

Los Modelos Logísticos

El modelo binomial, multinomial y ordinal

Los tipos de modelos



Gaussianos
 (*) Exponencial
 Gamma
 Poisson
 Binomial Negativa

Logístico Binomial

Logístico Multinomial

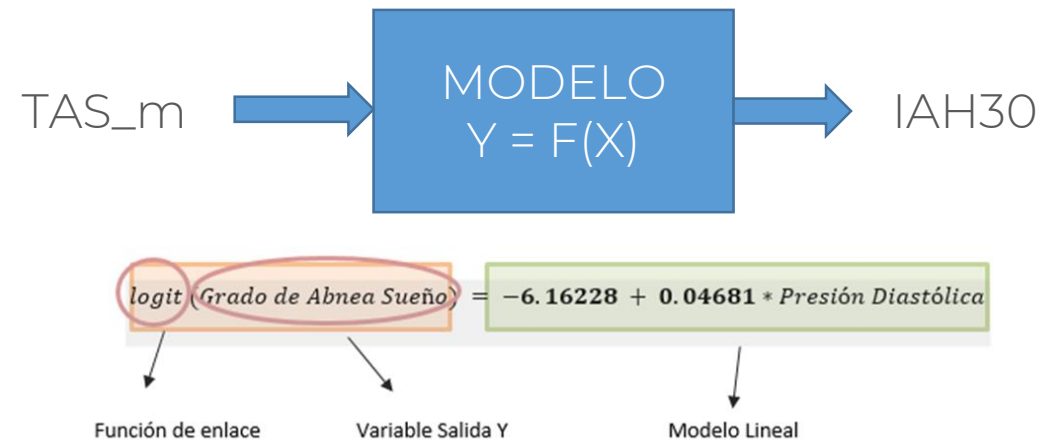
Logístico Ordinal

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 cualitativas BINOMIAL – DOS GRUPOS

OBJETIVO:
Explicar los datos como una relación lineal entre 1 entrada y 1 salida de medidas continuas

<p>Variables de Entrada Variables Independientes</p>	<p>VARIABLE CONTÍNUA</p> <p>Ej: Grado de Apnea del Sueño, IAH30:</p> <ul style="list-style-type: none"> IAH30 = 0 → Los que hacen menos de 30 abneas por hora → Pacientes Leves IAH30 = 1 → Los que hacen más de 30 abneas por hora → Pacientes Graves
<p>Variable de Salida Variable Dependiente</p>	<p>VARIABLE CONTÍNUA</p> <p>Ej: Presión Diastólica, TAS_m</p>
<p>Modelo Función matemática</p>	<p>$\text{Logit}(Y) = a \cdot X1 + b$</p> <p>Ej: $\text{logit}(\text{IAH30}) = a \cdot \text{TAS_m} + b + \text{error}$</p>
<p>Restricciones</p>	<p>Los residuos normales con media 0</p> <p>Varianzas de los residuos constantes</p> <p>No presentan colinealidad . VIF < 4</p>



Call:
glm(formula = IAH30 ~ TAS_m, family = "binomial", data = varEstudio)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.9949	-1.0487	0.6106	1.0163	1.7737

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-6.16228	1.47788	-4.170	3.05e-05 ***
TAS_m	0.04681	0.01093	4.284	1.84e-05 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 218.12 on 157 degrees of freedom
Residual deviance: 196.42 on 156 degrees of freedom
(1 observation deleted due to missingness)
AIC: 200.42

TAS_m	IAH30
160	1
130	1
150	1
132	0
166	1
147	0
138	1
143	0
165	1
165	0

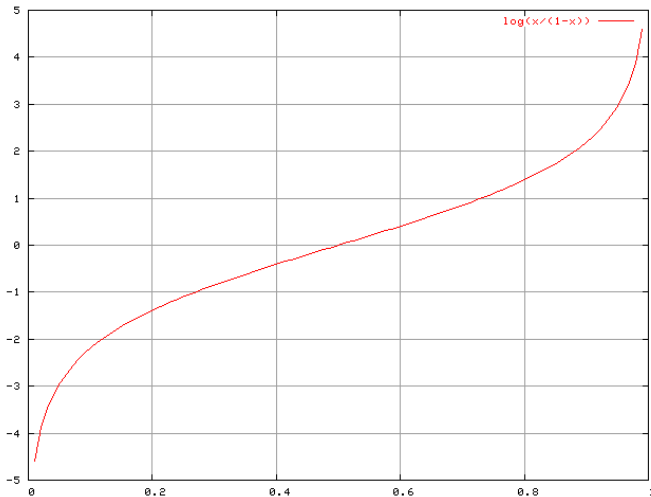
GLM con variables cualitativas BINOMIAL – DOS GRUPOS

OBJETIVO:

Explicar los datos como una relación lineal entre 1 entrada cualitativa y más de una variable de salida de diferentes tipos

$$\text{logit}(p(\text{IAH30})) = \ln \left(\frac{p(\text{IAH30})}{1 - p(\text{IAH30})} \right) = 0.04681 \cdot \text{TAS}_m - 6.16228 + \text{error}$$

$$p(\text{IAH30}) = \frac{1}{1 + e^{0.04681 \cdot \text{TAS}_m - 6.16228}}$$



- Utilidad: sabiendo la presión sistólica del paciente podemos saber la probabilidad de ser un paciente grave o no.
- La ecuación de la recta logística:

$$\text{logit}(\text{IAH30}) = 0.0468 \cdot \text{TAS}_m - 6.16$$

$$\text{logit}(\text{Probabilidad Grave}) = -6.16228 + 0.04681 \cdot 170 = 0.8576$$

$$\text{logit}^{-1}(0.8576) = \frac{1}{1 + e^{-0.8576}} = 0.7021 = 70.21\%$$

La probabilidad de ser Grave si tienes una Presión sistólica de 170 es de 70.21 %

La regresión Logística Binomial como clasificador

OBJETIVO:
Explicar los datos como una relación lineal entre 1 entrada cualitativa y más de una variable de salida de diferentes tipos

Entrada
Variables –
hormonas
etc.



Salida
Probabilidad de
ser positivo

$P > 90\% \rightarrow$ Positivo

$P < 90\% \rightarrow$ Negativo

Modelo clasificador

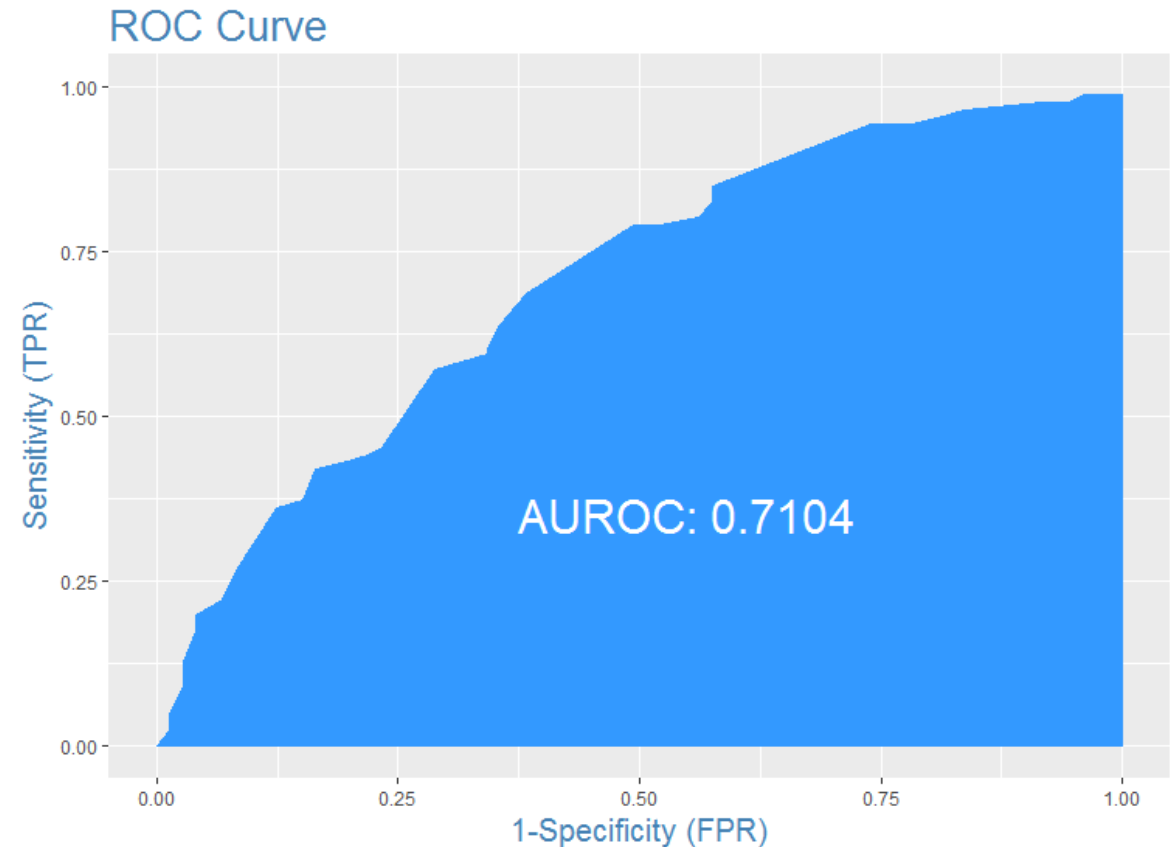
La regresión Logística Binomial como clasificador

- Un clasificador te permite poder saber en qué grupos te encuentras en función de una probabilidad
- Si ponemos un threshold = 50%
 - Si la probabilidad calculada $> 50\%$ → Es grave
 - Si la probabilidad calculada $< 50\%$ → NO es grave
- Podemos comparar los valores observados con los valores esperados y contar el número de aciertos y fracasos del clasificador
- Podremos contar el número de aciertos según el modelo

OBJETIVO:

Explicar los datos como una relación lineal entre 1 entrada cualitativa y más de una variable de salida de diferentes tipos

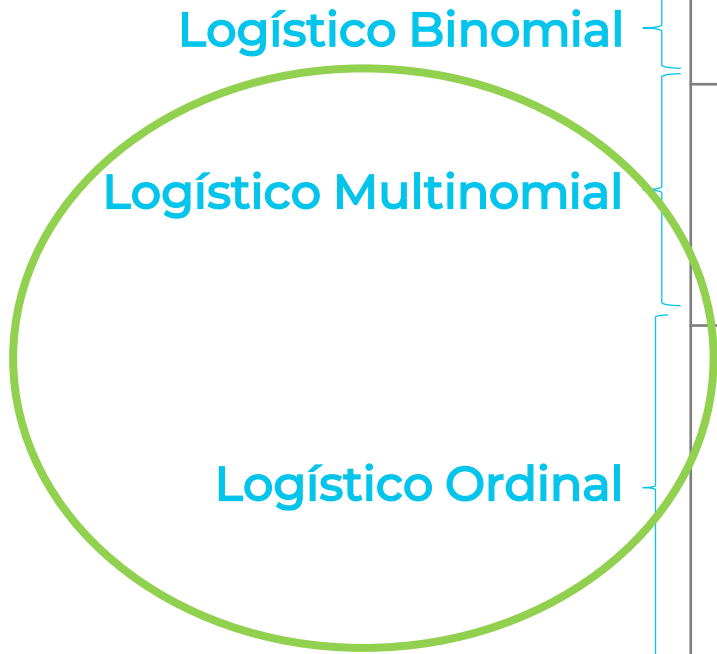
- Curva ROC



Los tipos de modelos



Gaussianos
(* Exponencial
Gamma
Poisson
Binomial Negativa



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+		

Take away

El resumen de la 1/2 lección

Lo más importante de la lección

- Un modelo logístico es un modelo que permite trabajar con variables cualitativas
 - Cualitativas Dicotómicas → Logístico Binomial
 - Cualitativas Politémicas → Logístico Multinomial
 - Cualitativas Ordinales → Logístico Ordinal
- La función de enlace es la logit
- La salida del modelo es una probabilidad
- La curva ROC nos permite analizar el rendimiento como clasificador en casos de modelos logísticos binomiales

Tú turno

A asentar esta lección con un ejercicio conceptual

A poner en práctica lo que has visto

- Descarga la hoja de trabajo
- Trabaja los ejercicios del modelo logístico y la curva ROC para familiarizarte con estos modelos
- ¡Te ayudará a entenderlo mejor!