

A6_Regresión_Poisson

Rodolfo Jesús Cruz Rebollar

2024-11-04

```
# Leer la base de datos de warpbreaks
```

```
datos = warpbreaks
```

```
head(datos)
```

breaks	wool	tension
26	A	L
30	A	L
54	A	L
25	A	L
70	A	L
52	A	L

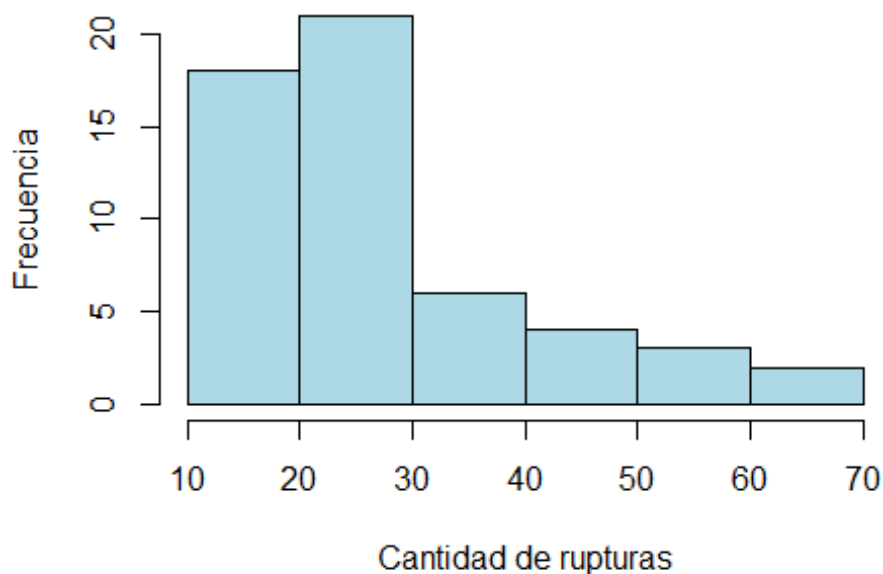
I. Análisis descriptivo

Histograma del número de rupturas

```
# Graficar histograma para la cantidad de rupturas
```

```
hist(datos$breaks, main = "Histograma de la cantidad de rupturas",  
      xlab = "Cantidad de rupturas", ylab = "Frecuencia",  
      col = "lightblue")
```

Histograma de la cantidad de rupturas



Media y varianza de la variable dependiente (cantidad de rupturas)

Calcular la media y la varianza de la variable dependiente (número de rupturas)

```
statistics = data.frame(c(mean(datos$breaks), var(datos$breaks)),  
                        row.names = c("Media", "Varianza"))
```

```
colnames(statistics) = c("Valor")
```

```
statistics
```

	Valor
Media	28.14815
Varianza	174.20405

Interpretación en el contexto de una regresión Poisson

En términos generales, en base al histograma y medidas estadísticas (media y varianza) previos, es posible afirmar que dado el hecho de que la media de las rupturas es igual a 28.14 y la varianza de dicha variable es a su vez igual a 174.2, entonces eso significa que por un lado, dentro del contexto de una regresión Poisson, se presentan aproximadamente 28.14 rupturas del urdimbre en promedio, además de que también dado que la varianza de la variable dependiente es igual a 174.2, lo cual es a su vez mayor a la media de dichos valores, por lo cual, eso es un indicativo de que los datos de la variable dependiente (cantidad de rupturas) presenta sobredispersión,

lo cual señala que en los datos de rupturas, existe una cantidad de varianza mayor que la que el modelo de regresión Poisson es capaz de explicar, por lo que una regresión Poisson podría no ser un modelo adecuado para representar a la cantidad de rupturas.

II. Ajustar 2 modelos de regresión Poisson

Regresión Poisson sin interacción

Ajustar modelo de regresión Poisson sin interacción a la cantidad de rupturas

```
poisson_NoInt = glm(breaks ~ wool + tension, data = datos, family =  
poisson(link = "log"))
```

Mostrar resumen del modelo de regresión Poisson sin interacción

```
s_PNI = summary(poisson_NoInt)
```

```
s_PNI
```

```
##
```

```
## Call:
```

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

```
##
```

```
## 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
```

Regresión Poisson con interacción

Ajustar modelo de regresión Poisson con interacción a los datos

```
poisson_Int = glm(breaks ~ wool * tension, data = datos, family =  
poisson(link = "log"))
```

Mostrar resumen del modelo de regresión Poisson con interacción

```
s_PI = summary(poisson_Int)

s_PI

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

Interpretación de variables dummy y modelos obtenidos

Modelo 1 (sin interacción)

En términos generales, en cuanto al modelo 1 (sin interacción), es posible afirmar que las variables dummy correspondientes al modelo de regresión de Poisson sin interacción son `woolB`, `tensionM` y `tensionH` y sus coeficientes son -0.2059, -0.3213 y -0.5184 respectivamente, por lo cual, el coeficiente de la variable dummy `woolB` hace referencia al hecho de que en caso de que el valor de `woolB` para una observación en particular sea igual a 1 (se utiliza lana tipo B), el valor del logaritmo natural de la variable de respuesta (`breaks`), experimentará un cambio de -0.2059 unidades más o menos con respecto a su valor actual, mientras que de forma similar, el coeficiente de la variable dummy `tensionM` indica que si para una observación en específico, se utiliza tensión tipo M (el valor de la variable `tensionM` para esa observación es 1), entonces el logaritmo natural de `breaks` experimentará un cambio de -0.3213 unidades respecto a su valor actual, y por último, el coeficiente de la variable `tensionH` señala que si para una observación en concreto, el valor de `tensionH` es 1 (se usa tensión tipo H), entonces el logaritmo natural de `breaks` tendrá un cambio de -0.5184 unidades respecto a su valor actual.

Ecuación del modelo de regresión sin interacción:

$$\ln(breaks) = 3.6919 - 0.2059 * woolB - 0.3213 * tensionM - 0.5184 * tensionH$$

Ecuación del modelo de interés sin interacción:

$$breaks = e^{3.6919 - 0.2059 * woolB - 0.3213 * tensionM - 0.5184 * tensionH}$$

Modelo 2 (con interacción)

De manera general, en relación al modelo 2 (con interacción), es posible apreciar que las variables dummy para el caso de este modelo son: *woolB*, *tensionM* y *tensionH* y sus coeficientes son -0.4566, -0.6186 y -0.5957 respectivamente, motivo por el cual, el coeficiente de *woolB* señala que si para una observación en específico, el valor de *woolB* es 1 (se utilizó lana tipo B), entonces el valor del logaritmo natural de *breaks* experimentará un aumento o decremento de -0.4566 unidades con respecto a su valor actual, mientras que de forma similar, en el caso del coeficiente de *tensionM*, éste se refiere a que si para una observación el valor de *tensionM* es 1 (se utilizó tensión tipo M), entonces el valor del logaritmo natural de *breaks* experimentará un cambio de -0.6186 unidades en relación a su valor actual, además de que en el caso del coeficiente de la variable *tensionH*, éste hace referencia a que si el valor de *tensionH* para alguna observación en particular es 1 (se usó tensión tipo H), entonces el valor del logaritmo natural de *breaks* experimenta un cambio de -0.5957 unidades respecto a su valor actual.

Ecuación del modelo de regresión con interacción:

$$\begin{aligned} \ln(breaks) \\ = 3.7967 - 0.4566 * woolB - 0.6186 * tensionM - 0.5957 * tensionH + 0.6381 \\ * woolB:tensionM + 0.1883 * woolB:tensionH \end{aligned}$$

Ecuación del modelo de interés con interacción:

$$\begin{aligned} breaks \\ = e^{3.7967 - 0.4566 * woolB - 0.6186 * tensionM - 0.5957 * tensionH + 0.6381 * woolB:tensionM + 0.1883 * woolB:tensionH} \end{aligned}$$

III. Selección del modelo

Prueba de hipótesis para la prueba de χ^2 :

H_0 : la desviación residual del modelo es igual a 0.

H_1 : la desviación residual del modelo es mayor a 0.

Calcular valor frontera de la zona de rechazo para el modelo sin interacción

```
gl_NoInt = s_PNI$df.null - s_PNI$df.residual
```

```

cat("Valor frontera de zona de rechazo para modelo sin interacción:",
    qchisq(0.05, gl_NoInt), "\n")

## Valor frontera de zona de rechazo para modelo sin interacción:
0.3518463

# Calcular valor frontera de la zona de rechazo para el modelo con
interacción

gl_Int = s_PI$df.null - s_PI$df.residual

cat("Valor frontera de zona de rechazo para modelo con interacción:",
    qchisq(0.05, gl_Int))

## Valor frontera de zona de rechazo para modelo con interacción:
1.145476

# Calcular el valor del estadístico de prueba y el valor p para modelo
sin interacción

statistic_NoInt = s_PNI$deviance

pvalor_NoInt = 1 - pchisq(statistic_NoInt, gl_NoInt)

cat("Estadístico de prueba para modelo sin interacción:",
    statistic_NoInt, "\n")

## Estadístico de prueba para modelo sin interacción: 210.3919

cat("Valor p para modelo sin interacción:", pvalor_NoInt)

## Valor p para modelo sin interacción: 0

# Calcular el valor del estadístico de prueba y el valor p para modelo
con interacción

statistic_Int = s_PI$deviance

pvalor_Int = 1 - pchisq(statistic_Int, gl_Int)

cat("Estadístico de prueba para modelo con interacción:", statistic_Int,
    "\n")

## Estadístico de prueba para modelo con interacción: 182.3051

cat("Valor p para modelo con interacción:", pvalor_Int)

## Valor p para modelo con interacción: 0

```

En los cálculos anteriores, es posible observar que el valor p tanto para el modelo con interacción como sin interacción, es igual a 0, lo cual al ser menor que 0.05, es posible afirmar que se rechaza la hipótesis nula H_0 de la prueba de hipótesis planteada con

anterioridad, lo cual implica que la desviación residual de ambos modelos (sin interacción y con interacción) resulta ser mayor a 0, motivo por el cual, eso implica que ambos modelos generados son adecuados para explicar los datos de rupturas en función del tipo de lana y del tipo de tensión utilizados en el urdimbre.

Comparación del AIC de ambos modelos:

```
# Índice AIC del modelo sin interacción

cat("AIC del modelo sin interacción: ", poisson_NoInt$aic, "\n")

## AIC del modelo sin interacción: 493.056

# Índice AIC del modelo con interacción

cat("AIC del modelo con interacción: ", poisson_Int$aic)

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

De acuerdo con el valor del AIC calculado para cada modelo, es posible apreciar que el modelo con interacción resulta tener el menor AIC de ambos, mismo que es igual a 468.9692089, en comparación con el AIC del modelo sin interacción, cuyo AIC es mayor, específicamente de 493.0559664, indicando que el modelo sin interacción es menos efectivo para explicar los datos que el modelo con interacción, por lo cual, el mejor modelo de acuerdo con el AIC es el modelo con interacción.

Comparación de coeficientes de los modelos:

```
# Generar una tabla con los valores de los coeficientes de ambos modelos

coefs_table = data.frame(Intercept = c(poisson_NoInt$coefficients[1],
                                       poisson_Int$coefficients[1]),
                        woolB = c(poisson_NoInt$coefficients[2],
                                  poisson_Int$coefficients[2]),
                        tensionM = c(poisson_NoInt$coefficients[3],
                                     poisson_Int$coefficients[3]),
                        tensionH = c(poisson_NoInt$coefficients[4],
                                     poisson_Int$coefficients[4]),
                        row.names = c("Modelo sin interacción", "Modelo
con interacción"))

colnames(coefs_table) = c("Intercept", "woolB", "tensionM", "tensionH")

coefs_table
```

	Intercept	woolB	tensionM	tensionH
Modelo sin interacción	3.691963	-0.2059884	-0.3213204	-0.5184885
Modelo con interacción	3.796737	-0.4566272	-0.6186830	-0.5957987

En la tabla anterior, mirando los coeficientes en valor absoluto de las 4 variables involucradas para ambos modelos, es posible apreciar que en general, los coeficientes pertenecientes al modelo 2 (con interacción), resultan ser ligeramente mayores a los del modelo 1 (sin interacción), lo cual es un indicativo de que al momento de calcular predicciones de las rupturas de urdimbre en base al modelo con interacción, las predicciones derivadas de dicho modelo, tendrán una mayor cantidad de variación o varianza que aquellas derivadas del modelo sin interacción, motivo por el cual, si se grafican los datos predichos a partir del modelo con interacción en un histograma o boxplot, se apreciará que en el caso del histograma las barras abarcarán una mayor región del eje horizontal del gráfico, mientras que en el caso del boxplot, se observará que la caja del mismo será más grande, esto a diferencia de éstos mismos gráficos pero para las predicciones realizadas en base al modelo sin interacción, en los cuales, tanto la caja del boxplot como la región abarcada por las barras del histograma serán menores.

Comparación de errores estándar de estimadores β_i de ambos modelos:

```
# Generar una tabla con los valores del error estándar para cada
# estimador beta_i
# para ambos modelos

std_error_models = data.frame(s_PNI$coefficients[, "Std. Error"],
                              s_PI$coefficients[, "Std. Error"][c(-5, -
6)],
                              row.names = c("Intercept", "woolB",
"tensionM", "tensionH"))

colnames(std_error_models) = c("Std. Error sin interacción",
                              "Std. Error con interacción")

std_error_models
```

	Std. Error sin interacción	Std. Error con interacción
Intercept	0.0454107	0.0499375
woolB	0.0515712	0.0801920
tensionM	0.0602658	0.0844001
tensionH	0.0639594	0.0837772

En la tabla previa se aprecia que para el caso del modelo sin interacción, dicho modelo resulta tener errores estándar por cada estimador β_i ligeramente menores que aquellos errores estándar correspondientes a los estimadores β_i del modelo con interacción, lo cual significa que al momento de realizar predicciones en base al modelo sin interacción, éstas mismas tendrán un grado de sesgo ligeramente menor con respecto a los datos reales junto con un nivel de precisión ligeramente mayor, que aquellas predicciones derivadas del modelo con interacción, esto principalmente debido a que es probable que al incluir las interacciones entre variables predictoras en el caso del modelo con interacción, se introduzca una ligera cantidad de sesgo al

modelo predictivo, por lo que las predicciones en base al modelo con interacción estarán ligeramente más alejadas de los datos reales que aquellas derivadas del modelo sin interacción.

Interpretación de coeficientes de ambos modelos:

Calcular e elevado a cada coeficiente del modelo sin interacción para facilitar su interpretación

```
print("Para el modelo sin interacción:")
```

```
## [1] "Para el modelo sin interacción:"
```

```
exp(poisson_NoInt$coefficients)
```

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

Calcular e elevado a cada coeficiente del modelo con interacción para facilitar su interpretación

```
print("Para el modelo con interacción:")
```

```
## [1] "Para el modelo con interacción:"
```

```
exp(poisson_Int$coefficients)
```

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

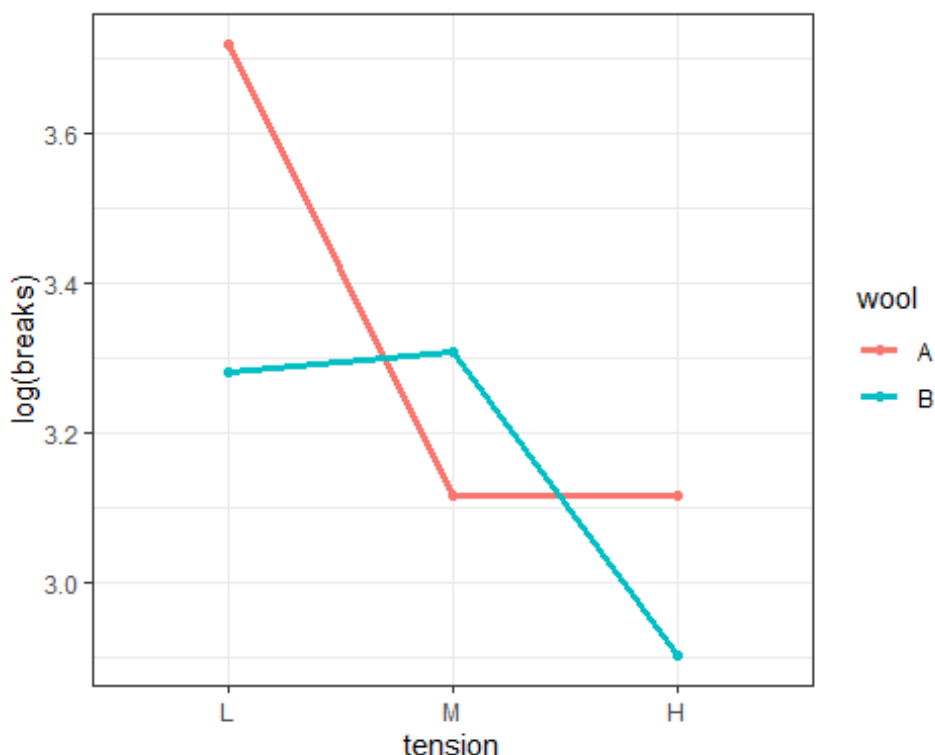
En primera instancia, en cuanto al modelo sin interacción, sus coeficientes son: 3.6919 para el intercepto, -0.2059 para la variable woolB, -0.3213 para tensionM y -0.5184 para tensionH, por lo cual, en el caso del intercepto del modelo, su coeficiente indica que al no utilizar ninguna clase de lana en específico y ningún tipo de tensión (M ni H), la tasa promedio de rupturas en el urdimbre será de 39.123538, mientras que en cuanto al coeficiente de woolB, este indica que por cada ruptura en el urdimbre producida por usar lana tipo B, la tasa promedio de rupturas en el urdimbre disminuye un 18.6157518%, mientras que de forma similar, para la interpretación para el coeficiente de tensionM, es posible afirmar que por cada ruptura adicional producida por utilizar tension tipo M, la tasa promedio de rupturas experimenta una disminución del 27.480916%, además de que el coeficiente de la variable tensionH, indica que por cada ruptura más producida al usar tensión tipo H, la tasa promedio de rupturas experimentará una disminución del 40.4580153%.

```
# Importar librería ggplot2 para realizar gráficos personalizados

library(ggplot2)

# Graficar el comportamiento de las interacciones del modelo con
# interacción para su
# mejor interpretación

ggplot(datos, 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"))
```



Por otro lado, en cuanto al modelo con interacción, sus coeficientes son: 3.7967 para el intercepto, -0.4566 para woolB, -0.6186 para tensionM, -0.5957 para tensionH, 0.6381 para la interacción de woolB con tensionM (woolB:tensionM) y 0.1883 para la interacción entre woolB y tensionH (woolB:tensionH), por lo que, el coeficiente del intercepto, hace referencia a que al no emplear ninguna clase de lana ni tension, la tasa promedio de rupturas del urdimbre será de 43.5555556, mientras que de forma similar, el coeficiente de la variable woolB indica que por cada ruptura más producida al usar lana tipo B, la tasa promedio de rupturas aumenta en un 36.6583541%, mientras que el coeficiente de tensionM señala a su vez que por cada ruptura de urdimbre, la tasa promedio de rupturas aumenta en un 46.1346633%, además, el coeficiente de tensionH hace referencia a que por cada ruptura de urdimbre, la tasa

promedio de rupturas aumenta en un 44.8877805%. Adicionalmente, en cuanto al coeficiente de la interacción `woolB:tensonM`, éste se refiere a que por cada unidad que aumente el valor de dicha interacción, la tasa promedio de rupturas de urdimbre disminuirá en un 89.3026393%, mientras que de forma similar, el coeficiente de la interacción `woolB:tensionH` señala que por cada unidad que aumente el valor de esa interacción, la tasa promedio de rupturas de urdimbre disminuirá en un 20.7271885%, tal como se puede observar en el gráfico anterior referente al comportamiento de la interacción entre el tipo de lana y el tipo de tensión empleados al momento de trabajar con el urdimbre.

Motivo por el cual, en resumen, el modelo con interacción es un mejor modelo, puesto que además de tener un AIC menor al del modelo sin interacción, junto con errores estándar bastante pequeños para cada uno de sus coeficientes β_i , también toma en cuenta los valores de la interacción entre el tipo de lana y el tipo de tensión, lo cual puede agregar un extra de precisión a las predicciones de las rupturas, ya que en ocasiones, ciertos tipos de lana no son capaces de resistir un determinado tipo de tensión aplicada sobre ellos, por lo que al no tener la suficiente resistencia, dichos tipos de lana se rompen fácilmente.

IV. Evaluación de los supuestos

Independencia

Prueba de hipótesis:

H_0 : los residuos no tienen autocorrelación (son independientes).

H_1 : los residuos sí tienen autocorrelación (no son independientes).

Librería lmtest para el test de independencia de Durbin Watson

```
library(lmtest)
```

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

```
##
```

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

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

```
##
```

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

Realizar test de independencia de los residuos del modelo sin interacción

```
dwtest(poisson_NoInt)
```

```
##
```

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

```
##
## data: poisson_NoInt
## DW = 2.0332, p-value = 0.3896
## alternative hypothesis: true autocorrelation is greater than 0
# Test de independencia de Los residuos del modelo con interacción

dwtest(poisson_Int)

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

En los tests de independencia de Durbin Watson aplicado a los residuos de ambos modelos, se puede apreciar que en el caso del modelo sin interacción, el p valor del test es igual a 0.3896, mientras que en el caso del modelo con interacción, el p valor del test es igual a 0.575, motivo por el cual, dado que para ambos modelos, el p valor resultó ser mayor a 0.05, entonces para los 2 modelos no se rechaza (se acepta) H_0 , lo cual significa que ambos modelos (sin y con interacción), poseen residuos que a su vez no presentan autocorrelación, lo que significa que en el caso de ambos modelos, sus residuos sí son independientes.

Sobredispersión de los residuos

Prueba de hipótesis:

H_0 : No hay una sobredispersión del modelo.

H_1 : Hay una sobredispersión del modelo.

```
# Importar Librería epiDisplay para realizar prueba de sobredispersión de 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:lmtest':
##
##      lrtest

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

# Realizar prueba de sobredispersión de residuos al modelo sin interacción

posgof_NI = poisgof(poisson_NoInt)

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

# Realizar prueba de sobredispersión de residuos al modelo con interacción

posgof_Int = poisgof(poisson_Int)

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

```

En los tests de sobredispersión de residuos, se aprecia que el p valor para el caso del modelo sin interacción es de $1.44e-21$, mientras que en el caso del modelo con interacción, el p valor es de $1.582e-17$, por lo que en ambos casos el p valor del test resulta ser mucho menor a 0.05, por lo cual, en ambos modelos, se rechaza H_0 , lo que significa que hay una sobredispersión de los residuos en ambos modelos (sin

interacción y con interacción), por lo que en última instancia, ambos modelos de regresión de Poisson no resultan ser suficientemente adecuados para explicar los datos analizados.

Modelos alternativos a la regresión Poisson

Ajustar un modelo cuasi Poisson a Los datos

```
cuasi_poisson = glm(breaks ~ wool + tension, data = datos,
                    family = quasipoisson(link = "log"))

summary_cuasi_poisson = summary(cuasi_poisson)

summary_cuasi_poisson

##
## Call:
## glm(formula = breaks ~ wool + tension, family = quasipoisson(link =
"log"),
##      data = datos)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.69196    0.09374  39.384 < 2e-16 ***
## woolB         -0.20599    0.10646  -1.935 0.058673 .
## tensionM      -0.32132    0.12441  -2.583 0.012775 *
## tensionH      -0.51849    0.13203  -3.927 0.000264 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 4.261537)
##
##      Null deviance: 297.37  on 53  degrees of freedom
## Residual deviance: 210.39  on 50  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

Ajustar un modelo binomial negativa a Los datos analizados

```
neg_binomial = glm.nb(breaks ~ wool + tension, data = datos,
                      control = glm.control(maxit = 1000))

summary_nb = summary(neg_binomial)

summary_nb

##
## Call:
## glm.nb(formula = breaks ~ wool + tension, data = datos, control =
glm.control(maxit = 1000),
```

```
##      init.theta = 9.944385436, link = log)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.6734     0.0979  37.520 < 2e-16 ***
## woolB         -0.1862     0.1010  -1.844  0.0651 .
## tensionM      -0.2992     0.1217  -2.458  0.0140 *
## tensionH      -0.5114     0.1237  -4.133 3.58e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(9.9444) family taken to be
1)
##
##      Null deviance: 75.464  on 53  degrees of freedom
## Residual deviance: 53.723  on 50  degrees of freedom
## AIC: 408.76
##
## Number of Fisher Scoring iterations: 1
##
##              Theta:  9.94
##              Std. Err.:  2.56
##
## 2 x log-likelihood:  -398.764
```

Prueba de χ^2 para comprobar desviación residual:

Hipótesis:

H_0 : la desviación residual del modelo es igual a cero.

H_1 : la desviación residual del modelo es mayor a cero.

```
# Prueba de chi cuadrado para comprobar que La desviación residual del
modelo cuasi poisson
# se aleje lo más posible del cero
```

```
gl_cuasi_poisson = summary_cuasi_poisson$df.null -
summary_cuasi_poisson$df.residual
```

```
val_front_cuasipoisson = qchisq(0.05, gl_cuasi_poisson)
```

```
statistic_cuasiPoisson = summary_cuasi_poisson$deviance
```

```
pvalue_cuasiPoisson = 1 - pchisq(statistic_cuasiPoisson,
gl_cuasi_poisson)
```

```
# Prueba de chi cuadrado para comprobar que La desviación residual del
modelo binomial
# negativa se aleje lo más posible del cero
```

```

gl_bn = summary_nb$df.null - summary_nb$df.residual

val_front_nb = qchisq(0.05, gl_bn)

statistic_nb = summary_nb$deviance

pvalue_nb = 1 - pchisq(statistic_nb, gl_bn)

# Mostrar resultados de la prueba de chi2 para ambos modelos

cat("Estadístico de prueba del modelo cuasi Poisson: ",
    statistic_cuasiPoisson, "\n")

## Estadístico de prueba del modelo cuasi Poisson: 210.3919

cat("p-value del modelo cuasi Poisson: ", pvalue_cuasiPoisson, "\n")

## p-value del modelo cuasi Poisson: 0

cat("Estadístico de prueba del modelo binomial negativa: ", statistic_nb,
    "\n")

## Estadístico de prueba del modelo binomial negativa: 53.72257

cat("p-value del modelo binomial negativa: ", pvalue_nb)

## p-value del modelo binomial negativa: 1.285816e-11

```

De acuerdo con los resultados de la prueba de χ^2 para los modelos cuasi Poisson y binomial negativa, se puede apreciar que el p valor del test para el caso del modelo cuasi Poisson es igual a cero, mientras que en el caso del modelo binomial negativa, dicho p valor es igual a 1.2858e-11, por lo cual, dado que para ambos modelos, el p-value resulta ser menor que 0.05, lo cual implica que tanto para el modelo cuasi Poisson como para el modelo binomial negativa, se rechaza la hipótesis nula H_0 , lo cual significa que la desviación residual de ambos modelos alternativos a la regresión Poisson resulta ser mayor a cero.

Comparación del AIC para cuasi Poisson y binomial negativa:

```

# Valor del indicador AIC para el modelo cuasi Poisson

cat("AIC del modelo cuasi Poisson: ", cuasi_poisson$aic, "\n")

## AIC del modelo cuasi Poisson: NA

# Valor del indicador AIC del modelo binomial negativa

cat("AIC del modelo binomial negativa: ", neg_binomial$aic)

## AIC del modelo binomial negativa: 408.7638

```


Como se puede observar en los resultados previos, el valor del AIC para el modelo cuasi Poisson no se encuentra definido, mientras que el AIC del modelo binomial negativa es igual a 408.7638, motivo por el cual, en base a ésta información, es posible afirmar que el modelo binomial negativa resulta ser más adecuado, dado que sí tiene un valor AIC claramente definido, por lo cual es posible comparar dicho valor con el valor AIC de los modelos de regresión Poisson sin interacción y con interacción generados con anterioridad, por lo que al realizar dicha comparación, es posible apreciar que el AIC del modelo binomial negativa resulta ser menor que el AIC del modelo sin interacción (493.056) y que el AIC del modelo con interacción (468.9692), por lo que de acuerdo al criterio de AIC, el modelo binomial negativa resulta ser el mejor de los 4 dado que posee el menor AIC de todos.

Comparación de coeficientes de cuasi Poisson y binomial negativa:

Tabla de Los coeficientes de cuasi Poisson y binomial negativa

```
cuasiPoisson_bn_table = data.frame(c(cuasi_poisson$coefficients[1],
neg_binomial$coefficients[1]), c(cuasi_poisson$coefficients[2],
neg_binomial$coefficients[2]), c(cuasi_poisson$coefficients[3],
neg_binomial$coefficients[3]), c(cuasi_poisson$coefficients[4],
neg_binomial$coefficients[4]), row.names = c("Cuasi Poisson", "Binomial
Negativa"))
```

```
colnames(cuasiPoisson_bn_table) = c("Intercept", "woolB", "tensionM",
"tensionH")
```

```
cuasiPoisson_bn_table
```

	Intercept	woolB	tensionM	tensionH
Cuasi Poisson	3.691963	-0.2059884	-0.3213204	-0.5184885
Binomial Negativa	3.673355	-0.1862111	-0.2992272	-0.5113955

En la tabla anterior, se puede observar que los coeficientes de ambos modelos poseen una ligera variación entre ellos, sin embargo, a pesar de que el modelo cuasi Poisson presenta coeficientes de magnitud ligeramente mayor a los del modelo binomial negativa, no obstante, el modelo cuasi Poisson no posee un valor AIC definido, además de que su p valor es cero, motivos por los cuales, aún resulta mejor el modelo binomial negativa para explicar los datos en cuestión.

Crear tabla con Los valores del error estándar para Los coeficientes de cada modelo

```
stdError_CP_bn = data.frame(cuasi_poisson$coefficients,
neg_binomial$coefficients,
row.names = c("Intercept", "woolB",
"tensionM", "tensionH"))
```

```
colnames(stdError_CP_bn) = c("Cuasi Poisson", "Binomial negativa")
```

stdError_CP_bn

	Cuasi Poisson	Binomial negativa
Intercept	3.6919631	3.6733546
woolB	-0.2059884	-0.1862111
tensionM	-0.3213204	-0.2992272
tensionH	-0.5184885	-0.5113955

Además de lo anterior, en la tabla de errores estándar por coeficiente de cada modelo, se puede apreciar que el modelo binomial negativa posee errores estándar para cada uno de sus coeficientes, ligeramente menores a los del modelo cuasi Poisson, lo cual indica que en general, las predicciones derivadas del modelo binomial negativa resultan tener un sesgo ligeramente menor que las provenientes del modelo cuasi Poisson, motivo por el cual, las predicciones del modelo binomial negativa son ligeramente más precisas que las del modelo cuasi Poisson, respaldando el hecho de que el modelo binomial negativa aún resulta ser mejor que el cuasi poisson para explicar los datos analizados.

Verificación de supuestos para cuasi Poisson y binomial negativa:

Independencia

H_0 : los residuos del modelo no presentan autocorrelación (son independientes).

H_1 : los residuos del modelo sí presentan autocorrelación (no son independientes).

Realizar test de independencia de Durbin Watson para comprobar que Los residuos del

modelo cuasi Poisson sean independientes

```
dwtest(cuasi_poisson)
```

```
##
```

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

```
##
```

```
## data:  cuasi_poisson
```

```
## DW = 2.0332, p-value = 0.3896
```

```
## alternative hypothesis: true autocorrelation is greater than 0
```

Test de independencia de Durbin Watson para Los residuos del modelo binomial negativa

```
dwtest(neg_binomial)
```

```
##
```

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

```
##
```

```
## data:  neg_binomial
```

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

En los tests de Durbin Watson anteriores, se aprecia que el p valor tanto para el modelo cuasi Poisson como para el modelo binomial negativa es igual a 0.3896, lo cual al ser mayor a 0.05, implica que no se rechaza H_0 , por lo cual, eso significa que los residuos de ambos modelos no presentan autocorrelación, en otras palabras, son independientes, sin embargo, el modelo de regresión Poisson con interacción tiene un p valor de 0.575, lo cual es superior a 0.3896, por lo que los residuos tienen mayor grado de independencia en el modelo poisson con interacción, que en el cuasi Poisson y en el binomial negativa, no obstante, dado que los modelos poisson y cuasi Poisson poseen un p valor de 0 y el modelo binomial negativa tiene un p valor de 1.2858e-11, el modelo binomial negativa aún resulta ser más adecuado para explicar los datos.

Sobredispersión de los residuos

Hipótesis:

H_0 : No hay una sobredispersión del modelo.

H_1 : Hay una sobredispersión del modelo.

```
# Realizar una prueba de sobredispersión del modelo binomial negativa
```

```
poisgof(neg_binomial)
```

```
## $results
## [1] "Goodness-of-fit test for Poisson assumption"
##
## $chisq
## [1] 53.72257
##
## $df
## [1] 50
##
## $p.value
## [1] 0.333697
```

```
# Nota: La prueba poisgof no es posible realizarla con el modelo cuasi Poisson, solamente con la binomial negativa
```

En los resultados de la prueba de sobredispersión de los residuos para el modelo binomial negativa, se puede observar que el p valor del test es igual a 0.333697, lo cual es mucho mayor a 0.05, lo cual implica que no se rechaza H_0 , lo cual significa que no hay una sobredispersión del modelo, por lo tanto, la prueba confirma que los residuos del modelo binomial negativa no presentan sobredispersión, motivo por el cual, dado que el modelo binomial negativa presenta fuerte independencia en cuanto a sus residuos, además de no presentar una sobredispersión de sus residuos, lo cual no se cumple en el caso del modelo cuasi Poisson (debido a que no es posible verificar la

sobredispersión de sus residuos), por lo cual, entre el modelo cuasi Poisson y el modelo binomial negativa, el mejor modelo es el modelo binomial negativa.

V. Conclusión final

A manera de conclusión final de todo el análisis realizado, es posible afirmar que el modelo que mejor satisface todos los criterios y supuestos necesarios para ser considerado un buen modelo para los datos, es el modelo binomial negativa, debido a que a diferencia de los otros 3 modelos (cuasi Poisson, Poisson sin interacción, Poisson con interacción) el modelo binomial negativa obtuvo un p valor de $1.2858e-11$ en la prueba de desviación residual (prueba de χ^2), a diferencia de los otros 3 modelos cuyo p valor en dicha prueba fue prácticamente de 0, además de que también, el modelo binomial negativa posee el menor valor del AIC (408.7638) entre todos los modelos probados, aunque también posee errores estándar ligeramente mayores para sus coeficientes que los modelos de Poisson sin y con interacción, no obstante, el modelo binomial negativa es el único de los 4 probados que cumple al mismo tiempo los supuestos de independencia y no sobredispersión de sus residuos, lo cual significa que las predicciones derivadas de dicho modelo serán las que tengan el mayor grado de precisión y confiabilidad posibles, motivo por el cual, en resumen, el modelo que mejor explica los datos es el modelo binomial negativa.