

Los Modelos de Poisson (o binomial negativa)

La regresión para las variables de conteo

Los tipos de modelos



Gaussianos

(* Exponencial

Gamma

Poisson
Binomial Negativa

Logístico Binomial

Logístico Multinomial

Logístico Ordinal

Variable Respuesta	Ejemplo	Rango	Función de enlace	Modelo
Cuantitativa Continua	Altura en cm	$(-\infty, +\infty)$	1	Modelo de regresión lineal generalizada
Cuantitativa Discreta tratada como una continua	Sueldo anual en €	$(0, +\infty)$	1	
Discreta como Contaje positivo	Número de centros comerciales por provincia	$[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, \dots]$	log	Modelo de Poisson
Cualitativa dicotómica	Lesionado o no	$[0, 1]$ o $[YES, NO]$	logit	Modelo logístico
Cualitativa Politémica	El trabajo en que estás puede ser: Fábrica, Oficina, Aire libre	Fábrica, Oficina, Aire Libre o $[0, 1, 2]$		
Cualitativa Ordinal	Rangos de Edad	25-34, 35-44, 45-54, 55-64, 65-74, 75+		

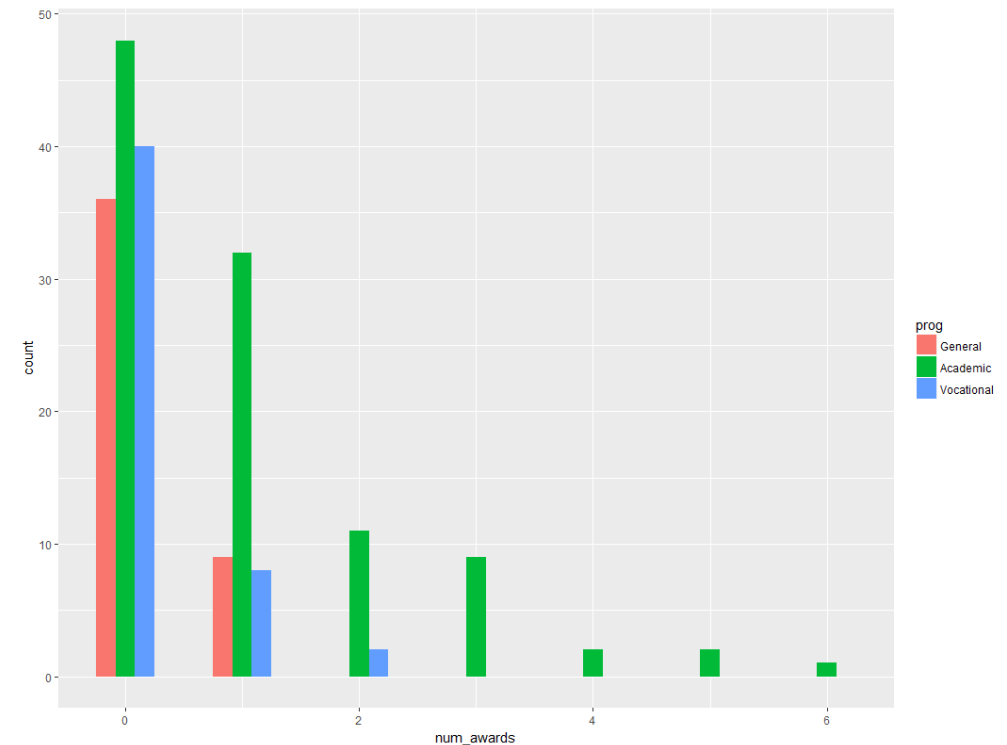
GLM con variables de conteo

Se han listado diferentes colegios dónde se han recogido las variables siguientes:

- **num_awards** es la variable respuesta (de salida) e indica la cantidad de premios obtenidos por los estudiantes en una escuela secundaria en un año
- **Math**: las matemáticas son una variable predictora continua y representan las notas de los estudiantes en su examen final de matemáticas
- **Prog** es una variable predictora categórica con tres niveles el tipo de programa en el que los estudiantes se inscribieron. Está codificada como 1 = "General", 2 = "Académico" y 3 = "Vocacional"

OBJETIVO:
Explicar los datos como una relación lineal entre 1 variable de conteo como entrada y variables de entrada (numéricas y cualitativas)

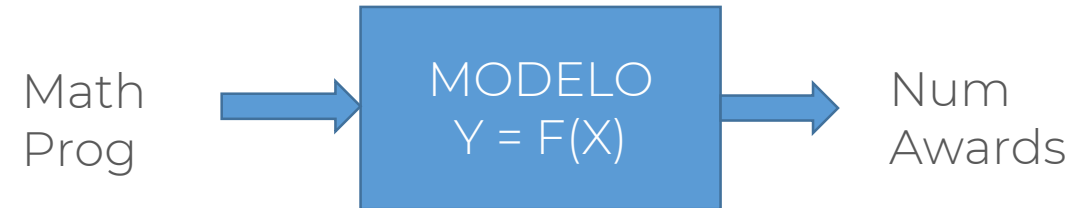
Queremos explicar como influye la nota en matemáticas, Math y el tipo de programa educativo, Prog



GLM con variables de conteaje

OBJETIVO:
Explicar los datos como una relación lineal entre 1 variable de conteaje como entrada y variables de entrada (numéricas y cualitativas)

Variables de Entrada Variables Independientes	<p>TODO TIPO DE VARIABLES</p> <p>Ej: Math y Prog: Prog: 0 = General 1 = Academico 2 = Vocacional</p>
Variable de Salida Variable Dependiente	<p>VARIABLE DISCRETA DE CONTAJE</p> <p>Ej: Num Awards</p>
Modelo Función matemática	<p>$\text{Log}(Y) = a_1 \cdot X_1 + a_2 \cdot X_2 + a_3 \cdot X_3 + a_4 \cdot X_4 + b$</p> <p>Ej: $\log(\text{num_awards}) = a_1 \cdot \text{math} + a_2 \cdot \text{ProgAcad} + a_3 \cdot \text{ProgVoc} + b + \text{error}$</p>
Restricciones	<p>Variable de conteaje</p> <p>Los residuos normales con media 0</p> <p>Varianzas de los residuos constantes</p> <p>No presentan colinealidad. VIF < 4</p> <p>No sobredispersión</p>



$$\text{Log}(\text{NumAwards}) = -5,24712 + 1,08386 \cdot \text{ProgAcad} + 0,36981 \cdot \text{ProgVoc} + 0,07015 \cdot \text{Math} + \text{Error}$$

Call:
glm(formula = num_awards ~ prog + math, family = "poisson", data = varEstudio)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.2043	-0.8436	-0.5106	0.2558	2.6796

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.24712	0.65845	-7.969	1.60e-15 ***
progAcademic	1.08386	0.35825	3.025	0.00248 **
progVocational	0.36981	0.44107	0.838	0.40179
math	0.07015	0.01060	6.619	3.63e-11 ***

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

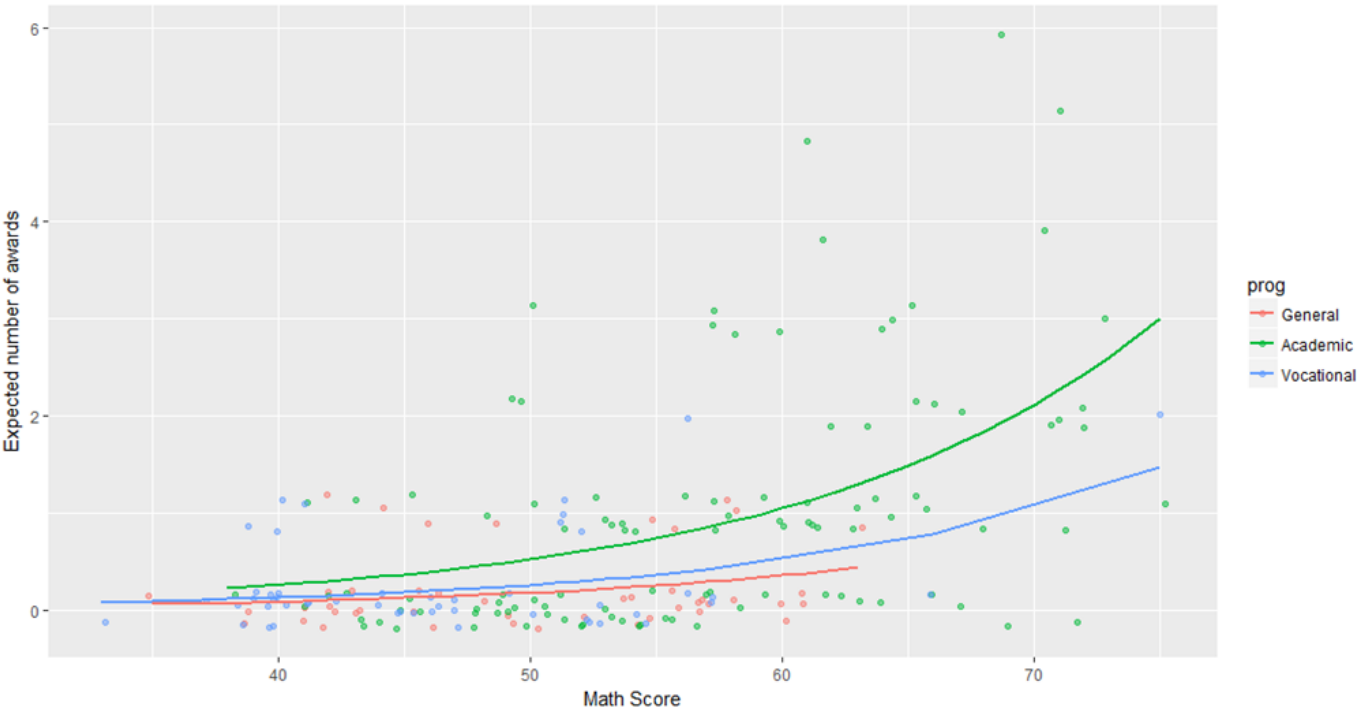
(Dispersion parameter for poisson family taken to be 1)

Null deviance: 287.67 on 199 degrees of freedom
Residual deviance: 189.45 on 196 degrees of freedom
AIC: 373.5

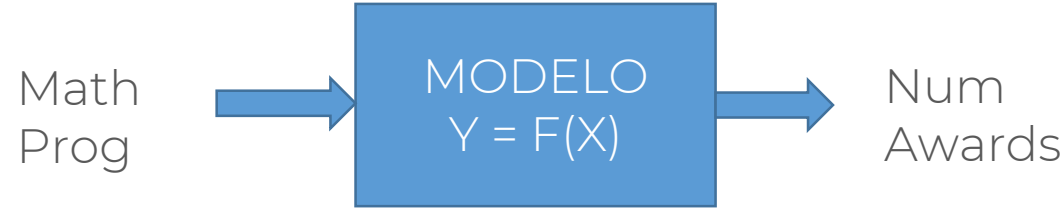
Number of Fisher scoring iterations: 6

GLM con variables de conteo

OBJETIVO:
Explicar los datos como una relación lineal entre 1 variable de conteo como entrada y variables de entrada (numéricas y cualitativas)



Coefficientes de dispersión >1 → binomial negativa



$$\text{Log}(\text{NumAwards}) = -5,24712 + 1,08386 \cdot \text{ProgAcad} + 0,36981 \cdot \text{ProgVoc} + 0,07015 \cdot \text{Math} + \text{Error}$$

```
Call:
glm(formula = num_awards ~ prog + math, family = "quasipoisson",
    data = varEstudio)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2043  -0.8436  -0.5106   0.2558   2.6796

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -5.24712    0.68503  -7.660 8.36e-13 ***
progAcademic   1.08386    0.37272   2.908 0.00406 **
progVocational  0.36981    0.45888   0.806 0.42128
math           0.07015    0.01103   6.362 1.38e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.082366)

Null deviance: 287.67  on 199  degrees of freedom
Residual deviance: 189.45  on 196  degrees of freedom
AIC: NA
```

Take away

El resumen de los modelos logísticos

Lo más importante de la lección

- Los modelos de Poisson son para variables de conteo
- La transformación es logarítmica en la respuesta (función de enlace)
- Tenemos que vigilar los residuos y la sobredispersión
 - Si el coeficiente de dispersión >1 → Binomial Negativa (Zona Tech Zen te lo cuenta)

Tú turno

A asentar esta lección con un ejercicio práctico

A poner en práctica lo que has visto

- Descarga la hoja de trabajo
- Rellena los ejercicios de los modelos de Poisson
- ¡Te ayudará a entender este interesante modelo!