

A6_A01742161

Rogelio Lizárraga

2024-11-04

Regresión Poisson

```
data<-warpbreaks
```

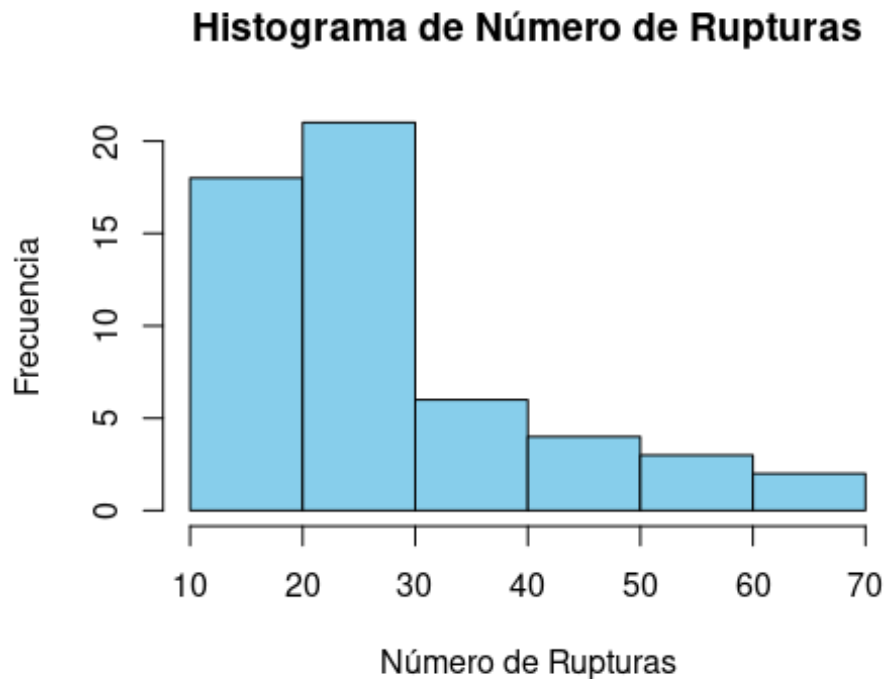
```
head(data,10)
```

```
##      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
```

I. Análisis Descriptivo

Histograma del número de rupturas

```
hist(data$breaks,
      main = "Histograma de Número de Rupturas",
      xlab = "Número de Rupturas",
      ylab = "Frecuencia",
      col = "skyblue",
      border = "black")
```



Como podemos observar, el histograma muestra la distribución del número de rupturas en distintos intervalos. La mayoría de las observaciones se encuentran en los valores bajos, lo que indica que las rupturas suelen ocurrir con menor frecuencia en la mayoría de los casos.

Interpretación en el contexto de una Regresión Poisson: En un modelo de regresión de Poisson, se espera que la media y la varianza de la variable dependiente sean aproximadamente iguales. Sin embargo, en este caso, la varianza es mucho mayor que la media, lo que sugiere la presencia de sobredispersión en los datos. Esto significa que un modelo de Poisson estándar podría no ser el mejor ajuste para estos datos. Para manejar esta sobredispersión, se podría considerar el uso de un modelo de Poisson negativo o explorar otras alternativas para mejorar el ajuste. ### Obtén la media y la varianza de la variable dependiente

```
media_breaks <- mean(data$breaks)
varianza_breaks <- var(data$breaks)
cat("Media de rupturas:", media_breaks, "\n")

## Media de rupturas: 28.14815

cat("Varianza de rupturas:", varianza_breaks, "\n")

## Varianza de rupturas: 174.2041
```

Media de 28.15 y Varianza de 174.2 ### Interpreta en el contexto de una Regresión Poisson

En un modelo de regresión de Poisson, se espera que la media y la varianza de la variable dependiente sean aproximadamente iguales. Sin embargo, podemos observar que en este caso, la varianza es mucho mayor que la media, por lo que es muy probable que haya una presencia de sobredispersión en los datos. Esto significa que un modelo de Poisson estándar podría no ser el mejor ajuste para estos datos.

II. Ajusta dos modelos de Regresión Poisson

Ajusta el modelo de regresión Poisson sin interacción

```
poisson_model <- glm(breaks ~ wool + tension, data = data, family =  
poisson(link = "log"))  
S <- summary(poisson_model)  
S  
  
##  
## Call:  
## glm(formula = breaks ~ wool + tension, family = poisson(link = "log"),  
##      data = data)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max   
## -3.6871  -1.6503  -0.4269   1.1902   4.2616   
##  
## 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 Obtenido

La ecuación para el modelo sin interacción es: $\log(E[\text{breaks}]) = 3.692 - 0.20(\text{woolB}) - 0.321(\text{tensionM}) - 0.519(\text{tensionH})$ woolB es una variable Dummy que toma el valor 1 si wool = B, y 0 si wool = A. tensionM es una variable Dummy que toma el valor 1 si tension = M, y 0 si tension = L. tensionH es una variable Dummy que toma el valor 1 si tension = H, y 0 si tension = L.

Interpretación

Intercepto (3.692): El intercepto es el logaritmo del número esperado de rupturas (breaks) para la categoría de referencia, es decir, wool = A y tension = L.

woolB (-0.206): Este coeficiente indica el cambio en el logaritmo del número esperado de rupturas cuando se usa wool = B en lugar de wool = A (la categoría de referencia). Al observar su valor, observamos que, si mantenemos constante la tensión, la lana tipo B tiene menos rupturas esperadas en comparación con la lana tipo A.

tensionM (-0.321): Este coeficiente indica el cambio en el logaritmo del número esperado de rupturas cuando se usa tension = M en lugar de tension = L (la categoría de referencia). Al observar su valor, observamos que, si mantenemos constante el tipo de lana, la tensión mediana reduce el número esperado de rupturas en comparación con la tensión baja o L.

tensionH (-0.519): Este coeficiente representa el cambio en el logaritmo del número esperado de rupturas cuando se usa tension = H en lugar de tension = L. Al observar su valor, observamos que, si mantenemos constante el tipo de lana, la tensión alta también reduce el número esperado de rupturas en comparación con la tensión baja.

Ajusta el modelo de regresión Poisson con interacción

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

```
S_interaction
```

```
##
## Call:
## glm(formula = breaks ~ wool * tension, family = poisson(link = "log"),
##      data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3383  -1.4844  -0.1291   1.1725   3.5153
##
## 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
```

Modelo Obtenido

La ecuación para el modelo con interacción es: $\log(E[breaks]) = 3.797 - 0.456(woolB) - 0.616(tensionM) - 0.595(tensionH) + 0.638(woolB:tensionM) + 0.188(woolB:tensionH)$

Interpretación

Intercepto (3.797): Este valor representa el logaritmo del número esperado de rupturas para la categoría de referencia (wool = A y tension = L).

woolB (-0.456): Cambio en el logaritmo del número esperado de rupturas cuando se usa wool = B en lugar de wool = A, manteniendo la tensión constante en L.

tensionM (-0.616): Cambio en el logaritmo del número esperado de rupturas cuando se usa tension = M en lugar de tension = L, manteniendo la lana constante en A.

tensionH (-0.595): Cambio en el logaritmo del número esperado de rupturas cuando se usa tension = H en lugar de tension = L, con la lana constante en A.

woolB:tensionM (0.638): Este coeficiente representa el efecto adicional cuando wool = B y tension = M en comparación con las categorías de referencia wool = A y tension = L.

woolB:tensionH(0.188): Este coeficiente representa el efecto adicional cuando wool = B y tension = H, en comparación con las categorías de referencia wool = A y tension = L.

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. H_0 : Deviance = 0 (el modelo no explica significativamente la variabilidad de los datos). H_1 : Deviance > 0 (el modelo explica significativamente la variabilidad de los datos).

Modelo Sin Interacción

Valor frontera de La zona de rechazo

```
gl_no_interaction <- S$df.null - S$df.residual
q_no_interaction <- qchisq(0.05, gl_no_interaction)
cat("Valor frontera (Modelo sin interacción) =", q_no_interaction, "\n")
```

```
## Valor frontera (Modelo sin interacción) = 0.3518463
```

Estadístico de prueba y valor p

```
dr_no_interaction <- S$deviance
```

```

cat("Estadístico de prueba (Modelo sin interacción) =", dr_no_interaction,
"\n")

## Estadístico de prueba (Modelo sin interacción) = 210.3919

vp_no_interaction <- 1 - pchisq(dr_no_interaction, gl_no_interaction)
cat("Valor p (Modelo sin interacción) =", vp_no_interaction, "\n")

## Valor p (Modelo sin interacción) = 0

## Modelo Con Interacción

# Valor frontera de la zona de rechazo
gl_interaction <- S_interaction$df.null - S_interaction$df.residual
q_interaction <- qchisq(0.05, gl_interaction)
cat("Valor frontera (Modelo con interacción) =", q_interaction, "\n")

## Valor frontera (Modelo con interacción) = 1.145476

# Estadístico de prueba y valor p
dr_interaction <- S_interaction$deviance
cat("Estadístico de prueba (Modelo con interacción) =", dr_interaction, "\n")

## Estadístico de prueba (Modelo con interacción) = 182.3051

vp_interaction <- 1 - pchisq(dr_interaction, gl_interaction)
cat("Valor p (Modelo con interacción) =", vp_interaction, "\n")

## Valor p (Modelo con interacción) = 0

```

Como podemos observar, ambos valores son muy cercanos, por lo que rechazamos la hipótesis nula de que la desviación es cero, lo que sugiere que ambos modelos tienen significancia.

AIC: Criterio de Akaike

```

cat("AIC (Modelo sin interacción) =", AIC(poisson_model), "\n")

## AIC (Modelo sin interacción) = 493.056

cat("AIC (Modelo con interacción) =", AIC(poisson_model_interaction), "\n")

## AIC (Modelo con interacción) = 468.9692

```

En este caso, el modelo con interacción tiene un AIC menor que el modelo sin interacción, lo que sugiere que el modelo con interacción es superior y proporciona un mejor ajuste a los datos, de acuerdo con el criterio de Akaike.

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

```

coef_no_interaction <- as.data.frame(summary(poisson_model)$coefficients)
coef_interaction <-
as.data.frame(summary(poisson_model_interaction)$coefficients)

```

```

coef_no_interaction$Coefficient <- rownames(coef_no_interaction)
coef_interaction$Coefficient <- rownames(coef_interaction)

colnames(coef_no_interaction) <- c("Estimate_No_Interaction",
"Std_Error_No_Interaction", "z_value_No_Interaction", "Pr_No_Interaction",
"Coefficient")
colnames(coef_interaction) <- c("Estimate_Interaction",
"Std_Error_Interaction", "z_value_Interaction", "Pr_Interaction",
"Coefficient")

comparison_table <- merge(coef_no_interaction, coef_interaction, by =
"Coefficient", all = TRUE)

print(comparison_table)

```

##	Coefficient	Estimate_No_Interaction	Std_Error_No_Interaction	
## 1	(Intercept)	3.6919631	0.04541069	
## 2	tensionH	-0.5184885	0.06395944	
## 3	tensionM	-0.3213204	0.06026580	
## 4	woolB	-0.2059884	0.05157117	
## 5	woolB:tensionH	NA	NA	
## 6	woolB:tensionM	NA	NA	
##	z_value_No_Interaction	Pr_No_Interaction	Estimate_Interaction	
## 1	81.301626	0.000000e+00	3.7967368	
## 2	-8.106520	5.209021e-16	-0.5957987	
## 3	-5.331721	9.728642e-08	-0.6186830	
## 4	-3.994256	6.489775e-05	-0.4566272	
## 5	NA	NA	0.1883632	
## 6	NA	NA	0.6381768	
##	Std_Error_Interaction	z_value_Interaction	Pr_Interaction	
## 1	0.04993753	76.029734	0.000000e+00	
## 2	0.08377723	-7.111702	1.146202e-12	
## 3	0.08440012	-7.330357	2.295399e-13	
## 4	0.08019202	-5.694172	1.239721e-08	
## 5	0.12989529	1.450115	1.470263e-01	
## 6	0.12215312	5.224400	1.747203e-07	

Interpretación de los coeficientes

Intercepto: En ambos modelos, el intercepto es significativamente positivo. Representa el logaritmo del número esperado de rupturas para la categoría de referencia (wool = A y tension = L). El valor del intercepto es ligeramente mayor en el modelo con interacción.

tensionH y tensionM: Los coeficientes de tensionH y tensionM son negativos en ambos modelos, indicando que la tensión media y alta tienden a reducir el número esperado de rupturas en comparación con la tensión baja. En el modelo con interacción, los coeficientes de tensionH y tensionM son ligeramente más negativos, sugiriendo un efecto más fuerte en la reducción de rupturas cuando se incluye la interacción. Los valores p son muy bajos (significativos) para ambas variables en ambos modelos.

woolB: El coeficiente de woolB es negativo en ambos modelos, indicando que el tipo de lana B tiende a tener menos rupturas que el tipo A. La magnitud del efecto de woolB es más fuerte en el modelo con interacción. El valor p es significativo en ambos modelos.

Interacciones (woolB:tensionH y woolB:tensionM) (solo en el modelo con interacción): woolB:tensionH tiene un valor positivo (aunque no significativo), lo que sugiere que, bajo tensión alta, el tipo de lana B podría aumentar levemente el número de rupturas en comparación con las combinaciones sin interacción. woolB:tensionM también es positivo y significativo, indicando que el efecto de tensionM en la reducción de rupturas es modificado cuando wool = B.

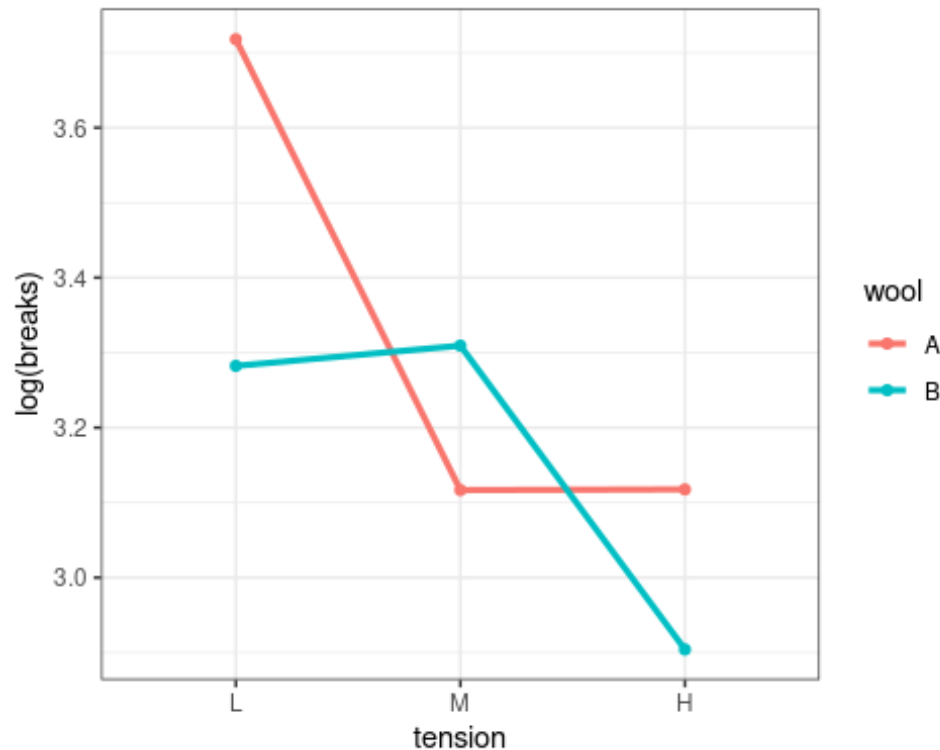
Interpretación de los errores estándar

Modelo sin interacción: Los errores estándar son menores en general para tensionH, tensionM, y woolB, lo que indica una mayor precisión en las estimaciones cuando no se considera la interacción.

Modelo con interacción: Los errores estándar son mayores en el modelo con interacción, lo que sugiere que la inclusión de los términos de interacción introduce más variabilidad en las estimaciones. Sin embargo, esto es esperado en modelos con términos de interacción, ya que captura relaciones más complejas entre las variables.

Visualización de la interacción

```
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(fill = "transparent"))
```

Lana Tipo A: Las rupturas disminuyen considerablemente con el aumento de la tensión (de baja a alta).

Lana Tipo B: Las rupturas son constantes de baja a media tensión, pero disminuyen notablemente a alta tensión.

La interacción muestra que la lana A responde más a los cambios de tensión en general, mientras que la lana B solo reduce las rupturas significativamente a alta tensión. Esto justifica el uso del modelo con interacción.

Conclusión

Aunque los errores estándar son mayores en el modelo con interacción, el beneficio de capturar la relación entre wool y tensión justifica esta variabilidad adicional, especialmente dado que el modelo con interacción tiene un menor AIC y proporciona una interpretación más detallada de cómo la tensión afecta cada tipo de lana. Por lo tanto, el modelo con interacción sigue siendo mejor.

IV. Evaluación de los supuestos

Independencia

H_0 = la autocorrelación es 0. H_1 = la autocorrelación es mayor a 0.

```
library(lmtest)
```

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

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

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

# Prueba de independencia (Durbin-Watson) en el modelo con interacción
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

# Prueba de independencia (Durbin-Watson) en el modelo sin interacción
dwtest(poisson_model)

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

Como nuestros valores $p > 0.05$, nuestra hipótesis inicial no se rechaza, por lo que no hay autocorrelación y los residuos son independientes.

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
```

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

# Prueba de sobredispersión en el modelo sin interacción
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

# Prueba de sobredispersión en el modelo con interacción
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
```

Como nuestros valores $p < 0.05$, nuestra hipótesis inicial se rechaza, por lo que hay una sobredispersión en ambos modelos.

Resultados de los modelos

Como podemos observar, los modelos utilizados no son adecuados, pues hay una gran sobredispersión en ambos. Por ello, implementaremos un modelo de binomial negativa con interacción.

IV.2 Implementación modelo de binomial negativa con interacción

```
library(MASS)
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)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -2.09611  -0.89383  -0.07212   0.65270   1.80646
##
## 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
```

La ecuación del modelo de regresión binomial negativa con interacción:

$\log(E[\text{breaks}]) = 3.7967 - 0.4566 \cdot \text{woolB} - 0.6187 \cdot \text{tensionM} - 0.5958 \cdot \text{tensionH} + 0.6382 \cdot (\text{woolB} \times \text{tensionM}) + 0.1884 \cdot (\text{woolB} \times \text{tensionH})$ ##### Interpretación
 ntercepto: El valor del intercepto (3.7967) representa el logaritmo del número esperado de rupturas para la combinación de referencia (wool = A y tension = L), y es altamente significativo ($p < 2e-16$).

woolB: El coeficiente de woolB (-0.4566) indica que, para la lana tipo B, el número de rupturas es menor en comparación con la lana tipo A, manteniendo la tensión en el nivel bajo (L). Este efecto es significativo ($p = 0.00375$).

tensionM y tensionH: Los coeficientes de tensionM (-0.6187) y tensionH (-0.5958) son negativos y significativos, lo que indica que las tensiones media y alta reducen el número de rupturas en comparación con la tensión baja, manteniendo el tipo de lana en A.

woolB

La interacción entre woolB y tensionM (0.6382) es positiva y significativa ($p = 0.00500$), lo que sugiere que cuando wool = B y tension = M, el efecto de la tensión media sobre las rupturas es diferente al de la lana tipo A. Específicamente, esta combinación reduce menos las rupturas en comparación con el efecto individual de tensionM.

woolB

La interacción entre woolB y tensionH (0.1884) no es significativa ($p = 0.4161$), lo que sugiere que la combinación de lana tipo B y tensión alta no tiene un efecto significativo en el número de rupturas más allá de los efectos individuales de woolB y tensionH. ### Desviación residual $H_0 = \text{Deviance} = 0$ (el modelo no explica significativamente la variabilidad de los datos). $H_1 = \text{Deviance} > 0$ (el modelo explica significativamente la variabilidad de los datos).

```
# Grados de Libertad
gl_bnm <- bnm$df.null - bnm$df.residual
# Valor frontera de la zona de rechazo
q_bnm <- qchisq(0.05, gl_bnm)
cat("Valor frontera (Modelo Binomial Negativa con interacción) =", q_bnm,
"\n")

## Valor frontera (Modelo Binomial Negativa con interacción) = 1.145476

# Estadístico de prueba y valor p
dr_bnm <- bnm$deviance
cat("Estadístico de prueba (Modelo Binomial Negativa con interacción) =",
dr_bnm, "\n")

## Estadístico de prueba (Modelo Binomial Negativa con interacción) =
53.50616

vp_bnm <- 1 - pchisq(dr_bnm, gl_bnm)
cat("Valor p (Modelo Binomial Negativa con interacción) =", vp_bnm, "\n")

## Valor p (Modelo Binomial Negativa con interacción) = 2.647427e-10
```

el valor $p < 0.05$, rechazamos H_0 , lo que sugiere que el modelo tiene significancia y es un buen ajuste para los datos. ### Criterio de Akaike

```
# AIC del modelo Binomial Negativa
aic_bnm <- AIC(bnm)
cat("AIC (Modelo Binomial Negativa con interacción) =", aic_bnm, "\n")

## AIC (Modelo Binomial Negativa con interacción) = 405.1248
```

El modelo Binomial Negativa con interacción tiene el menor AIC, indicando que proporciona el mejor ajuste entre los tres modelos. ### Comparaciones de coeficientes y errores estándar

```
# Extraer Los coeficientes y errores estándar del modelo de Binomial Negativa
coef_bnm <- as.data.frame(summary(bnm)$coefficients)
```

```

coef_bnm$Coefficient <- rownames(coef_bnm)

# Renombrar las columnas para facilitar la comparación
colnames(coef_bnm) <- c("Estimate_Binomial", "Std_Error_Binomial",
"z_value_Binomial", "Pr_Binomial", "Coefficient")

# Combinar los coeficientes y errores estándar de los tres modelos (Poisson
sin interacción, Poisson con interacción y Binomial Negativa con interacción)
comparison_table <- merge(comparison_table, coef_bnm, by = "Coefficient", all
= TRUE)
print(comparison_table)

##      Coefficient Estimate_No_Interaction Std_Error_No_Interaction
## 1      (Intercept)           3.6919631           0.04541069
## 2      tensionH           -0.5184885           0.06395944
## 3      tensionM           -0.3213204           0.06026580
## 4      woolB            -0.2059884           0.05157117
## 5 woolB:tensionH              NA              NA
## 6 woolB:tensionM              NA              NA
##      z_value_No_Interaction Pr_No_Interaction Estimate_Interaction
## 1           81.301626      0.000000e+00      3.7967368
## 2           -8.106520      5.209021e-16      -0.5957987
## 3           -5.331721      9.728642e-08      -0.6186830
## 4           -3.994256      6.489775e-05      -0.4566272
## 5              NA              NA           0.1883632
## 6              NA              NA           0.6381768
##      Std_Error_Interaction z_value_Interaction Pr_Interaction
Estimate_Binomial
## 1           0.04993753           76.029734  0.000000e+00
3.7967368
## 2           0.08377723           -7.111702  1.146202e-12      -
0.5957987
## 3           0.08440012           -7.330357  2.295399e-13      -
0.6186830
## 4           0.08019202           -5.694172  1.239721e-08      -
0.4566272
## 5           0.12989529           1.450115  1.470263e-01
0.1883632
## 6           0.12215312           5.224400  1.747203e-07
0.6381768
##      Std_Error_Binomial z_value_Binomial  Pr_Binomial
## 1           0.1081206      35.1157684 3.873382e-270
## 2           0.1594090      -3.7375471  1.858243e-04
## 3           0.1597372      -3.8731301  1.074465e-04
## 4           0.1575543      -2.8982209  3.752862e-03
## 5           0.2316419           0.8131652  4.161233e-01
## 6           0.2273908           2.8065200  5.007981e-03

```

Intercepto: Representa el logaritmo del número esperado de rupturas para la categoría de referencia (wool = A y tension = L). El intercepto en el modelo de Binomial Negativa es positivo y significativo.

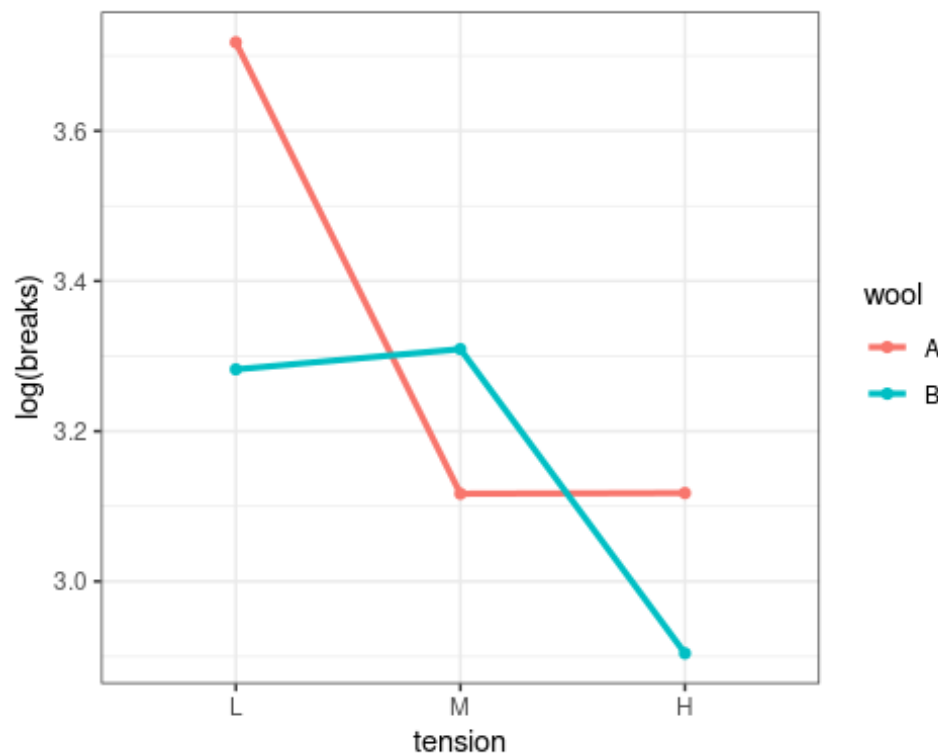
tensionH y tensionM: Estos coeficientes son negativos, lo que sugiere que tanto la tensión media como la alta reducen las rupturas en comparación con la tensión baja. Los valores son similares a los del modelo Poisson con interacción, pero ajustados a la sobredispersión.

woolB: El coeficiente negativo para woolB indica que la lana tipo B tiene menos rupturas en comparación con la lana tipo A. Este coeficiente es significativo en el modelo de Binomial Negativa.

Interacciones (woolB:tensionH y woolB:tensionM): En el modelo de Binomial Negativa, woolB:tensionM es positivo y significativo, sugiriendo que el efecto de tensionM en la reducción de rupturas cambia cuando wool = B

Gráfico de interacción

```
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(fill = "transparent"))
```



Lana A: Las rupturas disminuyen notablemente al pasar de tensión baja a media y se estabilizan de media a alta. Lana B: Las rupturas son constantes de baja a media tensión y disminuyen solo a alta tensión. Esto indica que la lana A responde más a cambios en la tensión,

mientras que la lana B solo reduce rupturas a alta tensión, justificando el modelo con interacción.

Validación de supuestos

Independencia

```
# Cargar el paquete necesario
library(lmtest)

# Prueba de independencia en el modelo de Binomial Negativa con interacción
dwtest(bnm, alternative = "two.sided")

##
## Durbin-Watson test
##
## data:  bnm
## DW = 2.2376, p-value = 0.8499
## alternative hypothesis: true autocorrelation is not 0
```

Como nuestros valores $p > 0.05$, nuestra hipótesis inicial no se rechaza, por lo que no hay autocorrelación y los residuos son independientes.

Sobredispersión

```
# Calcular el parámetro de dispersión para el modelo binomial negativa
dispersion_bnm <- sum(residuals(bnm, type = "deviance")^2) /
bnm$df.residual
cat("Parámetro de dispersión (binomial negativa):", dispersion_bnm, "\n")

## Parámetro de dispersión (binomial negativa): 1.114712
## Parámetro de dispersión (binomial negativa): 1.114712
```

Como podemos observar, obtenemos un parámetro de dispersión de 1.1147, lo cual es cercano a 1, por lo que el modelo maneja de manera adecuada la sobredispersión de los datos.

V. Mejor modelo

El mejor modelo es el de Binomial Negativa con interacción. Esto se debe a los siguientes factores: **Sobredispersión en los datos:** En el análisis inicial, los modelos Poisson presentaron una varianza mucho mayor que la media, lo que indica sobredispersión. La Regresión Binomial Negativa maneja de manera adecuada esta sobredispersión, realizando un ajuste exitoso de los datos.

Datos de conteo: Los datos representan conteos (número de rupturas), y los modelos de Regresión Binomial Negativa están diseñados específicamente para variables de respuesta de tipo conteo, especialmente cuando estos conteos no siguen una distribución Poisson perfecta.

Efecto de Interacción: Este conjunto de datos muestra una interacción entre el tipo de lana (wool) y el nivel de tensión (tension), afectando el número de rupturas. La Regresión Binomial Negativa permite incluir términos de interacción, lo que nos ayuda a capturar cómo los efectos de una variable dependen de los niveles de la otra.

Manejo adecuado de la variabilidad: El parámetro de dispersión calculado está cercano a 1, lo que indica que el modelo de Binomial Negativa ajusta bien la variabilidad observada sin sobreajuste.

En conclusión, el modelo de Regresión Binomial Negativa es bueno para este tipo de datos de conteo con sobredispersión y efectos de interacción, proporcionando un ajuste exitoso y preciso.