BIOESTATÍSTICA

Análise de Regressão Logística

MARIA DO ROSÁRIO D O LATORRE

GLEICE M S CONCEIÇÃO

FSP USP

Tipos de Variáveis

BIOESTATÍSTICA

Para escolher a medida ou o gráfico mais adequado devemos levar em conta o tipo de variável que está sendo analisada.

- variáveis qualitativas ou categóricas
  - qualitativa nominal (sexo, tipo de doença)
  - qualitativa ordinal (escolaridade)

- variáveis quantitativas ou numéricas
  - quantitativa discreta (número de filhos)
  - quantitativa contínua (peso, altura, anos de estudo)

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Quando as duas variáveis são qualitativas

BIOESTATÍSTICA

Estratégia

- ✓ Técnicas similares àquelas aprendidas na descrição de variáveis qualitativas:
- ✓ Tabelas de frequência conjunta
- ✓ Percentuais na linha e/ou coluna
- ✓ Gráfico de barras
- ✓ Medidas de associação
- ✓ Medidas de risco

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de associação e medidas de risco

BIOESTATÍSTICA

Variáveis qualitativas: Exposição e Doença

		Doença		
		Sim	Não	Total
Exposição	Sim	a	b	a+b
	Não	c	d	c+d
	Total	a+c	b+d	N

a + b = nº total de indivíduos expostos

c + d = nº total de indivíduos não expostos

a + c = nº total de indivíduos com a doença

b + d = nº total de indivíduos sem a doença

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

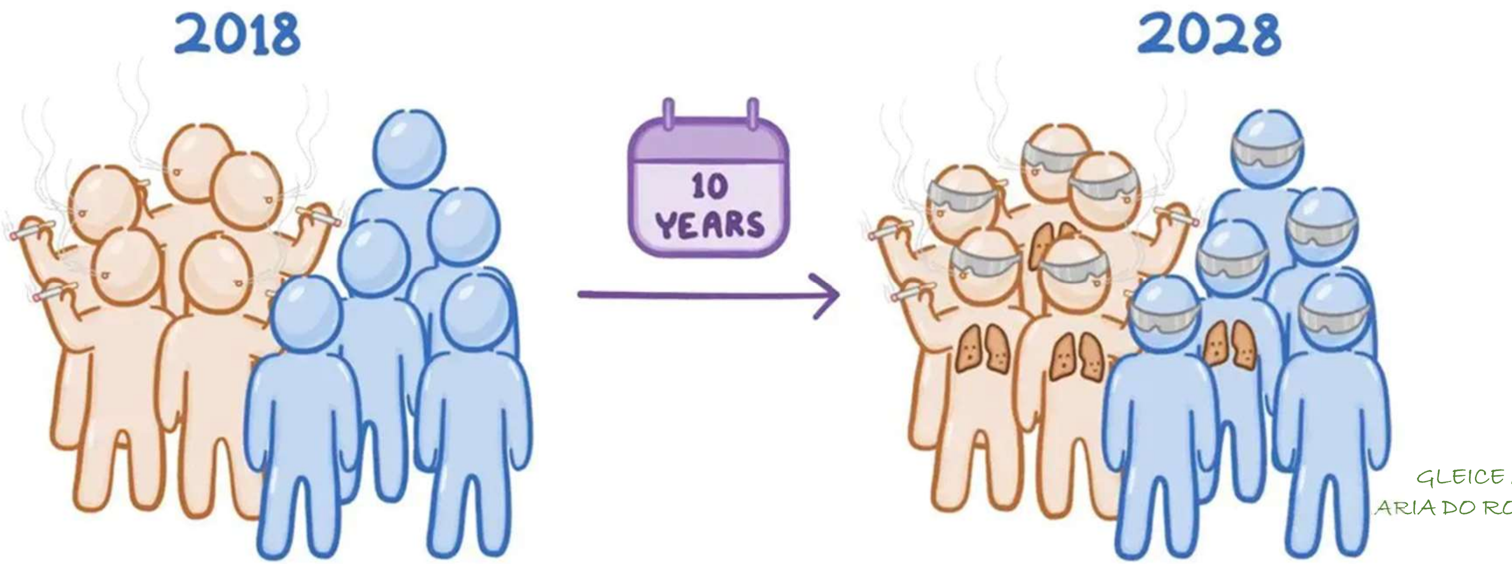
Alguns tipos de estudos epidemiológicos

BIOESTATÍSTICA

- ✓ Coorte
- ✓ Transversal
- ✓ Caso –controle

Coorte

Um grupo de indivíduos **expostos** e um grupo de **não expostos** a **fatores de risco** para a **doença em estudo** são seguidos ao longo de um período de tempo fixado, e verifica-se quem desenvolveu e quem não desenvolveu a doença de interesse.



GLEICE M S CONCEIÇÃO
ARIA DO ROSÁRIO D D LATORRE
FSP - USP

Coorte

BIOESTATÍSTICA

Variáveis qualitativas: Exposição e Doença

		Doença		
		Sim	Não	Total
Exposição	Sim	a	b	a+b
	Não	c	d	c+d
	Total	a+c	b+d	N

fixados

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Transversal

Um grupo de N indivíduos são investigados e cada cada indivíduo é classificado, ao mesmo tempo, como exposto ou não exposto e doente ou não doente.

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Transversal

BIOESTATÍSTICA

Variáveis qualitativas: Exposição e Doença

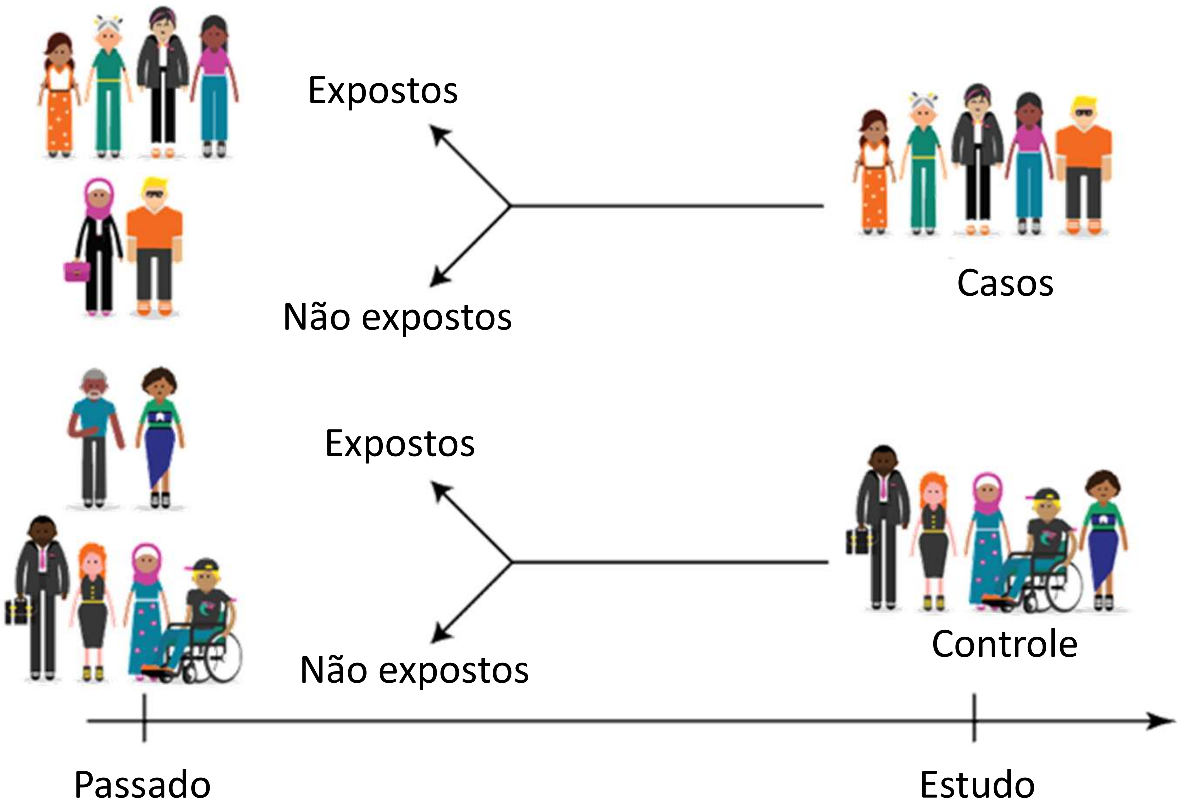
		Doença		
		Sim	Não	Total
Exposição	Sim	a	b	a+b
	Não	c	d	c+d
	Total	a+c	b+d	N

fixado

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Caso-controle

Um grupo de indivíduos **com a doença (casos)** e um grupo **sem a doença (controle)** são investigados e verifica-se quem foi exposto e quem não foi exposto, no **passado**, aos fatores de risco em **estudo**.



GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Caso-controle

BIOESTATÍSTICA

Variáveis qualitativas: Exposição e Doença

		Doença		
		Sim	Não	Total
Exposição	Sim	a	b	a+b
	Não	c	d	c+d
	Total	a+c	b+d	N

fixados

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de associação X medidas de risco

BIOESTATÍSTICA

- ✓ Medidas de associação

Expressam a existência (ou não) de associação entre as duas variáveis

Ex: Qui-quadrado, Fisher, etc.

Ho: não existe associação

Ha: existe associação

Obtidas da mesma forma nos três tipos de estudo.

Medidas de associação X medidas de risco

BIOESTATÍSTICA

✓ Medidas de risco

Expressam a magnitude da associação entre as duas variáveis.

Para cada tipo de estudo há uma medida de risco adequada:

- Coorte : Risco relativo
- Transversal: Razão de Prevalências
- Caso Controle: Razão de Chances

Ho: não existe associação (RR=1 ou OR=1 ou RP=1)

Ha: existe associação (RR≠1 ou OR ≠ 1 ou RP ≠ 1)

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos de coorte

BIOESTATÍSTICA

Doença

Exposição		Sim	Não	Total
	Sim	a	b	a+b
	Não	c	d	c+d
	Total	a+c	b+d	N

fixados

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos de coorte

BIOESTATÍSTICA

Doença

Exposição		Sim	Não	Total
	Sim	a	b	a+b
	Não	c	d	c+d
	Total	a+c	b+d	N

fixados

Sejam:

✓Incidência da doença (ou risco absoluto) em expostos:

$$I_e = \frac{a}{a+b}$$

✓Incidência (ou risco absoluto) em não expostos:

$$I_{ne} = \frac{c}{c+d}$$

Risco relativo (RR):

Razão entre a incidência da doença em expostos (I_e) e a incidência em não expostos (I_{ne}):

$$RR = \frac{I_e}{I_{ne}} = \frac{a/(a+b)}{c/(c+d)}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos de coorte

BIOESTATÍSTICA

Doença

Exposição		Sim	Não	Total
	Sim	a	b	a+b
	Não	c	d	c+d
	Total	a+c	b+d	N

fixados

Sejam:

✓Incidência da doença (ou risco absoluto) em expostos:

$$I_e = \frac{a}{a+b}$$

✓Incidência (ou risco absoluto) em não expostos:

$$I_{ne} = \frac{c}{c+d}$$

Risco Atribuível (RA):

Diferença entre a incidência da doença em expostos (I_e) e a incidência em não expostos (I_{ne}):

$$RA = I_e - I_{ne} = \frac{a}{a+b} - \frac{c}{c+d}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Coorte

Exemplo 1: Em um estudo de coorte para avaliar a associação entre o uso de contraceptivos e infarto do miocárdio, um grupo de mulheres que utilizava CO e um grupo que não utilizava foi seguido durante 30 anos e foram observados os seguintes resultados:

Tabela 1. Número de pacientes segundo o uso de contraceptivo oral e a ocorrência de infarto do miocárdio

		Infarto		
		Sim	Não	Total
Uso de CO	Sim	23	304	327
	Não	133	2816	2949
	Total	156	3120	3276

Medidas de risco em estudos de coorte

BIOESTATÍSTICA

Ex. 1: Infarto

Uso de CO		Sim	Não	Total
	Sim	23	304	327
	Não	133	2816	2949
	Total	156	3120	3276

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos de coorte

BIOESTATÍSTICA

Ex. 1: Infarto

Uso de CO	Sim	Não	Total
Sim	23	304	327
Não	133	2816	2949
Total	156	3120	3276

$$RR = \frac{I_e}{I_0} = \frac{a/(a+b)}{c/(c+d)}$$

$$RR = \frac{23/327}{133/2949} = \frac{0{,}0703}{0{,}0451} = 1{,}56$$

Interpretação:

- O risco de infarto em mulheres que usam contraceptivo oral é 1,56 vezes o risco em mulheres que não usam.

ou

- Mulheres que usam contraceptivo tem 56% mais risco de ter infarto do que as que não usam.

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos de coorte

BIOESTATÍSTICA

Ex. 1:

Uso de CO		Infarto Sim	Infarto Não	Total
	Sim	23	304	327
	Não	133	2816	2949
	Total	156	3120	3276

$$RA = I_e - I_0 = \frac{a}{a+b} - \frac{c}{c+d}$$

$$RA = \frac{23}{327} - \frac{133}{2949} = 0{,}0703 - 0{,}0451 = 0{,}0252$$

Interpretação:

- O risco adicional de infarto devido ao uso de contraceptivo oral é 0,0252.

ou

- O risco de infarto fica acrescido de 0,0252 se houver uso de contraceptivo oral.

Medidas de risco em estudos transversais

BIOESTATÍSTICA

Doença

Exposição

	Sim	Não	Total
Sim	a	b	a+b
Não	c	d	c+d
Total	a+c	b+d	N

fixado

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos transversais

BIOESTATÍSTICA

Doença

Exposição		Sim	Não	Total
	Sim	a	b	a+b
	Não	c	d	c+d
	Total	a+c	b+d	N

fixado

Sejam:

✓ Prevalência da doença em expostos:

$$P_e = \frac{a}{a+b}$$

✓ Prevalência da doença em não expostos:

$$P_{ne} = \frac{c}{c+d}$$

Razão de prevalências (RP):

Razão entre a prevalência da doença em expostos (P_e) e a prevalência em não expostos (P_{ne}):

$$RP = \frac{P_e}{P_{ne}} = \frac{a/(a+b)}{c/(c+d)}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Transversal

BIOESTATÍSTICA

Exemplo 2. Em um estudo transversal para investigar a associação entre sexo e ocorrência de doença coronariana, 500 indivíduos foram selecionados. Foi perguntado o sexo e feita uma avaliação médica sobre a presença de doença coronariana.

Tabela 2. Número de indivíduos segundo o sexo e a ocorrência de doença coronariana.

		Doença Coronariana		
		Sim	Não	Total
Sexo	Masculino	26	229	255
	Feminino	9	236	245
	Total	35	465	500

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos transversais

BIOESTATÍSTICA

Ex. 2:

Sexo	Doença Coronariana: Sim	Não	Total
Masculino	26	229	255
Feminino	9	236	245
Total	35	465	500

$$RP = \frac{a/(a+b)}{c/(c+d)}$$

$$RP = \frac{26/255}{9/245} = \frac{0{,}102}{0{,}0367} = 2{,}778$$

Interpretação:

- A prevalência de DC em homens é 2,8 vezes a prevalência em mulheres.

ou

- A prevalência de DC em homens é aproximadamente o triplo da prevalência em mulheres.

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos caso-controle

BIOESTATÍSTICA

Seja p a probabilidade de um evento ocorrer.

Chance é razão de duas probabilidades:

$$Chance = \frac{p}{1-p}$$

Probabilidade de

- ✓ face Ca em um lançamento de uma moeda honesta: 1/2
- ✓ Face 6 em um lançamento de um dado honesto: 1/6

Chance de

- ✓ face Ca em um lançamento de uma moeda honesta: (1/2) / (1/2) = 1/1 = 1
- ✓ Face 6 em um lançamento de um dado honesto: (1/6) / (5/6) = 1/5 = 0,20
- ✓ Sorteio do No. 345 um número em milhão: (1/1.000.000)/(999.999/1.000.000)= 1/999.999=0,0000001

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos caso-controle

BIOESTATÍSTICA

Doença

Exposição	Sim	Não	Total
Sim	a	b	a+b
Não	c	d	c+d
Total	a+c	b+d	N

fixados

Chance ou *Odds:* razão de duas probabilidades:

$$\text{Chance} = \frac{p}{1-p}$$

✓ Chance de doença em expostos:

Probabilidade de doença em expostos :

$$p_{d|e} = \frac{a}{a+b}$$

Probabilidade de não doença em expostos :

$$1 - p_{d|e} = \frac{b}{a+b}$$

✓ Chance de doença em expostos:

$$O_{d|e} = \frac{p_{d|e}}{1-p_{nd|e}} = \frac{a/(a+b)}{b/(a+b)} = \frac{a}{b}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos caso-controle

	Doença		
Exposição	Sim	Não	Total
Sim	a	b	a+b
Não	c	d	c+d
Total	a+c	b+d	N

fixados

Chance ou *Odds:* razão de duas probabilidades:

$$\text{Chance} = \frac{p}{1-p}$$

✓ Chance de doença em não expostos:

Probabilidade de doença em não expostos:

$$p_{d|ne} = \frac{c}{c+d}$$

Probabilidade de não doença em não expostos:

$$1 - p_{d|ne} = \frac{d}{c+d}$$

✓ Chance de doença em não expostos:

$$O_{d|ne} = \frac{p_{d|ne}}{1-p_{d|ne}} = \frac{c/(c+d)}{d/(c+d)} = \frac{c}{d}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos caso-controle

BIOESTATÍSTICA

Doença

Exposição	Sim	Não	Total
Sim	a	b	a+b
Não	c	d	c+d
Total	a+c	b+d	N

fixados

Chance ou *Odds:* razão de duas probabilidades:

$$\text{Chance} = \frac{p}{1-p}$$

- ✓ Chance de doença em expostos :

$$O_{d|e} = \frac{p_{d|e}}{1-p_{nd|e}} = \frac{a/(a+b)}{b/(a+b)} = \frac{a}{b}$$

- ✓ Chance de doença em não expostos :

$$O_{d|ne} = \frac{p_{d|ne}}{1-p_{d|ne}} = \frac{c/(c+d)}{d/(c+d)} = \frac{c}{d}$$

- ✓ Razão de Chances ou Odds Ratio (OR)

Razão entre a chance de doença em expostos ($O_{d|e}$) e a de doença em não expostos($O_{d|ne}$)

$$OR = \frac{O_{d|e}}{O_{d|ne}} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos caso-controle

BIOESTATÍSTICA

O que o R faz:

Doença

Exposição		Sim	Não	Total
	Sim	a	b	a+b
	Não	c	d	c+d
	Total	a+c	b+d	N

fixados

Chance ou *Odds:* razão de duas probabilidades:

$$\text{Chance} = \frac{p}{1-p}$$

✓ Chance de exposição em doentes:

$$O_{e|d} = \frac{a}{c}$$

✓ Chance de exposição em não doentes:

$$O_{e|nd} = \frac{b}{d}$$

✓ Razão de Chances ou Odds Ratio (OR)

Razão entre a chance de exposição em doentes ($O_{e|d}$)

e a chance de exposição em não doentes ($O_{e|nd}$)

$$OR = \frac{O_{e|d}}{O_{e|nd}} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Caso-controle

BIOESTATÍSTICA

Exemplo 3. Em um estudo do tipo caso controle para avaliar a associação entre o hábito de fumar e a ocorrência de câncer de pulmão, um grupo de indivíduos com câncer e um grupo saudável foi selecionado e foi perguntado a cada indivíduo sobre o hábito de fumar.

Tabela 3. Número de indivíduos segundo a ocorrência de câncer de pulmão e o hábito de fumar.

Hábito de fumar	Câncer: Sim	Câncer: Não	Total
Sim	35	65	100
Não	8	92	100
Total	43	157	200

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos caso-controle

Ex. 3:

Câncer

Hábito de fumar	Sim	Não	Total
Sim	35	65	100
Não	8	92	100
Total	43	157	200

$$OR = \frac{ad}{bc}$$

$$OR = \frac{35 * 92}{65 * 8} = 6,2$$

Interpretação:

- A chance de fumantes virem a ter câncer é 6,2 vezes a chance de não fumantes virem a ter câncer.

ou

- Fumantes têm 6,2 vezes a chance de ter câncer quando comparados a não fumantes.

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco em estudos caso-controle

BIOESTATÍSTICA

Situação 1

		Câncer		
		Sim	Não	Total
Hábito de fumar	Sim	70	30	100
	Não	30	70	100
	Total	100	100	200

$$OR = \frac{70 * 70}{30 * 30} = 5,4$$

$$\cancel{RR = \frac{70/100}{30/100} = 2,3}$$

Situação 2

		Câncer		
		Sim	Não	Total
Hábito de fumar	Sim	70	300	370
	Não	30	700	730
	Total	100	1000	1100

$$OR = \frac{70 * 700}{30 * 300} = 5,4$$

$$\cancel{RR = \frac{70/370}{30/730} = 4,6}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Medidas de risco

BIOESTATÍSTICA

Doença

Exposição

	Sim	Não	Total
Sim	a	b	a+b
Não	c	d	c+d
Total	a+c	b+d	N

Lembrando que:

- ✓ Em estudos de coorte:

  a + b e c + d são fixados: RR

- ✓ Em estudos de caso-controle

  a + c e b + d são fixados: OR

- ✓ Em estudos transversais

  somente N é fixado: RP

- ✓ χ^2 é adequado para os três tipos de estudo.

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Exercício

A Tabela abaixo apresenta os resultados de um estudo

Como obter o RR, a RP e a OR ?

Interprete-as.

Tabela 1. Número de crianças segundo o peso ao nascer e a ocorrência de déficit cognitivo.

		Déficit Cognitivo		
		Sim	Não	Total
Peso ao Nascer	Normal	2	53	55
	Baixo Peso	7	45	52
	Muito baixo peso	12	35	47
	Total	21	133	154

BIOESTATÍSTICA

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Déficit Cognitivo

Peso ao Nascer

	Sim	Não	Total
Normal	2	53	55
Baixo Peso	7	45	52
Muito baixo peso	12	35	47
Total	21	133	154

	Sim	Não	Total
Normal	2	53	55
Baixo Peso	7	45	52

	Sim	Não	Total
Normal	2	53	55
Muito baixo peso	12	35	47

BIOESTATÍSTICA

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística

BIOESTATÍSTICA

- ✓ Variável resposta é qualitativa dicotômica (presença ou ausência de um evento de interesse), também chamada de "desfecho".

Ex:

- Ocorrência de infarto do miocárdio (sim ou não)
- Ocorrência de câncer de pulmão (sim ou não)
- Cura (sim ou não)
- Óbito (sim ou não)

- ✓ Variáveis explicativas podem ser qualitativas ou quantitativas
- ✓ O objetivo é estudar os fatores associados à presença do evento de interesse

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

Exemplo

ID	Idade	DC	ID	Idade	DC	ID	Idade	DC	ID	Idade	DC	ID	Idade	DC
1	20	0	21	34	0	41	41	0	61	48	1	81	57	0
2	23	0	22	34	0	42	42	0	62	48	1	82	57	1
3	24	0	23	34	1	43	42	0	63	49	0	83	57	1
4	25	0	24	34	0	44	42	0	64	49	0	84	57	1
5	25	1	25	34	0	45	42	1	65	49	1	85	57	1
6	26	0	26	35	0	46	43	0	66	50	0	86	58	0
7	26	0	27	35	0	47	43	0	67	50	1	87	58	1
8	28	0	28	36	0	48	43	1	68	51	0	88	58	1
9	28	0	29	36	1	49	44	0	69	52	0	89	59	1
10	29	0	30	36	0	50	44	0	70	52	1	90	59	1
11	30	0	31	37	0	51	44	1	71	53	1	91	60	0
12	30	0	32	37	1	52	44	1	72	53	1	92	60	1
13	30	0	33	37	0	53	45	0	73	54	1	93	61	1
14	30	0	34	38	0	54	45	1	74	55	0	94	62	1
15	30	0	35	38	0	55	46	0	75	55	1	95	62	1
16	30	1	36	39	0	56	46	1	76	55	1	96	63	1
17	32	0	37	39	1	57	47	0	77	56	1	97	64	0
18	32	0	38	40	0	58	47	0	78	56	1	98	64	1
19	33	0	39	40	1	59	47	1	79	56	1	99	65	1
20	33	0	40	41	0	60	48	0	80	57	0	100	69	1

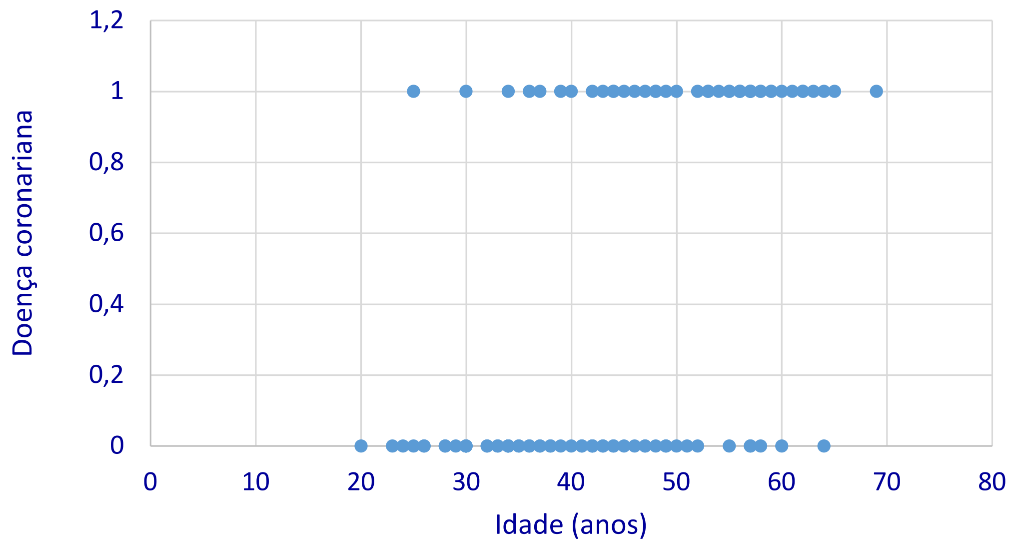
Fonte: Hosmer e Lemeshow, 2013

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

Exemplo

Y = doença coronariana



Fonte: Hosmer e Lemeshow, 2013

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

BIOESTATÍSTICA

Exemplo

Y = doença coronariana

Idade (anos)	DC Sim	DC Não	Total
20 - 29	1	9	10
30 - 34	2	13	15
35 - 39	3	9	12
40 - 44	5	10	15
45 - 49	6	7	13
50 - 54	5	3	8
55 - 59	13	4	17
60 - 69	8	2	10
Total	43	57	100

Fonte: Hosmer e Lemeshow, 2013

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

Exemplo

Y = doença coronariana

Idade (anos)	DC Sim	DC Não	Total	p_{Sim}
20 - 29	1	9	10	0.10
30 - 34	2	13	15	
35 - 39	3	9	12	
40 - 44	5	10	15	
45 - 49	6	7	13	
50 - 54	5	3	8	
55 - 59	13	4	17	
60 - 69	8	2	10	
Total	43	57	100	

Fonte: Hosmer e Lemeshow, 2013

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística

Se a variável resposta é dicotômica, não é possível adotar o modelo de regressão linear simples para este caso, porque suas suposições não estariam satisfeitas.

Lembrando ... o modelo de Regressão Linear Simples é dado por

$$E(Y_i) = \beta_0 + \beta_1 X_i \qquad \text{onde} \quad Y_i \sim Normal$$

Se a variável resposta é dicotômica e assume apenas dois valores, qual é a sua distribuição?

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Lembrando ...

Se Y é uma variável aleatória dicotômica, sua distribuição é Bernoulli.

$$Y = \begin{cases} 1 & (sucesso) & p \\ 0 & (fracasso) & 1-p \end{cases}$$

E sua função de probabilidades é dada por:

y	$P(Y=y)$
0	$1\text{-}p$
1	p

Ou, de forma resumida:

$$\mathrm{P}(Y=y) = p^{y}(1-p)^{1-y}\ , \quad Y = 0{,}1$$

Exemplos:

- ✓ cura (sim ou não),
- ✓ pressão elevada (sim ou não),
- ✓ óbito (sim ou não)
- ✓ ter uma determinada característica (sim ou não)

A esperança e a variância de Y são dadas por:

$$E(Y) = p$$

$$VAR(Y) = p(1-p)$$

Notação: $Y \sim Bernoulli(p)$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística

BIOESTATÍSTICA

Vários modelos foram desenvolvidos em que $Y \sim Bernoulli$:

- ✓Logístico
- ✓Probito
- ✓Complementar log-log
- ✓e outros

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

(uma única variável explicativa)

O modelo de regressão logística simples pode ser escrito como:

$$E(Y_i) = p_i = \frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}$$

onde

- ✓ Y_i é o valor da variável resposta para o *i-ésimo* indivíduo, Y_i = 0, 1
- ✓ $p_i = P(Y_i)$ é a probabilidade de ocorrência do evento de interesse para o i-ésimo indivíduo e é também a esperança de Y_i
- ✓ X_i é o valor da variável explicativa para o *i-ésimo* indivíduo
- ✓ β_0 e β_1 são os parâmetros a serem estimados

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

(uma única variável explicativa)

O modelo de regressão logística simples pode ser escrito como:

$$E(Y_i) = p_i = \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}}$$

Também podemos utilizar a forma

$$Y_i = p_i + \varepsilon_i$$ onde ε_i é um erro aleatório

Mas, em geral, vamos trabalhar com a anterior

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

BIOESTATÍSTICA

O modelo de regressão logística simples pode ser escrito como:

$$E(Y_i) = p_i = \frac{e^{\beta_0+\beta_1 X_i}}{1 + e^{\beta_0+\beta_1 X_i}}$$

Y_i pode assumir dois valores, 0 e 1. Então podemos obter

$$Prob(Y_i = 1) = p_i = \frac{e^{\beta_0+\beta_1 X_i}}{1 + e^{\beta_0+\beta_1 X_i}}$$

$$Prob(Y_i = 0) = 1 - p_i = 1 - \frac{e^{\beta_0+\beta_1 X_i}}{1 + e^{\beta_0+\beta_1 X_i}} = \frac{1 + e^{\beta_0+\beta_1 X_i} - e^{\beta_0+\beta_1 X_i}}{1 + e^{\beta_0+\beta_1 X_i}} = \frac{1}{1 + e^{\beta_0+\beta_1 X_i}}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

Por exemplo, no caso do banco de dados LOW, obtivemos

$$\beta_0 + \beta_1 X_i = -1.0871 + 0.7041 X_i$$

Sabendo que $P(Y_i = 1) = \dfrac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}}$

Obtenha:

$$P(Y_i = 1 | X_i = 1) = \frac{e^{-1.0871+0.7041}}{1 + e^{-1.0871+\ .7041}} = 0{,}4054$$

$$P(Y_i = 1 | X_i = 0) = \frac{e^{-1.0871}}{1 + e^{-1.0871}} = 0{,}2521$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

BIOESTATÍSTICA

$$Prob(Y_i = 1) = p_i = \frac{e^{\beta_0+\beta_1 X_i}}{1 + e^{\beta_0+\beta_1 X_i}} \quad \text{e} \quad Prob(Y_i = 0) = 1 - p_i = \frac{1}{1 + e^{\beta_0+\beta_1 X_i}}$$

$$\frac{p_i}{1-p_i} = \frac{\dfrac{e^{\beta_0+\beta_1 X_i}}{1 + e^{\beta_0+\beta_1 X_i}}}{\dfrac{1}{1 + e^{\beta_0+\beta_1 X_i}}} = \frac{e^{\beta_0+\beta_1 X_i}}{1 + e^{\beta_0+\beta_1 X_i}} \cdot \frac{1 + e^{\beta_0+\beta_1 X_i}}{1} = e^{\beta_0+\beta_1 X_i}$$

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

Então ....

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

BIOESTATÍSTICA

$$ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_i$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

O modelo de regressão logística simples

$$E(Y_i) = p_i = \frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}$$

Também pode ser escrito como

$$ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_i$$

A expressão do lado esquerdo é denominada logito ou log-odds.

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

BIOESTATÍSTICA

$$\frac{p}{1-p} = e^{\beta_0+\beta_1 X}$$

Interpretação dos parâmetros do modelo

1. Se a variável explicativa é categórica com 2 categorias (*dummie):*

- ✓ Para X = 0 (ausência do atributo), a chance de ter o desfecho é

$$\left(\frac{p}{1-p}\right)_{X=0} = e^{\beta_0}$$

- ✓ Para X = 1, a chance de ter o desfecho é

$$\left(\frac{p}{1-p}\right)_{X=1} = e^{\beta_0+\beta_1} = e^{\beta_0}\, e^{\beta_1} = \left(\frac{p}{1-p}\right)_{X=0} e^{\beta_1}$$

A chance de desenvolver o desfecho entre os indivíduos que têm o atributo é e^{β_1} vezes a chance indivíduos que não têm o atributo .

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

BIOESTATÍSTICA

Interpretação dos parâmetros do modelo

$$\frac{p}{1-p} = e^{\beta_0+\beta_1 \mathrm{X}}$$

1. Se a variável explicativa é categórica com 2 categorias (*dummie):*

✓ A odds ratio será

$$OR = \frac{\left(\frac{p}{1-p}\right)_{X=1}}{\left(\frac{p}{1-p}\right)_{X=0}} = \frac{e^{\beta_0}e^{\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

A exponencial do coeficiente β_1 é a OR

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

BIOESTATÍSTICA

Interpretação dos parâmetros do modelo

$$\frac{p}{1-p} = e^{\beta_0+\beta_1 X}$$

1. Se a variável explicativa é contínua:

- ✓ Para X = a, a chance de ter o desfecho é

$$\left(\frac{p}{1-p}\right)_{X=a} = e^{\beta_0+\beta_1 a}$$

- ✓ Para X = a+1, a chance de ter o desfecho é

$$\left(\frac{p}{1-p}\right)_{X=a+1} = e^{\beta_0+\beta_1(a+1)} = e^{\beta_0+\beta_1 a+\beta_1} = e^{\beta_0+\beta_1 a}e^{\beta_1}$$

Quando aumentamos X de uma unidade, a chance de ter o desfecho fica multiplicada por e^{β_1}

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

BIOESTATÍSTICA

Interpretação dos parâmetros do modelo

$$\frac{p}{1-p} = e^{\beta_0+\beta_1 \mathrm{X}}$$

1. Se a variável explicativa é contínua:

- ✓ Para X = a, a chance de ter o desfecho é

$$\left(\frac{p}{1-p}\right)_{X=a} = e^{\beta_0+\beta_1 a}$$

- ✓ Para X = a+1, a chance de ter o desfecho é

$$\left(\frac{p}{1-p}\right)_{X=a+2} = e^{\beta_0+\beta_1(a+2)} = e^{\beta_0+\beta_1 a+2\beta_1} = e^{\beta_0+\beta_1 a}e^{2\beta_1} = e^{\beta_0+\beta_1 a}e^{\beta_1}e^{\beta_1}$$

Quando aumentamos X de uma unidade, a chance de ter o desfecho fica multiplicada por e^{β_1}

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Regressão Logística Simples

BIOESTATÍSTICA

Interpretação dos parâmetros do modelo

$$\frac{p}{1-p} = e^{\beta_0+\beta_1 X}$$

1. Se a variável explicativa é contínua:

✓ A odds ratio será

$$OR = \frac{\left(\frac{p}{1-p}\right)_{X=a+1}}{\left(\frac{p}{1-p}\right)_{X=a}} = \frac{e^{\beta_0+\beta_1 a}e^{\beta_1}}{e^{\beta_0+\beta_1 a}} = e^{\beta_1}$$

A exponencial do coeficiente β_1 é a OR

Suposições do modelo de regressão logística

BIOESTATÍSTICA

1. Y é uma variável dicotômica (0,1).

   (a extensão para variáveis categóricas com mais de duas categorias não será vista neste curso)

2. Os valores de Y são independentes

3. A covariância entre dois erros é igual a zero

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Estimação dos parâmetros β_0 e β_1

BIOESTATÍSTICA

- ✓ Na regressão logística, os parâmetros são estimados pelo Método de Máxima Verossimilhança.
- ✓ De uma maneira genérica, pode-se dizer que o método da máxima verossimilhança fornece os valores para os parâmetros que maximizam a probabilidade de se obter o conjunto de dados existente.
- ✓ Para se aplicar este método, em primeiro lugar precisa-se definir a função de verossimilhança.

Método de Máxima Verossimilhança

BIOESTATÍSTICA

Na situação em que a variável resposta é dicotômica e tem distribuição de *Bernoulli*, tem-se:

$$Y_i = \begin{cases} 1 & P(Y_i = 1/X_i) = p_i = \dfrac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}} \\ 0 & P(Y_i = 0/X_i) = 1 - p_i = \dfrac{1}{1+e^{\beta_0+\beta_1 X_i}} \end{cases}$$

A função de probabilidades de Y é:

$$f(Y_i) = P(Y_i = y_i) = p^{y_i}(1-p)^{1-y_i} \qquad \text{onde } Y_i = 0{,}1, \quad i = 1{,}2{,}3, \ldots, n$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Método de Máxima Verossimilhança

BIOESTATÍSTICA

Assim, para aqueles pares $(x_i, 1)$, a contribuição para a função de verossimilhanca é p_i

e naqueles pares $(x_i, 0)$, a contribuição para a função de verossimilhanca é $1 - p_i$

A função de verssimilhança é definida pelo produto dos termos dados acima, isto é:

$$L(\beta) = \prod_{i=1}^{n} f(Y_i) \quad = \prod_{i=1}^{n} {p_i}^{y_i}(1 - p_i)^{1-y_i}$$ onde $Y_i = 0{,}1, \quad i = 1{,}2{,}3, \dots, n$

No entanto, é mais fácil maximizar a função $\ln[L(\beta)]$

$$\ln[L(\beta)] = \sum_{i=1}^{n} [y_i \ln(p_i) + (1 - y_i)\ln(1 - p_i)]$$ onde $Y_i = 0{,}1, \quad i = 1{,}2{,}3, \dots, n$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Método de Máxima Verossimilhança

Por exemplo, para o banco de dados LOW, vamos supor que a amostra seja:

Indivíduo	LOW (Y)	SMOKE (X)
1	0	0
2	1	1
3	1	0
4	0	0
...	...	...

Contribuição de cada indivíduo para a função de verossimilhança

$$L(\beta) = \prod_{i=1}^{n} {p_i}^{y_i}(1-p_i)^{1-y_i} \quad \text{onde } Y_i = 0{,}1, \quad i = 1{,}2{,}3, \dots, n$$

$$p_i = \frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}} \qquad \text{e} \qquad 1-p_i = \frac{1}{1+e^{\beta_0+\beta_1 X_i}}$$

$$L(\beta) = \underbrace{\frac{1}{1+e^{\beta_0+\beta_1 X_i}}}_{\text{indivíduo 1}} \cdot \underbrace{\frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}}_{\text{indivíduo 2}} \cdot \underbrace{\frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}}_{\text{indivíduo 3}} \cdot \underbrace{\frac{1}{1+e^{\beta_0+\beta_1 X_i}}}_{\text{indivíduo 4}} \cdot \dots$$

$$= \frac{1}{1+e^{\beta_0}} \cdot \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} \cdot \frac{e^{\beta_0}}{1+e^{\beta_0}} \cdot \frac{1}{1+e^{\beta_0}} \cdot \dots$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Método de Máxima Verossimilhança

BIOESTATÍSTICA

$$L(\beta) = \overbrace{\frac{1}{1+e^{\beta_0+\beta_1 X_i}}}^{\text{indivíduo 1}} \cdot \overbrace{\frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}}^{\text{indivíduo 2}} \cdot \overbrace{\frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}}}^{\text{indivíduo 3}} \cdot \overbrace{\frac{1}{1+e^{\beta_0+\beta_1 X_i}}}^{\text{indivíduo 4}} \cdot \ldots$$

$$= \frac{1}{1+e^{\beta_0}} \cdot \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} \cdot \frac{e^{\beta_0}}{1+e^{\beta_0}} \cdot \frac{1}{1+e^{\beta_0}} \cdot \ldots$$

$$\ln[L(\beta)] = ln\left(\frac{1}{1+e^{\beta_0}}\right) + ln\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right) + ln\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right) + ln\left(\frac{1}{1+e^{\beta_0}}\right) + \ldots$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Método de Máxima Verossimilhança

BIOESTATÍSTICA

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

Para encontrar os valores dos β_i que maximizam a função acima, deve-se derivar $\ln[L(\beta)]$em relação a cada um dos β_i e igualar a zero. Como estas equações não são lineares, são necessários métodos interativos e sua solução não é fácil! Porém os *softwares* fazem isso por nós !!!!

As equações são

$$\sum_{i=1}^{n} [y_i - p_i] = 0 \qquad \text{e} \qquad \sum_{i=1}^{n} x_i [y_i - p_i] = 0$$

e são chamadas equações de verossimilhança.

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Método de Máxima Verossimilhança

Estimativas dos parâmetros β_0 e β_1

Normalmente as saídas de computador fornecem não só os valores dos $\hat{\beta}_i$ ou , mas, também, seus respectivos erros padrão $SE(\hat{\beta}_i)$.

Tais valores serão utilizados para os testes de significância dos coeficientes e para o cálculos dos respectivos intervalos de confiança.

Método de Máxima Verossimilhança

Para o modelo sem nenhuma variável, só com β_0 , o logaritmo da função de verossimilhança pode ser calculado por:

$$\ln[L(\beta)] = n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)$$

onde

n_1 é o número de casos em que $Y=1$

n_0 é o número de casos em que $Y=0$

$n = n_0 + n_1$ é o número total de casos

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Testes de hipóteses

BIOESTATÍSTICA

Na regressão linear, utilizamos o resíduo do modelo $(Y_i - \hat{Y}_i)$ para fazer testes de hipóteses (Teste F da ANOVA) e para comparar modelos (Teste F parcial para comparar o modelo completo x modelo reduzido)

Na regressão logística, quem faz o papel do resíduo é a função desvio ou *deviance*, definida como $-2 * \ln[L(\beta)]$.

Testes de hipóteses

1. Teste da razão de verossimilhanças

- ✓ Compara a função de verossimilhança do modelo **ajustado** com o modelo **saturado**.
- ✓ O modelo **saturado** é aquele que contém tantos parâmetros quanto o número de observações da amostra, isto é, contém n parâmetros.
- ✓ O modelo **ajustado** contém menos parâmetros. Por exemplo, o modelo simples contém apenas 2 parâmetros, β_0 e β_1.
- ✓ Se as verossimilhanças dos dois modelos forem parecidas, significa que um modelo com menos parâmetros é tão bom para explicar a resposta quanto um modelo com n parâmetros.

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Testes de hipóteses

1. Teste da razão de verossimilhanças

Esta comparação é feita por meio da quantidade:

$$D = deviance(\text{modelo ajustado}) - deviance(\text{modelo saturado})$$

$$D = -2\ln[L(modelo\ ajustado)] - 2\ln[L(modelo\ saturado)]$$

$$D = -2ln\underbrace{\left[\frac{L(modelo\ ajustado)}{L(modelo\ saturado)}\right]}_{\text{razão de verossimilhanças}}$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Testes de hipóteses

1. Teste da razão de verossimilhanças

Para verificar a significância de uma variável independente, compara-se o valor de D dos modelos com e sem a variável. A mudança de D devido à inclusão da variável independente é:

$$G = D(modelo\ sem\ a\ variável) - D(modelo\ com\ a\ variável)$$

$$G = -2ln\left[\frac{L(modelo\ sem\ a\ variável)}{L(modelo\ saturado)}\right] - 2ln\left[\frac{L(modelo\ com\ a\ variável)}{L(modelo\ saturado)}\right]$$

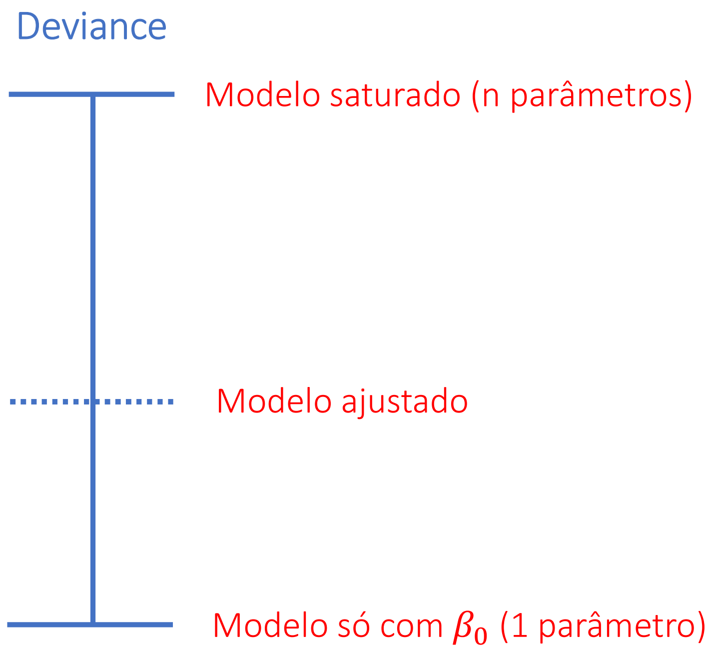
$$G = -2ln\left[\frac{L(modelo\ sem\ a\ variável)}{L(modelo\ com\ a\ variável)}\right]$$

$G \sim \chi^2_1$ para o teste de significância de 1 variável com duas categorias

No caso do modelo simples, $H_0: \beta_1 = 0$

Testes de hipóteses

1. Teste da razão de verssimilhanças



GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Testes de hipóteses

BIOESTATÍSTICA

2. Teste de Wald (baixo poder)

$$H_0\colon \beta_1 = 0 \iff H_0\colon OR = 1$$

$$W = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \qquad onde\ W_c \sim Normal(0,1)$$

3. Intervalo de confiança

$$IB(\beta_1, 1-\alpha) = \hat{\beta}_1 \pm z_{1-\alpha}\, SE(\hat{\beta}_1)$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Testes de hipóteses

4. Caso múltiplo

$$H_0\colon\ \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$H_a\colon existe\ pelo\ menos\ um\ \beta \neq 0$$

$G \sim \chi^2_k$ onde k é o número de parâmetros (βs) no modelo

BIOESTATÍSTICA

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Testes de hipóteses

BIOESTATÍSTICA

4. Caso múltiplo

Para testar a significância de cada coeficiente, utilizar o teste de Wald

$$H_0:\ \beta_i = 0 \qquad \Leftrightarrow \qquad H_0:\ OR(X_i) = 1 \qquad \Leftrightarrow \qquad H_0:\ RR(X_i) = 1$$

$$H_a:\ \beta_i \neq 0 \qquad \qquad H_a: OR(X_i) \neq 1 \qquad \qquad H_a: RR(X_i) \neq 1$$

$$W_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad onde\ W_i \sim Normal(0,1)$$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Risco Relativo em Regressão Logística

Sim!!!

$$P(Y=1) = p = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$$

$$RR = \frac{P(Y=1|X=1)}{P(Y=1|X=0)} = \frac{\dfrac{e^{\beta_0+\beta_1\cdot 1}}{1+e^{\beta_0+\beta_1\cdot 1}}}{\dfrac{e^{\beta_0+\beta_1\cdot 0}}{1+e^{\beta_0+\beta_1\cdot 0}}} = \frac{\dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}}{\dfrac{e^{\beta_0}}{1+e^{\beta_0}}} = \frac{e^{\beta_0}e^{\beta_1}}{1+e^{\beta_0+\beta_1}} \cdot \frac{1+e^{\beta_0}}{e^{\beta_0}} = \frac{e^{\beta_1}+e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$$

Se $\beta_1 = 0 \Rightarrow RR = \frac{1+e^{\beta_0}}{1+e^{\beta_0}} = 1$

Então: $H_0\colon \beta_1 = 0 \;\Leftrightarrow\; H_0\colon OR = 1 \;\Leftrightarrow\; H_0\colon RR = 1 \Leftrightarrow\; H_0\colon RP = 1$

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Risco Relativo em Regressão Logística

Mais de uma variável dependente

$$P(Y=1) = p = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2}}$$

$$RR = \frac{P(Y=1|X_1=1)}{P(Y=1|X_1=0)} = \frac{\dfrac{e^{\beta_0+\beta_1+\beta_2 X_2}}{1+e^{\beta_0+\beta_1+\beta_2 X_2}}}{\dfrac{e^{\beta_0+\beta_2 X_2}}{1+e^{\beta_0+\beta_2 X_2}}} = \frac{e^{\beta_0+\beta_1+\beta_2 X_2}}{1+e^{\beta_0+\beta_1+\beta_2 X_2}} \cdot \frac{1+e^{\beta_0+\beta_2 X_2}}{e^{\beta_0+\beta_2 X_2}}$$

$$= \frac{e^{\beta_1} e^{\beta_0+\beta_2 X_2}}{1+e^{\beta_0+\beta_2 X_2+\beta_1}} \cdot \frac{1+e^{\beta_0+\beta_2 X_2}}{e^{\beta_0+\beta_2 X_2}} = \frac{e^{\beta_1}}{1+e^{\beta_0+\beta_2 X_2+\beta_1}} \cdot \frac{1+e^{\beta_0+\beta_2 X_2}}{1}$$

$$= e^{\beta_1} \cdot \frac{1+e^{\beta_0+\beta_2 X_2}}{1+e^{\beta_0+\beta_1+\beta_2 X_2}} = \frac{e^{\beta_1} + e^{\beta_0+\beta_1+\beta_2 X_2}}{1+e^{\beta_0+\beta_1+\beta_2 X_2}}$$

Não dá pra cancelar X_2 !!!

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP

Risco Relativo em Regressão Logística

BIOESTATÍSTICA

Mais de uma variável dependente

Sejam

$p_0 = P(Y = 1|X = 0)$, isto é, a incidência do desfecho de interesse no grupo não exposto

$p_1 = P(Y = 1|X = 1)$, isto é, a incidência do desfecho de interesse no grupo exposto

A partir do modelo múltiplo, obtemos as estimativas das *odds ratio* para cada variável, ajustada para as demais (OR_{aj}). O risco relativo para cada variável, ajustado pelas demais (RR_{aj}), pode ser obtido a partir da OR_{aj} como:

$$RR_{aj} = \frac{OR_{aj}}{(1 - p_0) + (p_0 * OR_{aj})}$$

Jun Zhang, MB, PhD; Kai F. Yu, PhD. What's the Relative Risk? A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes. JAMA, November 18, 1998—Vol 280, No. 19.

GLEICE M S CONCEIÇÃO
MARIA DO ROSÁRIO D D LATORRE
FSP - USP