Introdução à Regressão Logística

Parte 4 – **Modelo de Regressão Logística Múltiplo**

Unidade III

Modelo de Regressão Logística Múltiplo

- Seja o conjunto de *p* variáveis explicativas ou regressoras para variável resposta binária *Y* designado por

$$\boldsymbol{x} = (x_1, x_2, \dots, x_p)$$

- O objetivo é modelar a probabilidade de sucesso como uma função de valores fixados das variáveis explicativas, isto é,

$$\pi(\boldsymbol{x}) = P(Y = 1 | x_1, x_2, \dots, x_p)$$

- Sejam $x_{i1}, x_{i2}, \dots, x_{ip}$ valores fixados das variáveis explicativas, *i=1,2,...,c* onde x_{ij} representa um valor referente à variável X_j .

Modelo de Regressão Logística Múltiplo

O modelo será

$$\pi_i = \frac{\exp(\beta_o + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_o + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}, \qquad i = 1, \dots, c$$

- ***Transformação logito:***

$$\log ito(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

- O parâmetro β_j refere-se ao efeito do acréscimo de uma unidade em X_j sobre o *ln* da *odds* de *Y=1*, mantendo as outras variáveis explicativas constantes.

- Assim, $exp(\beta_j)$ é o efeito multiplicativo na chance (*odds*) de *Y*=1 para o acréscimo de 1 unidade em X_j, mantendo as demais variáveis explicativas constantes.

PROBLEMA

Caranguejo Ferradura

Amostra de 173 fêmeas . As seguintes características foram investigadas para cada uma delas:

- Z - cor da carapaça
  (1- clara média 2 - média
  3 - escura média 4 - escura)
- X- largura da carapaça da fêmea *em cm;*
- Y - *se a fêmea tem pelo menos um satélite (1 – sim e 0- não)*

Cor é um indicador para idade dos caranguejos: caranguejos mais velhos tendem a ser mais escuros. A categoria de referência será a *cor escura .*

O modelo a ser ajustado será:

$$\log ito(\pi) = \beta_0 + \beta_{11} D_1 + \beta_{12} D_2 + \beta_{13} D_3 + \beta_2 x$$

Modelo com Variável Independente Qualitativa

- Uma variável independente qualitativa tendo l níveis requer l-*1* variáveis *dummy*.

- Suponha que a *j-ésima* variável independente, X_j, tenha l_j níveis. Designando as l_j-*1* variáveis *dummy* como D_{iju} e os seus respectivos coeficientes por β_{ju}, $u = 1,2,\ldots,l_j - 1$

$$\log ito(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \ldots + \sum_{u=1}^{l_j - 1} \beta_{ju} D_{iju} + \ldots + \beta_p x_{ip}$$

- Em termos matriciais temos:

$$\boldsymbol{\beta}_{px1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad \boldsymbol{X}_{px1} = \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_{p-1} \end{bmatrix}$$

- Assim:

$$\mathbf{X}'\boldsymbol{\beta} = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1}$$

- E portanto:

$$E\{Y\} = \frac{\exp(\mathbf{X}'\boldsymbol{\beta})}{1 + \exp(\mathbf{X}'\boldsymbol{\beta})}$$

Estimação dos Parâmetros do Modelo

→ *Método de máxima verossimilhança*

$$ln\ L(\boldsymbol{\beta}) = \sum_{i=1}^{n} Y_i(\boldsymbol{X}'_i\boldsymbol{\beta}) - \sum_{i=1}^{n} ln[1 + exp(\boldsymbol{X}'_i\boldsymbol{\beta})]$$

- Procedimentos numéricos são usados para encontrar os valores de $\beta_0, \beta_1, \ldots, \beta_{p-1}$ que maximizam $ln\ L(\boldsymbol{\beta})$. Essas estimativas de máxima verossimilhança serão designadas por $b_0, b_1, \ldots, b_{p-1}$.

Seja $\boldsymbol{b}$ o vetor das estimativas de máxima verossimilhança:

$$\boldsymbol{b}_{p\times 1} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix}$$

Inferência sobre os parâmetros do Modelo

- Os procedimentos de inferência dependem de amostras de tamanhos de grandes.

Sob condições geralmente aplicáveis, estimadores de máxima verossimilhança para regressão logística são:

- aproximadamente normalmente distribuídos;
- apresentam pouco ou nenhum viés;
- e tem as correspondentes variâncias e covariâncias aproximadas que são funções das derivadas parciais de segunda ordem do logaritmo da função de verossimilhança.

Inferência sobre os parâmetros do Modelo

Seja G a matriz de derivadas parciais de segunda ordem do *log* da função de verossimilhança, as derivadas sendo tomadas em relação aos parâmetros do modelo.

$$\underset{p\times p}{\mathbf{G}} = [g_{ij}] \qquad i = 0, 1, \ldots, p-1; \; j = 0, 1, \ldots, p-1$$

$$g_{00} = \frac{\partial^2 \log_e L(\boldsymbol{\beta})}{\partial \beta_0^2}$$

$$g_{01} = \frac{\partial^2 \log_e L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1}$$

etc.

matriz de *Hessiana*

$$S^2\{b\} = \left(| - g_{ij}|_{\beta=b}\right)^{-1}$$

Inferências sobre os coeficientes de regressão para o modelo de regressão logística simples são baseados nos seguintes resultados aproximados quando o tamanho da amostra é grande:

$$Z = \frac{\beta_k - b_k}{S\{b \quad k\}} \sim N(0,1)., \qquad k = 0,1, \ldots p-1.$$

Teste referente a um único parâmetro $\rightarrow$ *Teste de WALD*

Hipóteses:

$$H_0: \beta_k = 0$$
$$H_a: \beta_k \neq 0$$

Estatística do Teste:

$$z^* = \frac{b_k}{s\{b_k\}} \sim N(0,1)$$

Intervalode Confiança para os parâmetros

IC(1-α) para β_k:

$$b_k \pm z(1 - \alpha/2)s\{b_k\}$$

IC (1-α) *para* θ_k:

$$\exp[b_k \pm z(1 - \alpha/2)s\{b_k\}]$$

Teste da Razão de Verossimilhança para Significância Conjunta de Variáveis Explicativas

$$H_0)\ \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_{(m+1)}x_{i(m+1)} + \ldots + \beta_p x_{ip}$$

$$H_1)\ \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im} + \beta_{(m+1)}x_{i(m+1)} + \ldots + \beta_p x_{ip}$$

ou ainda:

$$H_0)\ \beta_1 = \beta_2 = \ldots = \beta_m = 0$$

A estatística do teste é definida por:

$$G^2 = -2(L_0 - L_1) \sim \chi^2 \quad (\text{sob } H_0)$$

com *(p+1)-(p+1-m)* = *m* graus de liberdade

Teste da Razão de Verossimilhança para Significância de cada Variável Preditora

$$H_0)\ \beta_j = 0$$

$$H_1)\ \beta_J \neq 0$$

- Caso particular do teste anterior, onde *m=1*.
- Pode-se usar também o **Teste de Wald**:

$$W_j^2 = \left(\frac{b_j}{A\hat{S}E(b_j)}\right)^2 \quad \sim \quad \chi^2 \text{ com } 1\ g.l.$$

Deviance

- Sejam

  L_M o máximo log da função de verossimilhança para o modelo de interesse;

  L_S o máximo do log da função de verossimilhança para o modelo mais complexo, que tem um parâmetro separado para cada logito e fornece um ajustamento perfeito aos logitos amostrais. Este modelo é designado *modelo saturado*.

Relembrando...

Estatística G^2 do Teste da Razão de Verossimilhança

- **Modelo Completo:**

$$E(Y_{ij}) = \pi_j \qquad j = 1,2,\ldots,c$$

onde π_j é a probabilidade de sucesso para cada valor distinto de X, j=1,2,...,c.

O modelo completo no caso da regressão logística é usualmente chamado de ***modelo saturado***.

- **Modelo Restrito (sob H_0):**

$$E[Y_{ij}] = \pi(x) = \frac{exp(\beta_0 + \beta_1 x)}{1 + exp(\beta_0 + \beta_1 x)}$$

- **Estatística do Teste:**

Considerando a estatística do teste da razão de verossimilhança:

$$G^2 = -2\,ln\left(\frac{L(R)}{L(F)}\right) = -2\left[ln\big(L(R)\big) - ln\big(L(F)\big)\right]$$

$\pi_j,$

- As estimativas de máxima verossimilhança para as ***c*** probabilidades $,\pi_j$ modelo completo são dadas pelas proporções amostrais:

$$p_j = \frac{Y_{.j}}{n_j} \qquad j = 1,2,\ldots,c$$

- $\hat{\pi}_j$ *estimativa de* π_j *pelo modelo reduzido para cada Xj, j=1,2,..., c,*

Assim a estatística do teste é dada por:

$$G^2 = -2\sum_{j=1}^{c}\left[Y_j\,ln\left(\frac{\hat{\pi}_j}{p_j}\right) + (n_j - Y_j)\,ln\left(\frac{1-\hat{\pi}_j}{1-p_j}\right)\right]$$

ou então,

Deviance

$$G^2 = -2\sum_{j=1}^{c}\left[Y_j\,ln\left(\frac{\hat{Y}_j}{Y_j}\right) + (n_j - Y_j)\,ln\left(\frac{n_j-\hat{Y}_j}{n_j-Y_j}\right)\right]$$

que sob a hipótese H_0, tem distribuição aproximadamente Qui-quadrado com $c - k$ graus de liberdade quando n_j for grande e $k < c$

k - número de parâmetros do modelo

c - número conjuntos de valores distintos das variáveis explicativas

Deviance

- A *deviance* do modelo é definida como

Deviance = - 2 [L_M – L_S].

Tem distribuição ~ χ^2 com ν g.l.

- Compara o modelo M com o modelo S (saturado)
- Testa a hipótese de que todos os parâmetros que estão no modelo saturado, mas não estão no modelo M são nulos

Deviance Parcial

- Sejam dois modelos, M_0 e M_1 tal que M_0 é um caso especial de M_1
- Dado que o modelo mais complexo (M_1) se ajusta, deseja-se testar que o modelo mais simples (M_0)se ajusta

$$G^2(M_0|M_1) = -2\,[L_0 - L_1] =$$
$$= -2\,[L_0 - L_S] - \{-2\,[L_1 - L_S]\} =$$
$$= \text{deviance}_0 - \text{deviance}_1$$

Tem distribuição ~ χ^2 com g.l. igual ao número de parâmetros adicionais não redundantes que existem em M_1 mas não em M_0

Etapas na Construção de Modelos de Regressão

1. Coleta e preparação dos dados
2. Redução das variáveis explicativas ou preditoras (*para estudos observacionais exploratórios*)
3. Refinamento e seleção do modelo
4. Validação do modelo

Seleção de Variáveis

- **Critérios de Seleção de Variáveis:**
  - ***Critério da Informação de Akaike***

$$AIC_p = -2ln\,L(b) + 2p$$

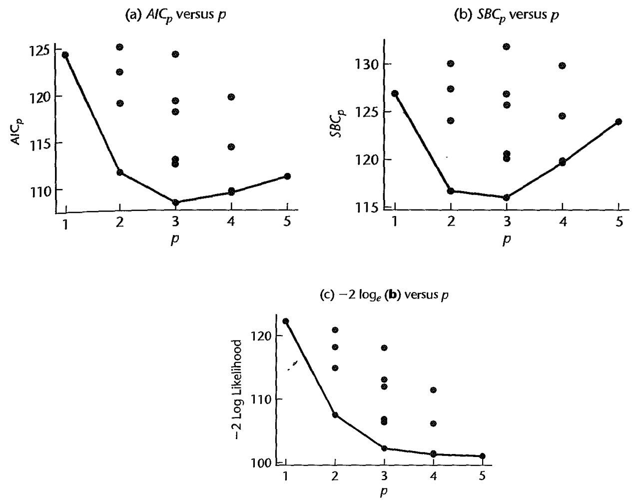
  - ***Critério Bayesiano de Schwarz***

$$SBC_p = -2ln\,L(b) + p ln(n)$$

  - ***Critério Log da verossimilhança***

$$-2ln\,L(b)$$

Modelos promissores renderão valores relativamente pequenos para esses critérios.



Seleção de Variáveis

- **Métodos de Seleção Automáticos:**
  - ***Backward***
  - ***Forward***
  - ***Stepwise***
- **Comparação entre modelos:** ➔ *deviance*

- Existem estratégias para selecionar variáveis explicativas que levam em conta questões como:
  - os objetivos do estudo,
  - significância estatística relativa,
  - multicolinearidade e
  - potencial confusão.

- Um processo de seleção proposto por Hosmer et al. (2013, Capítulo 4) utiliza diversas etapas para construir um modelo. De forma abreviada:
  - Construir um modelo inicial de efeitos principais usando variáveis explicativas que incluam as variáveis importantes conhecidas e outras que mostrem qualquer evidência de serem relevantes quando usadas como únicos preditores (por exemplo, tendo p-valor< 0,20).
  - Conduzir a eliminação backward, mantendo uma variável se for significativa em algum nível mais rigoroso ou mostrar evidências de ser um fator de confusão relevante, no sentido de que o efeito estimado de uma variável-chave muda substancialmente quando ela é removida.
  - Adicione ao modelo quaisquer variáveis que não foram incluídas na etapa 1, mas que sejam significativas ao ajustar as variáveis no modelo após a etapa 2, uma vez que uma variável pode não estar significativamente associada a Y mas pode dar uma contribuição importante na presença de outras variáveis.
  - Verifique se há interações plausíveis entre variáveis no modelo após a etapa 3, usando testes de significância em níveis convencionais como 0,05.
  - Realizar investigações diagnósticas de acompanhamento.

Table 5.1 Results of fitting several logistic regression models to predict horseshoe crab satellites.

Model	Explanatory Variables	Deviance	*df*	AIC	Models Compared	Deviance Difference
1	None	225.8	172	227.8		
2	C	212.1	169	220.1	(2) - (1)	13.7 ($df = 3$)
3	S	223.2	170	229.2	(3) - (1)	2.5 ($df = 2$)
4	W	194.5	171	198.5	(4) - (1)	31.3 ($df = 1$)
5	$C + W$	187.5	168	197.5	(5) - (2)	24.6 ($df = 1$)
					(5) - (4)	7.0 ($df = 3$)
6	$C + W + S$	186.6	166	200.6	(6) - (5)	0.9 ($df = 2$)
7	$C + W + C{*}W$	183.1	165	199.1	(7) - (5)	4.4 ($df = 3$)

Note: C = color, S = spine condition, W = width.

Resíduos para o Modelo de Regressão Logísitica

- Resíduo :
  - A análise de resíduos para regressão logística é mais difícil do que para modelos de regressão linear porque as respostas Y_i assumem apenas os valores 0 e 1. Consequentemente, o *i-ésimo* resíduo ordinário (comum), e_i assumirá um de dois valores:

$$e_i = \begin{cases} 1 - \hat{\pi}_i & \text{if } Y_i = 1 \\ -\hat{\pi}_i & \text{if } Y_i = 0 \end{cases}$$

Resíduos para o Modelo de Regressão Logísitica

- Resíduo de Pearson:

O resíduo de Pearson para o ajuste no ponto i é dado por:

$$e_i = \frac{f_{1i} - n_i \hat{\pi}_i}{\sqrt{[n_i \hat{\pi}_i (1 - n_i \hat{\pi}_i)]}}$$

onde: f_{1i} - número de sucessos para n_i repetições do *i-ésimo* conjunto de valores das variáveis explicativas.

$n_i\hat{\pi}_i$ - número de sucessos ajustado ou esperado.

Resíduos para o Modelo de Regressão Logísitica

- Resíduo de Pearson studentizado:

O resíduo de Pearson para o ajuste no ponto i é dado por:

$$r_{SP_i} = \frac{r_{P_i}}{\sqrt{1 - h_{ii}}}$$

$$\mathbf{H} = \widehat{\mathbf{W}}^{\frac{1}{2}}\mathbf{X}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{W}}^{\frac{1}{2}}$$

Resíduos para o Modelo de Regressão Logística

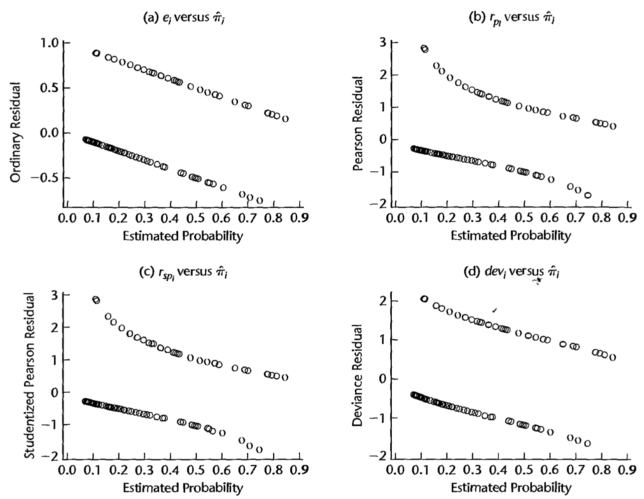
- ***Deviance* Residual:**

O resíduo de Pearson para o ajuste no ponto i é dado por:

$$dev_i = sign(Y_i - \hat{\pi}_i)\sqrt{-2[Y_i \log_e(\hat{\pi}_i) + (1 - Y_i)\log_e(1 - \hat{\pi}_i)]}$$

onde: f_{1i} - número de sucessos para n_i repetições do i-ésimo conjunto de valores das variáveis explicativas.

n_i - número de sucessos ajustado ou esperado.



Resumindo Poder Preditivo

- **Poder preditivo** - quão bem podemos prever o resultado da variável de resposta usando o ajuste do modelo.
- **Tabelas de Classificação**

Classifica o resultado binário Y com uma previsão de se $Y = 0$ *ou* 1.

A previsão para a observação *i* é :

$$\hat{Y}_i = 1 \quad quando\ \hat{\pi}_i > \pi_0$$
$$\hat{Y}_i = 0 \quad quando\ \hat{\pi}_i \leq \pi_0$$

- Uma possibilidade é tomar $\boldsymbol{\pi_0 = 0,5}$.

  Entretanto, se uma proporção baixa (alta) de observações tiver $Y = 1$, o ajuste do modelo pode nunca (sempre) ter $\hat{\pi}_i > 0{,}5$, caso em que nunca (sempre) se prevê $\hat{Y}_i = 1$.
- Outra possibilidade toma $\hat{\pi}_0$ como a proporção amostral de 1

	Prediction, $\pi_0 = 0.6416$		Prediction, $\pi_0 = 0.50$		
Actual	$\hat{y} = 1$	$\hat{y} = 0$	$\hat{y} = 1$	$\hat{y} = 0$	Total
$y = 1$	75	36	96	15	111
$y = 0$	19	43	31	31	62

- **Sensibilidade:** é a probabilidade de o teste ser positivo, sabendo-se que o evento de interesse ocorre.
- **Especificidade:** é a probabilidade de o teste ser negativo, sabendo-se que o evento de interesse não ocorre.

- Probabilidade de Classificações corretas

$$\boldsymbol{P(Classificação\ correta) = P(Y = 1\ e\ \hat{Y} = 1) + P(Y = 0\ e\ \hat{Y} = 0) =}$$
$$\boldsymbol{= P(\hat{Y} = 1 | Y = 1) P(Y = 1) + P(\hat{Y} = 0 | Y = 0) =}$$
$$\boldsymbol{= sensibilidade\ P(Y = 1) + especificidade\ P(Y = 0)}$$

que é uma média ponderada de sensibilidade e especificidade.

Curva ROC

- Uma curva ROC (*Receive Operating Characteristic*) é um gráfico que mostra a sensibilidade e a especificidade das previsões para todos os cortes possíveis π_0.
- Essa curva é mais informativa do que uma tabela de classificação, pois resume o poder preditivo para todos os π_0 possíveis.