

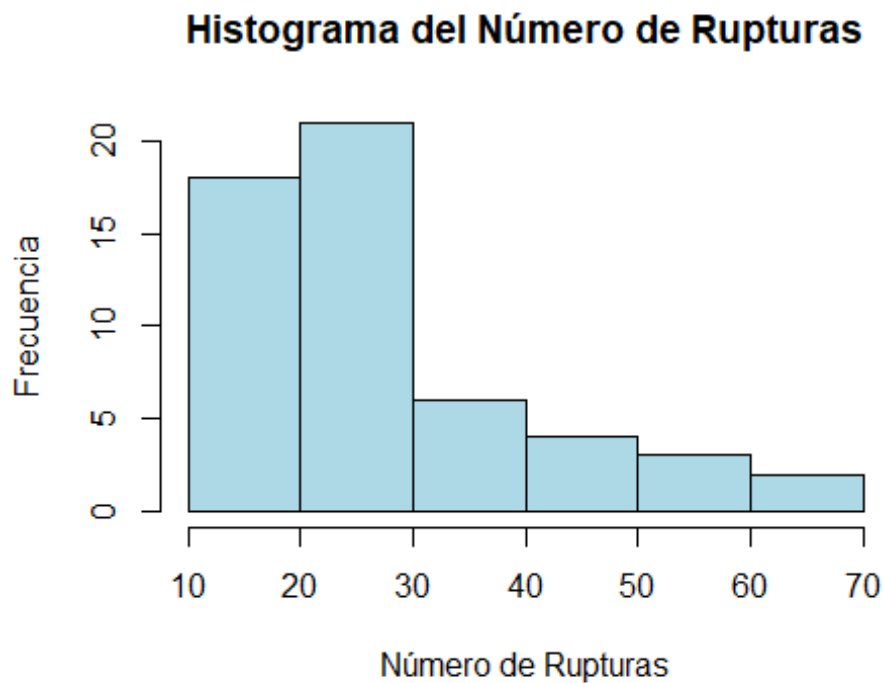
A6_A01571214

A01571214 - Lautaro Coteja

2024-11-04

A6 - Regresion Poisson

```
data <- warpbreaks  
hist(data$breaks, main = "Histograma del Número de Rupturas", xlab = "Número  
de Rupturas", ylab = "Frecuencia", col = "lightblue", border = "black")
```



```
mean_breaks <- mean(data$breaks)  
var_breaks <- var(data$breaks)  
mean_breaks  
## [1] 28.14815  
var_breaks  
## [1] 174.2041
```

Interpretacion en el contexto de una Regresion Poisson

La media de rupturas de urdimbre es de 28.15 y la varianza es de 174.20, lo cual indica una varianza notablemente mayor que la media. Esto sugiere una posible sobredispersión en los datos, una condición común en modelos de conteo, y es importante evaluarla para decidir si el modelo Poisson es el más adecuado o si sería preferible un enfoque como el modelo cuasi-Poisson o de Binomial Negativa.

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

##
## 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

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

##
## 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
```

Interpretacion

Sin interaccion

En este modelo, observamos que el tipo de lana y el nivel de tensión tienen efectos significativos en el número de rupturas. El coeficiente negativo de woolB implica que el uso de este tipo de lana reduce las rupturas comparado con woolA. Similarmente, los niveles de tensión media (tensionM) y alta (tensionH) disminuyen el número de rupturas en comparación con la tensión baja (tensionL). Estos coeficientes muestran el cambio en el logaritmo del número de rupturas asociado a cada variable.

Con Interaccion

Interpretación de los Coeficientes y la Interacción: El modelo con interacción revela no solo el efecto individual de cada factor, sino también cómo los diferentes niveles de tensión afectan de manera distinta según el tipo de lana. La interacción significativa entre woolB y tensionM indica que la reducción de rupturas para woolB es más marcada cuando la tensión es media. Este modelo sugiere una relación más compleja que permite capturar mejor la variabilidad en las rupturas de urdimbre.

```
aic_no_interaction <- AIC(poisson_model_no_interaction)
aic_interaction <- AIC(poisson_model_interaction)
aic_no_interaction

## [1] 493.056

aic_interaction

## [1] 468.9692
```

Interpretacion

Al comparar los modelos, el modelo con interacción presenta un AIC más bajo, lo que sugiere un mejor ajuste. Además, la desviación residual de este modelo es más baja, indicando una mejor explicación de la variabilidad de los datos. Sin embargo, el bajo valor p

de la desviación residual sugiere que ambos modelos pueden presentar limitaciones para capturar toda la variabilidad observada en los datos.

```
# Desviación residual y grados de libertad para el modelo sin interacción
dr_no_interaction <- deviance(poisson_model_no_interaction)
gl_no_interaction <- poisson_model_no_interaction$df.residual
p_value_no_interaction <- 1 - pchisq(dr_no_interaction, gl_no_interaction)

# Desviación residual y grados de libertad para el modelo con interacción
dr_interaction <- deviance(poisson_model_interaction)
gl_interaction <- poisson_model_interaction$df.residual
p_value_interaction <- 1 - pchisq(dr_interaction, gl_interaction)

dr_no_interaction
## [1] 210.3919

p_value_no_interaction
## [1] 0

dr_interaction
## [1] 182.3051

p_value_interaction
## [1] 0
```

Interpretación

La prueba de bondad de ajuste revela una posible sobredispersión en los datos, evidenciada por un valor p extremadamente bajo. Esto sugiere que el modelo Poisson puede no ser adecuado por completo y que un modelo cuasi-Poisson podría manejar mejor esta característica de los datos. La sobredispersión refleja que la variabilidad en las rupturas de urdimbre es mayor de lo que el modelo Poisson estándar puede capturar.

```
library(epiDisplay)

## Cargando paquete requerido: foreign
## Cargando paquete requerido: survival
## Cargando paquete requerido: MASS
## Cargando paquete requerido: nnet

poisgof(poisson_model_interaction)

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

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

poisson_model_quasi <- glm(breaks ~ wool + tension, data = data, family =
quasipoisson(link = "log"))
summary(poisson_model_quasi)

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

Interpretacion

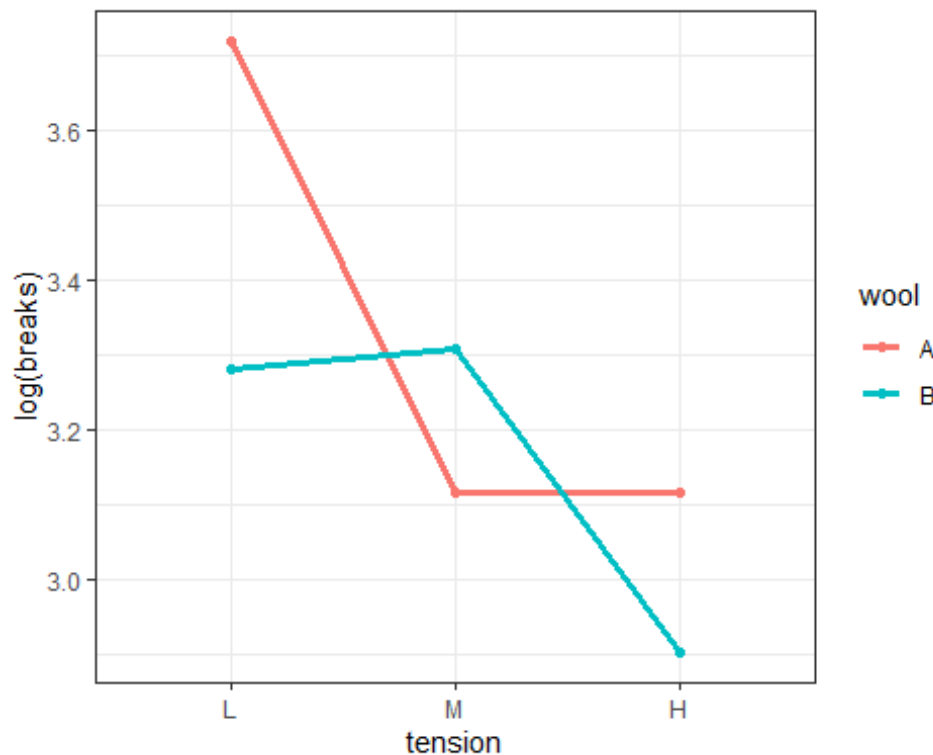
Este modelo ajustado permite evaluar si la sobredispersión es manejada adecuadamente mediante un enfoque de Quasi-Poisson.

```
library(ggplot2)

##
## Adjuntando el paquete: 'ggplot2'

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

```
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 La

gráfica de interacción permite visualizar la relación entre el tipo de lana, la tensión, y la cantidad de rupturas.

Conclusion

Elección del Mejor Modelo:

Basado en los valores de AIC y las pruebas de sobredispersión, el modelo con interacción o el modelo cuasi-Poisson representan las opciones más robustas. El modelo Poisson con interacción proporciona un ajuste más detallado al considerar cómo el tipo de lana y la tensión se combinan para afectar las rupturas. No obstante, dado el posible problema de sobredispersión, el modelo cuasi-Poisson podría ofrecer estimaciones más confiables, especialmente en escenarios donde la variabilidad es alta. En conclusión, para estos datos, se recomendaría utilizar el modelo Poisson con interacción o el cuasi-Poisson dependiendo de la naturaleza y la importancia de la sobredispersión en el análisis final.