

Regresión Poisson

Catherine Rojas

2024-10-29

Trabajaremos con el paquete `dataset`, que incluye la base de datos `warpbreaks`, que contiene datos del hilo (yarn) para identificar cuáles variables predictoras afectan la ruptura de urdimbre.

```
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
```

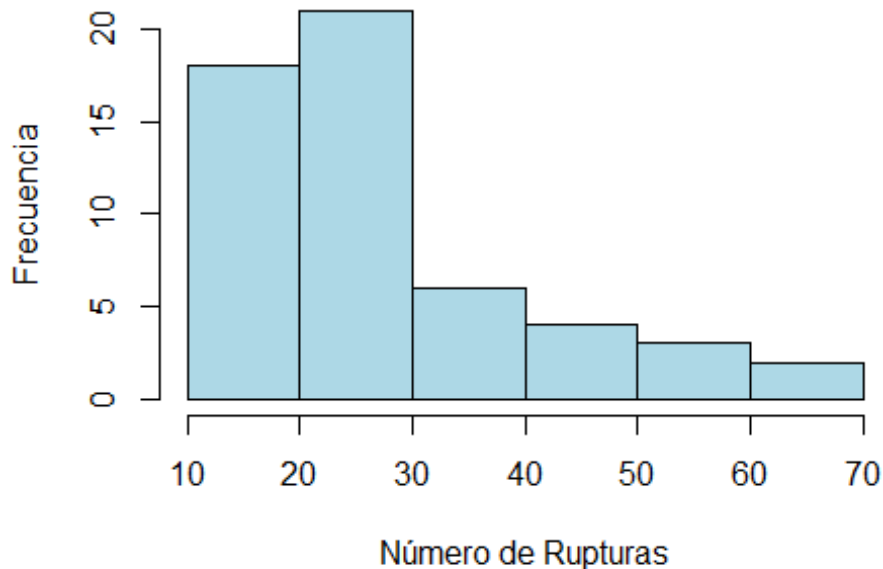
Este conjunto de datos indica cuántas roturas de urdimbre ocurrieron para diferentes tipos de telares por telar, por longitud fija de hilo:

- `breaks`: número de rupturas
- `wool`: tipo de lana (A o B)
- `tension`: el nivel de tensión (L, M, H)

I. Análisis Descriptivo

```
# Histograma del número de rupturas  
hist(data$breaks, main = "Histograma del Número de Rupturas",  
      xlab = "Número de Rupturas", ylab = "Frecuencia", col = "lightblue",  
      border = "black")
```

Histograma del Número de Rupturas



```
# Calcular La media de La variable dependiente (número de rupturas)
media_breaks <- mean(data$breaks)
cat("\n Media de 'breaks': \n", media_breaks)

##
## Media de 'breaks':
## 28.14815

# Calcular La varianza de La variable dependiente (número de rupturas)
varianza_breaks <- var(data$breaks)
cat("\n Varianza de 'breaks': \n", varianza_breaks)

##
## Varianza de 'breaks':
## 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 deben ser aproximadamente iguales. Esto se debe a que la distribución de Poisson asume que la varianza es igual a la media.
- La varianza es significativamente mayor que la media, lo cual indica una sobredispersión en los datos, es decir, que la variabilidad en el número de rupturas es mayor de lo que la distribución de Poisson podría captar.

- Dado que hay sobredispersión en los datos, una Regresión de Poisson estándar podría no ser adecuada para este conjunto de datos, ya que este modelo subestimaría la varianza de los datos, lo que podría llevar a errores en las inferencias.

II. Ajusta dos modelos de Regresión Poisson

Ajustar el modelo de Regresión Poisson sin interacción

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

Resumen de Los modelos

```
S_sin_interaccion <- summary(modelo_poisson_sin_interaccion)
S_sin_interaccion
```

```
##
## 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
```

Ajustar el modelo de Regresión Poisson con interacción

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

```
S_con_interaccion <- summary(modelo_poisson_con_interaccion)
S_con_interaccion
```

```
##
## 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.

Modelo sin interacción

- Este modelo estima el número de rupturas (breaks) en función de las variables wool y tension, sin considerar la interacción entre ellas.
- **Intercepto (3.69196)** Representa el logaritmo del número esperado de rupturas cuando la lana es del tipo A y la tensión es baja (L). En términos de la tasa de rupturas, es $e^{3.69196} \approx 40.26$.
- **woolB (-0.20599)** El coeficiente indica que al cambiar de lana tipo A a tipo B, el logaritmo de la tasa de rupturas disminuye en 0.20599, manteniendo constante la tensión. En términos de la tasa de rupturas, esto representa un factor de $e^{-0.20599} \approx 0.814$, es decir, una reducción aproximada del 18.6%.
- **tensionM (-0.32132)** Al cambiar la tensión de baja (L) a media (M), la tasa de rupturas disminuye en un factor de $e^{-0.32132} \approx 0.725$, manteniendo constante el tipo de lana.
- **tensionH (-0.51849)** Al cambiar la tensión de baja (L) a alta (H), la tasa de rupturas disminuye en un factor de $e^{-0.51849} \approx 0.596$, lo que equivale a una reducción del 40.4%.
- **Modelo matemático**

$$\log(E(\text{breaks})) \\ = 3.69196 - 0.20599 \cdot \text{woolB} - 0.32132 \cdot \text{tensionM} - 0.51849 \cdot \text{tensionH}$$

Modelo con interacción

Este modelo incluye la interacción entre wool y tension, lo que permite evaluar cómo la combinación de ambas variables afecta el número de rupturas.

Intercepto (3.79674) Representa el logaritmo del número esperado de rupturas cuando la lana es del tipo A y la tensión es baja (L).

woolB (-0.45663) Al cambiar de lana tipo A a tipo B, el logaritmo de la tasa de rupturas disminuye en 0.45663 cuando la tensión es baja (L).

tensionM (-0.61868) Al cambiar la tensión de baja (L) a media (M) en lana tipo A, el logaritmo de la tasa de rupturas disminuye en 0.61868.

tensionH (-0.59580) Al cambiar la tensión de baja (L) a alta (H) en lana tipo A, el logaritmo de la tasa de rupturas disminuye en 0.59580.

woolB:tensionM (0.63818) Representa la interacción entre lana tipo B y tensión media (M). Incrementa el logaritmo de la tasa de rupturas en 0.63818 cuando ambos factores están presentes.

woolB:tensionH (0.18836) Representa la interacción entre lana tipo B y tensión alta (H). Sin embargo, su p-valor indica que no es estadísticamente significativo, por lo cual su efecto podría no ser concluyente.

Modelo matemático

$$\log(E(\text{breaks})) \\ = 3.79674 - 0.45663 \cdot \text{woolB} - 0.61868 \cdot \text{tensionM} - 0.59580 \cdot \text{tensionH} + 0.63818 \\ \cdot (\text{woolB:tensionM}) + 0.18836 \cdot (\text{woolB:tensionH})$$

Conclusión

El modelo con interacción parece capturar mejor la relación entre wool y tension debido a su menor devianza residual (182.31 vs 210.39) y menor AIC (468.97 vs 493.06), lo que indica un mejor ajuste. Las interacciones significativas sugieren que el efecto de la lana depende del nivel de tensión, lo cual es importante para la interpretación

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.

- **AIC:** Criterio de Akaike.
- **Comparación entre los coeficientes y los errores estándar de ambos modelos:**

Desviación residual (Prueba de χ^2)

- Si el modelo nulo explica a los datos, entonces la desviación nula será pequeña. Lo mismo ocurre con la desviación residual. Puesto que se supone que el modelo contiene variables significativas, lo importante es que la desviación residual del modelo sea suficientemente pequeña.
- La prueba de χ^2 mide qué tan lejana está del cero la desviación residual del modelo. Cuanto más lejos esté de cero, mejor será el modelo; cuanto más cerca, peor será el modelo para explicar la variabilidad de los datos.

El modelo plantea las siguientes hipótesis:

$H_0 \quad \{\text{Deviance}\} = 0$

$H_1 \quad \{\text{Deviance}\} > 0$

donde los grados de libertad gl se calculan como:

$$gl = gl_{desviaciónresidual} = n - (p + 1)$$

```
# Calcular el valor frontera de La zona de rechazo para el modelo sin
interacción
gl_sin_interaccion <- S_sin_interaccion$df.null -
S_sin_interaccion$df.residual

valor_frontera_sin_interaccion <- qchisq(0.05, df = gl_sin_interaccion)
cat("Valor frontera (sin interacción):", valor_frontera_sin_interaccion)

## Valor frontera (sin interacción): 0.3518463

# Calcular el valor frontera de La zona de rechazo para el modelo con
interacción
gl_con_interaccion <- S_con_interaccion$df.null -
S_con_interaccion$df.residual
valor_frontera_con_interaccion <- qchisq(0.05, df = gl_con_interaccion)
cat("Valor frontera (con interacción):", valor_frontera_con_interaccion)

## Valor frontera (con interacción): 1.145476
```

Interpretación de los valores frontera

- Debemos considerar los valores frontera calculados para la zona de rechazo en ambos modelos, sin y con interacción. Estos valores frontera se basan en la distribución chi-cuadrado y permiten evaluar si la inclusión de la interacción mejora significativamente el ajuste del modelo.

- El valor frontera para el modelo sin interacción es 0.3518463, y para el modelo con interacción es 1.145476. Esto indica que el modelo con interacción tiene un ajuste significativamente mejor, ya que su valor frontera más alto sugiere una mayor tolerancia a la devianza, capturando mejor las interacciones entre “wool” y “tension”. Además, los valores menores de AIC y devianza residual en el modelo con interacción refuerzan su preferencia frente al modelo sin interacción.

```
# Calcular estadístico de prueba y valor p para Le modelo sin interaccion
dr_sin_interaccion <- S_sin_interaccion$deviance
cat("Estadístico de prueba =", dr_sin_interaccion, "\n")

## Estadístico de prueba = 210.3919

vp_sin_interaccion <- 1 - pchisq(dr_sin_interaccion, gl_sin_interaccion)
cat("Valor p =", vp_sin_interaccion)

## Valor p = 0

# Calcular estadístico de prueba y valor p para Le modelo con interaccion
dr_con_interaccion <- S_con_interaccion$deviance
cat("Estadístico de prueba =", dr_con_interaccion, "\n")

## Estadístico de prueba = 182.3051

vp_con_interaccion = 1 - pchisq(dr_con_interaccion, gl_con_interaccion)
cat("Valor p =", vp_con_interaccion)

## Valor p = 0
```

Interpretación de los estadísticos y el valor p

- Los resultados indican que el **modelo con interacción** se ajusta mejor a los datos que el modelo sin interacción. Ambos modelos tienen un valor p de 0, lo que sugiere que aún presentan cierta devianza respecto a los datos observados; sin embargo, el menor estadístico de prueba en el modelo con interacción muestra que la inclusión de la interacción entre “wool” y “tension” mejora el ajuste, capturando mejor las relaciones que influyen en el número de rupturas.

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

```
# Comparar Los AIC de cada modelo
AIC_sin_interaccion <- AIC(modelo_poisson_sin_interaccion)
cat("AIC sin interacción:", AIC_sin_interaccion)

## AIC sin interacción: 493.056

AIC_con_interaccion <- AIC(modelo_poisson_con_interaccion)
cat("AIC con interacción:", AIC_con_interaccion)
```

```
## AIC con interacción: 468.9692
```

Interpretación de los AIC

- El modelo con interacción tiene un AIC menor (468.9692) que el modelo sin interacción (493.056), lo que indica un mejor ajuste a los datos. Esto sugiere que incluir la interacción entre “woolW y”tension” mejora la capacidad predictiva del modelo, siendo la opción preferible.

Compara los coeficientes

```
# Extraer coeficientes y errores estándar de ambos modelos
coef_sin_interaccion <- S_sin_interaccion$coefficients[, 1]
se_sin_interaccion <- S_sin_interaccion$coefficients[, 2]

coef_con_interaccion <- S_con_interaccion$coefficients[, 1]
se_con_interaccion <- S_con_interaccion$coefficients[, 2]

# Crear una tabla para comparar Los coeficientes y errores estándar

terminos <- union(rownames(S_sin_interaccion$coefficients),
rownames(S_con_interaccion$coefficients))

tabla_comparacion <- data.frame(
  Termino = terminos,
  Coeficiente_Sin_Interaccion = coef_sin_interaccion[match(terminos,
rownames(S_sin_interaccion$coefficients))],
  Coeficiente_Con_Interaccion = coef_con_interaccion[match(terminos,
rownames(S_con_interaccion$coefficients))],
  Error_Estandar_Sin_Interaccion = se_sin_interaccion[match(terminos,
rownames(S_sin_interaccion$coefficients))],
  Error_Estandar_Con_Interaccion = se_con_interaccion[match(terminos,
rownames(S_con_interaccion$coefficients))]
)

tabla_comparacion

##               Termino Coeficiente_Sin_Interaccion
## (Intercept)      (Intercept)          3.6919631
## woolB           woolB          -0.2059884
## tensionM        tensionM          -0.3213204
## tensionH        tensionH          -0.5184885
## woolB:tensionM woolB:tensionM              NA
## woolB:tensionH woolB:tensionH              NA
##               Coeficiente_Con_Interaccion
Error_Estandar_Sin_Interaccion
## (Intercept)          3.7967368
0.04541069
## woolB          -0.4566272
```



```

0.05157117
## tensionM -0.6186830
0.06026580
## tensionH -0.5957987
0.06395944
## woolB:tensionM 0.6381768
NA
## woolB:tensionH 0.1883632
NA
## Error_Estandar_Con_Interaccion
## (Intercept) 0.04993753
## woolB 0.08019202
## tensionM 0.08440012
## tensionH 0.08377723
## woolB:tensionM 0.12215312
## woolB:tensionH 0.12989529

```

Interpreta los coeficientes de ambos modelos.

- **Intercepto:** Representa el logaritmo del número esperado de rupturas cuando wool es de tipo A y tension es baja (L). El valor es ligeramente mayor en el modelo con interacción (3.7967 vs. 3.6919), indicando un ajuste algo distinto al incluir las interacciones.
- **woolB:** Cambiar de lana tipo A a tipo B reduce el logaritmo de la tasa de rupturas en ambos modelos. Sin embargo, el efecto es mayor en el modelo con interacción (-0.4566) en comparación con el modelo sin interacción (-0.2060), lo que sugiere que el tipo de lana tiene un efecto más fuerte cuando se considera la interacción con tension.
- **tensionM y tensionH:** Ambos niveles de tensión (media y alta) también tienen un impacto negativo en la tasa de rupturas. En el modelo con interacción, los efectos de tensionM (-0.6168) y tensionH (-0.5958) son mayores que en el modelo sin interacción, lo cual indica que la interacción entre wool y tension influye en el efecto de la tensión en las rupturas.

Interacciones (woolB y woolB):

- woolB (0.6382): Indica que cuando la lana es tipo B y la tensión es media, el logaritmo de la tasa de rupturas aumenta, sugiriendo un efecto combinado positivo que compensa parcialmente la reducción en los coeficientes individuales de woolB y tensionM.
- woolB (0.1884): Aunque tiene un coeficiente positivo, su valor es menor y no es significativo, indicando un efecto de interacción más débil en esta combinación.

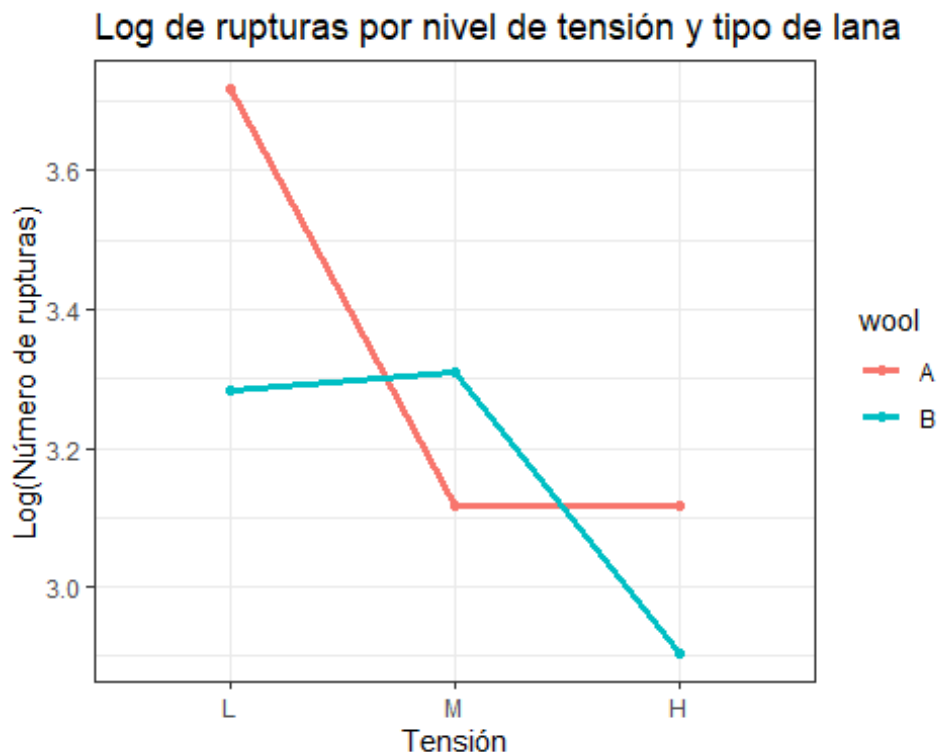
Conclusión * La tabla muestra que el modelo con interacción captura mejor la relación entre wool y tension. Al incluir la interacción, los efectos de woolB y tensionM son más fuertes, indicando que el tipo de lana y la tensión tienen un impacto mayor en

conjunto. Las interacciones woolB:tensionM y woolB:tensionH revelan efectos adicionales, aunque el impacto de woolB:tensionH es débil y no significativo. Los errores estándar son mayores en el modelo con interacción, reflejando su mayor complejidad.

```
# Crear la gráfica
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.3

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")) +
  labs(title = "Log de rupturas por nivel de tensión y tipo de lana",
       x = "Tensión", y = "Log(Número de rupturas)")
```



Interpretación

La gráfica muestra que la **lana A** tiene más rupturas a baja tensión, pero disminuyen al aumentar la tensión. En contraste, la **lana B** es más resistente a cambios de tensión baja y media, reduciendo rupturas solo en alta tensión. Esto refleja una interacción entre el tipo de lana y la tensión.

Define cuál de los dos es un mejor modelo

- Considerando todos los resultados previos (estadísticos de prueba, valores p, AIC y análisis de interacción), el **modelo con interacción** es el mejor. Este modelo no solo presenta un AIC menor, lo que indica un mejor ajuste, sino que también captura la interacción significativa entre “wool” y “tension”, reflejando mejor las variaciones en el número de rupturas. La inclusión de la interacción permite un modelo más preciso y representativo de los datos.

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 *poisgof*, que es una prueba χ^2 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:

H_0 : No hay una sobredispersión del modelo

H_1 : Hay una sobredispersión del modelo

```
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:ggplot2':
##
##     alpha

# Definir el modelo con interacción como 'pm'
pm <- modelo_poisson_con_interaccion

# Prueba de Sobredispersión con poisgof
poisgof(pm)
```

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

Interpretación

- La prueba de sobredispersión muestra un valor p muy bajo, lo que indica que el modelo de Poisson no cumple con el supuesto de que la varianza es igual a la media; hay sobredispersión en los datos.

```
library(lmtest)
```

```
## Warning: package 'lmtest' was built under R version 4.3.3
## Loading required package: zoo
## Warning: package 'zoo' was built under R version 4.3.3
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
##
## Attaching package: 'lmtest'
## The following object is masked from 'package:epiDisplay':
##
##   lrtest
```

Realizar la prueba de Durbin-Watson en el modelo de Poisson con interacción

```
dwtest(pm)
```

```
##
## Durbin-Watson test
##
## data: pm
## DW = 2.2376, p-value = 0.575
## alternative hypothesis: true autocorrelation is greater than 0
```

Interpretación

- La prueba de Durbin-Watson, sin embargo, muestra independencia de los residuos (DW = 2.2376, p = 0.575). En resumen, debido a la sobredispersión, sería recomendable considerar un modelo de Binomial Negativa.

Dado que los resultados indican sobredispersión en el modelo de Poisson y que se observó un mejor ajuste con el modelo con interacción, optaremos por definir los modelos con interacción en ambas alternativas: Quasi-Poisson y Binomial Negativa.

Dado que los resultados indican sobredispersión en el modelo de Poisson y que se observó un mejor ajuste con el modelo con interacción, optaremos por definir los modelos con interacción en ambas alternativas: Quasi-Poisson y Binomial Negativa.

```
# Modelo Quasi-Poisson con interacción
poisson.model3 <- glm(breaks ~ wool * tension, data = data, family =
quasipoisson(link = "log"))
summary_poisson_model3 <- summary(poisson.model3)
summary_poisson_model3

##
## Call:
## glm(formula = breaks ~ wool * tension, family = quasipoisson(link =
"log"),
##     data = data)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.79674    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
```

Interpretación

- El modelo Quasi-Poisson con interacción ajusta bien la sobredispersión y muestra que los factores “woolB”, “tensionM”, “tensionH” y la interacción “woolB:tensionM” son significativos, afectando el número de rupturas. La

interacción “woolB:tensionH” no es significativa. Este modelo es adecuado para capturar la relación entre “wool”, “tension” y las rupturas ajustando la varianza correctamente.

```
# Modelo Binomial Negativa con interacción
library(MASS)

bnm <- glm.nb(breaks ~ wool * tension, data = data, control =
glm.control(maxit = 1000))
summary_bnm <- summary(bnm)
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
```

Interpretación

- El modelo Binomial Negativa con interacción ajusta bien la sobredispersión. Los factores “woolB”, “tensionM”, “tensionH”, y la interacción “woolB:tensionM” son significativos, afectando el número de rupturas. La interacción “woolB:tensionH” no es significativa. Con un AIC de 405.12, este

modelo es adecuado para capturar las relaciones entre “wool”, “tension”, y las rupturas.

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

```
# Cálculo del valor frontera para la prueba de chi-cuadrado

# Quasi-Poisson
gl_poisson3 <- summary_poisson_model3$df.null -
summary_poisson_model3$df.residual
valor_frontera_poisson3 <- qchisq(0.05, df = gl_poisson3)
cat("Valor frontera (Quasi-Poisson):", valor_frontera_poisson3, "\n")

## Valor frontera (Quasi-Poisson): 1.145476

# Binomial Negativa
gl_bnm <- summary_bnm$df.null - summary_bnm$df.residual
valor_frontera_bnm <- qchisq(0.05, df = gl_bnm)
cat("Valor frontera (Binomial Negativa):", valor_frontera_bnm, "\n")

## Valor frontera (Binomial Negativa): 1.145476

# Cálculo del estadístico de prueba y valor p
# Quasi-Poisson
dr_poisson3 <- summary_poisson_model3$deviance
cat("Estadístico de prueba (Quasi-Poisson):", dr_poisson3, "\n")

## Estadístico de prueba (Quasi-Poisson): 182.3051

vp_poisson3 <- 1 - pchisq(dr_poisson3, gl_poisson3)
cat("Valor p (Quasi-Poisson):", vp_poisson3, "\n")

## Valor p (Quasi-Poisson): 0

# Binomial Negativa
dr_bnm <- summary_bnm$deviance
cat("Estadístico de prueba (Binomial Negativa):", dr_bnm, "\n")

## Estadístico de prueba (Binomial Negativa): 53.50616

vp_bnm <- 1 - pchisq(dr_bnm, gl_bnm)
cat("Valor p (Binomial Negativa):", vp_bnm, "\n")

## Valor p (Binomial Negativa): 2.647427e-10

# Comparar Los AIC de cada modelo (si está disponible)

# Quasi-Poisson (puede que no esté disponible en todos los casos)
AIC_poisson3 <- AIC(poisson.model3)
cat("AIC Quasi-Poisson:", AIC_poisson3, "\n")

## AIC Quasi-Poisson: NA
```

```

# Binomial Negativa
AIC_bnm <- AIC(bnm)
cat("AIC Binomial Negativa:", AIC_bnm, "\n")

## AIC Binomial Negativa: 405.1248

# Extraer coeficientes y errores estándar de ambos modelos
coef_poisson3 <- summary_poisson_model3$coefficients[, 1]
se_poisson3 <- summary_poisson_model3$coefficients[, 2]

coef_bnm <- summary_bnm$coefficients[, 1]
se_bnm <- summary_bnm$coefficients[, 2]

terminos <- union(rownames(summary_poisson_model3$coefficients),
rownames(summary_bnm$coefficients))
tabla_comparacion <- data.frame(
  Termino = terminos,
  Coeficiente_Quasi_Poisson = coef_poisson3[match(terminos,
rownames(summary_poisson_model3$coefficients))],
  Coeficiente_Binomial_Negativa = coef_bnm[match(terminos,
rownames(summary_bnm$coefficients))],
  Error_Estandar_Quasi_Poisson = se_poisson3[match(terminos,
rownames(summary_poisson_model3$coefficients))],
  Error_Estandar_Binomial_Negativa = se_bnm[match(terminos,
rownames(summary_bnm$coefficients))]
)

tabla_comparacion

##               Termino Coeficiente_Quasi_Poisson
## (Intercept)      (Intercept)          3.7967368
## woolB              woolB             -0.4566272
## tensionM           tensionM          -0.6186830
## tensionH           tensionH          -0.5957987
## woolB:tensionM woolB:tensionM          0.6381768
## woolB:tensionH woolB:tensionH          0.1883632
##               Coeficiente_Binomial_Negativa
Error_Estandar_Quasi_Poisson
## (Intercept)          3.7967368
0.09688254
## woolB              -0.4566272
0.15557852
## tensionM          -0.6186830
0.16374255
## tensionH          -0.5957987
0.16253410
## woolB:tensionM          0.6381768
0.23698620
## woolB:tensionH          0.1883632
0.25200659

```



```
##          Error_Estandar_Binominal_Negativa
## (Intercept)          0.1081206
## woolB              0.1575543
## tensionM           0.1597372
## tensionH           0.1594090
## woolB:tensionM      0.2273908
## woolB:tensionH      0.2316419
```

Interpretación general de los resultados

- **Prueba de Chi-cuadrado y Devianza:**

Quasi-Poisson

- La devianza del modelo Quasi-Poisson es significativamente mayor que el valor frontera, y el valor p es 0, indicando que el modelo se ajusta bien a los datos y que hay una alta variabilidad explicada por el modelo.

Binomial Negativa:

- La devianza del modelo Binomial Negativa también es significativamente mayor que el valor frontera y el valor p es muy bajo, indicando un buen ajuste del modelo y una variabilidad bien explicada.
- **Comparación de AIC**
- El modelo Binomial Negativa tiene un AIC disponible y bajo, lo que sugiere un buen ajuste comparativo en ausencia del AIC del modelo Quasi-Poisson.
- **Comparación de Coeficientes y Errores Estándar:**
- Coeficientes: Son casi idénticos entre ambos modelos para cada término, lo que indica que ambos modelos capturan de manera similar el efecto de cada variable en el número de rupturas.
- Errores estándar: Los errores estándar son ligeramente mayores en el modelo Binomial Negativa, pero las diferencias son mínimas, lo cual indica estabilidad en las estimaciones de ambos modelos.

Conclusión * Ambos modelos muestran un buen ajuste a los datos. Sin embargo, el modelo Binomial Negativa es preferible ya que maneja bien la sobredispersión, tiene un AIC bajo, y proporciona estimaciones precisas y estables.

```
# No se aplica poisgof para Binomial Negativa, pero se puede observar el
# parámetro de dispersión (theta) en el resumen de glm.nb
cat("Parámetro de dispersión (theta) para el modelo Binomial Negativa con
interacción:\n")
```

```
## Parámetro de dispersión (theta) para el modelo Binomial Negativa con
interacción:
```

```
summary(bnm)$theta
```

```
## [1] 12.08216
```

Interpretación * El valor de Theta es 12.08216, lo cual indica que el modelo Binomial Negativa maneja bien la sobredispersión presente en los datos. Un valor alto de theta muestra que el modelo ajusta adecuadamente la variabilidad adicional en los datos, lo cual es una ventaja en comparación con el modelo Poisson estándar, que asume que la varianza es igual a la media.

```
# Prueba de Durbin-Watson para el modelo Binomial Negativa con
interacción
cat("Prueba de Durbin-Watson para el modelo Binomial Negativa con
interacción:\n")
```

```
## Prueba de Durbin-Watson para el modelo Binomial Negativa con
interacción:
```

```
dwtest(bnm)
```

```
##
## Durbin-Watson test
##
## data: bnm
## DW = 2.2376, p-value = 0.575
## alternative hypothesis: true autocorrelation is greater than 0
```

Interpretación

- El estadístico de Durbin-Watson es 2.2376 y el valor p es 0.575, lo cual está cerca de 2 y tiene un valor p alto. Esto indica que no hay autocorrelación significativa en los residuos del modelo, cumpliendo con el supuesto de independencia.

Conclusión Después de revisar los supuestos, modelo Binomial Negativa con interacción es preferible porque proporciona un mejor ajuste a los datos con sobredispersión, capturando la variabilidad de manera más precisa y cumpliendo con los supuestos estadísticos necesarios.

V. Define cuál es tu mejor modelo

- Después de realizar un análisis exhaustivo de los diferentes modelos (Poisson sin interacción, Poisson con interacción, Quasi-Poisson con interacción y Binomial Negativa con interacción), la mejor opción es el modelo Binomial Negativa con interacción.
- **Ajuste a la Sobredispersión:** La sobredispersión en el dataset se maneja mejor con el modelo de Binomial Negativa, ya que su parámetro de dispersión alto (theta = 12.08216) captura la variabilidad extra.
- **Menor AIC:** El modelo de Binomial Negativa tiene un AIC de 405.12, el más bajo entre los modelos, indicando un mejor ajuste equilibrado.

- **Significancia de los Coeficientes:** Los coeficientes para “woolB”, “tensionM”, “tensionH” y la interacción woolB:tensionM son significativos, lo que demuestra que el modelo representa efectivamente los efectos de estas variables y sus interacciones en las rupturas.
- **Prueba de Chi-cuadrado y Devianza:** La devianza alta y el valor p bajo confirman un buen ajuste del modelo de Binomial Negativa.
- **Supuestos de Independencia:** La prueba de Durbin-Watson muestra ausencia de autocorrelación en los residuos, cumpliendo con el supuesto de independencia.

Conclusión

El modelo Binomial Negativa con interacción es el modelo preferible debido a su capacidad para manejar la sobredispersión, su menor AIC, la significancia de los coeficientes, y su cumplimiento de los supuestos estadísticos clave. Este modelo proporciona el mejor ajuste y una representación precisa de la relación entre wool, tension y el número de rupturas, capturando efectivamente la variabilidad y la interacción entre los factores.