

RegresionPoisson

Carlos David Contreras Chacon
2022-11-03

Regresión Poisson

1

```
data<-warpbreaks
head(data,10)
```

```
##      breaks wool tension
## 1         26    A       L
## 2         30    A       L
## 3         54    A       L
## 4         25    A       L
## 5         70    A       L
## 6         52    A       L
## 7         51    A       L
## 8         26    A       L
## 9         67    A       L
## 10        18    A       M
```

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

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

2

```
dim(data)
```

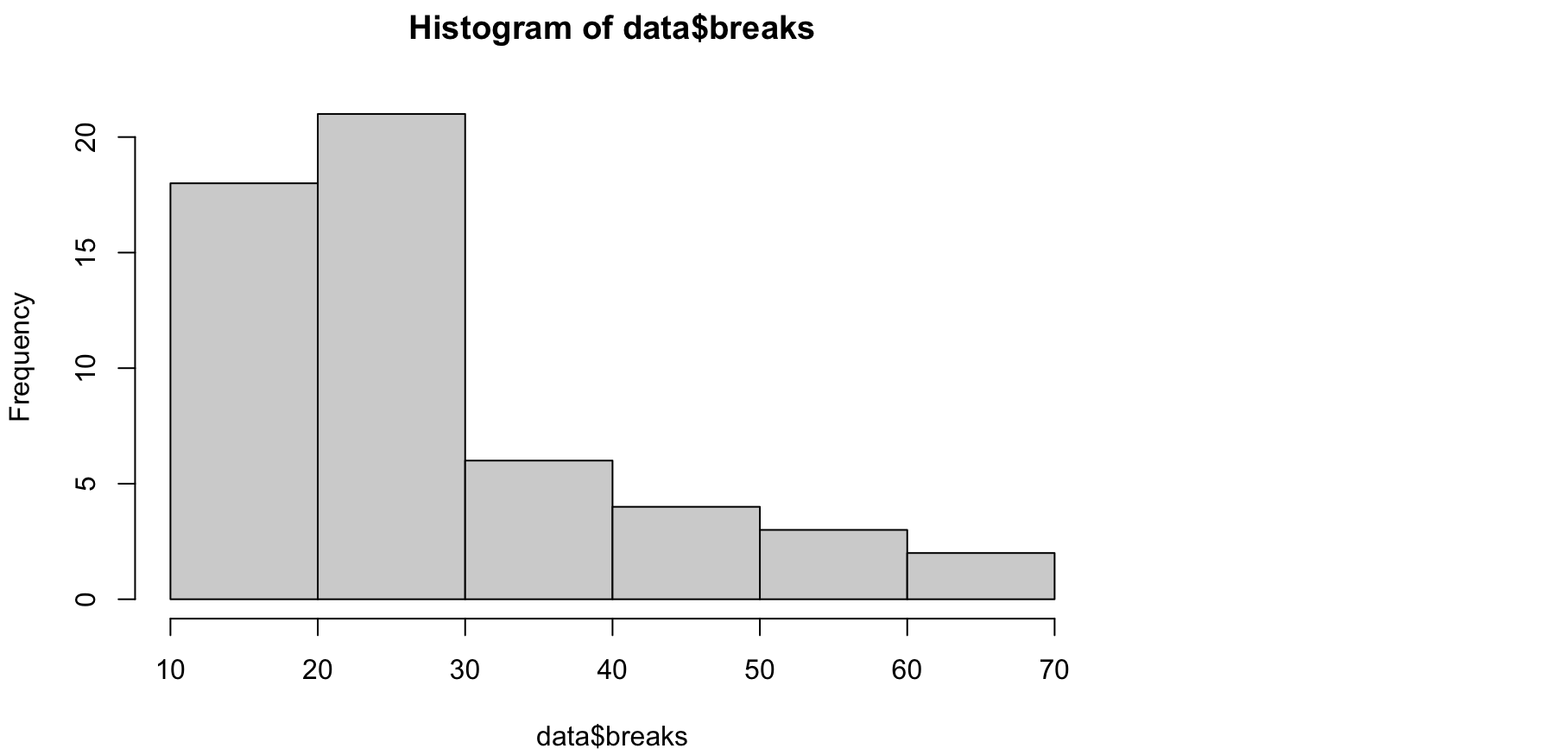
```
## [1] 54  3
```

```
summary(data)
```

```
##      breaks      wool  tension
##  Min.   :10.00  A:27  L:18
## 1st Qu.:18.25  B:27  M:18
##  Median :26.00          H:18
##   Mean   :28.15
## 3rd Qu.:34.00
##   Max.   :70.00
```

Tenemos un dataframe de 54 entradas, cada uno con 3 variables diferentes. La variable breaks es numérica y se refiere al número de pausas realizadas. Wool es una variable categorica que describe el tipo de lana entre A y B. Tension tambien es una variable categorica que mide el nivel de tensión en el telar.

```
hist(data$breaks)
```



Se puede ver que la distribución parece a simple vista una de tipo poisson.

```
mean(data$breaks)
```

```
## [1] 28.14815
```

```
var(data$breaks)
```

```
## [1] 174.2041
```

Se observa una gran dispersión entre los datos con las medidas que tenemos.

3

```
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)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6871  -1.6503  -0.4269   1.1902   4.2616
##
## 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
```

Se observan valores p inferiores a 0.05, lo que nos dice que neustras variables categoricas transformadas en dummies (wool y tension) tienen efecto significativo sobre las roturas.

```
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)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6871  -1.6503  -0.4269   1.1902   4.2616
##
## 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
```

Comparando los modelos se observa que el unico cambio corresponde a los errores estandar. Como la media y varianza no son iguales, se está realizando una sobreestimación.