cin.ufpe.br

Centro de Informática

U · F · P · E

Regressão Logística

Daniel Araújo Melo - dam2@cin.ufpe.br

Universidade Federal de Pernambuco

Introdução

- ➔Objetivo
  - Encontrar o melhor modelo para descrever a relação entre variável de saída (variável dependente) e variáveis independentes (preditoras ou explanatórias)
- ➔Variável de saída binária ou dicotômica
  - (1 ou 0)
  - Reflete no modelo paramétrico
- ➔Segue os mesmos princípios da regressão linear

Exemplo

Table 1.1 Age and Coronary Heart Disease (CHD) Status of 100 Subjects

ID	AGE	AGRP	CHD	ID	AGE	AGRP	CHD
1	20	1	0	51	44	4	1
2	23	1	0	52	44	4	1
3	24	1	0	53	45	5	0
4	25	1	0	54	45	5	1
5	25	1	1	55	46	5	0
6	26	1	0	56	46	5	1
7	26	1	0	57	47	5	0
8	28	1	0	58	47	5	0
9	28	1	0	59	47	5	1
10	29	1	0	60	48	5	0
11	30	2	0	61	48	5	1
12	30	2	0	62	48	5	1
13	30	2	0	63	49	5	0
14	30	2	0	64	49	5	0
15	30	2	0	65	49	5	1
16	30	2	1	66	50	6	0
17	32	2	0	67	50	6	1
18	32	2	0	68	51	6	0
19	33	2	0	69	52	6	0
20	33	2	0	70	52	6	1
21	34	2	0	71	53	6	1
22	34	2	0	72	53	6	1
23	34	2	1	73	54	6	1
24	34	2	0	74	55	7	0
25	34	2	0	75	55	7	1

Table 1.1 Age and Coronary Heart Disease (CHD) Status of 100 Subjects

ID	AGE	AGRP	CHD	ID	AGE	AGRP	CHD
26	35	3	0	76	55	7	1
27	35	3	0	77	56	7	1
28	36	3	0	78	56	7	1
29	36	3	1	79	56	7	1
30	36	3	0	80	57	7	0
31	37	3	0	81	57	7	0
32	37	3	1	82	57	7	1
33	37	3	0	83	57	7	1
34	38	3	0	84	57	7	1
35	38	3	0	85	57	7	1
36	39	3	0	86	58	7	0
37	39	3	1	87	58	7	1
38	40	4	0	88	58	7	1
39	40	4	1	89	59	7	1
40	41	4	0	90	59	7	1
41	41	4	0	91	60	8	0
42	42	4	0	92	60	8	1
43	42	4	0	93	61	8	1
44	42	4	0	94	62	8	1
45	42	4	1	95	62	8	1
46	43	4	0	96	63	8	1
47	43	4	0	97	64	8	0
48	43	4	1	98	64	8	1
49	44	4	0	99	65	8	1
50	44	4	0	100	69	8	1

Scatterplot

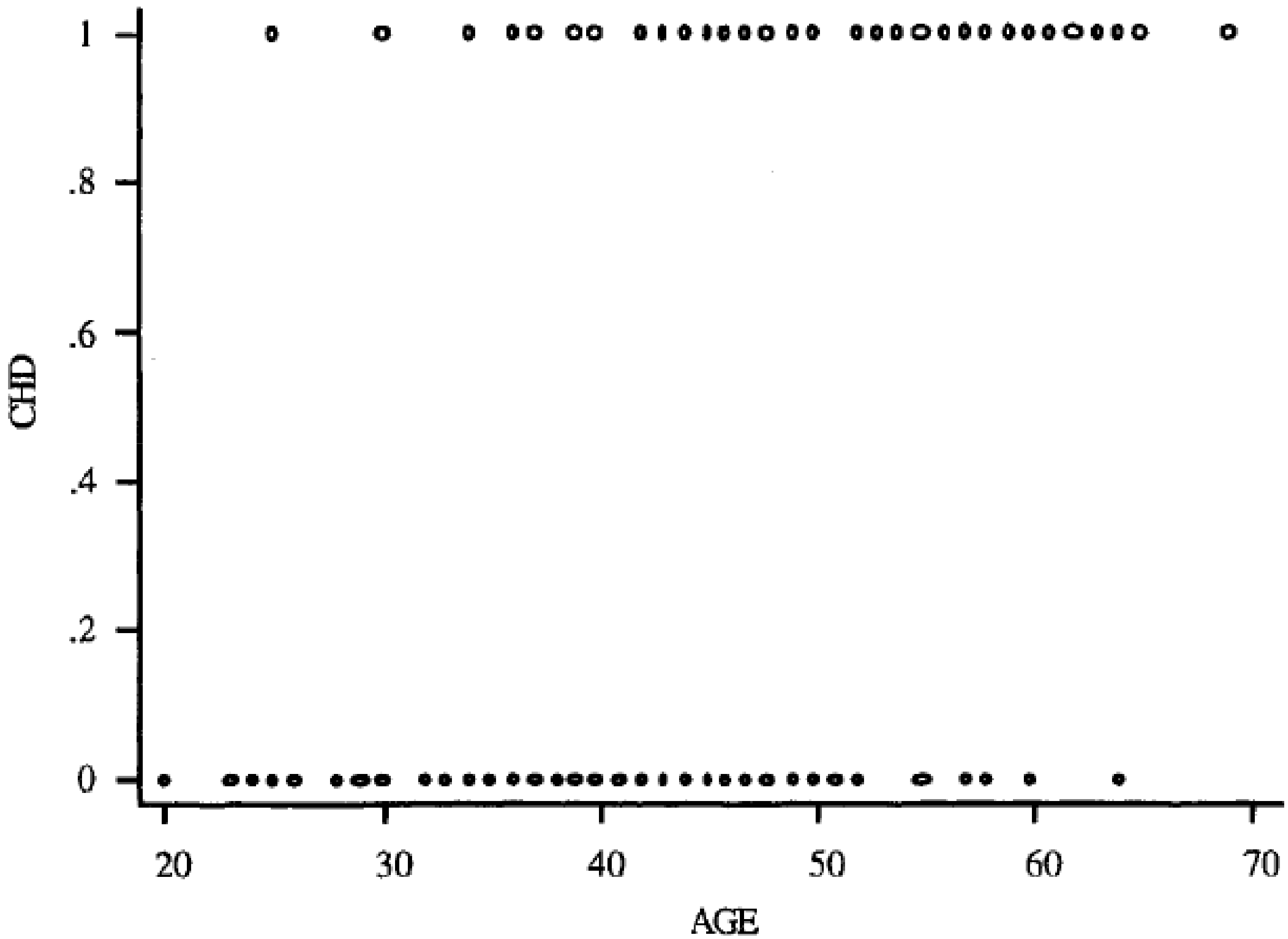


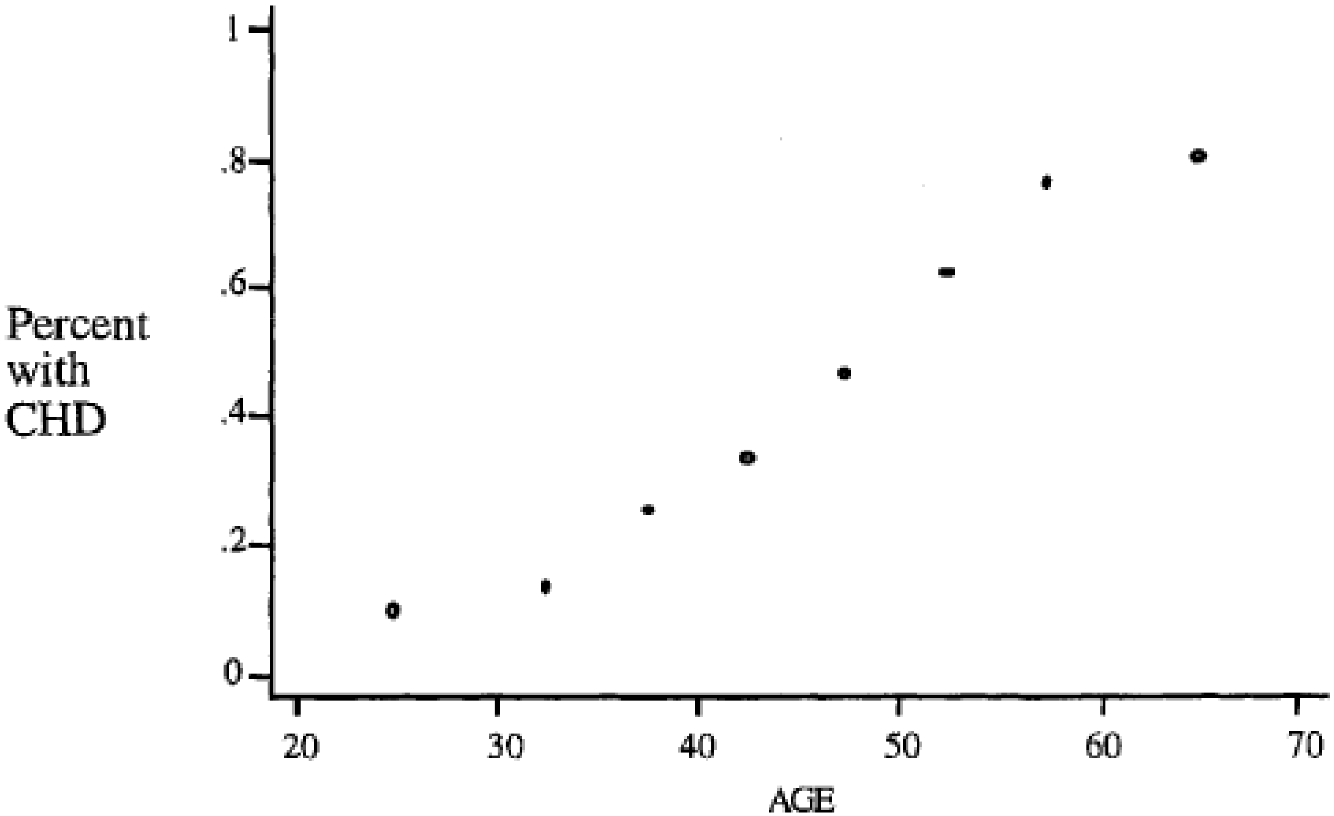
Tabela de Frequência

Table 1.2 Frequency Table of Age Group by CHD

Age Group	n	CHD Absent	CHD Present	Mean (Proportion)
20 – 29	10	9	1	0.10
30 – 34	15	13	2	0.13
35 – 39	12	9	3	0.25
40 – 44	15	10	5	0.33
45 – 49	13	7	6	0.46
50 – 54	8	3	5	0.63
55 – 59	17	4	13	0.76
60 – 69	10	2	8	0.80
Total	100	57	43	0.43

Relação entre CHD e Idade

➔S-Shaped –> distribuição cumulativa



Função de Resposta

➔Considerando Modelo de Regressão Simples:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$Y_i = \begin{cases} 1 \\ 0 \end{cases}$$

➔Valor esperado: $E(Y_i) = \beta_0 + \beta_1 X_i \qquad (1)$

➔Considere Y uma v.a. Bernoulli com distribuição de probabilidade

$$Y_i = 1 \rightarrow P(Y_i = 1) = \pi_i$$
$$Y_i = 0 \rightarrow P(Y_i = 0) = 1 - \pi_i$$

Função de Resposta(2)

➔Pela definição de valor esperado, obtemos::

$$E(Y_i) = \pi_i \qquad (2)$$

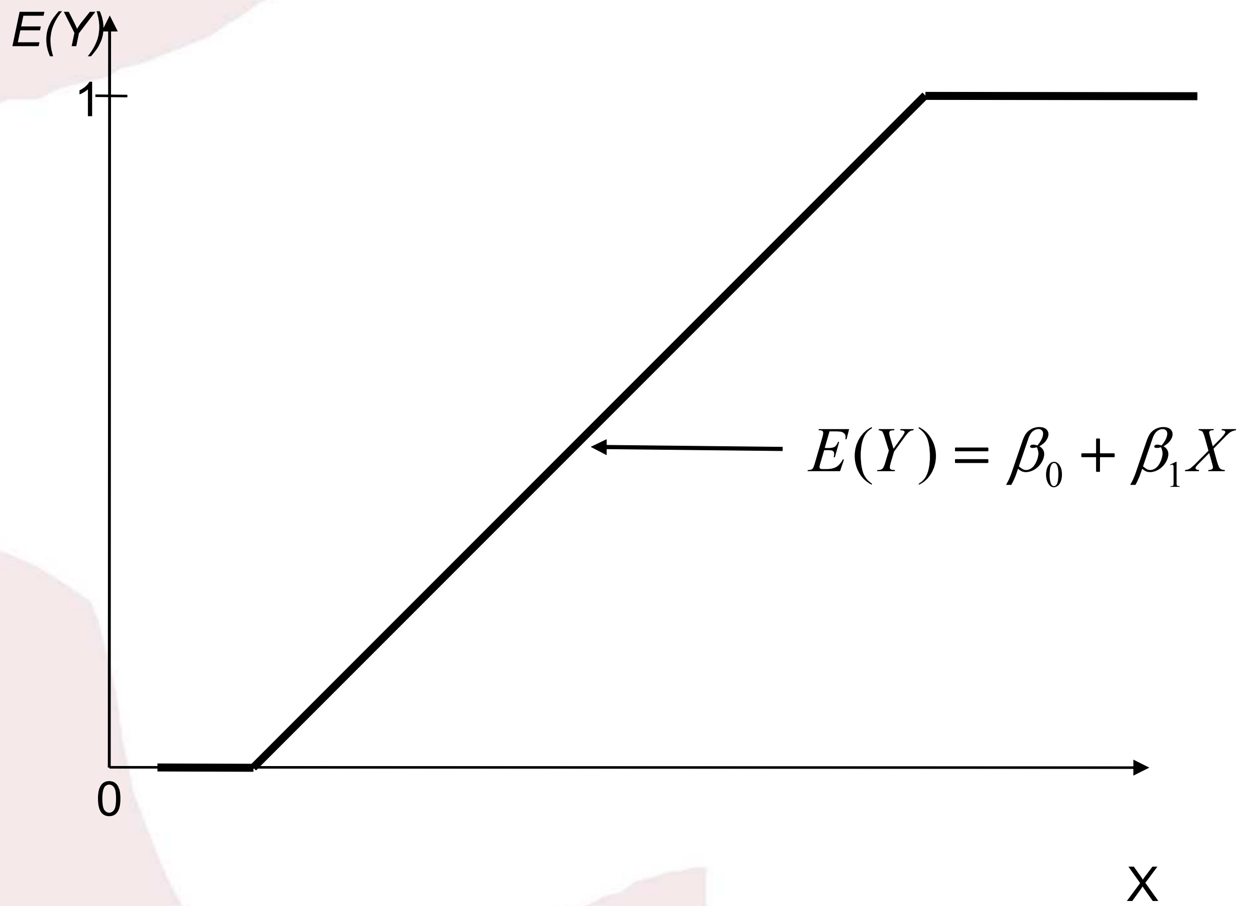
➔Igualando (1) e (2)

$$E(Y_i) = \beta_0 + \beta_1 X_i = \pi_i \qquad (3)$$

- A resposta média, quando a variável resposta é uma variável binária, sempre representa a probabilidade de Y = 1, para o nível da variável preditora x.

$$\boxed{E(Y \mid x) = \beta_0 + \beta_1 x\,.}$$

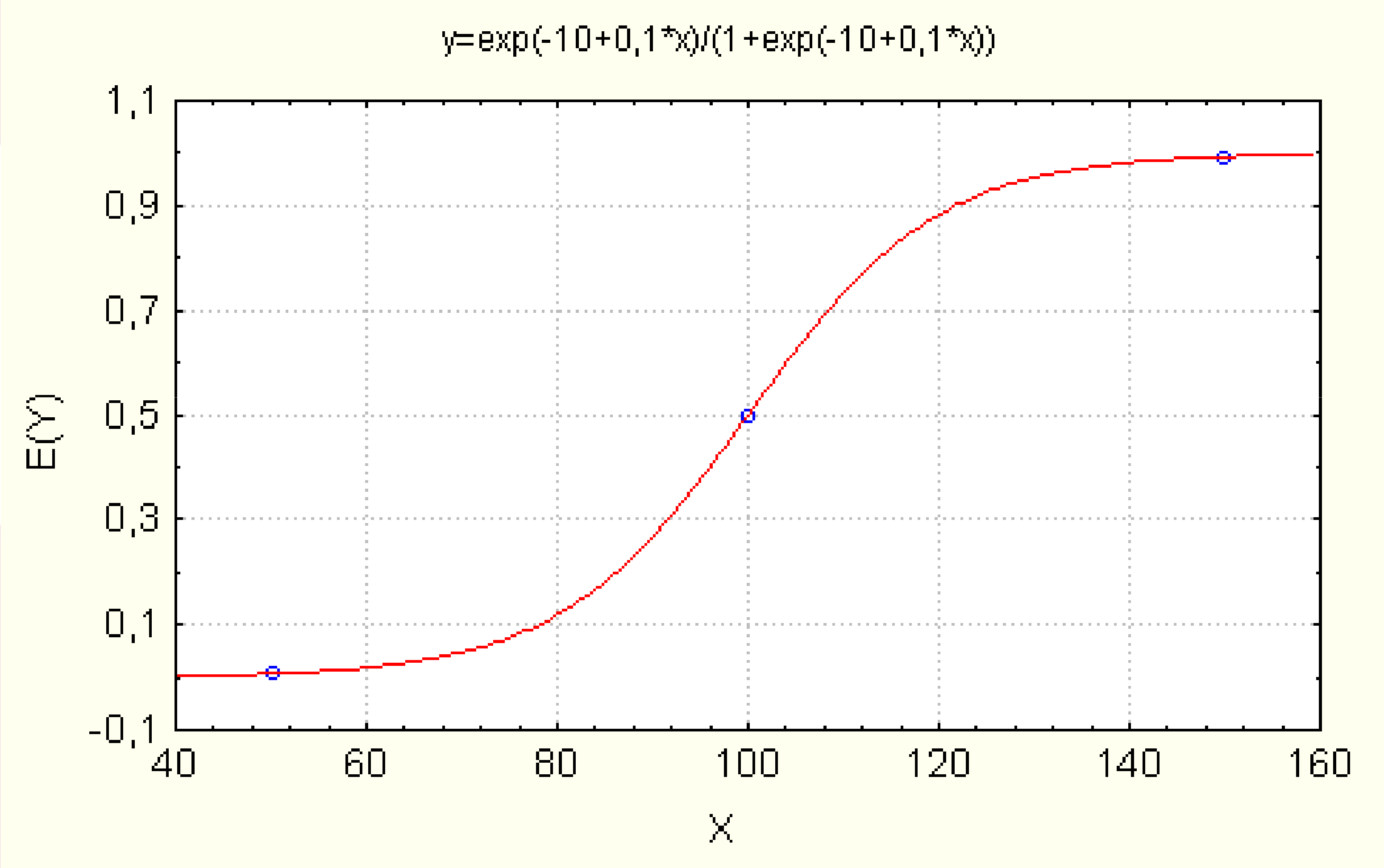
Função de Resposta



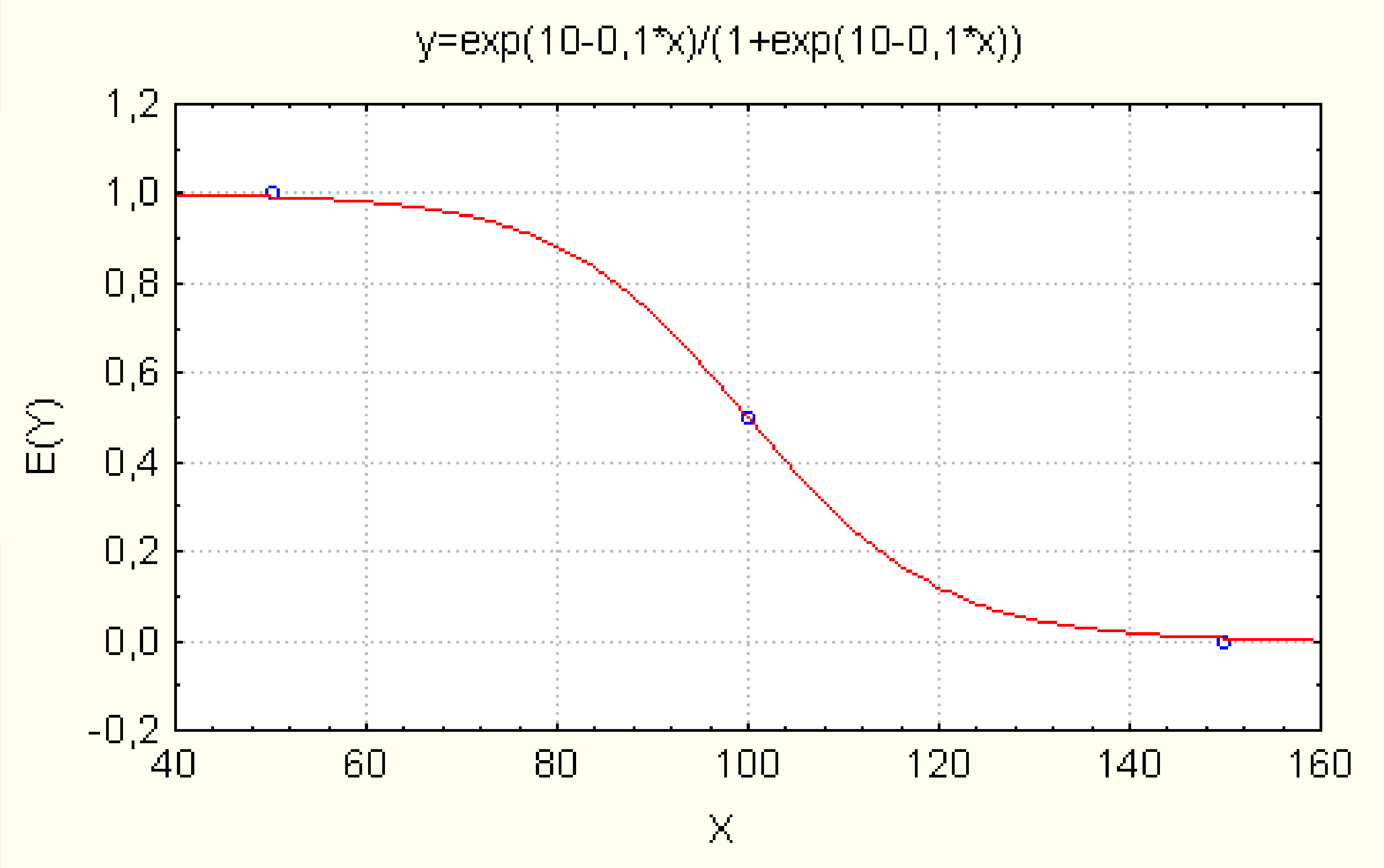
Função Resposta logística

➔ Este modelo freqüentemente não representa bem a situação em estudo. Ao invés, um modelo onde as probabilidades 0 e 1 são encontradas assintoticamente, como mostra a figura a seguir, é, de modo geral, mais apropriada.

Função Resposta Logística



Função Resposta logística



Modelo de Regressão

➔ Razões em utilizar distribuição logística

- Função de fácil utilização e extremamente flexível
- Interpretação razoável

➔ Transformação logit

- Possui parâmetros lineares;
- Pode ser contínua;
- Varia de -∞ a +∞ de acordo com valores de x

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} .$$

$$g(x) = \ln\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \beta_0 + \beta_1 x .$$

Modelo de Regressão

➔Erros não seguem distribuição normal

$$y = \pi(x) + \varepsilon.$$

- P/ Y=1
  - $\varepsilon=1-\pi(x)$ com probabilidade $\pi(x)$
- P/ Y=0
  - $\varepsilon=-\pi(x)$ com probabilidade $1-\pi(x)$
  - ε possui distribuição com média 0 e variância igual a $\pi(x)[1-\pi(x)]$
- Variância heterogênea;
- Variável de saída segue distribuição binomial com probabilidade dada pela média condicional $\pi(x)$;
- Erro segue distribuição binomial

Adequação do Modelo

- Quando método dos mínimos quadrados é aplicado a um modelo com saída dicotômica os estimadores não possuem mesmas propriedades estatísticas como na Regressão Linear

Estimação de Parâmetros

➔ Função de vizinhança (FV)

- Baseada na máxima verossimilhança usada na Regressão Linear
- Expressa a probabilidade do dado observado como função dos parâmetros desconhecidos
- Para pares(x_i,y_i), onde:
  - $y_i=1$, contribuição para FV é $\pi(x_i)$;
  - $y_i=0$,contribuição para FV é $1-\pi(x_i)$.
- Contribuição de um par(x_i,y_i): $\pi(x_i)^{y_i}\left[1-\pi(x_i)\right]^{1-y_i}$
- Como observações independentes:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi(x_i)^{y_i}\left[1-\pi(x_i)\right]^{1-y_i}$$

Estimação de Parâmetros

➔Log Verossimilhança

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \sum_{i=1}^{n} \{ y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)] \}$$

- Para encontrar valores de **β** que maximiza *L*(**β**) diferenciamos L(**β**) por β_0 e β_1:

$$\sum [y_i - \pi(x_i)] = 0 \qquad \sum x_i [y_i - \pi(x_i)] = 0.$$

- **β** encontrado é chamado estimador da maximo-verossimilhança ^**β**

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{\pi}(x_i).$$

Modelo Exemplo

Table 1.3 Results of Fitting the Logistic Regression Model to the Data in Table 1.1

Variable	Coeff.	Std. Err.	z	P>|z|
AGE	0.111	0.0241	4.61	<0.001
Constant	−5.309	1.1337	−4.68	<0.001

Log likelihood = −53.67656

$$\hat{\pi}(x) = \frac{e^{-5.309+0.111\times AGE}}{1+e^{-5.309+0.111\times AGE}} \qquad \hat{g}(x) = -5.309 + 0.111 \times \mathrm{AGE}.$$

➔Z:

- Razão entre coeficientes estimados e erros padrão estimados

Significância dos Coeficientes

- O modelo que inclui uma determinada variável diz mais sobre sobre variável resposta do que o modelo sem a variável?
  - Goodness-of-fit
- Variação não explicada:
  - Soma dos Quadrados dos Resíduos $SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$
- Variação explicada
  - $$SSR = \left[\sum_{i=1}^{n} (y_i - \bar{y}_i)^2\right] - \left[\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right]$$

Significância dos Coeficientes

➔Teste daTaxa de verossimilhança (Desvio)

- Comparação entre valores observados e preditos;
- Baseada na log verossimilhança
  - Modelo saturado

$$D = -2\ln\left[\frac{\text{(likelihood of the fitted model)}}{\text{(likelihood of the saturated model)}}\right]$$

    - *Contém tantos parâmetros quanto observações.*
  - Utiliza-se 2 vezes o log para obter quantidade cuja distribuição seja conhecida e possa ser utilizada em um testes de hipóteses,onde $\pi_i = \pi(x_i)$

$$D = -2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1-y_i)\ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right]$$

- Mesmo papel da soma dos quadrados dos resíduos na regressão linear.

Significância dos Coeficientes

➔Teste daTaxa de verossimilhança (Desvio)

- Em modelos saturados verossimilhança =1 $\hat{\pi}_i = y_i$

$$l(\text{saturated model}) = \prod_{i=1}^{n} y_i^{y_i} \times (1 - y_i)^{(1-y_i)} = 1$$

- Logo:

$$D = -2\ln(\text{likelihood of the fitted model})$$

- Para testar significância de variável:

$$\boxed{G = D(\text{model without the variable}) - D(\text{model with the variable})}$$

  - *G* possui mesmo papel do que numerador do teste parcial de *F* na regressão linear. Como a verossimilhança do modelo saturado é a mesma nos 2 modelos:

$$G = -2\ln\left[\frac{(\text{likelihood without the variable})}{(\text{likelihood with the variable})}\right]$$

Significância dos Coeficientes

➔Teste daTaxa de verossimilhança (Desvio)

- Para caso de única variável regressora (independente), quando a variável não está no modelo:
  - $\hat{}\beta_0=\ln(n_1/n_0)$, onde $n_1=Ey_i$, $n_0=E(1-y_i)$, e valor predito é constante, n_1/n.

$$G = -2\ln\left[\frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^{n}\hat{\pi}_i^{y_i}\left(1-\hat{\pi}_i\right)^{(1-y_i)}}\right]$$

$$G = 2\left\{\sum_{i=1}^{n}\left[y_i\ln(\hat{\pi}_i)+(1-y_i)\ln(1-\hat{\pi}_i)\right]-\left[n_1\ln(n_1)+n_0\ln(n_0)-n\ln(n)\right]\right\}$$

- Na hipótese que $\beta_1=0$, G segue chi-quadrado com 1 grau de liberdade

Significância dos Coeficientes

➔Desvio do exemplo:

$$G = 2\{-53.677 - [43\ \ln(43) + 57\ \ln(57) - 100\ \ln(100)]\}$$
$$= 2[-53.677 - (-68.331)] = 29.31.$$

$$P[\chi^2(1) > 29.31] < 0.001$$

- Rejeira Hipótese que β_1 é igual a zero - há evidências que a variável AGE é significante para o modelo.

Significância dos Coeficientes

➔Testes equivalentes:

- Teste de Wald

$$W = \frac{\hat{\beta}_1}{\widehat{\mathrm{SE}}\left(\hat{\beta}_1\right)} = \frac{0.111}{0.024} = 4.61$$

  - Resultado sobre a hipótese que $\beta_1=0$ segue distribuição normal
  - Baixa precisão, pode falhar em rejeitar hipótese nula quando variável é significante.

- Teste Score
  - Baseado na teoria de distribuição das derivadas de log verossimilhança

$$\mathrm{ST} = \frac{\sum_{i=1}^{n} x_i\left(y_i - \bar{y}\right)}{\sqrt{\bar{y}(1-\bar{y})\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}} \qquad \mathrm{ST} = \frac{296.66}{\sqrt{3333.742}} = 5.14$$

Intervalo de Confiança

➔Baseados no Teste Wald

$$\hat{\beta}_1 \pm z_{1-\alpha/2}\widehat{\mathrm{SE}}\left(\hat{\beta}_1\right), \qquad \boxed{\hat{\beta}_0 \pm z_{1-\alpha/2}\widehat{\mathrm{SE}}\left(\hat{\beta}_0\right)}$$

$$\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1 x\,.$$

$$\widehat{\mathrm{Var}}[\hat{g}(x)] = \widehat{\mathrm{Var}}\left(\hat{\beta}_0\right) + x^2\widehat{\mathrm{Var}}\left(\hat{\beta}_1\right) + 2x\widehat{\mathrm{Cov}}\left(\hat{\beta}_0, \hat{\beta}_1\right).$$

$$\hat{g}(x) \pm z_{1-\alpha/2}\widehat{\mathrm{SE}}[\hat{g}(x)],$$

Intervalo de Confiança

➔Exemplo:

Table 1.4 Estimated Covariance Matrix of the Estimated Coefficients in Table 1.3

	AGE	Constant
AGE	0.000579	
Constant	–0.026677	1.28517

$$\hat{g}(50) = -5.31 + 0.111 \times 50 = 0.240.$$

$$\widehat{\text{Var}}[\hat{g}(50)] = 1.28517 + (50)^2 \times 0.000579 + 2 \times 50 \times (-0.026677) = 0.0650$$

$$\widehat{\text{SE}}[\hat{g}(50)] = 0.2549 \qquad 0.240 \pm 1.96 \times 0.2550 = (-0.260,\ 0.740)$$

$$\hat{\pi}(50) = \frac{e^{\hat{g}(50)}}{1 + e^{\hat{g}(50)}} = \frac{e^{-5.31+0.111\times 50}}{1 + e^{-5.31+0.111\times 50}} = 0.560 \qquad \frac{e^{\hat{g}(x) \pm z_{1-\alpha/2}\widehat{\text{SE}}[\hat{g}(x)]}}{1 + e^{\hat{g}(x) \pm z_{1-\alpha/2}\widehat{\text{SE}}[\hat{g}(x)]}}.$$

$$\frac{e^{-0.260}}{1 + e^{-0.260}} = 0.435, \qquad \frac{e^{0.740}}{1 + e^{0.740}} = 0.677.$$

Regressão Logística

Regressão Logística Múltipla

Modelo de Regressão Logística Múltipla

➔ Considere coleção de variáveis independentes denotadas pelo vetor:

- **x**' $=(x_1,x_2,\ldots x_p)$
- $g(\mathbf{x})=\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p$

$$\pi(\mathbf{x})=\frac{e^{g(\mathbf{x})}}{1+e^{g(\mathbf{x})}}\ .$$

  - Se variáveis dependentes são discretas, é inadequado incluí-las no modelo como se fossem variáveis escalares. Deve-se utilizar variáveis de design (ou dummy).
  - Se uma variável discreta possui k valores possíveis, serão necessárias k-1 variáveis dummy.

Modelo de Regressão Logística Múltipla

Table 2.1 An Example of the Coding of the Design Variables for Race, Coded at Three Levels

	Design Variable	
RACE	D_1	D_2
White	0	0
Black	1	0
Other	0	1

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \cdots + \sum_{l=1}^{k_j - 1} \beta_{jl} D_{jl} + \beta_p x_p .$$

Adequação do Modelo

- Método de estimação semelhante ao caso univariado
  - Existem p+1 equações de vizinhança obtidas pela diferenciação da função de log verossimilhança com respeito aos p+1 coeficientes.

$$\sum_{i=1}^{n}\left[y_i - \pi(\mathbf{x}_i)\right] = 0 \qquad \sum_{i=1}^{n} x_{ij}\left[y_i - \pi(\mathbf{x}_i)\right] = 0$$

  - Estimadores são obtidos através da matriz das segundas derivadas parciais da função de log verossimilhança

$$\boxed{\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = -\sum_{i=1}^{n} x_{ij}^2 \pi_i (1-\pi_i)} \qquad \frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_l} = -\sum_{i=1}^{n} x_{ij} x_{il} \pi_i (1-\pi_i)$$

Adequação do Modelo

➔Matriz de informação observada

- Matriz (p+1)(p+1), denominada $\mathbf{I}(\boldsymbol{\beta})$ contendo o negativo dos termos encontrados nas equações:

$$\boxed{\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = -\sum_{i=1}^{n} x_{ij}^2 \pi_i (1-\pi_i)} \qquad \frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_l} = -\sum_{i=1}^{n} x_{ij} x_{il} \pi_i (1-\pi_i)$$

- Variâncias e Covariâncias dos coeficientes estimados são obtidos através da matriz inversa Var($\boldsymbol{\beta}$)= $\mathbf{I}^{-1}(\boldsymbol{\beta})$

$$\boxed{\widehat{\mathrm{SE}}(\hat{\beta}_j) = \left[\widehat{\mathrm{Var}}(\hat{\beta}_j)\right]^{1/2}}$$

Adequação do Modelo

➔Matriz de informação observada

$$\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}}) = \mathbf{X}'\mathbf{V}\mathbf{X}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \qquad \mathbf{V} = \begin{bmatrix} \hat{\pi}_1(1-\hat{\pi}_1) & 0 & \cdots & 0 \\ 0 & \hat{\pi}_2(1-\hat{\pi}_2) & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \hat{\pi}_n(1-\hat{\pi}_n) \end{bmatrix}.$$

Modelo Exemplo

Table 2.2 Estimated Coefficients for a Multiple Logistic Regression Model Using the Variables AGE, Weight at Last Menstrual Period (LWT), RACE, and Number of First Trimester Physician Visits (FTV) from the Low Birth Weight Study

Variable	Coeff.	Std. Err.	z	P>|z|
AGE	–0.024	0.0337	–0.71	0.480
LWT	–0.014	0.0065	–2.18	0.029
RACE_2	1.004	0.4979	2.02	0.044
RACE_3	0.433	0.3622	1.20	0.232
FTV	–0.049	0.1672	–0.30	0.768
Constant	1.295	1.0714	1.21	0.227

Log likelihood = –111.286

$$\hat{g}(\mathbf{x}) = 1.295 - 0.024 \times AGE - 0.014 \times LWT + 1.004 \times RACE\_2 + 0.433 \times RACE\_3 - 0.049 \times FTV.$$

Significância dos Coeficientes

- De forma similar ao caso univariado utiliza-se a estatística de teste G
  - Os valores adequados, ^Pi, no modelo são baseados em um vetor contendo p+1 parâmetros, ^B
  - Sob a hiopótese nula que os p parâmetros são iguais a zero, G segue a chi-quadrado com p graus de liberdade.
  - Para o exemplo: $G = -2[(-117.336) - (-111.286)] = 12.099$

    $\alpha = 0.05$ level. $P[\chi^2(5) > 12.099] = 0.034$
  - Rejeina hipótese nula e conclui que ao menos um dos parâmetros é significante

Significância dos Coeficientes

➔ Teste Wald

$$W_j = \hat{\beta}_j / \widehat{\mathrm{SE}}(\hat{\beta}_j).$$

$$W = \hat{\boldsymbol{\beta}}' \left[\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) \right]^{-1} \hat{\boldsymbol{\beta}}$$
$$= \hat{\boldsymbol{\beta}}'(\mathbf{X}'\mathbf{V}\mathbf{X})\hat{\boldsymbol{\beta}},$$

- Sob a hipótese que um coeficiente individual é zero, segue distribuição normal

Table 2.3 Estimated Coefficients for a Multiple Logistic Regression Model Using the Variables LWT and RACE from the Low Birth Weight Study

Variable	Coeff.	Std. Err.	z	P>|z|
LWT	–0.015	0.0064	–2.36	0.018
RACE_2	1.081	0.4881	2.22	0.027
RACE_3	0.481	0.3567	1.35	0.178
Constant	0.806	0.8452	0.95	0.340

Log likelihood = –111.630

$$P[\chi^2(2) > 0.688] = 0.709.$$

$$G = -2[(-111.630) - (-111.286)] = 0.688,$$

Intervalo de Confiança

- Estimação para coeficientes similara ao caso univariado;

$$\hat{g}(\mathbf{x}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p.$$

$$\boxed{\hat{g}(\mathbf{x}) = \mathbf{x}'\hat{\boldsymbol{\beta}},}$$

$$\hat{\boldsymbol{\beta}}' = \left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p\right)$$

$$\widehat{\mathrm{Var}}[\hat{g}(\mathbf{x})] = \sum_{j=0}^{p} x_j^2 \widehat{\mathrm{Var}}\left(\hat{\beta}_j\right) + \sum_{j=0}^{p}\sum_{k=j+1}^{p} 2x_j x_k \widehat{\mathrm{Cov}}\left(\hat{\beta}_j, \hat{\beta}_k\right).$$

$$\boxed{\widehat{\mathrm{Var}}\left(\hat{\boldsymbol{\beta}}\right) = (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}.}$$

$$\begin{aligned}\widehat{\mathrm{Var}}[(\hat{g}(\mathbf{x}))] &= \mathbf{x}'\widehat{\mathrm{Var}}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{x} \\ &= \mathbf{x}'(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{x}\ .\end{aligned}$$

Intervalo de Confiança

➔Exemplo

$$\hat{g}(LWT=150, RACE=White) = 0.806 - 0.015 \times 150 + 1.081 \times 0 + 0.481 \times 0 = -1.444$$

$$\hat{\pi}(LWT=150, RACE=White) = \frac{e^{-1.444}}{1+e^{-1.444}} = 0.191.$$

$$\hat{\text{Var}}[\hat{g}(LWT=150, RACE=White)] = \hat{\text{Var}}(\hat{\beta}_0) + (150)^2 \times \hat{\text{Var}}(\hat{\beta}_1) +$$
$$(0)^2 \times \hat{\text{Var}}(\hat{\beta}_2) + (0)^2 \times \hat{\text{Var}}(\hat{\beta}_3) + 2 \times 150 \times \hat{\text{Cov}}(\hat{\beta}_0, \hat{\beta}_1)$$
$$+ 2 \times 0 \times \hat{\text{Cov}}(\hat{\beta}_0, \hat{\beta}_2) + 2 \times 0 \times \hat{\text{Cov}}(\hat{\beta}_0, \hat{\beta}_3) + 2 \times 150 \times 0 \times \hat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2)$$
$$+ 2 \times 150 \times 0 \times \hat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_3) + 2 \times 0 \times 0 \times \hat{\text{Cov}}(\hat{\beta}_2, \hat{\beta}_3)$$

$$= 0.7143 + (150)^2 \times 0.000041 + 0 \times 0.2382 + 0 \times 0.1272$$
$$+ 2 \times 150 \times (-0.0052) + 2 \times 0 \times 0.0226 + 2 \times 0 \times (-0.1035)$$
$$+ 2 \times 150 \times 0 \times (-0.000647) + 2 \times 150 \times 0 \times 0.000036$$
$$+ 2 \times 0 \times 0 \times 0.0532 = 0.0768$$

Intervalo de Confiança

➔Exemplo

$$\widehat{\mathrm{SE}}\left[\hat{g}(LWT=150, RACE=White)\right]=0.2771.$$

$$\boxed{-1.444 \pm 1.96 \times 0.2771=(-1.988,\ -0.901).}$$

Table 2.4 Estimated Covariance Matrix of the Estimated Coefficients in Table 2.3

	LWT	RACE_2	RACE_3	Constant
LWT	0.000041			
RACE_2	−0.000647	0.2382		
RACE_3	0.000036	0.0532	0.1272	
Constant	−0.005211	0.0226	−0.1035	0.7143

Interpretação do Modelo

Table 3.1 Values of the Logistic Regression Model When the Independent Variable Is Dichotomous

Outcome Variable (Y)	Independent Variable (X) $x = 1$	$x = 0$
$y = 1$	$\pi(1) = \dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$\pi(0) = \dfrac{e^{\beta_0}}{1+e^{\beta_0}}$
$y = 0$	$1-\pi(1) = \dfrac{1}{1+e^{\beta_0+\beta_1}}$	$1-\pi(0) = \dfrac{1}{1+e^{\beta_0}}$
Total	1.0	1.0

Bibliografia

➔ Applied Logistic Regression

- Hosmes, David W.
- Lemeshow, Stanley

➔ Material de Aula

- Pinto, Rogério de M. C.
- www.inf.ufsc.br/~ogliari/arquivos/