

4-regresion-poisson

October 10, 2023

1 4. Regresión Poisson

Francisco Mestizo Hernández A01731549

1.1 Inicio del problema

Cargamos los datos del dataset

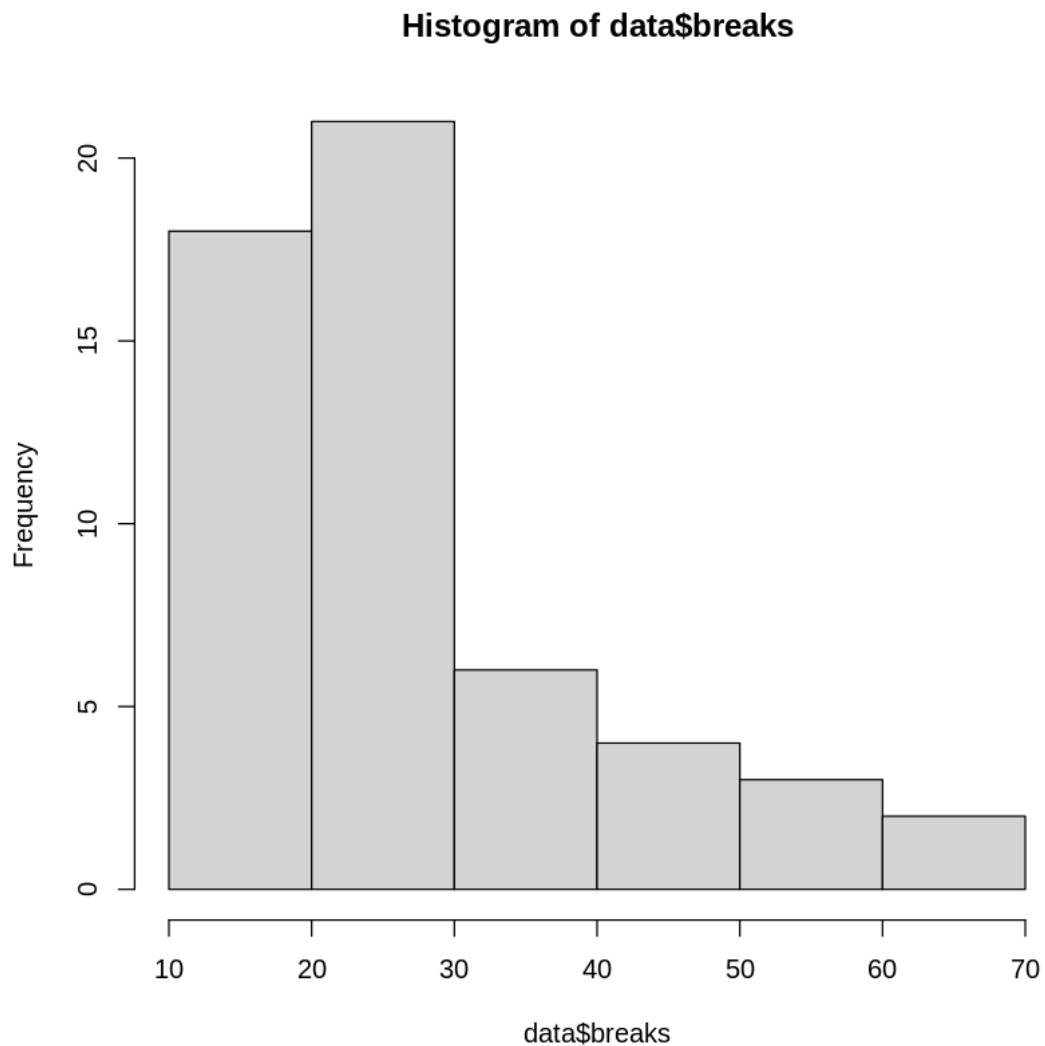
```
[1]: data<-warpbreaks  
head(data,10)
```

A data.frame: 10 × 3

	breaks <dbl>	wool <fct>	tension <fct>
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

Obtenemos el histograma para las rupturas

```
[2]: hist(data$breaks)
```



1.2 Calculo de media y varianza

Obtenemos la media y la varianza de los breaks

```
[9]: cat("Media:", mean(data$breaks), '\n')  
      cat("Varianza:", var(data$breaks))
```

Media: 28.14815
Varianza: 174.2041

1.3 Regresión Poisson

Comenzamos probando con el modelo poisson normal.

```
[6]: poisson.model<-glm(breaks ~ wool + tension, data, family = poisson(link =
  ↪"log"))
summary(poisson.model)
```

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

El modelo tiene una desviación residual muy alta, lo que quiere decir que tenemos demasiada variación en los residuos. De hecho, desde que vemos que la varianza y la media son muy diferentes, tenemos que recurrir directamente al modelo quasipoisson.

```
[7]: poisson.model2<-glm(breaks ~ wool + tension, data = data, family =
  ↪quasipoisson(link = "log"))
summary(poisson.model2)
```

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

Este modelo es el que tuvimos que usar desde el inicio, pero el tipo de lana parece no ser muy significativo. Podemos probar con el modelo sin la variable del tipo de lana.

```
[10]: poisson.model2<-glm(breaks ~ tension, data = data, family = quasipoisson(link =  
  ↪"log"))  
summary(poisson.model2)
```

Call:

```
glm(formula = breaks ~ tension, family = quasipoisson(link = "log"),  
    data = data)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.59426	0.08382	42.881	< 2e-16 ***
tensionM	-0.32132	0.12928	-2.485	0.016260 *
tensionH	-0.51849	0.13721	-3.779	0.000414 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 4.601927)

Null deviance: 297.37 on 53 degrees of freedom
Residual deviance: 226.43 on 51 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4

Podemos concluir con que el modelo es bueno, porque después de eliminar la el tipo de lana, las betas casi no cambian, por lo que el modelo es consistente. Además, tenemos una desviación nula alta, que indica que el modelo es muchísimo mejor que un modelo sin parámetros. Aún así, lo único malo sería que tenemos una desviación residual alta, porque quiere decir que hay mucha variación en los residuos. Por lo tanto, este último modelo sería el mejor.