

EL MODELO DE POISSON

Si la variable respuesta es una variable de conteo el mejor modelo es el modelo de Poisson.

Las variables de conteo tienen la particularidad que son siempre positivas y son discretas.

Un modelo gaussiano o de regresión como el que has visto no es suficiente.

En este ejercicio quiero que trabajes el modelo que has visto en los vídeos para que lo entiendas paso a paso.

¡Empezamos!

EL OBJETIVO DEL MODELO

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".

La variable de salida es el num_awards

Las variables de entrada son Math y Prog

¿Según este modelo qué objetivo perseguimos? ¿Qué podremos ver?

Queremos calcular un modelo que nos haga de predictor del número de premios en función de la nota de matemáticas el programa en el que estaban los estudiantes
Podremos calcular el número de premios en función de estas variables

Calcula el modelo de Poisson y comprueba las restricciones del mismo.

```
Log(numAward)=0.07*math+1.08(progacademic-5.24
```

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

EL MODELO DE POISSON

(Dispersion parameter for poisson family taken to be 1)

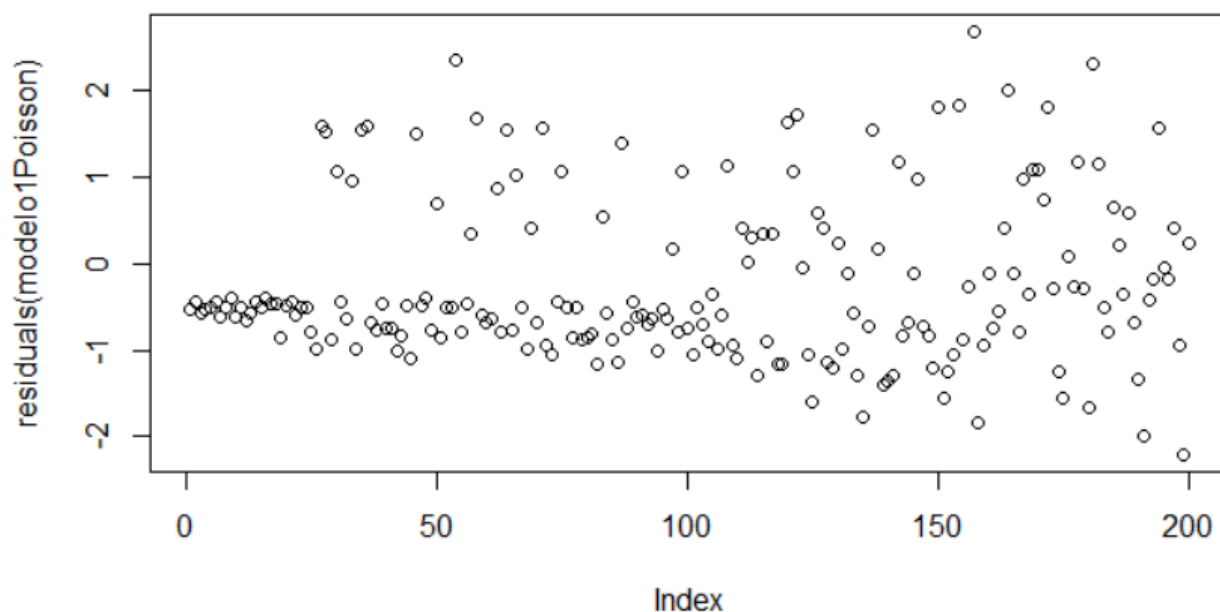
	Estimate	Robust SE	Pr(> z)	LL	UL
(Intercept)	-5.2471244	0.64599839	4.566630e-16	-6.51328124	-3.98096756
progAcademic	1.0838591	0.32104816	7.354745e-04	0.45460476	1.71311353
progVocational	0.3698092	0.40041731	3.557157e-01	-0.41500870	1.15462716
math	0.0701524	0.01043516	1.783975e-11	0.04969947	0.09060532

Los IC de Math y prog academic no incluyen el 0

Aplicando el test de CHI cuadrado

	res.deviance	df	p
[1,]	189.4496	196	0.6182274

La p es grande por lo que el modelo es aceptable



¿Tenemos sobre dispersión?

Comprueba si el modelo tiene sobre dispersión

(Dispersion parameter for quasipoisson family taken to be 1.082366)

La dispersión es muy cercana a 1 por lo que no hay sobredispersión

EL MODELO DE POISSON

En ambos casos (con sobre dispersión o no) Calcula el modelo de binomial negativa o quasipoisson. Este modelo nos permite ser más finos en caso de sobre dispersión (>1):

Copia el resultado del modelo

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.29331	0.70416	-7.517	5.60e-14 ***
progAcademic	1.07507	0.36699	2.929	0.0034 **
progVocational	0.36696	0.45230	0.811	0.4172
math	0.07107	0.01153	6.165	7.05e-10 ***

Son significativos las variables: math que afecta positivamente y si el programa es c academico, no influye si el programa es vocational o similar a normal.

	Estimate	Robust SE	Pr(> z)	LL	UL
(Intercept)	-5.29331242	0.64911459	3.501475e-16	-6.56557703	-4.02104782
progAcademic	1.07507108	0.32138089	8.223459e-04	0.44516453	1.70497764
progVocational	0.36695930	0.40391002	3.636049e-01	-0.42470434	1.15862294 (Incluye 0)
math	0.07106875	0.01044062	9.969108e-12	0.05060514	0.09153236

El programa vocational incluye el 0 en el intervalo de confianza . el resto de los factores no

Aplicando el test de Chi cuadrado:

	res.deviance	df	p
[1,]	169.7625	196	0.9123467

La P es grande por lo que el modelo es aceptable

Modelo quasipoisson

Coefficients:

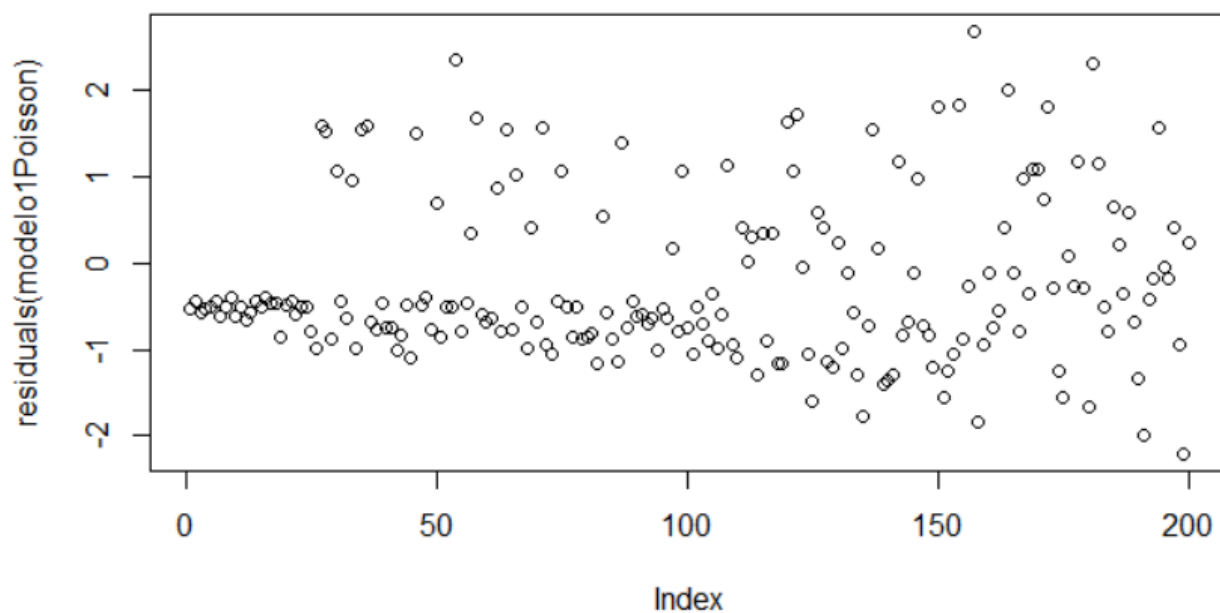
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.24712	0.68503	-7.660	8.36e-13 ***
progAcademic	1.08386	0.37272	2.908	0.00406 **
progVocational	0.36981	0.45888	0.806	0.42128
math	0.07015	0.01103	6.362	1.38e-09 ***

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

EL MODELO DE POISSON

(Dispersion parameter for quasipoisson family taken to be 1.082366)

El coeficiente de dispersion es 1.08, lo que significa que tenemos un poquito de sobredispersión



¿Hay diferencias entre los coeficientes del modelo de Poisson y la binomial negativa o quasi poisson?

Describe las diferencias en los coeficientes que ves

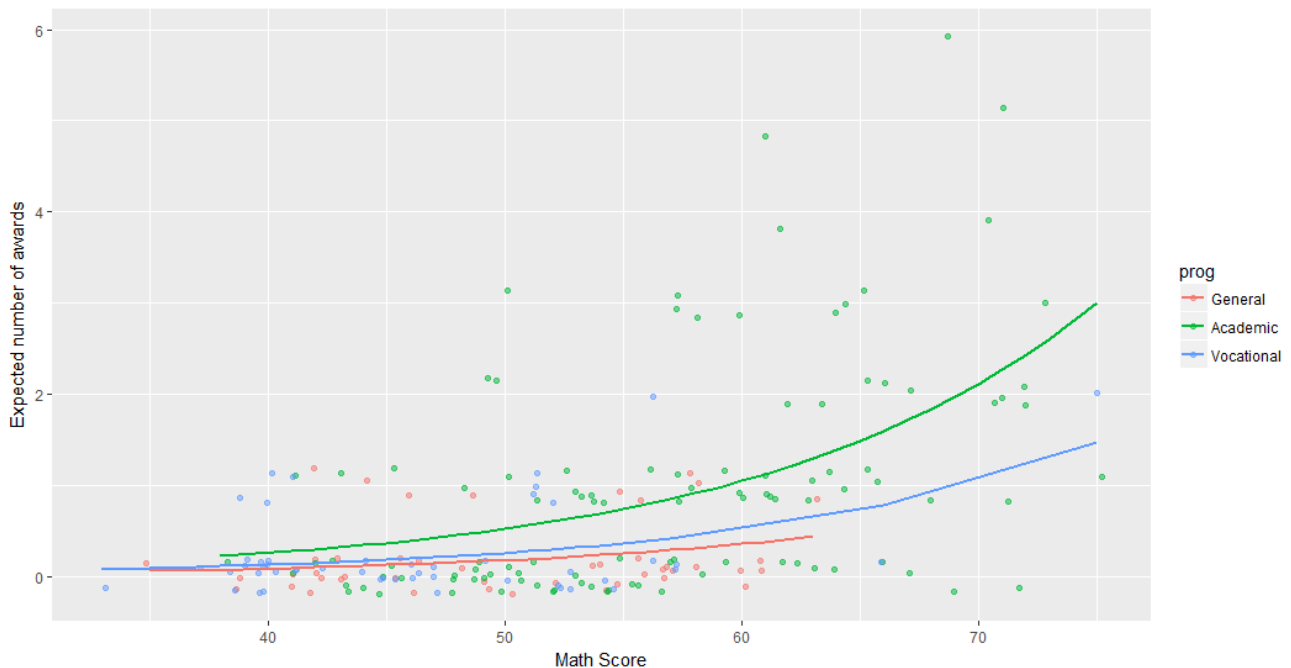
No existen apenas diferencias en los coeficientes de los 2 modelos, solo el de quasipoisson los coeficientes son ligeramente mayores

INTERPRETA EL MODELO

Ahora que has calculado el modelo es momento de interpretarlo.

Te resumo los resultados del modelo para que te sea más fácil esta parte:

En el siguiente gráfico puedes ver el número de premios que predice el modelo en función de la nota en matemáticas y el tipo de programa.



Y este es el resultado del modelo:

```
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

Number of Fisher Scoring iterations: 6
```



EL MODELO DE POISSON

¿Qué puedes afirmar sobre la influencia de la variable math en el modelo? ¿Cómo lo interpretas?

Explica lo que ves

Math influye positivamente en el numero de premios, a mayor math mayor numero de premios, por eso el factor positivo 0.07(el factor es significativo)

La interpretacion del modelo podría ser que a mayor math mayor numero de premios en el programa académico

El programa general y el vocacional son muy parecidos (el factor académico no influye de forma en el numero de premio(no significativa))

¿Qué puedes afirmar sobre la influencia de la variable prog en el modelo? ¿Cómo lo interpretas?

Explica si hay diferencias en los tres programas y por qué

Hay diferencias en el programa académico respecto al general y al vocacional, ya que este ultimo no influye en el modelo y es muy similar al programa general.

El programa académico obtiene mayor numero de premios que el resto de los programas (coef 1.08 significativo)

Ahora ya has aplicado por primera vez un modelo de Poisson.

Estos ejercicios te están ayudando a ver las posibilidades de los modelos y su flexibilidad.

¡Sigue así!