

Introdução à Regressão Logística

Parte 2 – Modelos Logitos

Unidade III

Modelos Logitos para Variáveis Qualitativas

Exemplo 1 - *Uso de Cartão de Crédito*

Em uma pesquisa de mercado para uma certa firma, deseja-se analisar os dados sobre uso de cartão de crédito e renda anual.

Coletou-se dados sobre Renda em 1000 u.m. e uso de cartão de crédito (sim / não)

- Categorias de renda: < 25 000 u.m.;
≥ 25,000 u.m.;

Variável
Resposta

Variável Explicativa

Renda	Uso		Total
	Não	Sim	
< 25	200 (67)	100 (33)	300 (100)
≥ 25	100 (20)	400 (80)	500 (100)
Total	300 (37,5)	500 (62,5)	800 (100)

Variável Explicativa Qualitativa Binária

- O uso de variáveis indicadoras (“*dummy*”) permite a inclusão de variáveis qualitativas, denominadas fatores, no modelo.

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}$$

sendo que: $X_i = 0$ se renda < 25
 $X_i = 1$ se renda ≥ 25

- Resposta média logito:

$$\pi^*(x_i) = \ln \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right) = \beta_0 + \beta_1 x_i$$

- Note que se $X_i = 0$ então:

$$\ln \left(\frac{\pi(0)}{1 - \pi(0)} \right) = \beta_0 + \beta_1 \times 0 = \beta_0 \Rightarrow \frac{\pi(0)}{1 - \pi(0)} = e^{\beta_0}$$

e se $X_i = 1$ então:

$$\ln \left(\frac{\pi(1)}{1 - \pi(1)} \right) = \beta_0 + \beta_1 \times 1 = \beta_0 + \beta_1 \Rightarrow \frac{\pi(1)}{1 - \pi(1)} = e^{\beta_0 + \beta_1}$$

- Consequentemente:

$$\frac{\left(\frac{\pi(1)}{1 - \pi(1)} \right)}{\left(\frac{\pi(0)}{1 - \pi(0)} \right)} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1} = \theta$$

que mede o quanto mais provável (ou improvável) é o “sucesso” entre os com nível $x = 1$ do que entre os com nível $x = 0$.

- Cabe ressaltar que este resultado só vale quando a variável for codificada como 0 ou 1.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1060.501	886.311
SC	1065.186	895.680
-2 Log L	1058.501	882.311

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	176.1903	1	<.0001
Score	174.2222	1	<.0001
Wald	157.2385	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.6931	0.1225	32.0302	<.0001
renda	>=25	2.0794	0.1658	157.2385	<.0001

$$\hat{\beta}_1 = 2,0794 \Rightarrow \hat{\theta} = e^{2,0794} = 7,9997$$

Number of Observations Read	4
Number of Observations Used	4
Sum of Frequencies Read	800
Sum of Frequencies Used	800

Response Profile		
Ordered Value	cartao	Total Frequency
1	n	300
2	s	500

Probability modeled is cartao='s'.

Class Level Information		
Class	Value	Design Variables
renda	<25	0
	>=25	1

Efeito da codificação e o papel das variáveis *dummy*

X	Sucesso (Y=1)	Insucesso (Y=0)	odds	Logito ln(odds)
a	$\pi(a) = \frac{e^{\beta_0 + \beta_1 a}}{1 + e^{\beta_0 + \beta_1 a}}$	$1 - \pi(a) = \frac{1}{1 + e^{\beta_0 + \beta_1 a}}$	$\frac{\pi(a)}{1 - \pi(a)} = e^{\beta_0 + \beta_1 a}$	$\beta_0 + \beta_1 a$
b	$\pi(b) = \frac{e^{\beta_0 + \beta_1 b}}{1 + e^{\beta_0 + \beta_1 b}}$	$1 - \pi(b) = \frac{1}{1 + e^{\beta_0 + \beta_1 b}}$	$\frac{\pi(b)}{1 - \pi(b)} = e^{\beta_0 + \beta_1 b}$	$\beta_0 + \beta_1 b$

$$\ln(\theta) = \ln \left(\frac{\pi(a)}{1 - \pi(a)} \right) = \ln \left(\frac{\pi(a)}{1 - \pi(a)} \right) - \ln \left(\frac{\pi(b)}{1 - \pi(b)} \right) = \beta_0 + \beta_1 a - (\beta_0 + \beta_1 b) = \beta_1(a - b)$$

$$\ln(\theta) = \beta_1(a - b) \Rightarrow \theta = e^{\beta_1(a - b)}$$

Métodos de Construção de Variáveis *dummy*

- **Método parcial** – célula de referência – atribui 0 ao nível a ser considerado como referência e 1 ao outro nível.
- **Método marginal** – desvio da média – atribui -1 a um dos níveis e 1 ao outro nível.

Usando Método Marginal

Response Profile		
Ordered Value	cartao	Total Frequency
1	n	300
2	s	500

Probability modeled is cartao='s'.

Class Level Information		
Class	Value	Design Variables
renda	<25	-1
	>=25	1

b

a

Model Convergence Status	
Convergence criterion (GCONV=1E-8) satisfied.	

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1060.501	886.311
SC	1065.186	895.680
-2 Log L	1058.501	882.311

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	176.1903	1	<.0001
Score	174.2222	1	<.0001
Wald	157.2385	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.3466	0.0829	17.4707	<.0001
renda	>=25	1.0397	0.0829	157.2385	<.0001

$$\hat{\beta}_1 = 1.0397 \Rightarrow \hat{\theta} = e^{1.0397(1 - (-1))} = e^{2.0794} = 7.9997$$

EXEMPLO 2 - Coração x Idade

Em uma pesquisa deseja-se estudar se idade é um fator de risco para doença coronariana.

Coletou-se dados para uma amostra de 100 pacientes sobre:

- Doença coronariana (0 – não/ 1 – sim).
- Idade (< 55 anos / >=55 anos).

e os seguintes resultados foram obtidos:

Idade	Doença Coronariana		Total
	sim	não	
< 55 anos	22	51	73
>=55 anos	21	6	27
Total	43	57	100

Response Profile		
Ordered Value	coracao	Total Frequency
1	s	43
2	n	57

Probability modeled is coracao='s'.

Class Level Information		
Class	Value	Design Variables
idade	<55	0
	>=55	1

Model Convergence Status	
Convergence criterion (GCONV=1E-8) satisfied.	

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	138.663	121.959
SC	141.268	127.169
-2 Log L	136.663	117.959

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	18.7039	1	<.0001
Score	18.2516	1	<.0001
Wald	15.6898	1	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
idade	1	15.6898	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.8408	0.2551	10.8652	0.0010
idade	>=55	2.0935	0.5285	15.6898	<.0001

Odds Ratio Estimates		
Effect	Point Estimate	95% Wald Confidence Limits
idade >=55 vs <55	8.114	2.880 22.861

Variável Explicativa Qualitativa Politômica

- Em geral, se uma variável qualitativa tem ***l*** níveis possíveis, então ***l-1*** variáveis *dummy* ou de delineamento são necessárias.
- As ***l-1*** variáveis *dummy* serão designadas como D_u e os coeficientes destas variáveis serão designados por $\beta_u, u = 1, \dots, l-1$.

O modelo expresso em termos da função *logito* será:

$$\text{logito}(\pi) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \sum_{u=1}^{l-1} \beta_u D_u$$

Métodos de Construção de Variáveis *dummy*

Método parcial – célula de referência –

Este método faz todas as variáveis *dummy* igual a zero para o nível considerado de referência e então aloca uma única variável *dummy* igual a 1 para cada um dos outros níveis de X .

Níveis de X	Variáveis <i>Dummy</i>		
	D_1	D_2	D_3
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

Nível de referência

EXEMPLO 3 - Coração x Raça Variável qualitativa politômica

Em uma pesquisa deseja-se estudar se existe associação entre raça e doença coronariana.

Coletou-se dados para uma amostra de 100 pacientes sobre:

- Doença coronariana (0 – não/ 1 – sim).
- Raça (branca / negra / hispânica / outra)

e os resultados seguintes resultados foram obtidos:

Raça	Doença Coronariana		Total
	sim	nao	
Branca	5	20	25
Negra	20	10	30
Hispânica	15	10	25
Outra	10	10	20
Total	50	50	100

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3$$

Response Profile		
Ordered Value	coracao	Total Frequency
1	s	50
2	n	50

Probability modeled is coracao='s'.

Class Level Information				
Class	Value	Design Variables		
raca	branca	-1	-1	-1
	hispanic	1	0	0
	negra	0	1	0
	outra	0	0	1

Model Convergence Status	
Convergence criterion (GCONV=1E-8) satisfied.	

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.629	132.587
SC	143.235	143.008
-2 Log L	138.629	124.587

Testing Global Null Hypothesis: BETA=0				
Test	Chi-Square	DF	Pr > ChiSq	
Likelihood Ratio	14.0420	3	0.0028	
Score	13.3333	3	0.0040	
Wald	11.7715	3	0.0082	

Type 3 Analysis of Effects				
Effect	DF	Wald Chi-Square	Pr > ChiSq	
raca	3	11.7715	0.0082	

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.0719	0.2189	0.1079	0.7425
raca	hispanic	1	0.4774	0.3623	1.7363	0.1876
raca	negra	1	0.7651	0.3506	4.7619	0.0291
raca	outra	1	0.0719	0.3846	0.0350	0.8517

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
raca hispanic vs branca	6.000	1.693	21.261
raca negra vs branca	8.000	2.316	27.633
raca outra vs branca	4.000	1.074	14.895