Prof. Luiz Paulo Lopes Fávero

PRINTS DURANTE A AULA 20/08/2024

Obs	Dummy (Y)	X_1	X_2	...	X_K
M	1 SIM				
J	0 NÃO				
P	1 SIM				
A	1 SIM				
.	0 NÃO				
.	0 NÃO				
.	1 SIM				
.	.				
n	1 SIM				

Y:
1: evento (SIM)
0: NÃO evento (NÃO)

2 dicotômica
2 reg. logística binária!

p: probabilidade do evento.

1 - p: probabilidade do NÃO evento.

$$\text{chance (odds)} = \frac{p}{1-p} \quad \begin{matrix}\text{evento}\\ \text{NÃO evento}\end{matrix}$$

$p = 0,80 \rightarrow \text{chance} = \frac{4}{1} = 4$

$p = 0,25 \rightarrow \text{chance} = \frac{0,25}{0,75} = \frac{1}{3}$

$p = 0,50 \rightarrow \text{chance} = \frac{1}{1} = 1$

MQO (OLS):

$$\hat{y}_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki}$$

Logística Binária: (Hosmer & Lemeshow, Communications in Statistics, 1980)

$$\ln(\text{chance}_i) = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki}$$

z: logito

$$\ln\left(\frac{p}{1-p}\right) = z$$

$$\frac{p}{1-p} = e^z$$

$$p(1+e^z) = e^z$$

$$p = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$ Prob. evento

$$p_i = \frac{1}{1+e^{-(\alpha + \beta_1 X_{1i} + \cdots + \beta_K X_{Ki})}}$$

Origem: Juros Compostos (Mat. Financeira)

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n \cong 2{,}71828$$

(no começo séc 18)

Número Euler

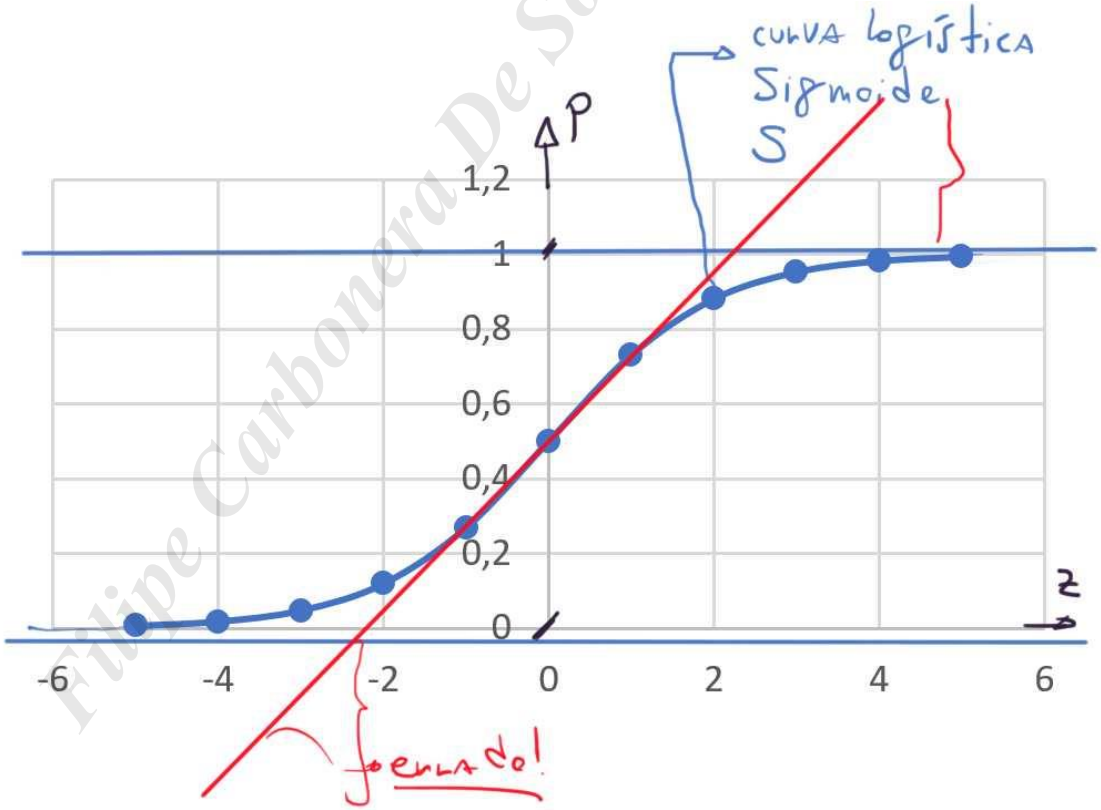
Número Napier (no final séc 16) } "o descobridor de logaritmos".

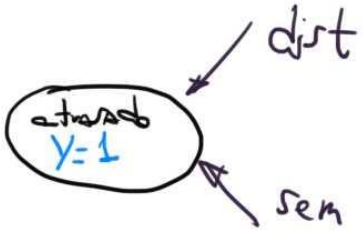
$$\ln = \log_e$$

$$P_{\substack{\text{evento}\\ y=1}} = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

$$P_{\substack{\text{não evento}\\ y=0}} = 1 - \frac{e^z}{1+e^z} = \frac{1}{1+e^z}$$

$$\text{Soma} = 1.$$





$$P_{\substack{atrasado=sim \\ y=1 \quad i}} = \frac{1}{1 + e^{-(\alpha + \beta_1 . dist_i + \beta_2 . sem_i)}}$$

$$p(Y_i) = p_i^{Y_i} . (1 - p_i)^{1-Y_i}$$

Distribuição Bernoulli:

$$p(1) = p^1 . (1-p)^0 = p$$

$$p(0) = p^0 . (1-p)^1 = 1 - p$$

α	-26,16
β_1	0,19
β_2	2,36

Somatória LL_i	-50,46638

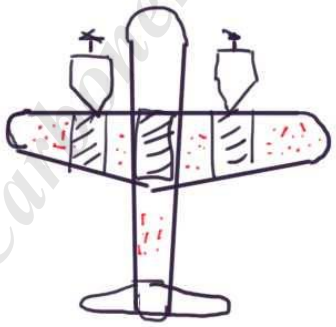
Generalized Linear Model Regression Results

Dep. Variable:	atrasado	No. Observations:	100
Model:	GLM	Df Residuals:	97
Model Family:	Binomial	Df Model:	2
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-50.466
Date:	Tue, 20 Aug 2024	Deviance:	100.93
Time:	21:34:51	Pearson chi2:	86.7
No. Iterations:	7	Pseudo R-squ. (CS):	0.2913
Covariance Type:	nonrobust		

	coef	std err	z (WALD)	P>|z|	[0.025	0.975]
Intercept (α)	-26.1665	8.442	-3.100	0.002	-42.712	-9.621
dist (β_1)	0.1904	0.076	2.493	0.013	0.041	0.340
sem (β_2)	2.3629	0.795	2.972	0.003	0.804	3.921

$$P_{atrasado_i\ (Y=1)} = \frac{1}{1 + e^{-(-26,16 + 0,19 \cdot dist_i + 2,36 \cdot sem_i)}}$$

Z de WALD: (1902 - 1950)



Logit Regression Results

Dep. Variable:	atrasado	No. Observations:	100
Model:	Logit	Df Residuals:	97
Method: (maximum likelihood est.)	MLE	Df Model:	2
Date:	Tue, 20 Aug 2024	Pseudo R-squ.:	0.2544
Time:	22:17:05	Log-Likelihood:	-50.466
converged:	True	LL-Null:	-67.686
Covariance Type:	nonrobust	LLR p-value: (Análogo p-value F)	3.324e-08 (χ^2)

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept (α)	-26.1665	8.442	-3.100	0.002	-42.713	-9.620
dist (β_1)	0.1904	0.076	2.493	0.013	0.041	0.340
sem (β_2)	2.3629	0.795	2.972	0.003	0.804	3.921

$$\chi^2 = -2.(LL_0 - LL_m)$$ ↑ melhor

$$\text{pseudo } R^2 \text{ (McFadden)} = \frac{-2LL_0 - (-2LL_m)}{-2LL_0}$$ ↓ melhor

$$AIC = -2LL_m + 2.(k+1) \quad (k+1 = \beta's)$$

$$BIC = -2.LL_m + (k+1).\ln(n) \quad (k+1 = \beta's)$$

Cutoff:

Se phat $\geq$ cutoff $\Rightarrow$ Evento

Se phat $<$ cutoff $\Rightarrow$ NÃO Evento

Matriz de Confusão: (Classificação Cruzada)

Cutoff = 0,5

	True (Real) 1	True (Real) 0
classified (previsto) 1	46	16
classified (previsto) 0	13	25
	59	41

Sensitividade (Recall) = $\frac{46}{59}$ | Especificidade = $\frac{25}{41}$ | Acurácia (EGN) = $\frac{46+25}{100}$

Log Lik (LLf)
pseudo R^2
AIC
BIC

} Indicador performance do modelo (independem do cutoff).

Sensitividade
Especificidade
Acurácia

} Dependem do cutoff.