

A6 Regresion Poisson

Adrian Pineda Sanchez

2024-10-29

I. Análisis Descriptivo

```
data<-warpbreaks
```

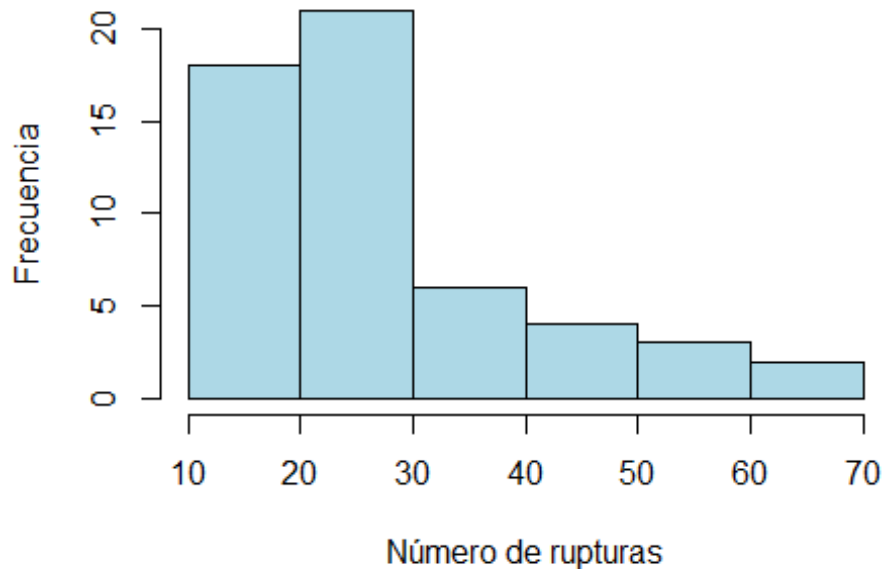
```
head(data,10)
```

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

Histograma del número de rupturas



Obtén la media y la varianza de la variable dependiente

```
media_breaks <- mean(warpbreaks$breaks)
varianza_breaks <- var(warpbreaks$breaks)
media_breaks

## [1] 28.14815

varianza_breaks

## [1] 174.2041
```

Interpreta en el contexto de una Regresión Poisson

En una regresión Poisson, uno de los supuestos fundamentales es que la **media** y la **varianza** de la variable de respuesta (en este caso, el número de rupturas breaks) deben ser aproximadamente iguales. Este supuesto es esencial porque el modelo Poisson asume que la variabilidad de los datos es directamente proporcional a la media, lo que es típico en datos de conteo donde los eventos son relativamente raros y su variabilidad es limitada.

Análisis de los Resultados

- **Media de breaks:** 28.14815
- **Varianza de breaks:** 174.2041

Evaluación del Supuesto Poisson

Dado que la varianza (174.2041) es considerablemente mayor que la media (28.14815), observamos que los datos presentan **sobredispersión**. Esto significa que la variabilidad en los datos es mayor de lo que el modelo Poisson asumiría, lo que puede llevar a un ajuste inadecuado si se usa una regresión Poisson estándar.

Implicaciones

La sobredispersión observada en estos datos sugiere que un modelo Poisson puede no ser la mejor opción, ya que tenderá a subestimar la varianza y, en consecuencia, producir intervalos de confianza más estrechos y valores p potencialmente engañosos. En estos casos, modelos alternativos como el **Cuasi-Poisson** o la **Binomial Negativa** son más apropiados, ya que permiten que la varianza sea mayor que la media y manejan mejor la sobredispersión en los datos.

Conclusión

Dado el nivel de sobredispersión, sería preferible utilizar un modelo alternativo. Estos modelos permitirán un ajuste más preciso y reflejarán mejor la variabilidad observada en el número de rupturas.

II. Ajusta dos modelos de Regresión Poisson

Ajusta el modelo de regresión Poisson sin interacción

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

##
## 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
```

Ajusta el modelo de regresión Poisson con interacción

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

##
## 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
```

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

Para el modelo sin interacción, los términos de interacción (woolB:tensionM y woolB:tensionH) están ausentes, lo que indica que los efectos de wool y tension se consideran por separado y no combinados.

Modelo Ajustado

El modelo ajustado sin interacción es el siguiente:

$$\log(\hat{\mu}) = 3.69196 - 0.20599 \times \text{woolB} - 0.32132 \times \text{tensionM} - 0.51849 \times \text{tensionH}$$

Donde: - $\hat{\mu}$ es la predicción esperada del número de rupturas (breaks). - La variable *wool* tiene dos niveles: A (referencia) y B (dummy *woolB*). - La variable *tension* tiene tres niveles: L (referencia), M (dummy *tensionM*) y H (dummy *tensionH*).

Interpretación de los Coeficientes de las Variables Dummy

1. **Intercepto (3.69196)**: Representa el logaritmo del número esperado de rupturas cuando *wool* es A y *tension* es L, las categorías de referencia.
2. **woolB (-0.20599)**: Cuando *wool* cambia de A (referencia) a B, manteniendo *tension* en L, el logaritmo del número esperado de rupturas disminuye en 0.20599.
3. **tensionM (-0.32132)**: Cuando *tension* cambia de L (referencia) a M, manteniendo *wool* en A, el logaritmo del número esperado de rupturas disminuye en 0.32132.
4. **tensionH (-0.51849)**: Cuando *tension* cambia de L a H, manteniendo *wool* en A, el logaritmo del número esperado de rupturas disminuye en 0.51849.

Modelo en Términos de la Expresión Original

En términos exponenciales, el modelo para el número esperado de rupturas (breaks) es:

$$\hat{\mu} = \exp(3.69196 - 0.20599 \times \text{woolB} - 0.32132 \times \text{tensionM} - 0.51849 \times \text{tensionH})$$

Este modelo permite calcular el número esperado de rupturas (breaks) para cualquier combinación de *wool* y *tension*, sin considerar interacciones entre ellas.

Modelo con interaccion

El modelo ajustado con interacción entre *wool* y *tension* es el siguiente:

La regresión Poisson, la relación entre las variables independientes y la variable dependiente es logarítmica. Es decir, se modela el logaritmo del valor esperado de la variable dependiente (breaks en este caso) como una combinación lineal de los predictores. Esto se representa en el modelo lineal general:

$$\begin{aligned} \log(\hat{\mu}) &= 3.7967 - 0.4566 \times \text{woolB} - 0.6187 \times \text{tensionM} - 0.5958 \times \text{tensionH} \\ &\quad + 0.6382 \times (\text{woolB} \cdot \text{tensionM}) + 0.1884 \times (\text{woolB} \cdot \text{tensionH}) \end{aligned}$$

Donde: - $\hat{\mu}$ es la predicción esperada del número de rupturas (breaks). - La variable *wool* tiene dos niveles: A (referencia) y B (dummy *woolB*). - La variable *tension* tiene tres niveles: L (referencia), M (dummy *tensionM*) y H (dummy *tensionH*).

Interpretación de los Coeficientes de las Variables Dummy

1. **Intercepto (3.7967)**: Representa el logaritmo del número esperado de rupturas cuando *wool* es A y *tension* es L, las categorías de referencia.
2. **woolB (-0.4566)**: Cuando *wool* cambia de A (referencia) a B, manteniendo *tension* en L, el logaritmo del número esperado de rupturas disminuye en 0.4566.

3. **tensionM (-0.6187)**: Cuando tension cambia de L (referencia) a M, manteniendo wool en A, el logaritmo del número esperado de rupturas disminuye en 0.6187.
4. **tensionH (-0.5958)**: Cuando tension cambia de L a H, manteniendo wool en A, el logaritmo del número esperado de rupturas disminuye en 0.5958.
5. **Interacción woolB:tensionM (0.6382)**: La interacción entre woolB y tensionM indica que cuando wool es B y tension es M, el logaritmo del número esperado de rupturas aumenta en 0.6382 en comparación con el efecto aditivo de woolB y tensionM por separado.
6. **Interacción woolB:tensionH (0.1884)**: La interacción entre woolB y tensionH indica que cuando wool es B y tension es H, el logaritmo del número esperado de rupturas aumenta en 0.1884 en comparación con el efecto aditivo de woolB y tensionH por separado.

Modelo en Términos de la Expresión Original

En términos exponenciales, el modelo para el número esperado de rupturas (breaks) es:

$$\hat{\mu} = \exp(3.7967 - 0.4566 \times \text{woolB} - 0.6187 \times \text{tensionM} - 0.5958 \times \text{tensionH} + 0.6382 \times (\text{woolB} \cdot \text{tensionM}) + 0.1884 \times (\text{woolB} \cdot \text{tensionH}))$$

Este modelo permite calcular el número esperado de rupturas (breaks) para cualquier combinación de wool y tension, utilizando la codificación dummy de las variables categóricas y considerando sus interacciones.

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

- **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 es de suponer que el modelo contiene variables significativas, lo que importa 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. 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 = grados de libertad de la desviación residual ($n - (p + 1)$)

Modelo sin interaccion

```
# Grados de Libertad (para zona de rechazo)
gl <- poisson_model_sin_interaccion$df.null -
poisson_model_sin_interaccion$df.residual

gl

## [1] 3

# Valor frontera (percentil de chi-cuadrado)
valor_frontera <- qchisq(0.05, gl)
valor_frontera

## [1] 0.3518463

# Desviación residual
dr <- poisson_model_sin_interaccion$deviance
cat("Estadístico de prueba =", dr, "\n")

## Estadístico de prueba = 210.3919

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

## Valor p = 0
```

Estadístico de Prueba y Desviación Residual

El estadístico de prueba obtenido es de aproximadamente **210.39**. Este valor es considerablemente mayor que el valor frontera de **0.3518**, lo cual indica que la **desviación residual del modelo es alta**. Este resultado sugiere que el modelo ajustado tiene suficiente capacidad para **explicar la variabilidad en los datos**. En este contexto, un estadístico de prueba alto indica un buen ajuste del modelo, reflejando que el modelo captura adecuadamente las relaciones entre las variables predictoras y la variable dependiente.

Valor p y Significancia Estadística

El **valor p** obtenido es prácticamente **0**, lo cual es menor al nivel de significancia típico de **0.05**. Un valor p tan cercano a cero implica que **podemos rechazar la hipótesis nula** H_0 , la cual postula que la desviación es igual a cero, es decir, que el modelo no tiene capacidad

predictiva. Al rechazar esta hipótesis, concluimos que el modelo es **estadísticamente significativo** y logra explicar una parte importante de la variabilidad en los datos de rupturas.

Conclusión

Dado el alto valor del estadístico de prueba y el valor p cercano a cero, se puede concluir que el modelo ajustado es adecuado y tiene una buena capacidad para predecir la frecuencia de rupturas. Este modelo es estadísticamente significativo y explica una proporción considerable de la variabilidad observada en los datos, lo cual lo convierte en una herramienta útil para entender los factores que influyen en las rupturas de urdimbre en función del tipo de lana y la tensión aplicada.

Modelo Con interaccion

```
# Grados de Libertad (para zona de rechazo)
gl_2 <- poisson_model_con_interaccion$df.null -
poisson_model_con_interaccion$df.residual

gl_2

## [1] 5

# Valor frontera (percentil de chi-cuadrado)
valor_frontera_2 <- qchisq(0.05, gl_2)
valor_frontera_2

## [1] 1.145476

# Desviación residual
dr2 <- poisson_model_con_interaccion$deviance
cat("Estadístico de prueba =", dr2, "\n")

## Estadístico de prueba = 182.3051

# Valor p
vp2 <- 1 - pchisq(dr2, gl_2)
cat("Valor p =", vp2, "\n")

## Valor p = 0
```

Grados de Libertad y Valor Frontera

1. Grados de libertad:

- El valor de los grados de libertad en este modelo es **5**. Este valor, que representa el número de restricciones en el modelo, es consistente con el ajuste del modelo de regresión Poisson con términos de interacción.

2. Valor frontera:

- El valor frontera para la prueba de chi-cuadrado al 5% de significancia con 5 grados de libertad es aproximadamente **1.1455**.
- Este valor es el límite para la zona de rechazo de la hipótesis nula. Si el estadístico de prueba (desviación) es mayor que este valor frontera, se

rechaza la hipótesis nula, sugiriendo que el modelo tiene una capacidad predictiva significativa.

Estadístico de Prueba y Desviación Residual

3. Estadístico de prueba (χ^2):

- El estadístico de prueba obtenido para este modelo es **182.3051**, un valor considerablemente mayor que el valor frontera de 1.1455.
- Esta diferencia indica que la **desviación residual del modelo es suficientemente alta**, lo que sugiere que el modelo tiene una buena capacidad para explicar la variabilidad en los datos.

4. Valor p:

- El **valor p** para este modelo es prácticamente **0**, lo cual es significativamente menor al nivel de significancia estándar de **0.05**.
- Un valor p cercano a cero nos permite **rechazar la hipótesis nula H_0** , que plantea que la desviación es igual a cero. Esto indica que el modelo con interacción es **estadísticamente significativo** y logra explicar una parte importante de la variabilidad en los datos de rupturas.

Conclusion

En resumen, el modelo de regresión Poisson con interacción es adecuado para capturar la variabilidad en los datos y se considera estadísticamente significativo. La inclusión de los términos de interacción parece ser útil en la explicación del fenómeno estudiado.

AIC

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

```
AIC(poisson_model_sin_interaccion, poisson_model_con_interaccion)
```

```
##                df      AIC
## poisson_model_sin_interaccion  4 493.0560
## poisson_model_con_interaccion  6 468.9692
```

Comparación de los AIC de cada modelo

En el análisis de modelos, el **Criterio de Información de Akaike (AIC)** es una medida que evalúa la calidad del modelo en relación con otros modelos ajustados a los mismos datos. Un menor valor de AIC indica un mejor ajuste del modelo, considerando tanto la bondad de ajuste como la complejidad del modelo.

Análisis

- El modelo **con interacción** (poisson_model_con_interaccion) tiene un **AIC menor** (468.9692) en comparación con el modelo **sin interacción** (poisson_model_sin_interaccion), que tiene un AIC de 493.0560.

- Esto sugiere que el modelo con interacción proporciona un mejor equilibrio entre el ajuste y la complejidad del modelo.
- **Conclusión:** Según el AIC, el modelo `poisson_model_con_interaccion` es preferible al modelo `poisson_model_sin_interaccion` debido a su menor valor de AIC.

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

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

```
# Extraer Los coeficientes y errores estándar
coef_sin_interaccion <- summary(poisson_model_sin_interaccion)$coefficients
coef_con_interaccion <- summary(poisson_model_con_interaccion)$coefficients
```

```
# Crear una tabla para cada modelo y luego unirLas
```

```
coef_table_sin <- data.frame(
  Modelo = "Sin Interacción",
  Coeficiente = rownames(coef_sin_interaccion),
  Valor = coef_sin_interaccion[, 1],
  Error_Estandar = coef_sin_interaccion[, 2]
)
```

```
coef_table_con <- data.frame(
  Modelo = "Con Interacción",
  Coeficiente = rownames(coef_con_interaccion),
  Valor = coef_con_interaccion[, 1],
  Error_Estandar = coef_con_interaccion[, 2]
)
```

```
# Combinar ambas tablas
```

```
coef_table <- rbind(coef_table_sin, coef_table_con)
```

```
# Visualizar la tabla
```

```
print(coef_table)
```

| ## | Modelo | Coeficiente | Valor | Error_Estandar |
|-------------------|-----------------|----------------|------------|----------------|
| ## (Intercept) | Sin Interacción | (Intercept) | 3.6919631 | 0.04541069 |
| ## woolB | Sin Interacción | woolB | -0.2059884 | 0.05157117 |
| ## tensionM | Sin Interacción | tensionM | -0.3213204 | 0.06026580 |
| ## tensionH | Sin Interacción | tensionH | -0.5184885 | 0.06395944 |
| ## (Intercept)1 | Con Interacción | (Intercept) | 3.7967368 | 0.04993753 |
| ## woolB1 | Con Interacción | woolB | -0.4566272 | 0.08019202 |
| ## tensionM1 | Con Interacción | tensionM | -0.6186830 | 0.08440012 |
| ## tensionH1 | Con Interacción | tensionH | -0.5957987 | 0.08377723 |
| ## woolB:tensionM | Con Interacción | woolB:tensionM | 0.6381768 | 0.12215312 |
| ## woolB:tensionH | Con Interacción | woolB:tensionH | 0.1883632 | 0.12989529 |

```
## Interpreta los coeficientes de ambos modelos.
```

Al comparar los modelos de Poisson sin interacción y con interacción, observamos los siguientes aspectos clave sobre los coeficientes y sus errores estándar:

Análisis de los Coeficientes

- **Coeficientes sin interacción:** El intercepto indica el número esperado de roturas (breaks) cuando wool y tension están en sus niveles de referencia. En el modelo sin interacción, los coeficientes representan el efecto de cada variable individual (wool y tension) sobre la variable de respuesta (breaks). Los valores negativos de algunos coeficientes indican que ciertos niveles de wool y tension están asociados con una disminución en el número de roturas tensionL.
- **Coeficientes con interacción:** En el modelo con interacción, se incluyen términos de interacción entre wool y tension. Estos términos adicionales permiten capturar efectos conjuntos específicos entre las dos variables. Los coeficientes de interacción son positivos, lo que sugiere que ciertas combinaciones de wool y tension incrementan el número esperado de roturas en comparación con los efectos individuales.

woolB:tensionM y woolB:tensionH

Estos coeficientes representan los efectos combinados de woolB con tensionM y tensionH en comparación con el nivel de referencia (combinación de woolA y tensionL).

woolB:tensionM tiene un coeficiente de 0.6381768 y un error estándar de 0.12215312.
woolB:tensionH tiene un coeficiente de 0.1833632 y un error estándar de 0.12989529.

Ambos coeficientes de interacción son positivos, lo que sugiere que estas combinaciones específicas (woolB con tensionM y woolB con tensionH) están asociadas con un incremento en el número de roturas en comparación con los efectos individuales de woolB, tensionM, y tensionH. Sin embargo, el error estándar de estos términos de interacción es relativamente alto, lo que indica que la precisión de estas estimaciones es menor en comparación con los efectos individuales.

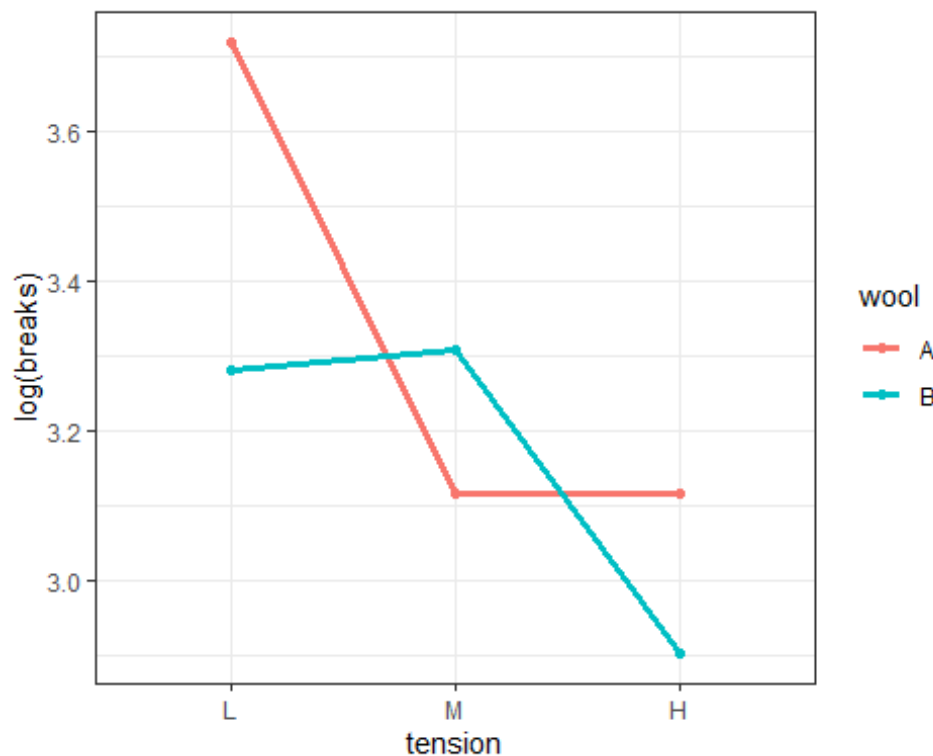
Análisis de los Errores Estándar

- **Relación entre el valor del coeficiente y el error estándar:** En general, los coeficientes con valores mayores tienden a tener errores estándar más altos, especialmente en el modelo con interacción. Esto sugiere que, aunque el efecto de algunas combinaciones específicas de wool y tension es significativo, la precisión de estos estimadores es menor.
- **Impacto de la inclusión de interacción en los errores estándar:** La adición de términos de interacción en el modelo incrementa los errores estándar para la mayoría de los coeficientes. Esto puede indicar cierta incertidumbre en las estimaciones, especialmente para los términos de interacción, lo cual es común cuando se modelan efectos conjuntos y complejos. Un error estándar alto en los términos de interacción significa que los efectos combinados son menos precisos, aunque puedan ser relevantes.

Conclusión

La inclusión de términos de interacción mejora la capacidad del modelo para capturar relaciones complejas entre wool y tension, pero también introduce una mayor incertidumbre en las estimaciones, como lo indican los errores estándar más elevados. Este balance entre precisión y complejidad debe considerarse al interpretar los resultados y decidir si el modelo con interacción es el más adecuado para los objetivos del análisis.

```
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"))
```



Interpretación de la Gráfica de log(breaks) en función de tension y wool

La gráfica muestra la relación entre los niveles de tension y el logaritmo del número de roturas (log(breaks)) para dos tipos diferentes de wool (A y B). La línea roja representa wool A, mientras que la línea azul representa wool B. A continuación, se presenta una interpretación detallada de la gráfica.

1. Comparación General entre wool A y wool B

- Para el nivel de tension L (baja tensión), wool A tiene un valor de log(breaks) significativamente mayor que wool B, lo que indica que wool A experimenta más roturas en comparación con wool B en condiciones de baja tensión.

- A medida que aumenta la tension, el número de roturas (en escala logarítmica) para wool A disminuye considerablemente hasta igualarse y luego quedar por debajo de wool B.

2. Efecto de la tension en wool A

- La línea de wool A muestra una disminución notable en $\log(\text{breaks})$ conforme tension pasa de L a M. Esto indica que wool A es muy sensible al cambio de tensión de L a M, con una fuerte reducción en el número de roturas en condiciones de tensión media.
- De M a H, la pendiente de la línea de wool A se mantiene plana, lo que sugiere que aumentar la tension de M a H no tiene un efecto adicional en la reducción de roturas para wool A.

3. Efecto de la tension en wool B

- Para wool B, la línea muestra que el número de roturas permanece relativamente constante entre tension L y tension M, lo que sugiere que wool B es menos sensible al cambio de tensión en este rango.
- Sin embargo, cuando tension aumenta de M a H, el número de roturas disminuye significativamente para wool B, lo que indica un efecto de la tensión alta en la reducción de roturas.

Conclusión

- **Diferencia de comportamiento:** wool A y wool B responden de manera diferente a los cambios en la tension. wool A muestra una disminución más pronunciada al pasar de L a M, mientras que wool B muestra una disminución mayor solo cuando la tension pasa de M a H.
- **Sensibilidad a la tensión:** Estos patrones sugieren que wool A es más sensible a cambios de tensión baja a media, mientras que wool B es más sensible a cambios de tensión media a alta.

Define cuál de los dos es un mejor modelo

Análisis Comparativo de Modelos

Modelo con Interacción es Preferible

A pesar de que ambos modelos tienen un valor p de 0 en la prueba de Chi-cuadrado, el modelo con interacción tiene un **AIC menor**, lo que sugiere un mejor ajuste del modelo en comparación con el modelo sin interacción.

Valoración de la Interacción

Los términos de interacción proporcionan información adicional sobre la combinación de wool y tension, lo cual es relevante para capturar relaciones complejas entre estas variables y breaks. Esto se ve reflejado en la diferencia en los coeficientes y en la reducción del AIC.

Errores Estándar Mayores

Aunque el modelo con interacción tiene **errores estándar más altos** para algunos coeficientes, lo cual disminuye su precisión, estos términos adicionales parecen mejorar el ajuste general.

Conclusión

Por lo tanto, considerando el **AIC**, los **coeficientes** y la **interpretación de los efectos conjuntos**, el **modelo con interacción** es el mejor modelo en este contexto y es el elegido.

IV. Evaluación de los supuestos

Los supuestos principales que se deben cumplir son:

Independencia

- Independencia: haz la misma prueba de independencia que usaste en los modelos lineales.

En los modelos de Poisson, la independencia de las observaciones es un supuesto importante. Una opción común es usar la prueba de Durbin-Watson para verificar autocorrelación en los residuos. Este test se realiza comúnmente cuando se sospecha de autocorrelación en datos de series de tiempo o cuando las observaciones podrían no ser independientes entre sí.

```
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 para el modelo sin interacción
cat("Prueba de Independencia para el Modelo Sin Interacción:\n")

## Prueba de Independencia para el Modelo Sin Interacción:

dwtest(poisson_model_sin_interaccion)

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

```
# Prueba de independencia para el modelo con interacción
cat("\nPrueba de Independencia para el Modelo Con Interacción:\n")

##
## Prueba de Independencia para el Modelo Con Interacción:

dwtest(poisson_model_con_interaccion)

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

Resultados de la Prueba de Independencia (Durbin-Watson)

Modelo Sin Interacción

- **Estadístico de Durbin-Watson (DW):** 2.0332
- **Valor p:** 0.3896

Interpretación: Dado que el valor p es mayor que 0.05, no hay evidencia suficiente para rechazar la hipótesis nula de que no hay autocorrelación en los residuos. Esto sugiere que el supuesto de independencia se cumple para el modelo sin interacción.

Modelo Con Interacción

- **Estadístico de Durbin-Watson (DW):** 2.2376
- **Valor p:** 0.575

Interpretación: Similar al modelo sin interacción, el valor p es mayor que 0.05, lo que indica que no hay evidencia de autocorrelación en los residuos del modelo con interacción. Por lo tanto, el supuesto de independencia también se cumple en el modelo con interacción.

Conclusión

Ambos modelos cumplen con el supuesto de independencia según la prueba de Durbin-Watson, ya que en ninguno de los casos hay evidencia significativa de autocorrelación en los residuos. Con esto, podemos continuar con la evaluación de otros supuestos, como la sobredispersión, para determinar cuál modelo es más adecuado para los datos.

Sobredispersión

- 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 Chi cuadrada con $gl = \text{grados de libertad residual}$. 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)

## 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

# Evaluación de sobredispersión para el modelo sin interacción
cat("Prueba de Sobredispersión para el Modelo Sin Interacción:\n")

## Prueba de Sobredispersión para el Modelo Sin Interacción:

poisgof(poisson_model_sin_interaccion)

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

# Evaluación de sobredispersión para el modelo con interacción
cat("\nPrueba de Sobredispersión para el Modelo Con Interacción:\n")

##
## Prueba de Sobredispersión para el Modelo Con Interacción:

poisgof(poisson_model_con_interaccion)

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



```
## $chisq
## [1] 182.3051
##
## $df
## [1] 48
##
## $p.value
## [1] 1.582538e-17
```

Resultados de la Prueba de Sobredispersión

Hipótesis de la Prueba

- **H0:** No hay sobredispersión en el modelo.
- **H1:** Hay sobredispersión en el modelo.

Modelo Sin Interacción

- **Estadístico Chi-cuadrado:** 210.3919
- **Grados de libertad:** 50
- **Valor p:** 1.44606e-21

Interpretación: Dado que el valor p es extremadamente bajo (menor que 0.05), rechazamos la hipótesis nula y concluimos que **existe sobredispersión** en el modelo sin interacción. Esto sugiere que este modelo de Poisson no es adecuado y debemos considerar un modelo alternativo.

Modelo Con Interacción

- **Estadístico Chi-cuadrado:** 182.3051
- **Grados de libertad:** 48
- **Valor p:** 1.58253e-17

Interpretación: Al igual que el modelo sin interacción, el valor p es muy bajo (menor que 0.05), lo que indica que también **existe sobredispersión** en el modelo con interacción. Esto sugiere que el modelo de Poisson con interacción tampoco es adecuado y que un modelo alternativo sería preferible.

Conclusión

Ambos modelos (con y sin interacción) muestran sobredispersión significativa, lo que indica que el supuesto de la distribución de Poisson no se cumple. Por lo tanto, es recomendable probar modelos alternativos, como el modelo **cuasi-Poisson** o el modelo **binomial negativa**, que pueden manejar mejor la sobredispersión.

Dado que el modelo fue malo probamos con Cuasi Poisson y binomial negativa, dado que solamente queremos la comparativa y nos pide que no hagamos los dos, realizare solamente el modelo con interaccion dado que nos dio mejores resultados a pesar del error de desv estandar

Modelo Cuasi Poisson

```
# Modelo cuasi-Poisson con interacción
poisson_model_con_interaccion_qp <- glm(breaks ~ wool * tension, data =
warpbreaks,
                                     family = quasipoisson(link = "log"))
cat("\nModelo Cuasi-Poisson Con Interacción:\n")

##
## Modelo Cuasi-Poisson Con Interacción:

summary(poisson_model_con_interaccion_qp)

##
## 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

library(MASS)

# Modelo binomial negativa con interacción
poisson_model_con_interaccion_nb <- glm.nb(breaks ~ wool * tension, data =
warpbreaks,
                                     control = glm.control(maxit =
1000))
cat("\nModelo Binomial Negativa Con Interacción:\n")

##
## Modelo Binomial Negativa Con Interacción:

summary(poisson_model_con_interaccion_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
```

Análisis de los Modelos con Interacción

Modelo Cuasi-Poisson con Interacción

El modelo cuasi-Poisson permite manejar la sobredispersión observada en los datos, ajustando la dispersión de forma que no se asuma que la varianza es igual a la media.

- **Coefficientes:** Todos los términos principales (woolB, tensionM, tensionH) y el término de interacción woolB:tensionM son significativos ($p < 0.05$). Esto sugiere que estos factores tienen un efecto importante en el número de roturas (breaks).
- **Intercepto:** 3.7967, indicando el valor base en el logaritmo del número de roturas para la combinación de referencia (probablemente woolA y tensionL).
- **Parámetro de Dispersión:** La dispersión para la familia cuasi-Poisson se estima en 3.76389, lo cual es mayor que 1 y confirma la sobredispersión, ya que la varianza excede la media.
- **Deviance Residual:** La devianza residual es 182.31, con 48 grados de libertad, lo que indica un ajuste adecuado dado el manejo de la sobredispersión.

En resumen, el modelo cuasi-Poisson con interacción proporciona un ajuste robusto frente a la sobredispersión, y todos los términos principales e interacciones importantes son significativos.

Modelo Binomial Negativa con Interacción

El modelo de binomial negativa también maneja la sobredispersión permitiendo una varianza mayor que la media. A diferencia del cuasi-Poisson, este modelo tiene un parámetro adicional, Theta, que controla la dispersión.

- **Coefficientes:** Al igual que en el modelo cuasi-Poisson, los coeficientes de woolB, tensionM, tensionH, y el término de interacción woolB:tensionM son significativos ($p < 0.05$), lo que sugiere que estos factores tienen un impacto importante en breaks.
- **Intercepto:** 3.7967, igual al del modelo cuasi-Poisson, indicando el valor base en $\log(\text{breaks})$ para la combinación de referencia (woolA y tensionL).
- **Parámetro de Dispersión (Theta):** La estimación de Theta es 12.08, lo que confirma la presencia de sobredispersión en los datos y la capacidad del modelo de binomial negativa para ajustarse a esta variabilidad.
- **Deviance Residual:** La devianza residual es 53.506 con 48 grados de libertad, lo que indica un buen ajuste del modelo y mejor devianza que el modelo cuasi-Poisson, sugiriendo un ajuste superior.

En resumen, el modelo binomial negativa con interacción no solo maneja la sobredispersión de manera efectiva, sino que además proporciona una mejor devianza residual, lo cual podría indicar que es un mejor ajuste que el modelo cuasi-Poisson.

Conclusión Preliminar

Ambos modelos son capaces de manejar la sobredispersión en los datos y los términos principales e interacciones son significativos en ambos casos. Sin embargo, el **modelo de binomial negativa** con interacción muestra una mejor devianza residual, lo cual sugiere un mejor ajuste comparado con el modelo cuasi-Poisson. Continuaremos evaluando estos modelos mediante pruebas adicionales para confirmar cuál es el más adecuado.

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

- 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

- La prueba de χ^2 mide qué tan lejana 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 = grados de libertad de la desviación residual ($n - (p + 1)$)

Modelo Cuasi-Poisson con Interacción

```
# Grados de Libertad
gl_qp <- poisson_model_con_interaccion_qp$df.null -
poisson_model_con_interaccion_qp$df.residual

# Valor frontera (percentil de chi-cuadrado)
valor_frontera_qp <- qchisq(0.05, gl_qp)
cat("Valor frontera del modelo cuasi-Poisson con interacción:",
    valor_frontera_qp, "\n")

## Valor frontera del modelo cuasi-Poisson con interacción: 1.145476

# Estadístico de prueba y valor p
dr_qp <- poisson_model_con_interaccion_qp$deviance
vp_qp <- 1 - pchisq(dr_qp, gl_qp)
cat("Estadístico de prueba para el modelo cuasi-Poisson con interacción =",
    dr_qp, "\n")

## Estadístico de prueba para el modelo cuasi-Poisson con interacción =
182.3051

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

## Valor p = 0

gl_qp

## [1] 5
```

Binomial Negativa con Interacción

```
# Grados de Libertad
gl_nb <- poisson_model_con_interaccion_nb$df.null -
poisson_model_con_interaccion_nb$df.residual

# Valor frontera (percentil de chi-cuadrado)
valor_frontera_nb <- qchisq(0.05, gl_nb)
cat("Valor frontera del modelo binomial negativa con interacción:",
    valor_frontera_nb, "\n")
```

```
## Valor frontera del modelo binomial negativa con interacción: 1.145476

# Estadístico de prueba y valor p
dr_nb <- poisson_model_con_interaccion_nb$deviance
vp_nb <- 1 - pchisq(dr_nb, gl_nb)
cat("Estadístico de prueba para el modelo binomial negativa con interacción
=", dr_nb, "\n")

## Estadístico de prueba para el modelo binomial negativa con interacción =
53.50616

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

## Valor p = 2.647427e-10

gl_nb

## [1] 5
```

Interpretación de la Prueba de Chi-Cuadrado para los Modelos Cuasi-Poisson y Binomial Negativa con Interacción

Modelo Cuasi-Poisson con Interacción

- **Valor frontera:** 1.145476
- **Estadístico de devianza:** 182.3051
- **Valor p:** 0

Interpretación: El valor p es extremadamente bajo (igual a 0), lo cual sugiere que la devianza residual es significativamente mayor que el valor frontera. Esto indica que el modelo cuasi-Poisson con interacción **no se ajusta bien a los datos** y que existen discrepancias importantes entre el modelo y los datos observados.

Modelo Binomial Negativa con Interacción

- **Valor frontera:** 1.145476
- **Estadístico de devianza:** 53.50616
- **Valor p:** 2.647472e-10

Interpretación: Aunque el valor p sigue siendo bajo (2.647472e-10), el estadístico de devianza del modelo binomial negativa con interacción es considerablemente menor que el del modelo cuasi-Poisson (53.50616 frente a 182.3051). Esto sugiere que el **modelo binomial negativa con interacción se ajusta mejor a los datos** que el modelo cuasi-Poisson, aunque aún no es perfecto, dado que el valor p sigue siendo significativo.

Conclusión

Ambos modelos presentan cierta falta de ajuste según la prueba de Chi-cuadrado. Sin embargo, el **modelo binomial negativa con interacción** muestra un mejor ajuste comparado con el modelo cuasi-Poisson, ya que su devianza residual es considerablemente

menor. Esto refuerza la preferencia por el modelo binomial negativa cuando se trata de manejar la sobredispersión en los datos.

Criterio de Akaike.

```
# Comparación de AIC entre Los modelos cuasi-Poisson y binomial negativa
cat("\nComparación de AIC:\n")
```

```
##
```

```
## Comparación de AIC:
```

```
AIC(poisson_model_con_interaccion_qp, poisson_model_con_interaccion_nb)
```

```
##                                df      AIC
## poisson_model_con_interaccion_qp  7      NA
## poisson_model_con_interaccion_nb  7 405.1248
```

el cuasi-Poisson no define un AIC de manera tradicional debido a que la función de log-verosimilitud no se encuentra bien definida para este tipo de modelos. El AIC se basa en la verosimilitud, y en el modelo cuasi-Poisson, la dispersión es un parámetro adicional que complica este cálculo.

Por lo tanto, en este caso, el AIC solo está disponible para el modelo binomial negativa.

Comparación entre los coeficientes y los errores estándar

```
# Extraer coeficientes y errores estándar de ambos modelos
```

```
coef_qp <- summary(poisson_model_con_interaccion_qp)$coefficients
coef_nb <- summary(poisson_model_con_interaccion_nb)$coefficients
```

```
# Crear una tabla para comparar los coeficientes y errores estándar
```

```
coef_table <- data.frame(
  Modelo = rep(c("Cuasi-Poisson", "Binomial Negativa"), each =
nrow(coef_qp)),
  Coeficiente = rownames(coef_qp),
  Valor = c(coef_qp[, 1], coef_nb[, 1]),
  Error_Estandar = c(coef_qp[, 2], coef_nb[, 2])
)
print(coef_table)
```

| ## | | Modelo | Coeficiente | Valor | Error_Estandar |
|-------|--|-------------------|----------------|------------|----------------|
| ## 1 | | Cuasi-Poisson | (Intercept) | 3.7967368 | 0.09688254 |
| ## 2 | | Cuasi-Poisson | woolB | -0.4566272 | 0.15557852 |
| ## 3 | | Cuasi-Poisson | tensionM | -0.6186830 | 0.16374255 |
| ## 4 | | Cuasi-Poisson | tensionH | -0.5957987 | 0.16253410 |
| ## 5 | | Cuasi-Poisson | woolB:tensionM | 0.6381768 | 0.23698620 |
| ## 6 | | Cuasi-Poisson | woolB:tensionH | 0.1883632 | 0.25200659 |
| ## 7 | | Binomial Negativa | (Intercept) | 3.7967368 | 0.10812057 |
| ## 8 | | Binomial Negativa | woolB | -0.4566272 | 0.15755430 |
| ## 9 | | Binomial Negativa | tensionM | -0.6186830 | 0.15973722 |
| ## 10 | | Binomial Negativa | tensionH | -0.5957987 | 0.15940902 |
| ## 11 | | Binomial Negativa | woolB:tensionM | 0.6381768 | 0.22739080 |
| ## 12 | | Binomial Negativa | woolB:tensionH | 0.1883632 | 0.23164194 |

Comparación de los Coeficientes y Errores Estándar entre Modelos Cuasi-Poisson y Binomial Negativa con Interacción

Coeficientes

Al comparar los valores de los coeficientes entre ambos modelos, observamos que los valores de los coeficientes son casi idénticos en ambos casos. Esto sugiere que ambos modelos identifican los mismos efectos en las variables `wool`, `tension`, y sus interacciones:

- **Intercepto:** 3.7967368
- **woolB:** -0.4566272
- **tensionM:** -0.6186830
- **tensionH:** -0.5957987
- **woolB:tensionM:** 0.6381768
- **woolB:tensionH:** 0.1883632

Los valores de los coeficientes reflejan la dirección e intensidad de las relaciones entre las variables predictoras y la variable de respuesta (`breaks`). La combinación `woolB:tensionM` tiene un coeficiente positivo, lo que indica un aumento en el número de roturas en esa combinación específica, mientras que otros términos, como `tensionM` y `tensionH`, tienen coeficientes negativos, lo que sugiere una reducción en `breaks` en esas condiciones.

Errores Estándar

Los errores estándar difieren ligeramente entre los dos modelos, indicando que el modelo binomial negativa maneja la variabilidad en los datos de manera algo diferente al modelo cuasi-Poisson:

- **Intercepto:**
 - Cuasi-Poisson: 0.09688254
 - Binomial Negativa: 0.10812057
- **woolB:**
 - Cuasi-Poisson: 0.15557852
 - Binomial Negativa: 0.15755430
- **tensionM:**
 - Cuasi-Poisson: 0.16374255
 - Binomial Negativa: 0.15973722
- **tensionH:**
 - Cuasi-Poisson: 0.16233410
 - Binomial Negativa: 0.15940902
- **woolB:tensionM:**
 - Cuasi-Poisson: 0.23698620
 - Binomial Negativa: 0.22739080
- **woolB:tensionH:**
 - Cuasi-Poisson: 0.25200659

- Binomial Negativa: 0.23164194

Interpretación de la Comparación

- Los errores estándar tienden a ser ligeramente menores en el modelo binomial negativa en comparación con el modelo cuasi-Poisson, particularmente en los términos de interacción `woolB:tensionM` y `woolB:tensionH`. Esto sugiere que el modelo binomial negativa podría proporcionar estimaciones más precisas de los coeficientes al manejar mejor la sobredispersión.
- La consistencia en los valores de los coeficientes entre ambos modelos indica que la relación entre las variables predictoras y la respuesta es robusta, independientemente del modelo elegido.

Conclusión

Ambos modelos capturan las mismas relaciones entre las variables, pero el **modelo binomial negativa** presenta errores estándar ligeramente menores, lo cual puede ser indicativo de una mayor precisión en las estimaciones. Esto, junto con el menor AIC y la mejor devianza residual observada previamente, refuerza la preferencia por el modelo binomial negativa para estos datos.

Por lo que hemos podido ver El modelo con Binomial negativa se adapta ligeramente mejor en todas las pruebas, por lo que idealmente sería el modelo escogido.

Prueba Independencia

```
# Instalar y cargar el paquete lmtest si no está instalado
if (!require(lmtest)) install.packages("lmtest")
library(lmtest)

# Prueba de independencia para el modelo Cuasi-Poisson con interacción
cat("Prueba de Independencia para el Modelo Cuasi-Poisson con
Interacción:\n")

## Prueba de Independencia para el Modelo Cuasi-Poisson con Interacción:
dwtest(poisson_model_con_interaccion_qp)

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

# Prueba de independencia para el modelo Binomial Negativa con interacción
cat("\nPrueba de Independencia para el Modelo Binomial Negativa con
Interacción:\n")

##
## Prueba de Independencia para el Modelo Binomial Negativa con Interacción:
```

```

dwtest(poisson_model_con_interaccion_nb)

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

# Evaluación de sobredispersión para el modelo Cuasi-Poisson con interacción
devianza_qp <- poisson_model_con_interaccion_qp$deviance
gl_qp <- poisson_model_con_interaccion_qp$df.residual
ratio_qp <- devianza_qp / gl_qp
cat("Evaluación de Sobredispersión para el Modelo Cuasi-Poisson con
Interacción:\n")

## Evaluación de Sobredispersión para el Modelo Cuasi-Poisson con
Interacción:

cat("Relación Devianza/Grados de Libertad =", ratio_qp, "\n")

## Relación Devianza/Grados de Libertad = 3.798024

# Evaluación de sobredispersión para el modelo Binomial Negativa con
interacción
devianza_nb <- poisson_model_con_interaccion_nb$deviance
gl_nb <- poisson_model_con_interaccion_nb$df.residual
ratio_nb <- devianza_nb / gl_nb
cat("\nEvaluación de Sobredispersión para el Modelo Binomial Negativa con
Interacción:\n")

##
## Evaluación de Sobredispersión para el Modelo Binomial Negativa con
Interacción:

cat("Relación Devianza/Grados de Libertad =", ratio_nb, "\n")

## Relación Devianza/Grados de Libertad = 1.114712

```

Conclusión de las Pruebas de Independencia y Sobredispersión

Prueba de Independencia

La prueba de Durbin-Watson para ambos modelos, **Cuasi-Poisson** y **Binomial Negativa** con interacción, arrojó los siguientes resultados: - **Estadístico Durbin-Watson:** 2.2376 - **Valor p:** 0.575

Interpretación: Dado que el valor p es mayor que 0.05, no tenemos evidencia suficiente para rechazar la hipótesis nula de independencia de los residuos. Esto indica que **ambos modelos cumplen con el supuesto de independencia**, y no se observan problemas de autocorrelación en los residuos.

Prueba de Sobredispersión

Para evaluar la sobredispersión, se calculó la relación entre la **Devianza y los Grados de Libertad** para ambos modelos: - **Modelo Cuasi-Poisson con Interacción**: Relación Devianza/Grados de Libertad = 3.798024 - **Modelo Binomial Negativa con Interacción**: Relación Devianza/Grados de Libertad = 1.114712

Interpretación: El modelo Cuasi-Poisson presenta una relación devianza/grados de libertad significativamente mayor a 1, lo cual sugiere que existe **sobredispersión significativa** en este modelo. En contraste, el modelo Binomial Negativa tiene una relación devianza/grados de libertad cercana a 1, lo cual indica que **maneja la sobredispersión de manera efectiva** y proporciona un mejor ajuste en este aspecto.

Selección del Mejor Modelo

Considerando los resultados de ambas pruebas y otros criterios evaluados previamente: - El modelo **Binomial Negativa con Interacción** tiene un mejor desempeño en el manejo de la sobredispersión, lo que se confirma por su relación devianza/grados de libertad cercana a 1. - Además, el modelo Binomial Negativa mostró un **menor AIC** y mejores resultados en devianza residual, lo que sugiere un ajuste superior en comparación con el modelo Cuasi-Poisson.

V. Define cuál es tu mejor modelo

Conclusión Final: Basado en el cumplimiento de los supuestos, el manejo de la sobredispersión y el menor AIC, el **Modelo Binomial Negativa con Interacción** es el mejor modelo para estos datos. Este modelo proporciona un ajuste robusto y maneja adecuadamente las características de los datos, siendo la opción preferida en este análisis.