

- Modelo de Regresión Poisson

## ● Caso de implementacion



### Objetivo del caso

Elaborar un modelo de regresión Poisson que permita establecer un conteo de la cantidad de Homicidios ocurridos los días del mes, en un periodo de un año

### Origen de datos

Los datos utilizados son tomados de <https://www.datos.gov.co/Seguridad-y-Defensa>, donde se registran los Homicidios ocurridos en el país.

Para la construcción del modelo se toman los datos de Homicidios en el año 2017. Para la validación, se contrastará con los Homicidios que fueron registrados en el año 2018.

## Construcción



Para construir los conjuntos de datos se toman los **días** de todos los **meses**, el **sexo**, la **edad** promedio, y la **cantidad** de homicidios registrados. A continuación se muestra la estructura de los datos para las primeras filas:

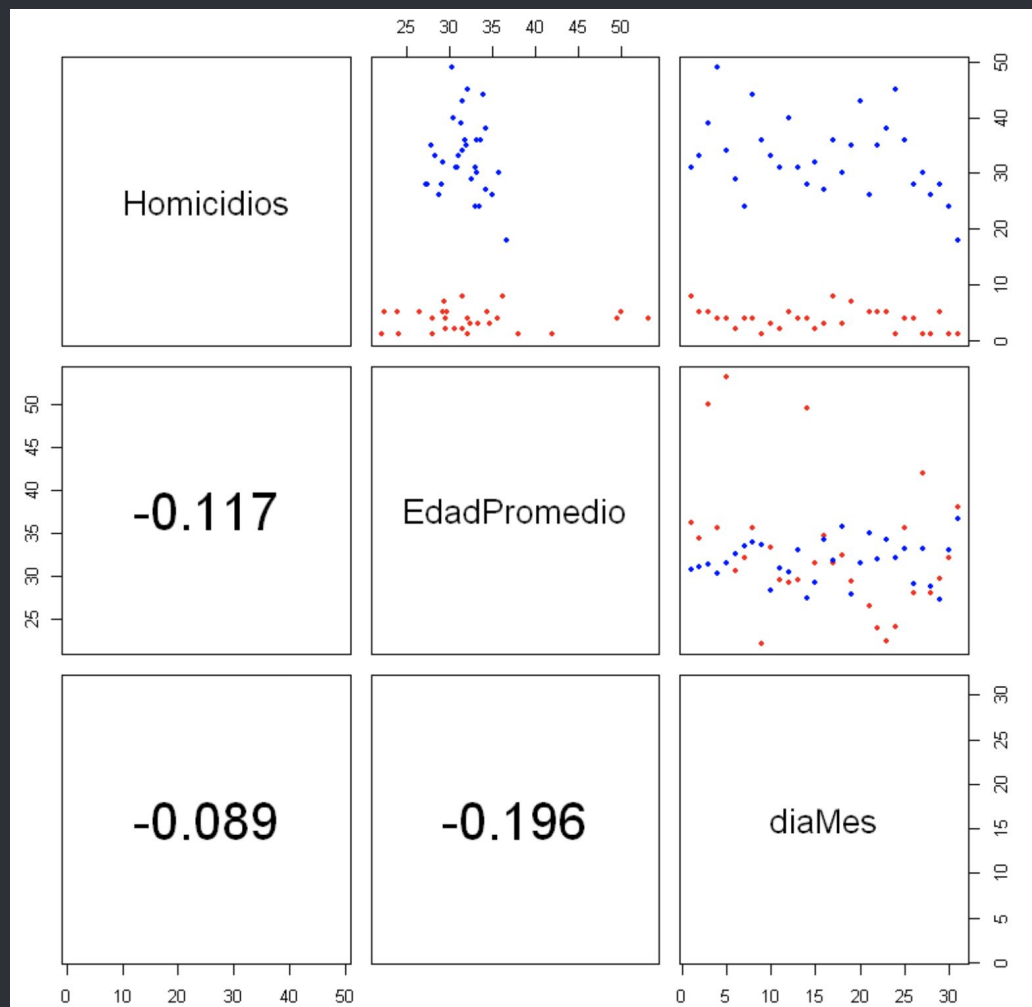
<b>diaMes</b>	<b>Sexo</b>	<b>Edad Promedio</b>	<b>Homicidios</b>
1	F	36.25000	8
1	M	30.74194	31
2	F	34.40000	5
2	M	31.00000	33

## ● Entendimiento



La gráfica muestra la correlación entre la cantidad de Homicidios, la edad promedio y los días de los meses.

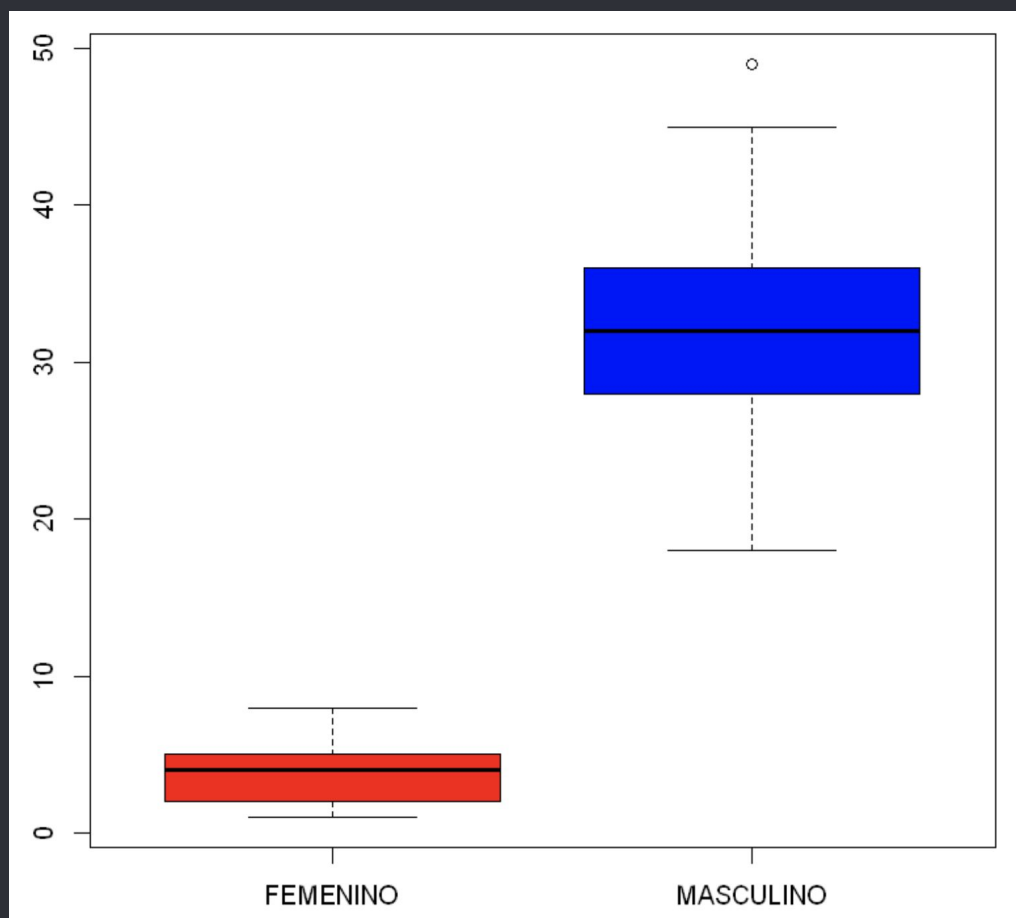
En la gráfica de dispersión entre se, identifican con **rojo** las víctimas de sexo **femenino** y en **azul** las de sexo **masculino**.



## ● Entendimiento



En el diagrama de caja se muestra la **cantidad** de homicidios respecto al sexo de la víctima. Es claro que la cantidad de víctimas hombres es mucho **mayor** que la cantidad de víctimas mujeres.



# Modelo



```
modelo.poisson=glm(Homicidios~.,family = poisson(log),data=Homicidios2017)
summary(modelo.poisson)
```

Call:

```
glm(formula = Homicidios ~ ., family = poisson(log), data = Homicidios2017)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.04142	-0.91276	-0.03448	0.61020	2.46462

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	1.568056	0.322155	4.867	1.13e-06	***
diaMes	-0.009038	0.003330	-2.714	0.00665	**
SexoMASCULINO	2.180402	0.100693	21.654	< 2e-16	***
EdadPromedio	-0.003617	0.009185	-0.394	0.69369	

---

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 886.762 on 60 degrees of freedom  
Residual deviance: 68.066 on 57 degrees of freedom  
AIC: 331.8

Number of Fisher Scoring iterations: 4

## Modelo

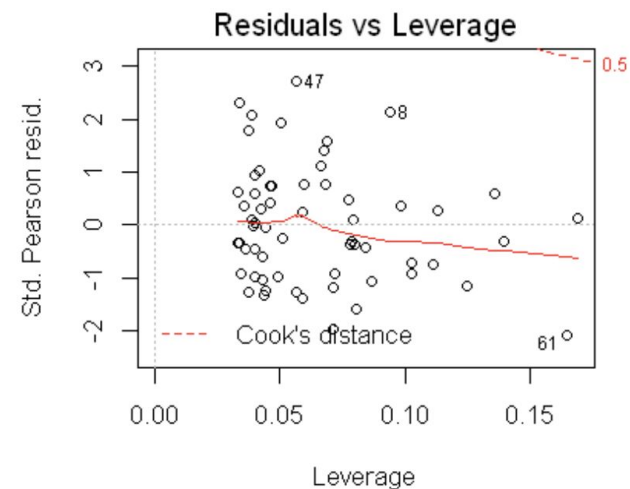
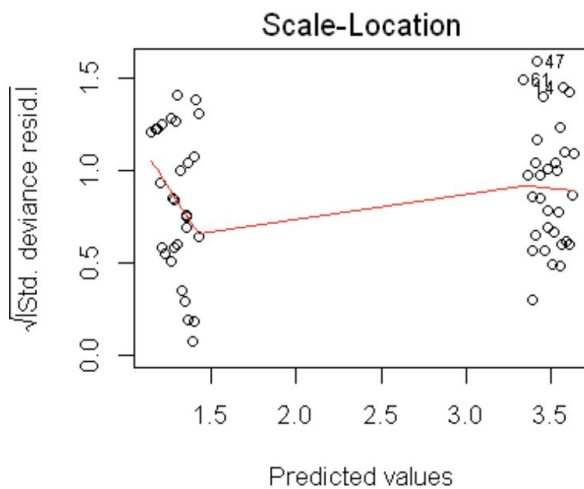
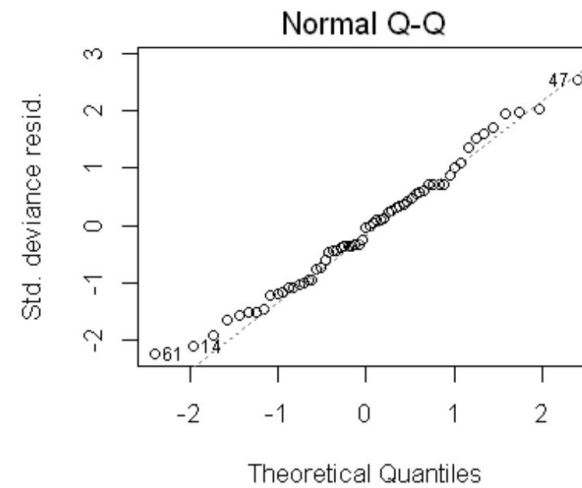
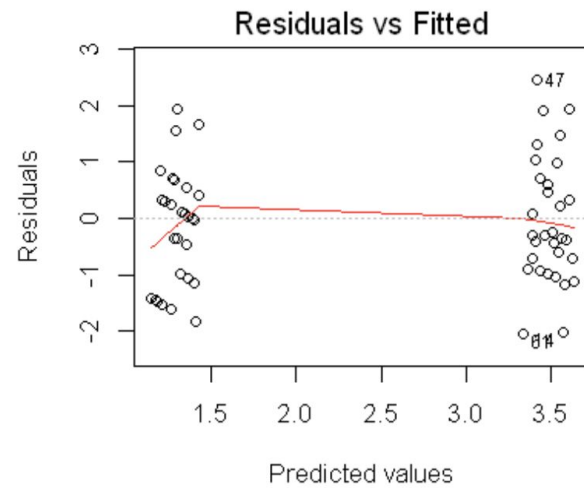


La ecuación para el **conteo** de homicidios queda de la siguiente manera:

$$Homicidios_M = e^{1.568 - 0.009diaMes + 2.18Sexo[Masculino] - 0.003EdadPromedio}$$

$$Homicidios_F = e^{1.568 - 0.009diaMes - 0.003EdadPromedio}$$

# ● Análisis





## Validación



Para la validación del modelo se cargan los datos de Homicidios ocurridos en el año 2018.

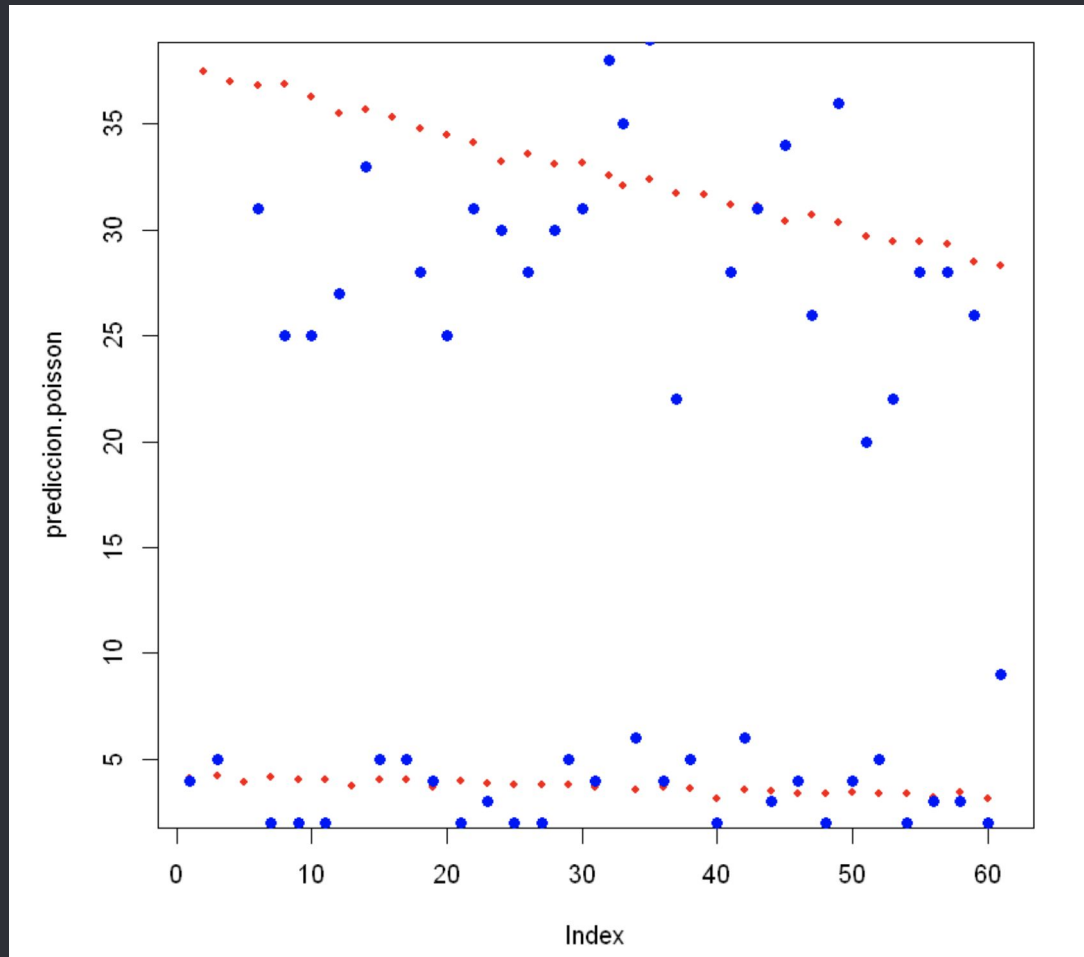
En el formato inicial:

diaMes	Sexo	Edad Promedio	Homicidios
1	F	41.75000	4
1	M	31.87500	48
2	F	31.80000	5
2	M	32.81818	44

## Resultados



En la gráfica se muestra en rojo la **predicción** realizada por el modelo de Homicidios en 2018, y en azul los homicidios ocurridos realmente en ese año.





## *An Introduction to Statistical Learning with Applications in R*

*Datos:* <https://www.datos.gov.co/Seguridad-y-Defensa/Homicidios-2017/mkw6-468s>

<https://www.datos.gov.co/Seguridad-y-Defensa/Homicidios-2018/ndtu-iqtk>

*Código:* <https://github.com/jhalep/ProyectoFinalHM/blob/master/Proyecto%20Final.ipynb>