Actividad-18-A6

Saúl Francisco Vázquez del Río

2024-10-29

I. Análisis Descriptivo

Histograma del número de rupturas Obtén la media y la varianza de la variable dependiente Interpreta en el contexto de una Regresión Poisson

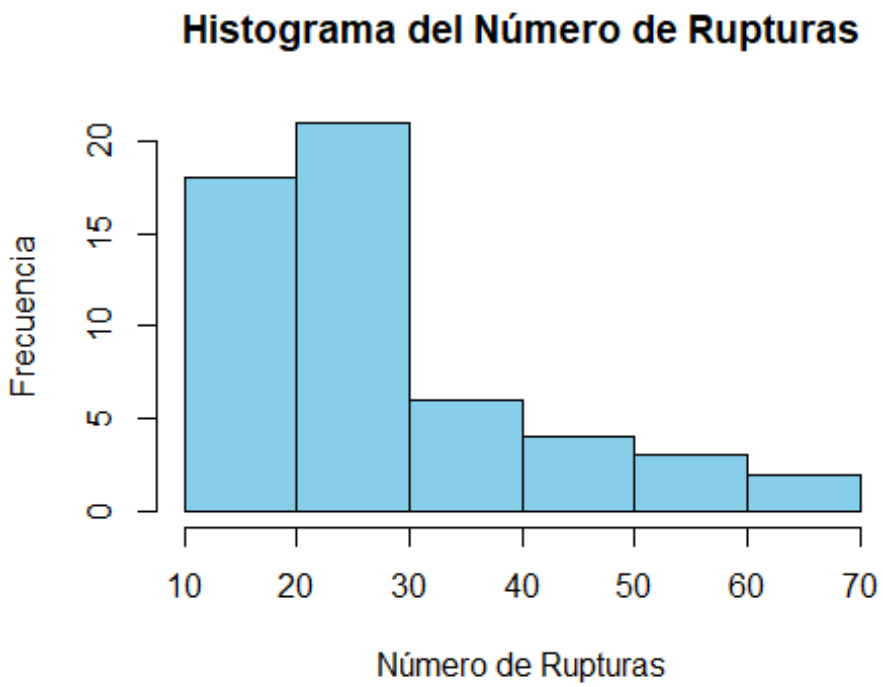
```
data<-warpbreaks
head(data,10)
```

```
##    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
```

```
# Cargar el conjunto de datos warpbreaks y mostrar las primeras 10 filas
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 número de rupturas
hist(data$breaks, main = "Histograma del Número de Rupturas", xlab =
"Número de Rupturas", ylab = "Frecuencia", col = "skyblue", border =
"black")
```




```
# Cálculo de la media y varianza de la variable dependiente (breaks)
media_breaks <- mean(data$breaks)
varianza_breaks <- var(data$breaks)

media_breaks

## [1] 28.14815

varianza_breaks

## [1] 174.2041
```

Se observa en el histrograma de que los datos estan mayormente ubicados en el inicio de este teniendo un cremiento en el inicio y conforme se aumenta el numero de rupturas en el eje x la frecuencia de estas baja.

II. Ajusta dos modelos de Regresión Poisson

Ajusta el modelo de regresión Poisson sin interacción 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) 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.

```
# Modelo Poisson sin interacción
poisson_model <- glm(breaks ~ wool + tension, data = data, family =
```

```r
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)  3.69196    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

# Modelo Poisson con interacción
poisson_model_inter <- glm(breaks ~ wool * tension, data = data, family =
poisson(link = "log"))
S_inter <- summary(poisson_model_inter)
S_inter

##
## 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
```

Un vez analizadas las variables podemos observar que la mejor lana es la B ya que esta no tiene tantas rupturas con tension baja o medida, ademas que cuando esta tiene tension alta es muy improvable que esta se rompa.

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 Aikaike 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 = $Snull.deviance - S$df.residual qchisq(0.05,gl) Estadístico de prueba y valor p: dr = S$deviance cat("Estadístico de prueba =",dr, "") vp = 1-pchisq(dr,gl) cat("Valor p =",vp)

```
# Grados de libertad para el modelo sin interacción
gl <- S$df.null - S$df.residual
valor_frontera <- qchisq(0.05, gl)

# Estadístico de prueba y valor p para el modelo sin interacción
dr <- S$deviance
cat("Estadístico de prueba =", dr, "\n")

## Estadístico de prueba = 210.3919

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

## Valor p = 0

# Grados de libertad para el modelo con interacción
gl_inter <- S_inter$df.null - S_inter$df.residual
valor_frontera_inter <- qchisq(0.05, gl_inter)

# Estadístico de prueba y valor p para el modelo con interacción
dr_inter <- S_inter$deviance
cat("Estadístico de prueba =", dr_inter, "\n")
```

```
## Estadístico de prueba = 182.3051

vp_inter <- 1 - pchisq(dr_inter, gl_inter)
cat("Valor p =", vp_inter, "\n")

## Valor p = 0
```

Compara los AIC de cada modelo. Recuerda que un menor AIC indica un mejor modelo. 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 de ambos modelos (haz una tabla para que se facilite la comparación)

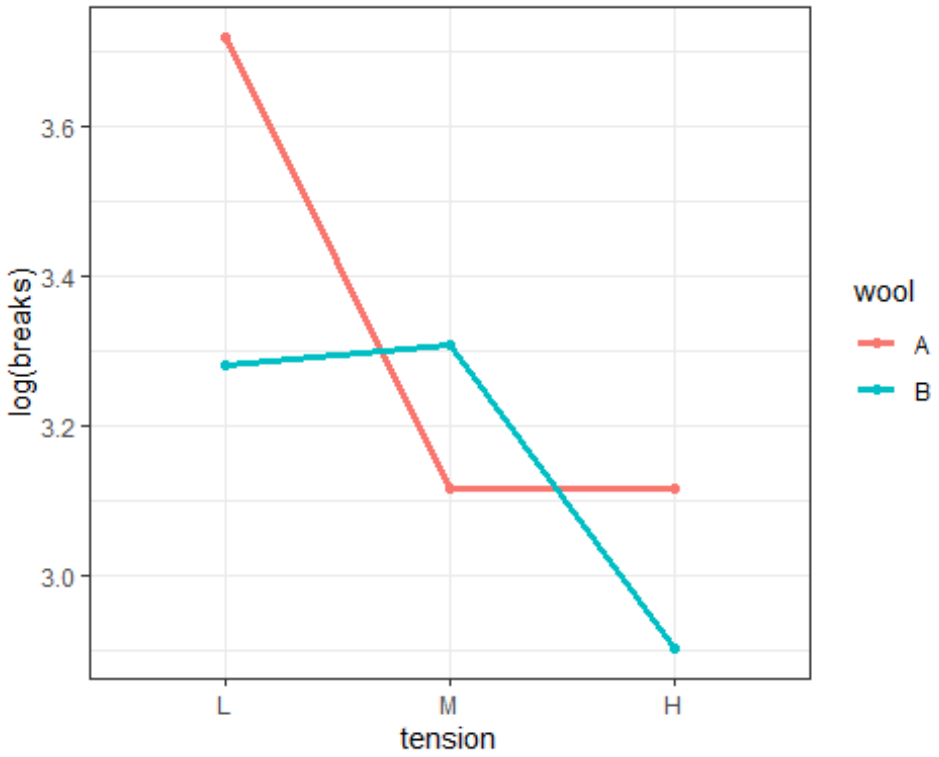
```
AIC_sin_interaccion <- AIC(poisson_model)
AIC_con_interaccion <- AIC(poisson_model_inter)
cat("AIC Modelo sin Interacción =", AIC_sin_interaccion, "\n")

## AIC Modelo sin Interacción = 493.056

cat("AIC Modelo con Interacción =", AIC_con_interaccion, "\n")

## AIC Modelo con Interacción = 468.9692
```

Interpreta los coeficientes de ambos modelos. Para interpretar mejor la interacción gráficala con el siguiente código: library(ggplot2) ggplot(data, 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(fil=“transparent”)) Define cuál de los dos es un mejor modelo

```
# Extraer los coeficientes y errores estándar de ambos modelos
coef_sin_inter <- coef(S)[, "Estimate"]
std_err_sin_inter <- coef(S)[, "Std. Error"]

coef_con_inter <- coef(S_inter)[, "Estimate"]
std_err_con_inter <- coef(S_inter)[, "Std. Error"]

# Crear la tabla de comparación
tabla_comparacion <- data.frame(
  Modelo = c(rep("Sin Interacción", length(coef_sin_inter)), rep("Con
Interacción", length(coef_con_inter))),
  Coeficiente = c(names(coef_sin_inter), names(coef_con_inter)),
  Estimación = c(coef_sin_inter, coef_con_inter),
  `Error Estándar` = c(std_err_sin_inter, std_err_con_inter)
)
tabla_comparacion

##             Modelo    Coeficiente Estimación Error.Estándar
## 1  Sin Interacción    (Intercept)  3.6919631     0.04541069
## 2  Sin Interacción          woolB -0.2059884     0.05157117
## 3  Sin Interacción       tensionM -0.3213204     0.06026580
## 4  Sin Interacción       tensionH -0.5184885     0.06395944
```

```
## 5  Con Interacción    (Intercept)  3.7967368     0.04993753
## 6  Con Interacción          woolB -0.4566272     0.08019202
## 7  Con Interacción       tensionM -0.6186830     0.08440012
## 8  Con Interacción       tensionH -0.5957987     0.08377723
## 9  Con Interacción woolB:tensionM  0.6381768     0.12215312
## 10 Con Interacción woolB:tensionH  0.1883632     0.12989529
```

```
library(ggplot2)
ggplot(data, 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(fil="transparent"))
```



Comprarando los dos modelos realizados podemos llegar a la conclusion que el modelo con interaccion es el mejor que el modelo sin interaccion, esto lo podemo saber que el modelo con interracion tiene un mejor AIC este siendo de 468.97 y su coeficientes de woolB, tensionM y tensionH son mayores a los otros coeficientes del modelo sin interaccion.

##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) poisgof(pm) 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.model2) 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) Define si usas defines tus modelos con interacción o sin interacción (no hagas los dos) Define el mejor modelo usando las mismas pruebas y crtierios que usaste en los modelos Poisson

```
library(lmtest)
```

```
## Cargando paquete requerido: zoo
```

```
##
## Adjuntando el paquete: 'zoo'
```

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

```
library(epiDisplay)
```

```
## Cargando paquete requerido: foreign
```

```
## Cargando paquete requerido: survival
```

```
## Cargando paquete requerido: MASS
```

```
## Cargando paquete requerido: nnet
```

```
##
## Adjuntando el paquete: 'epiDisplay'
```

```
## The following object is masked from 'package:lmtest':
##
##     lrtest
```

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

```
library(MASS)
dwtest(poisson_model)
```

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

```
## data:  poisson_model
## DW = 2.0332, p-value = 0.3896
## alternative hypothesis: true autocorrelation is greater than 0
```

```
poisgof(poisson_model)
```

```
## $results
## [1] "Goodness-of-fit test for Poisson assumption"
##
## $chisq
## [1] 210.3919
##
## $df
## [1] 50
##
## $p.value
## [1] 1.44606e-21
```

```
# 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 con interacción
bnm <- glm.nb(breaks ~ wool * tension, data = 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
```

```
cat("AIC Modelo Poisson =", AIC(poisson_model_inter), "\n")
```

```
## AIC Modelo Poisson = 468.9692
```

```
cat("AIC Modelo Cuasi-Poisson =", AIC(poisson_model3), "\n")
```

```
## AIC Modelo Cuasi-Poisson = NA
```

```
cat("AIC Modelo Binomial Negativa =", AIC(bnm), "\n")
```

```
## AIC Modelo Binomial Negativa = 405.1248
```

V. Define cuál es tu mejor modelo

En conclusion el mejor modelo fue el modelo binomial negativo ya que su AIC fue el menor de los tres modelos probados, esto se debe a que se escojio al modelo con interacion dandonos un mejor resultado que el modelo sin intereaccion, ademas que H0 no se rechaza no habiendo una sobredispersion en el modelo, haciendo que el modelo se ajuste correctamente.