

# Métodos Estatísticos Avançados em Epidemiologia

Modelo de Regressão Logística

Enrico A. Colosimo

Departamento de Estatística  
Universidade Federal de Minas Gerais  
<http://www.est.ufmg.br/~enricoc>

# Regressão Logística Binária

- 1 Característica Básica: Desfecho BINÁRIO
- 2 Objetivo:
  - Identificar Fatores de Risco ou Prognóstico;
  - Comparar grupos, controlando por fatores de confusão;
  - Predição.
- 3 Referências Bibliográficas Clássicas:
  - Applied Logistic Regression, Hosmer & Lemeshow (2000), 2nd edição;
  - Modelling Binary Data, Collett (2002), 2nd edição;
  - Giolo, Introdução à Análise de Dados Categóricos com Aplicações (2011).

## ORIGINAL ARTICLE

## Probability of Cancer in Pulmonary Nodules Detected on First Screening CT

Annette McWilliams, M.B., Martin C. Tammemagi, Ph.D., John R. Mayo, M.D., Heidi Roberts, M.D., Geoffrey Liu, M.D., Kam Soghrati, M.D., Kazuhiro Yasufuku, M.D., Ph.D., Simon Martel, M.D., Francis Laberge, M.D., Michel Gingras, M.D., Sukhinder Atkar-Khattra, B.Sc., Christine D. Berg, M.D., Ken Evans, M.D., Richard Finley, M.D., John Yee, M.D., John English, M.D., Paola Nasute, M.D., John Goffin, M.D., Serge Pukta, M.D., Lori Stewart, M.D., Scott Tsai, M.D., Michael R. Johnston, M.D., Daria Manos, M.D., Garth Nicholas, M.D., Glenwood D. Goss, M.D., Jean M. Seely, M.D., Kayvan Amjadi, M.D., Alain Tremblay, M.D.C.M., Paul Burrows, M.D., Paul MacEachern, M.D., Rick Bhatia, M.D., Ming-Sound Tsao, M.D., and Stephen Lam, M.D.

## ABSTRACT

## BACKGROUND

Major issues in the implementation of screening for lung cancer by means of low-dose computed tomography (CT) are the definition of a positive result and the management of lung nodules detected on the scans. We conducted a population-based prospective study to determine factors predicting the probability that lung nodules detected on the first screening low-dose CT scans are malignant or will be found to be malignant on follow-up.

## METHODS

We analyzed data from two cohorts of participants undergoing low-dose CT screening. The development data set included participants in the Pan-Canadian Early Detection of Lung Cancer Study (PanCan). The validation data set included participants involved in chemoprevention trials at the British Columbia Cancer Agency (BCCA), sponsored by the U.S. National Cancer Institute. The final outcomes of all nodules of any size that were detected on baseline low-dose CT scans were tracked. Parsimonious and fuller multivariable logistic-regression models were prepared to estimate the probability of lung cancer.

## RESULTS

In the PanCan data set, 1871 persons had 7008 nodules, of which 102 were malignant, and in the BCCA data set, 1090 persons had 5021 nodules, of which 42 were malignant. Among persons with nodules, the rates of cancer in the two data sets were 5.5% and 3.7%, respectively. Predictors of cancer in the model included older age, female sex, family history of lung cancer, emphysema, larger nodule size, location of the nodule in the upper lobe, part-solid nodule type, lower nodule count, and spiculation. Our final parsimonious and full models showed excellent discrimination and calibration, with areas under the receiver-operating-characteristic curve of more than 0.90, even for nodules that were 10 mm or smaller in the validation set.

## CONCLUSIONS

Predictive tools based on patient and nodule characteristics can be used to accurately estimate the probability that lung nodules detected on baseline screening low-dose CT scans are malignant. (Funded by the Terry Fox Research Institute and others; ClinicalTrials.gov number, NCT00751660.)

From Vancouver General Hospital (A.M., J.R.M., K.E., R.F., J.Y., J.E., S.L.) and the British Columbia Cancer Agency (A.M., S.A.-K., S.L.), Vancouver, BC; the Department of Community Health Sciences, Brock University, St. Catharines, ON (M.C.T.); University Health Network—Princess Margaret Cancer Centre and Toronto General Hospital, Toronto (R.R., G.L., K.S., K.Y., M.-S.T.); Juravinski Hospital and Cancer Center, Hamilton, ON (J.G., S.P., L.S., S.T.); Ottawa Hospital Cancer Centre, Ottawa (G.N., G.D.G., J.M.S., K.A.); Institut Universitaire de Cardiologie et de Pneumologie de Québec, Québec City, QC (S.M., F.L., M.C.); Dalhousie University, Halifax, NS (M.R., D.M.); University of Calgary, Calgary, AB (A.T., P.B., P.M.); and Memorial University of Newfoundland, St. John's (R.B.)—all in Canada; the National Cancer Institute, National Institutes of Health, Bethesda, MD (C.D.B.); and Hospital Universitario Austral, Pilar, Buenos Aires (P.N.). Address reprint requests to Dr. Lam at the Department of Integrative Oncology, British Columbia Cancer Agency, 675 W. 10th Ave., Vancouver, BC V5Z 1L3, Canada.

N Engl J Med 2013;369:910-8.  
DOI: 10.1056/NEJMoa1214726  
Copyright © 2013 Massachusetts Medical Society.

## Tipos de Estudos

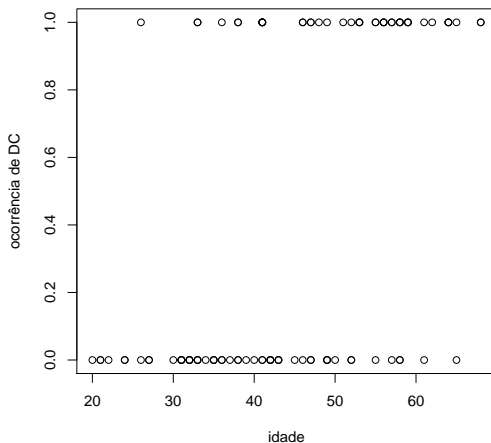
- Estudo Transversal: Regressão Logística usada com frequência.
- Estudo Longitudinal: Regressão Logística pouco, ou raramente, utilizada neste desenho.

## Exemplo - Texto Profa. Suely - pag. 98.

- Uma amostra de 100 indivíduos acompanhados por um período pré-determinado.
- Resposta: incidência de doença coronariana.
- Resposta para cada indivíduo foi sim (1) ou não (0).
- Covariável de interesse: 8 faixas etárias (idade): 20-29, ..., 60-69.
- Dados aparecem na pag. 98 do texto da Profa. Suely.
- 43 ocorrências de doença coronariana.

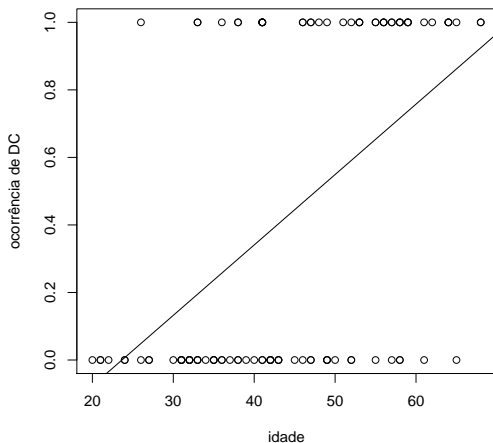
## Gráfica de Dispersão

Desfecho: 43 casos



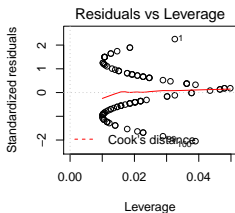
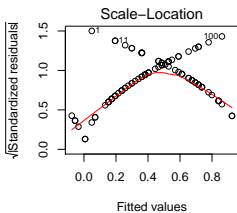
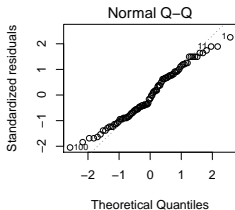
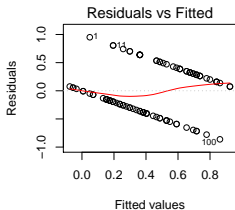
## Regressão Linear Simples

Desfecho: 43 casos



# Regressão Linear Simples- Análise de Resíduos

Desfecho: 43 casos





## Regressão Linear - Análise de Resíduos

- 1 Teste Homocedasticidade: valor absoluto dos resíduos vs preditos  
valor-p = 0,0158
- 2 Teste Normalidade: Shapiro-Wilk  
valor-p = 0,06034

## Descrever os Dados Agrupados

Faixa Etária	Sim	Não	Prop. DC
20-29 (25)	1	9	0,10
30-34 (32)	2	13	0,15
35-39 (38)	3	9	0,25
40-44 (43)	5	10	0,33
45-49 (47)	6	7	0,46
50-54 (53)	5	3	0,63
55-59 (57)	13	4	0,76
60-69 (65)	8	2	0,80

## Entrada dos Dados

Existem duas formas de entrada dos dados para resposta binária.

- Uma linha para cada indivíduo:

indivíduo	faixa etária	resposta
1	1 (25)	0
.....	..	.
100	5 (47)	1
Total	...	43

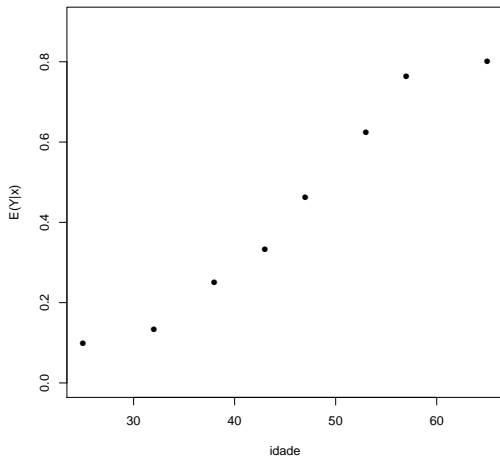
## Entrada dos Dados

Existem duas formas de entrada dos dados para resposta binária.

- Uma linha para cada combinação de covariáveis.

Faixa Etária	Sim	Não
20-29 (25)	1	9
30-34 (32)	2	13
35-39 (38)	3	9
40-44 (43)	5	10
45-49 (47)	6	7
50-54 (53)	5	3
55-59 (57)	13	4
60-69 (65)	8	2

## Descrição Gráfica por Faixa Etária



Porque não usar o modelo de regressão linear?

média (Y:desfecho) =  $E[Y] = P[Y = 1/X]$ .

Ou seja, vamos modelar a probabilidade de ocorrência do evento de interesse.

Limitações:

- 1  $DP(Y) \propto E[Y]$  (viola homocedasticidade);
- 2  $0 \leq P[Y = 1/X] \leq 1$ .
- 3 desfecho não tem distribuição normal.

## Regressão Logística Binária

Vamos considerar inicialmente o caso mais simples, com somente uma covariável.

Candidatas Naturais (Distribuição Acumulada):

- $\pi(x) = \frac{\exp(x)}{1+\exp(x)}$  (logit)
- $\pi(x) = \Phi(x)$  (probit/normal)

$$\pi(x) = P[Y = 1/x] = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

Transformação LOGIT

$$\text{logit}(x) = \log \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x$$

# Regressão Logística Binária

- 1 Forma do Modelo
  - Logit (Regressão Logística).
  - Outras formas menos utilizadas: Probit, Complemento log log.
- 2 Inferência para  $\beta$ 
  - Estimador de Máxima Verossimilhança;
  - Propriedades dos Estimadores;
- 3 Técnicas de Adequação do modelo.
- 4 Interpretação do modelo (razão de chances).



## Estimador de Máxima Verossimilhança

- Inferência Estatística: (Modelo + Método) + Dados
- Método: Máxima Verossimilhança (Sir. Ronald Fisher)
- Função de Verossimilhança: medida da informação fornecida pelos dados para um ou mais parâmetros de um modelo probabilístico.
- Ideia: valores dos parâmetros que tornam os dados amostrais mais prováveis devem ser preferidos àqueles que os fazem menos provável.

## Função de Verossimilhança

- Modelo: Regressão Logística
- Dados
  - $Y$ : desfecho (43 eventos e 57 não eventos)
  - $X$ : idade dos indivíduos
- Função de Verossimilhança:
  - Um evento contribui com:

$$P[Y = 1/x] = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

- Um não-evento contribui com:

$$P[Y = 0/x] = 1 - P[Y = 1/x] = \frac{1}{1 + \exp(\beta_0 + \beta_1 x)}$$

## Modelo Linear Logístico - Um único regressor.

Considere uma amostra de respostas binárias  $y_1, \dots, y_n$  e uma covariável  $x_1, \dots, x_n$ . A função de verossimilhança é

$$\begin{aligned} L(\beta_0, \beta_1) &= \prod_{i=1}^n \left( \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \right)^{y_i} \left( \frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)} \right)^{1-y_i} \\ &= \prod_{i=1}^n \frac{\exp(\beta_0 + \beta_1 x_i)^{y_i}}{1 + \exp(\beta_0 + \beta_1 x_i)} \end{aligned}$$

O estimador de máxima verossimilhança é o valor de  $(\beta_0, \beta_1)$  que maximiza a função de verossimilhança.

## Propriedades do EMV

O EMV tem, para grandes amostras, distribuição normal;

Outra estatística: Razão de verossimilhanças.

Estas estatísticas são utilizadas para testar hipóteses e construir intervalos de confiança.

## Resultados do Exemplo

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-5.12300	1.11111	-4.611	4.01e-06	***
idade	0.10578	0.02337	4.527	5.99e-06	***

Number of Fisher Scoring iterations: 4

## Resultados do Ajuste

Y: presença ou não de doença coronariana;

X: idade (em anos);

$n = 100$ .

Variável	Estimativa	E.P.	Wald
Idade	0,106	0,023	4,53 ( $p < 0,001$ )
Constante	-5,123	1,11	-4,61 ( $p < 0,001$ )

$$\hat{\pi}(x) = \frac{\exp(-5,12 + 0,106 \text{ idade})}{1 + \exp(-5,12 + 0,106 \text{ idade})}$$

$$\widehat{\text{logit}}(x) = -5,12 + 0,106 \text{ idade}$$

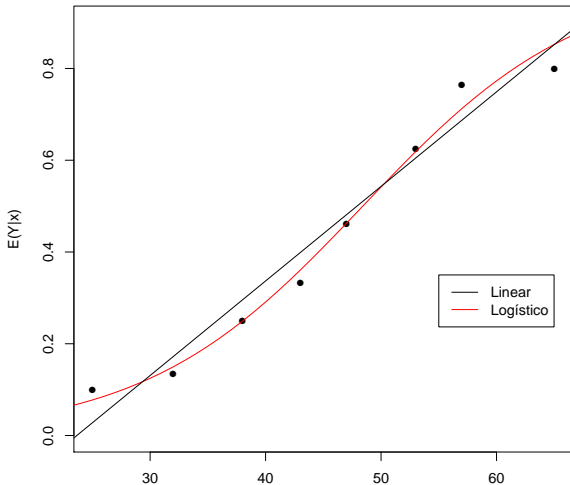
$$\log(\text{verossimilhança}) = \log L(\hat{\beta}_0, \hat{\beta}_1) = -10,86$$

Sob  $H_0 : \beta_1 = 0$ ,  $\log L(\hat{\beta}_0) = -24,92$ .

$$\text{TRV} = 2(-10,86 + 24,92) = \text{Null Deviance} - \text{Residual Deviance} = 28,118 = 5.3^2$$

## Resultados do Ajuste

Interpretação: Razão de chances =  $\exp(0,1058) = 1,11$  (1,06;1,16), isto significa que para o aumento de um ano na idade a chance de doença coronariana aumenta em 11%.



## Verificando a Adequação do Ajuste

- $n$ : tamanho da amostra
- $N$ : número de possíveis combinações da covariável.
- $n_i$ : é o tamanho de cada combinação de covariável e  $y_i$  é o número de "sucessos".
- Exemplo ( $n=100$ ,  $N=8$ )

Idade	$y_i$	$n_i$
25	1	10
32	2	15
38	3	12
43	5	15
47	6	13
53	5	8
57	13	17
65	8	10



## Verificando a Adequação do Ajuste

- Resíduo de Pearson;

$$Q_P = \sum_{i=1}^N \frac{(y_i - n_i \hat{\pi}_i)^2}{n_i \hat{\pi}_i}$$

- Resíduo do desvio (deviance).

$$Q_D = 2 \sum_{i=1}^N n_i \log(y_i / (n_i \hat{\pi}_i))$$

- Estes dois resíduos têm, para grandes amostras, uma distribuição qui-quadrado com  $N-p$  graus de liberdade ( $p$ : número de parâmetros do modelo).
- Estes dois resíduos somente têm validade se  $N$  for bem menor que  $n$ . Em caso contrário, devemos utilizar o teste de Hosmer e Lemeshow.

## Testes de Adequação do Ajuste

- Desvio

$$Q_D = 0,584; \quad \text{valor-p} = 0,997$$

- Pearson

$$Q_P = 0,597; \quad \text{valor-p} = 0,996$$

## Teste de Hosmer e Lemeshow

- Quando  $N \approx n$  os testes de Pearson e do Desvio não podem ser utilizados.
- Hosmer e Lemeshow (1980) propuseram agrupar os dados baseado nas probabilidades estimadas. Usualmente utilizamos no máximo  $g = 10$  grupos. Duas formas de agrupar:
  - $(0; 0, 10), \dots, (0, 9; 1)$ .
  - utilizando percentis  $(0.1, \dots, 0.9)$  das probabilidades estimadas.
- A estatística teste é do tipo Pearson:

$$HL = \sum_{i=1}^g \frac{(y_i - n_i \bar{\pi}_i)^2}{n_i (\bar{\pi}_i) (1 - \bar{\pi}_i)}$$

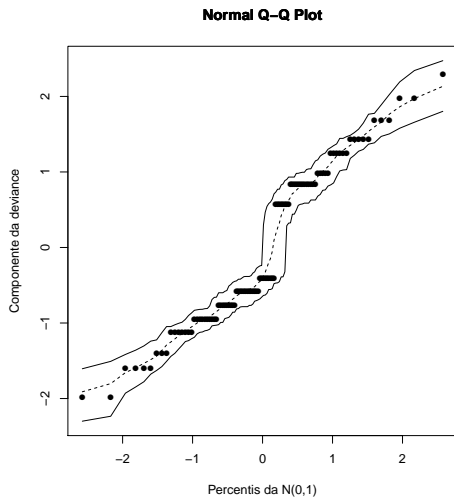
sob  $H_0$  (modelo é adequado), HL tem uma distribuição qui-quadrado com  $g - 2$  graus de liberdade.

## Técnica Gráfica: Envelope simulado

Uma técnica simples chamada de "envelope simulado" permite avaliar a adequação do modelo a partir de resíduos padronizados simulados.

- Obtenha os resíduos padronizados para o modelo de regressão logística.
- Gere valores simulados para estes resíduos;
- Construa bandas de confiança para comparar os resíduos observados com percentis da distribuição normal padrão.
- Se o modelo estiver adequado, os resíduos padronizados do modelo devem estar dentro das bandas de confiança.

## Exemplo: Envelope simulado



## Regressão Logística Múltipla

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1})}{1 + \exp(-\beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1})}$$

$$\text{logit}(x) = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1}$$

### EXEMPLO

- $Y$ : mortalidade infantil;
- $X_1$ : educação da mãe;
- $X_2$ : número de uniões da mãe;
- $X_3$ : região geográfica (urbana ou rural);
- $X_4$ : idade da mãe.
- $p = 5$ .

## Interpretando os Coeficientes Estimados

### 1- Regressor Dicotômico

	Y=1	Y=0
X = 1	$\pi(1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$	$1 - \pi(1) = \frac{1}{1 + \exp(\beta_0 + \beta_1)}$
X = 0	$\pi(0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$	$1 - \pi(0) = \frac{1}{1 + \exp(\beta_0)}$

$$RC = \frac{\pi(1)/1 - \pi(1)}{\pi(0)/1 - \pi(0)} = \exp(\beta_1)$$

## Interpretando os Coeficientes Estimados

### 1- Regressor Dicotômico

EXEMPLO:

Y: mortalidade infantil (0 - não e 1- morte)

X: número de uniões da mãe (0 - > 1 e 1 - 1 união))

RESULTADOS:  $\hat{\beta}_1 = 0,765$  e portanto  $\widehat{RC} = \exp(0,765) = 2,2$ .

INTERPRETAÇÃO: a chance de mortalidade infantil entre mulheres com uma união é cerca de duas vezes (2,2) à daquelas com mais de uma união.



## Interpretando os Coeficientes Estimados

### 2- Regressor Categórico

Y: mortalidade infantil

X: raça (branca, parda ou preta)

Raça	$X_1$	$X_2$
0-Branca	0	0
1-Parda	1	0
2-Preta	0	1

Resultado:  $\widehat{\beta}_1 = 0,40$ ,  $\widehat{\beta}_2 = 1,1$  e, portanto,  $\widehat{RC}_{10} = \exp(0,40) = 1,5$ ,  $\widehat{RC}_{20} = \exp(1,1) = 3$  e  $\widehat{RC}_{21} = \exp(1,1 - 0,4) = 2$

INTERPRETAÇÃO: a chance de mortalidade infantil entre mulheres pardas é cerca de 1,5 vezes à das brancas.

## Interpretando os Coeficientes Estimados

### 3- Regressor Contínuo

Vai depender da forma como o regressor entrou no modelo e, principalmente, na sua unidade de medida.

Y: Mortalidade Infantil

X: idade da mãe (em anos)

Resultado:

$$\widehat{\text{logit}}(X) = -1,8 - 0,05X$$

$$\widehat{RC}(1) = \exp(-0,05) = 0,95$$

INTERPRETAÇÃO: Isto indica que cada aumento de um ano na idade da mãe reduz a mortalidade infantil em 5%.

## Interpretando os Coeficientes Estimados

INCONVENIENTES do Regressor Contínuo.

- 1 Interpretação em termos de acréscimos de um ano pode não ter interesse clínico. Por exemplo, interpretação em termos de acréscimo de quatro anos:

$$\widehat{RC}(4) = \exp(4\widehat{\beta}_1) = 0,82$$

Isto significa que um aumento de 4 anos na educação da mãe reduz a mortalidade infantil em cerca de 18%. E um intervalo de aproximadamente 95% de confiança é dado por:

$$\exp(c\widehat{\beta}_1 \pm 1,96(c) E.P.(\widehat{\beta}_1))$$

NO EXEMPLO:

$$\exp(4(-0,05) \pm 1,96(4)(0,015)) \text{ ou } (0,73; 0,92)$$

é um intervalo aproximado de 95% de confiança para  $RC(4)$ .

### INCONVENIENTES do Regressor Contínuo.

#### 2 Faixa de Comparação

- O modelo assume que a redução de mortalidade infantil quando comparamos uma mãe analfabeta com uma com 4 anos de educação é igual a comparar uma com 20 anos com outra de 24 anos de educação.
- Provavelmente a afirmação acima não é verdadeira.
- Esta é restrita do modelo estatístico.

## Interpretando os Coeficientes Estimados

4 - Caso: Mais de uma covariável.

EXEMPLO:

Y: Mortalidade Infantil

$X_1$ : região (urbana e rural)

$X_2$ : educação da mãe (em anos).

	Região	
	rural	urbano
No. mortes	9	4
Média de Ed. Mãe	3	10
No. Crianças	100	100

$$\widehat{RC}(\text{não ajustado por educação}) = \frac{9/91}{4/96} = 2,4$$

## Interpretando os Coeficientes Estimados

4 - Caso: Mais de uma covariável.

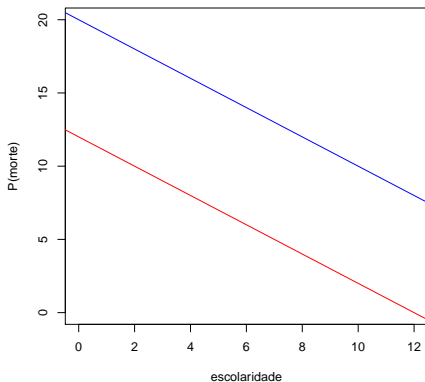
PERGUNTA: Esta diferença é realmente devido aos grupos ou também à educação da mãe?

- Em outras palavras, a educação da mãe pode estar confundido o efeito de região (meio)?
- Utilizando o modelo de regressão logística podemos ajustar por região e educação da mãe.

- $$\widehat{RC}(\text{ajustado por educação}) = 1,5$$

## Ideia de Interação

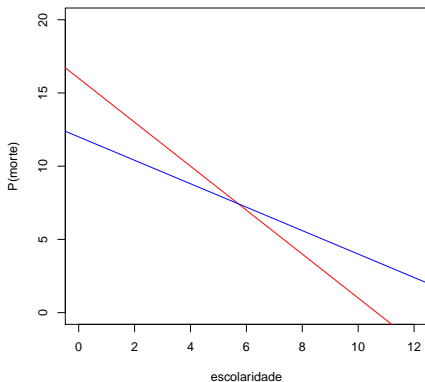
- Sem Interação



- $Y$ : mortalidade infantil (0/1);
  - $x$ : escolaridade (anos);
- grupos: 1 e 2 (exemplo: urbano e rural)

## Ideia de Interação

- Com Interação



- $Y$ : logit ( $x$ );
- $x$ : escolaridade (anos);
- grupos: 1 e 2 (exemplo: urbano e rural)



## Interpretando os Coeficientes Estimados - Com Interação

- $Y$ : mortalidade infantil (0- não/ 1- sim);
- $x_1$ : escolaridade da mãe (anos);
- $x_2$ : meios : 0-urbano e 1-rural)

### Modelo Logístico Ajustado

$$\widehat{\text{logit}}(x) = 2,2 - 0,05\text{escolaridade} + 0,41\text{meio} - 0,02(\text{escolaridade} * \text{meio})$$

- Modelo para Meio Urbano

$$\widehat{\text{logit}}(x) = 2,2 - 0,05 \text{ escolaridade}$$

- Modelo para Meio Rural

$$\begin{aligned}\widehat{\text{logit}}(x) &= 2,2 - 0,05 \text{ escolaridade} + 0,41 - 0,02 \text{ escolaridade} \\ &= 2,61 - 0,07 \text{escolaridade}\end{aligned}$$

## Interpretando os Coeficientes Estimados - Com Interação

### 1 Interpretando Escolaridade

- Meio Urbano

$$\widehat{\text{logit}}(x) = 2,2 - 0,05 \text{ escolaridade}$$

$$\widehat{RC} = \exp(-0,05) = 0,95.$$

A chance de mortalidade infantil reduz em 5% para cada ano a mais de escolaridade da mãe no meio urbano.

- Meio Rural

$$\begin{aligned}\widehat{\text{logit}}(x) &= 2,2 - 0,05 \text{ escolaridade} + 0,41 - 0,02 \text{ escolaridade} \\ &= 2,15 - 0,07 \text{ escolaridade}\end{aligned}$$

$$\widehat{RC} = \exp(-0,07) = 0,93.$$

A chance de mortalidade infantil reduz em 7% para cada ano a mais de escolaridade da mãe no meio rural.

## Interpretando os Coeficientes Estimados - Com Interação

### 2 Interpretando Meio

- Mães com 1 ano de educação

$$\begin{aligned}\hat{\logit}(x) &= 2,2 - 0,05 + 0,41 \text{ meio} - 0,02 \text{ meio} \\ &= 2,15 + 0,39 \text{ meio}\end{aligned}$$

$$\hat{RC} = \exp(0,39) = 1,48$$

A chance de mortalidade infantil no meio rural é 1,5 vezes a do urbano para mães com 1 ano de educação.

- Mães com 8 anos de educação (ensino fundamental completo)

$$\begin{aligned}\hat{\logit}(x) &= 2,2 - 0,40 + 0,41 \text{ meio} - 0,16 \text{ meio} \\ &= 1,80 + 0,25 \text{ meio}\end{aligned}$$

$$\hat{RC} = \exp(0,25) = 1,28$$

A chance de mortalidade infantil no meio rural é 1,3 vezes a do urbano para mães com 8 anos de educação.

- Mães com 15 anos de educação (ensino superior)

$$\hat{RC} = \exp(0,11) = 1,12$$

A chance de mortalidade infantil no meio rural é 1,12 vezes a do urbano para mães com curso superior.

### Exemplo 7.3.1 - pag. 114, Giolo (2011)

Y: presença ou não de doença coronariana;

$X_1$ : ECG (0/1);

$X_2$ : sexo (0/1)

$n = 78$ .

Variável	Estimativa	E.P.	Wald
sexo	1,227	0,498	2,56 ( $p = 0,0103$ )
ecg	1,055	0,498	2,12 ( $p = 0,034$ )
Constante	-1,1747	0,485	2,118 ( $p = 0,0155$ )

$$\widehat{\text{logit}}(x) = -1,17 + 1,28 \text{ sexo} + 1,06 \text{ ecg}$$

## Testes de Adequação do Ajuste

- `ajust$fitted.values`

1 2 3 4 0.2360103 0.4699914 0.5255469 0.7607465

- `ajust$y`

1 2 3 4 0.2666667 0.4444444 0.5000000 0.7777778

- `ajust$residuals`

1 2 3 4 0.17002058 -0.10255715 -0.10245520 0.09357272

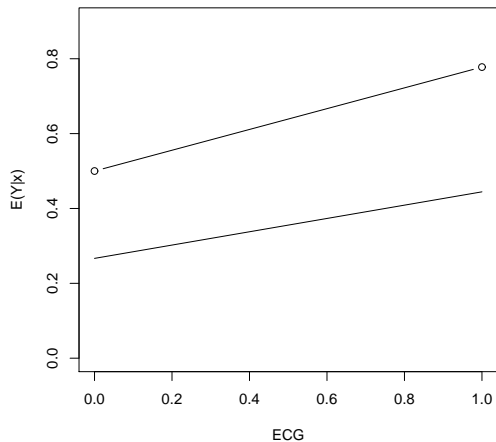
- Deviance:

$$Q_D = 0.5837769; \quad \text{valor} - p = 0.4448357$$

- Pearson:

$$Q_P = 0.2154859; \quad \text{valor} - p = 0.6425012$$

## Gráfico de Interação



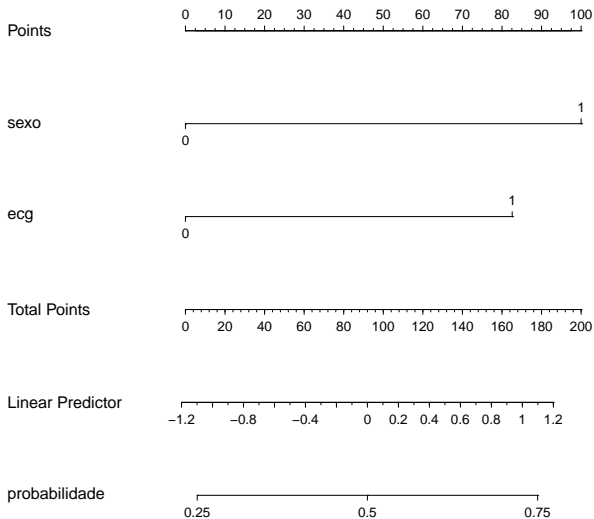
$TRV = 0.214 - 0 = 0.214$  valor-p = 0,644.

## Resultados do Ajuste

Interpretação: Razão de chances

- Sexo:  $RC = \exp(1,227) = 3,4$ , isto significa a chance de doença coronariana entre os homens é 3,4 vezes a chance entre as mulheres.
- ECG:  $RC = \exp(1,054) = 2,9$ , isto significa a chance de doença coronariana entre aqueles com ECG alterado é cerca de 3 vezes a chance entre os com ECG normal.

## Apresentação dos Resultados: Nomograma





## Breve Roteiro para a Construção de Modelos

- 1 Descrever o estudo e seus objetivos.
- 2 Desenho do estudo.
- 3 Descrição das variáveis.
- 4 Análise univariada (avaliar a forma das covariáveis quantitativas).
- 5 Modelo final obtido por um método de seleção de variáveis.
- 6 Incluir e testar os termos de interação.
- 7 Verificar a adequação do modelo final.
- 8 Interpretar o modelo final.

## Aplicação: Mini Avaliação Nutricional (MAN) para Idosos

### 1 Motivação

- A desnutrição é uma condição que se inicia com o baixo consumo de nutrientes podendo evoluir para estados mais graves;
- Desafios para os geriatras é identificar os idosos que necessitam de uma intervenção dietética.
- O ideal seria que uma avaliação nutricional completa mas fica restrita aos seus custos e ao tempo demandado para tal.
- A Mini Avaliação Nutricional (MAN) foi desenvolvida com o objetivo de proporcionar um rápido diagnóstico do estado nutricional.

- 2 O score final da MAN classifica:  $\leq 24$  - risco de desnutrição,  
 $> 24$  - bem nutrido.

## Mini Nutritional Assessment MNA®

Nestlé  
Nutrition Institute

Apelido:		Nome:			
Sexo:	Idade:	Peso, kg:	Altura, cm:	Data:	

Responda à secção "Triagem", preenchendo as caixas com os números adequados. Some os números da secção "Triagem". Se a pontuação obtida for igual ou menor que 11, continue o preenchimento do questionário para obter a pontuação indicadora de desnutrição.

Triagem	
<b>A Nos últimos três meses houve diminuição da ingestão alimentar devido a perda de apetite, problemas digestivos ou dificuldade para mastigar ou deglutir?</b>	
0 = diminuição grave da ingestão	<input type="checkbox"/>
1 = diminuição moderada da ingestão	
2 = sem diminuição da ingestão	
<b>B Perda de peso nos últimos 3 meses</b>	
0 = superior a três quilos	<input type="checkbox"/>
1 = não sabe informar	
2 = entre um e três quilos	
3 = sem perda de peso	
<b>C Mobilidade</b>	
0 = não anda ao longo de 4 cadeiras de rodas	<input type="checkbox"/>
1 = deambula mas não é capaz de sair de casa	
2 = normal	
<b>D Passou por algum stress psicológico ou doença aguda nos últimos três meses?</b>	
0 = sim	<input type="checkbox"/>
2 = não	
<b>E Problemas neuropsíquicos</b>	
0 = demência ou depressão graves	<input type="checkbox"/>
1 = demência ligeira	
2 = sem problemas psicológicos	
<b>F Índice de Massa Corporal = peso em kg / (estatura em m)<sup>2</sup></b>	
0 = IMC < 19	<input type="checkbox"/>
1 = 19 ≤ IMC < 21	
2 = 21 ≤ IMC < 23	
3 = IMC ≥ 23	
<b>Pontuação da Triagem (subtotal, máximo de 14 pontos)</b>	<input type="checkbox"/> <input type="checkbox"/>
12-14 pontos: estado nutricional normal	
8-11 pontos: sob risco de desnutrição	
0-7 pontos: desnutrido	
Para uma avaliação mais detalhada, continue com as perguntas G-R	
Avaliação global	
<b>G O doente vive na sua própria casa (ou em instituição geriátrica ou hospital)</b>	<input type="checkbox"/>
0 = sim	0 = não
<b>H Utiliza mais de três medicamentos diferentes por dia?</b>	<input type="checkbox"/>
0 = sim	1 = não
<b>I Lesões de pele ou escaras?</b>	<input type="checkbox"/>
0 = sim	1 = não

**Referências:**  
 1. Vellas, Vilain, H., Rolland, G., et al. Overview of the MNA® - Its History and Current Status. *Jour. Health Aging*, 2006, 19:446-466.  
 2. Rubenstein, L.J., Hawker, J.D., Savva, A., Guigoz, Y., Vellet, B. Screening for Undernutrition in Geriatric Practice: Developing the Short-Form Mini Nutritional Assessment (MNA-SF). *J. Gerontol.* 2001; 56A: 1038-1077.  
 3. Guigoz, Y. The Mini-Nutritional Assessment (MNA)®: Review of the Literature - What does it tell us? *J. Nutr Health Aging*, 2006, 10:404-410.  
 © Société des Produits Nestlé, S.A., Vevey, Switzerland, Trademark Owners.  
 © Nestlé, Vevey, Switzerland. Nestlé logo: 1989-1994.  
 Para maiores informações: [www.mna-nutrition.com](http://www.mna-nutrition.com)

<b>J Quantas refeições faz por dia?</b>	
0 = uma refeição	<input type="checkbox"/>
1 = duas refeições	
2 = três refeições	
<b>K O doente consome:</b>	
+ pelo menos uma porção diária de leite ou derivados (leite, queijo, iogurte)?	sim <input type="checkbox"/> não <input type="checkbox"/>
+ duas ou mais porções semanais de leguminosas ou ovos?	sim <input type="checkbox"/> não <input type="checkbox"/>
+ carne, peixe ou aves todos os dias?	sim <input type="checkbox"/> não <input type="checkbox"/>
0.0 = nenhuma ou uma resposta "sim"	<input type="checkbox"/> <input type="checkbox"/>
0.5 = duas respostas "sim"	
1.0 = três respostas "sim"	
<b>L O doente consome duas ou mais porções diárias de fruta ou produtos hortícolas?</b>	
0 = não	<input type="checkbox"/>
1 = sim	
<b>M Quantos copos de líquidos (água, sumo, café, chá, leite) o doente consome por dia?</b>	
0.0 = menos de três copos	<input type="checkbox"/> <input type="checkbox"/>
0.5 = três a cinco copos	
1.0 = mais de cinco copos	
<b>N Modo de se alimentar</b>	
0 = não é capaz de se alimentar sozinho	<input type="checkbox"/>
1 = alimenta-se sozinho, porém com dificuldade	
2 = alimenta-se sozinho sem dificuldade	
<b>O O doente acredita ter algum problema nutricional?</b>	
0 = acredita estar desnutrido	<input type="checkbox"/>
1 = não sabe dizer	
2 = acredita não ter um problema nutricional	
<b>P Em comparação com outras pessoas da mesma idade, como considera o doente a sua própria saúde?</b>	
0 = pior	<input type="checkbox"/> <input type="checkbox"/>
0.5 = não sabe	
1.0 = igual	
2.0 = melhor	
<b>Q Perímetro braquial (PB) em cm</b>	
0.0 = PB < 21	<input type="checkbox"/> <input type="checkbox"/>
0.5 = 21 ≤ PB ≤ 22	
1.0 = PB > 22	
<b>R Perímetro da perna (PP) em cm</b>	
0 = PP < 31	<input type="checkbox"/>
1 = PP ≥ 31	
<b>Avaliação global (máximo 16 pontos)</b>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<b>Pontuação da triagem</b>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<b>Pontuação total (máximo 30 pontos)</b>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

**Avaliação do Estado Nutricional**

de 24 a 30 pontos  estado nutricional normal

de 17 a 23,5 pontos  sob risco de desnutrição

menor de 17 pontos  desnutrido

## Aplicação: Mini Avaliação Nutricional (MAN) para Idosos

- 3 Descrição do Estudo: Um estudo transversal foi conduzido na FM da UFMG com 33 idosos para verificar a relação entre as variáveis bioquímicas (hemoglobina e ferritina) e antropométricas (ângulo de fase e percentual de gordura corporal) e o escore obtido por meio da aplicação da Mini Avaliação Nutricional.
- 4 Objetivo: avaliar se a MAN seria uma boa ferramenta para prever alterações bioquímicas e antropométricas características da desnutrição.
- 5 Covariáveis:
  - Categóricas: gênero;
  - Contínuas: idade (anos), ângulo de fase (o), percentual de gordura (%) e ferritina (ng/ml)

## MAN para Idosos: Análise Descritiva

```
> summary(dados)
```

MAN	Sexo	Idade	Angulo	GC
<24 :15	Feminino :16	Min. :60.00	Min. :4.400	Min. :17.23
>=24:18	Masculino:17	1st Qu.:68.00	1st Qu.:4.830	1st Qu.:21.44
		Median :75.00	Median :5.500	Median :27.25
		Mean :73.88	Mean :5.785	Mean :26.85
		3rd Qu.:78.00	3rd Qu.:6.160	3rd Qu.:31.50
		Max. :90.00	Max. :8.820	Max. :35.12

Ferritina

```
Min. : 10.10  
1st Qu.: 34.50  
Median : 62.90  
Mean : 67.42  
3rd Qu.: 99.40  
Max. :175.10
```

## MAN para Idosos: Multicolinearidade

```
> cor(dados[3:7])
```

	Idade	Angulo	GC	Ferritina
Idade	1.0000000	-0.482143419	0.146190678	-0.2352079
Angulo	-0.4821434	1.000000000	0.004429318	0.2243951
GC	0.1461907	0.004429318	1.000000000	-0.1695677
Ferritina	-0.2352079	0.224395065	-0.169567725	1.0000000

## MAN para Idosos: Análise Bivariada

```
> summary(glm(MAN ~ Angulo, family = binomial , data = dados))

Call: glm(formula = MAN ~ Angulo, family = binomial(link = "logit"),
  data = dados)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4776  -1.1098   0.4329   1.0747   1.4811

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.8729     2.2198  -1.745   0.0810 .
      Angulo    0.7152     0.3954   1.809   0.0705 .
---

(Dispersion parameter for binomial family taken to be 1)

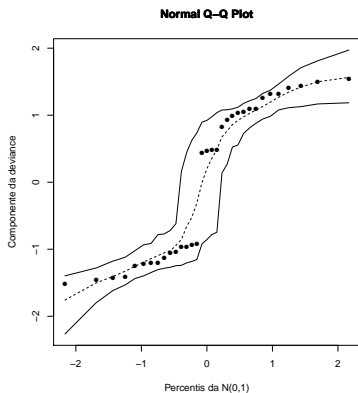
    Null deviance: 45.475  on 32  degrees of freedom
Residual deviance: 40.968  on 31  degrees of freedom AIC: 44.968
```

## MAN para Idosos: Adequação do modelo

```
> hosmerlem(y-1, ajuste4$fitted.values, g = 10) ## Ok
$chisq
[1] 3.637034

$p.value
[1] 0.8882982

$df
[1] 8
```





## MAN para Idosos: Interpretação do modelo

- Modelo Final  $\rightarrow$  Ângulo
- Interpretação

$$\widehat{RC} = \exp(0.7152) = 2 \text{ (IC 90\%; 1.2; 4.4)}$$

A chance de eutrófico (bem nutrido) aumenta em 100% para cada uma unidade de aumento do ângulo de fase.

- A análise foi muito pobre por conta do (pequeno) tamanho da casuística.