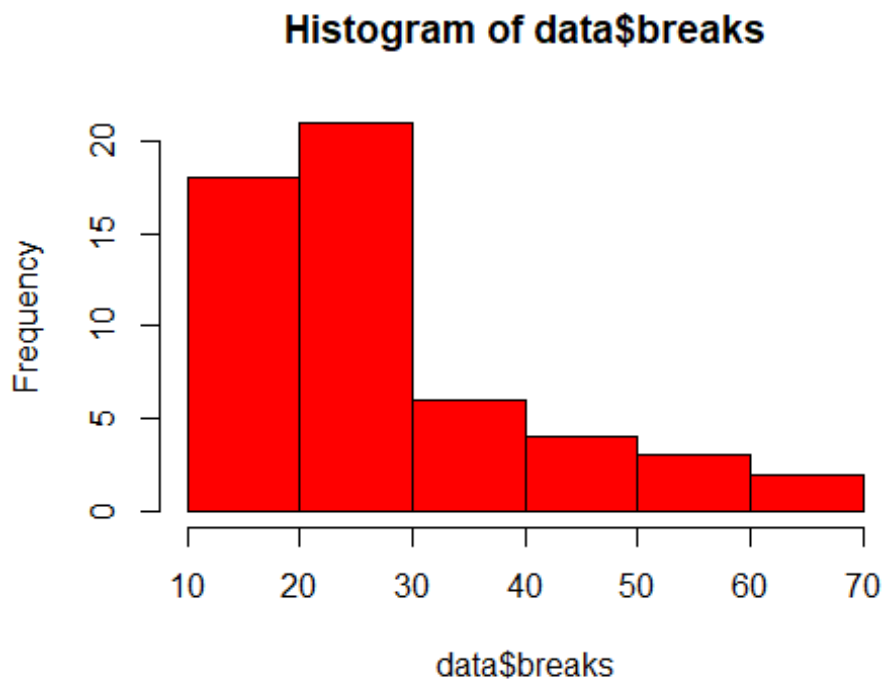
```
data<-warpbreaks
```

```
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 numero de rupturas

```
hist(data$breaks, col="red")
```



Obtén la media y la varianza de la variable dependiente

```
media <- mean(data$breaks)
var <- var(data$breaks)

cat("Media:", media, "\n")
## Media: 28.14815

cat("Varianza:", var, "\n")
## Varianza: 174.2041
```

Interpreta en el contexto de una Regresión Poisson

En una regresión de Poisson, uno de los supuestos clave es que la media y la varianza de la variable dependiente (breaks) sean aproximadamente iguales. Si la varianza es considerablemente mayor que la media, esto indica sobredispersión, un fenómeno común en datos de conteo que puede requerir ajustes adicionales.

Ajusta el modelo de regresión Poisson sin interacción

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

S <- summary(poisson_model)
S
```

```
##
## Call:
## glm(formula = breaks ~ wool + tension, family = poisson(link = "log"),
##      data = warpbreaks)
##
## 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

exp(coef(poisson_model))

## (Intercept)      woolB      tensionM      tensionH
## 40.1235380    0.8138425    0.7251908    0.5954198
```

Ajuste el modelo de regresión con interacción

```
# Ajustar el modelo con interacción
poisson_model_interaction <- glm(breaks ~ wool * tension, data =
warpbreaks, family = poisson(link = "log"))

# Resumen del modelo con interacción
S_interaction <- summary(poisson_model_interaction)
S_interaction

##
## Call:
## glm(formula = breaks ~ wool * tension, family = poisson(link = "log"),
##      data = warpbreaks)
##
## 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

exp(coef(poisson_model_interaction))

##      (Intercept)          woolB          tensionM          tensionH
## woolB:tensionM
##      44.5555556      0.6334165      0.5386534      0.5511222
##      1.8930264
## woolB:tensionH
##      1.2072719
```

El modelo con interacción presenta un AIC menor (468.97 frente a 493.06), lo que indica un mejor ajuste. La inclusión de la interacción wool:tension muestra efectos adicionales en el número de rupturas que no se capturaban en el modelo sin interacción, especialmente con combinaciones específicas de lana y tensión.

Cálculo de la desviación residual y prueba de Chi-cuadrado

Sin interacción

```
# Grados de Libertad y valor frontera para el modelo sin interacción
gl_sin_interaccion <- S$df.null - S$df.residual
q_chi_sin_interaccion <- qchisq(0.05, gl_sin_interaccion)

# Estadístico de prueba y valor p
dr_sin_interaccion <- S$deviance
vp_sin_interaccion <- 1 - pchisq(dr_sin_interaccion, gl_sin_interaccion)

cat("Modelo sin interacción:\n")

## Modelo sin interacción:

cat("Grados de libertad:", gl_sin_interaccion, "\n")

## Grados de libertad: 3

cat("Valor frontera (Chi-cuadrado):", q_chi_sin_interaccion, "\n")

## Valor frontera (Chi-cuadrado): 0.3518463

cat("Estadístico de prueba =", dr_sin_interaccion, "\n")

## Estadístico de prueba = 210.3919

cat("Valor p =", vp_sin_interaccion, "\n\n")
```

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

Con interaccion

```
# Grados de libertad y valor frontera para el modelo con interacción
gl_con_interaccion <- S_interaction$df.null - S_interaction$df.residual
q_chi_con_interaccion <- qchisq(0.05, gl_con_interaccion)

# Estadístico de prueba y valor p
dr_con_interaccion <- S_interaction$deviance
vp_con_interaccion <- 1 - pchisq(dr_con_interaccion, gl_con_interaccion)

cat("Modelo con interacción:\n")

## Modelo con interacción:

cat("Grados de libertad:", gl_con_interaccion, "\n")

## Grados de libertad: 5

cat("Valor frontera (Chi-cuadrado):", q_chi_con_interaccion, "\n")

## Valor frontera (Chi-cuadrado): 1.145476

cat("Estadístico de prueba =", dr_con_interaccion, "\n")

## Estadístico de prueba = 182.3051

cat("Valor p =", vp_con_interaccion, "\n\n")

## Valor p = 0
```

Ambos modelos tienen un valor p de 0, lo que indica que ambos modelos son significativos en cuanto a su capacidad para explicar la variabilidad en los datos. El modelo con interacción tiene un estadístico de prueba menor, lo que indica que se ajusta mejor a los datos.

Comparacion AIC

```
AIC_sin_interaccion <- S$aic
AIC_con_interaccion <- S_interaction$aic

cat("AIC del modelo sin interacción:", AIC_sin_interaccion, "\n")

## AIC del modelo sin interacción: 493.056

cat("AIC del modelo con interacción:", AIC_con_interaccion, "\n")

## AIC del modelo con interacción: 468.9692
```

El AIC del modelo con interacción es menor (468.9692) que el del modelo sin interacción (493.056), lo cual sugiere que el modelo con interacción ofrece un mejor balance entre ajuste y complejidad.

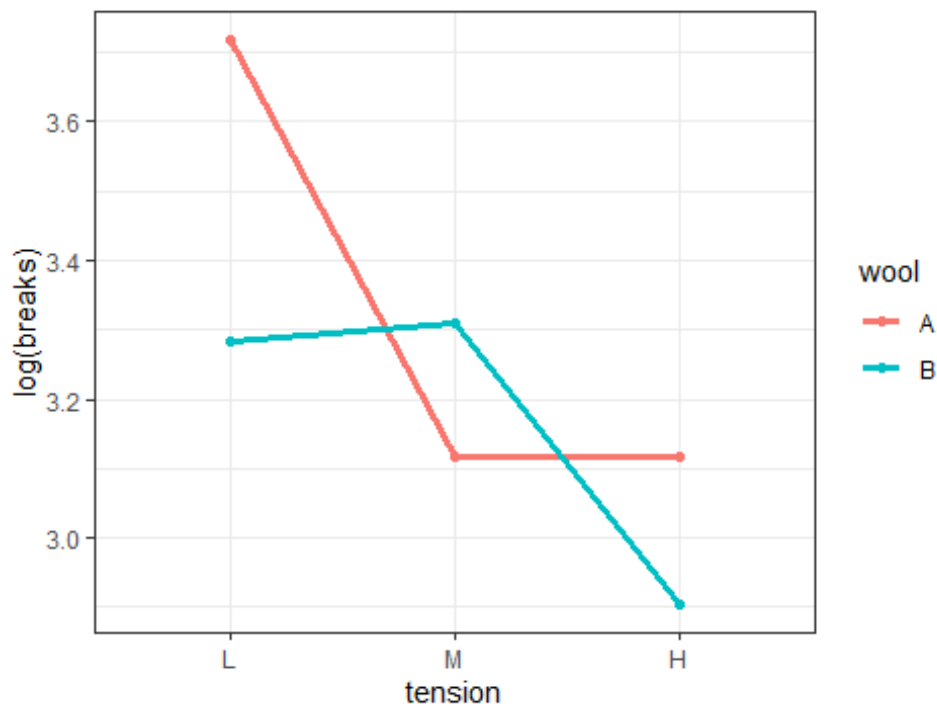
Comparación de coeficientes y errores estándar

```
coef_nombres <- c("(Intercept)", "woolB", "tensionM", "tensionH",  
"woolB:tensionM", "woolB:tensionH")  
  
coef_sin_interaccion <- coef(poisson_model)  
coef_con_interaccion <- coef(poisson_model_interaction)  
se_sin_interaccion <- sqrt(diag(vcov(poisson_model)))  
se_con_interaccion <- sqrt(diag(vcov(poisson_model_interaction)))  
  
comparacion <- data.frame(  
  Coeficiente_Sin_Interaccion = coef_sin_interaccion[coef_nombres],  
  SE_Sin_Interaccion = se_sin_interaccion[coef_nombres],  
  Coeficiente_Con_Interaccion = coef_con_interaccion[coef_nombres],  
  SE_Con_Interaccion = se_con_interaccion[coef_nombres]  
)  
  
comparacion  
  
##               Coeficiente_Sin_Interaccion SE_Sin_Interaccion  
## (Intercept)                3.6919631      0.04541069  
## woolB                   -0.2059884      0.05157117  
## tensionM                 -0.3213204      0.06026580  
## tensionH                 -0.5184885      0.06395944  
## woolB:tensionM                      NA              NA  
## woolB:tensionH                      NA              NA  
##               Coeficiente_Con_Interaccion SE_Con_Interaccion  
## (Intercept)                3.7967368      0.04993753  
## woolB                   -0.4566272      0.08019202  
## tensionM                 -0.6186830      0.08440012  
## tensionH                 -0.5957987      0.08377723  
## woolB:tensionM            0.6381768      0.12215312  
## woolB:tensionH            0.1883632      0.12989529
```

Gráfico de interaccion

```
library(ggplot2)  
ggplot(warpbreaks, 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")) +  
  labs(title = "Interacción entre Wool y Tension en el Número de  
Rupturas")
```

Interacción entre Wool y Tension en el Número de Rupturas



El gráfico muestra que el efecto de tension en breaks varía según el tipo de wool. Esto sugiere que la interacción entre wool y tension es relevante y, por lo tanto, respalda la inclusión de los términos de interacción en el modelo.

Conclusión mejor modelo

Basado en los criterios de AIC, desviación residual, y la interpretación de los coeficientes, el modelo con interacción es mejor que el de sin interacción

Evaluación de la independencia

```
library(lmtest)
```

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

```
##
```

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

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## as.Date, as.Date.numeric
```

```
dwtest(poisson_model_interaction)
```

```
##
```

```
## Durbin-Watson test
```

```
##
```

```
## data: poisson_model_interaction
```

```
## DW = 2.2376, p-value = 0.575
## alternative hypothesis: true autocorrelation is greater than 0
```

El valor de DW=2.2376 con un valor p de 0.575 sugiere que no hay autocorrelación significativa en los residuos del modelo. Esto indica que el supuesto de independencia de los residuos se cumple en el modelo con interacción.

Verificación de Sobredispersión

```
library(epiDisplay)
```

```
## Warning: package 'epiDisplay' was built under R version 4.3.3
```

```
## 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
```

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

```
##
```

```
##      alpha
```

```
poisgof(poisson_model_interaction)
```

```
## $results
```

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

```
##
```

```
## $chisq
```

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

```
##
```

```
## $df
```

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

```
##
```

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

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

El valor de Chi-cuadrado es 182.3051 con 48 grados de libertad y un valor p de 1.582538×10^{-17} , que es extremadamente pequeño. Dado el valor p muy bajo, rechazamos la hipótesis nula de que no hay sobredispersión. Esto indica que hay sobredispersión en el modelo, lo que sugiere que el modelo de Poisson estándar no es adecuado y que necesitamos un modelo alternativo.

Dado que existe sobredispersión en el modelo de Poisson, probaremos los siguientes modelos alternativos:

Modelo Quasi-Poisson

```
poisson_model_quasi <- glm(breaks ~ wool * tension, data = warpbreaks,  
family = quasipoisson(link = "log"))
```

```
summary(poisson_model_quasi)
```

```
##  
## Call:  
## glm(formula = breaks ~ wool * tension, family = quasipoisson(link =  
"log"),  
## data = warpbreaks)  
##  
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.79674 0.09688 39.189 < 2e-16 ***  
## woolB -0.45663 0.15558 -2.935 0.005105 **  
## tensionM -0.61868 0.16374 -3.778 0.000436 ***  
## tensionH -0.59580 0.16253 -3.666 0.000616 ***  
## woolB:tensionM 0.63818 0.23699 2.693 0.009727 **  
## woolB:tensionH 0.18836 0.25201 0.747 0.458436  
## ---  
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for quasipoisson family taken to be 3.76389)  
##  
## Null deviance: 297.37 on 53 degrees of freedom  
## Residual deviance: 182.31 on 48 degrees of freedom  
## AIC: NA  
##  
## Number of Fisher Scoring iterations: 4
```

Modelo Binomial Negativa

Ajuste del modelo binomial negativa con interacción

```
model_nb <- glm.nb(breaks ~ wool * tension, data = warpbreaks, control =  
glm.control(maxit = 1000))
```

Resumen del modelo binomial negativa

```
summary(model_nb)
```

```
##  
## Call:  
## glm.nb(formula = breaks ~ wool * tension, data = warpbreaks,  
## 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

```

Conclusión

Dado que el modelo binomial negativa tiene el menor AIC y ajusta bien los datos con una desviación residual aceptable, este es el modelo recomendado. Este modelo proporciona un ajuste superior en términos de balance entre precisión y complejidad, y maneja la sobredispersión de manera efectiva.

Podemos interpretar el modelo binomial negativa como el modelo final para estos datos. Este modelo indica que la interacción entre wool y tension tiene un efecto significativo en el número de rupturas de urdimbre, excepto en la combinación woolB:tensionH, que no es significativa.