

# Exemplo Preferência de Automóveis

Gilberto A. Paula

Departamento de Estatística  
IME-USP, Brasil  
giapaula@ime.usp.br

2<sup>o</sup> Semestre 2023

- 1 Preferência Automóveis
- 2 Análise de Dados Preliminar
- 3 Modelo Binomial
- 4 Resultados Modelo Ajustado
- 5 Curva ROC
- 6 Conclusões
- 7 Referências

## Descrição dos Dados

Como ilustração de exemplo de modelo binomial para dados não agrupados, vamos considerar os dados sobre

- **preferência de automóveis** (1: americano, 0: japonês)

de uma amostra aleatória de 263 consumidores (Foster, Stine e Waterman, 1998, pp. 338-339).

## Descrição dos Dados

A probabilidade de preferência por carro americano será relacionada com as seguintes variáveis explicativas do comprador(a):

- **idade** (em anos);
- **gênero** (0: masculino; 1: feminino);
- **estado civil** (0:casado(a), 1:solteiro(a)).

- 1 Preferência Automóveis
- 2 Análise de Dados Preliminar**
- 3 Modelo Binomial
- 4 Resultados Modelo Ajustado
- 5 Curva ROC
- 6 Conclusões
- 7 Referências

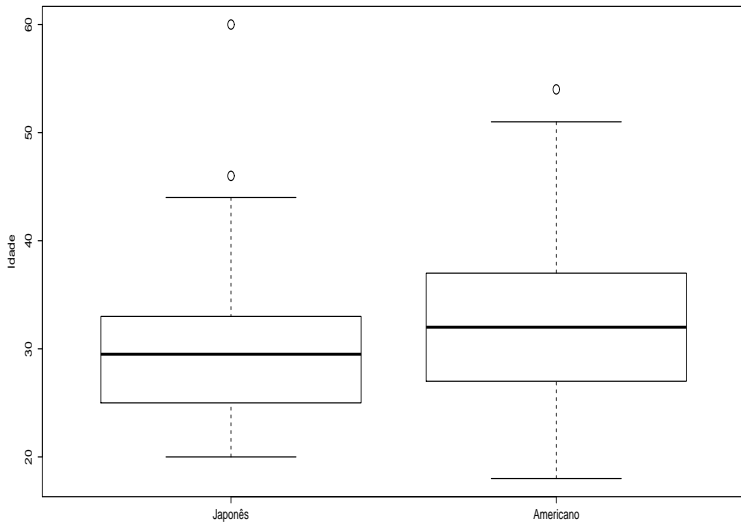
### Preferência segundo o Gênero

	Masculino	Feminino
Americano	61 (42,4%)	54 (45,4%)
Japonês	83 (57,6%)	65 (54,6%)
Total	144	119

### Preferência segundo o Estado Civil

	Casado	Solteiro
Americano	83 (48,8%)	32 (34,4%)
Japonês	87 (51,2%)	61 (65,6%)
Total	170	93

# Boxplots da Idade segundo a Preferência



- 1 Preferência Automóveis
- 2 Análise de Dados Preliminar
- 3 Modelo Binomial**
- 4 Resultados Modelo Ajustado
- 5 Curva ROC
- 6 Conclusões
- 7 Referências



## Descrição

Denotaremos por  $y_i$  a preferência do  $i$ -ésimo comprador

- 0: automóvel japonês;
- 1: automóvel americano.

Vamos propor o seguinte modelo binomial para explicar a probabilidade de preferência por automóvel americano:

- $y_i | x_i \stackrel{\text{ind}}{\sim} \text{Be}(\pi_i)$ ,
- $\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_1 + \beta_2 \times \text{Idade}_i + \beta_3 \times \text{Gênero}_i + \beta_4 \times \text{ECivil}_i$ ,

em que  $\pi_i$  denota a probabilidade do  $i$ -ésimo comprador preferir automóvel americano, para  $i = 1, \dots, 263$ ,

- 1 Preferência Automóveis
- 2 Análise de Dados Preliminar
- 3 Modelo Binomial
- 4 Resultados Modelo Ajustado**
- 5 Curva ROC
- 6 Conclusões
- 7 Referências

## Estimativas

Efeito	Estimativa	E.Padrão	E/E.Padrão	Valor-P
Constante	-1,653	0,708	-2,33	0,020
Idade	0,050	0,022	2,27	0,021
GêneroF	-0,094	0,256	-0,37	0,712
ECivilS	-0,518	0,272	-1,90	0,057

A variável gênero é não significativa.

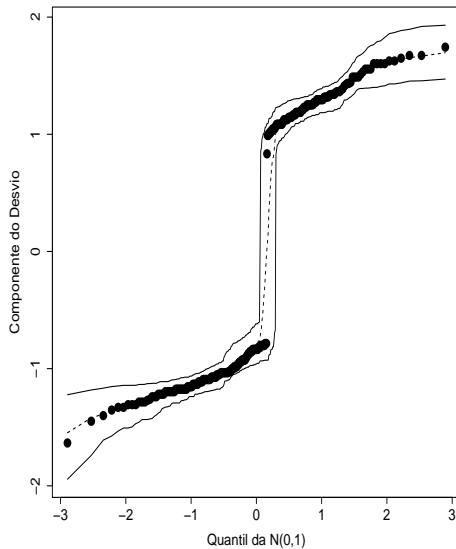
## Estimativas

Efeito	Estimativa	E.Padrão	E/E.Padrão	Valor-P
Constante	-1,600	0,692	-2,31	0,021
Idade	0,050	0,021	2,38	0,021
ECivilS	-0,526	0,272	-1,93	0,053

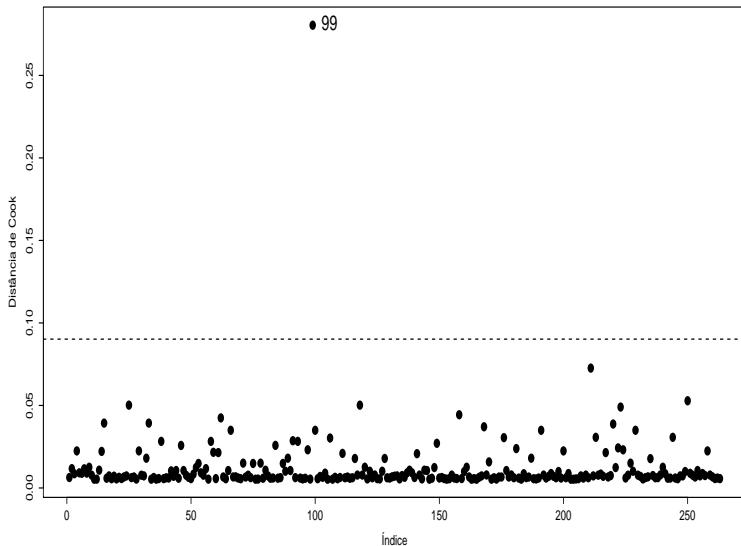
## Interação

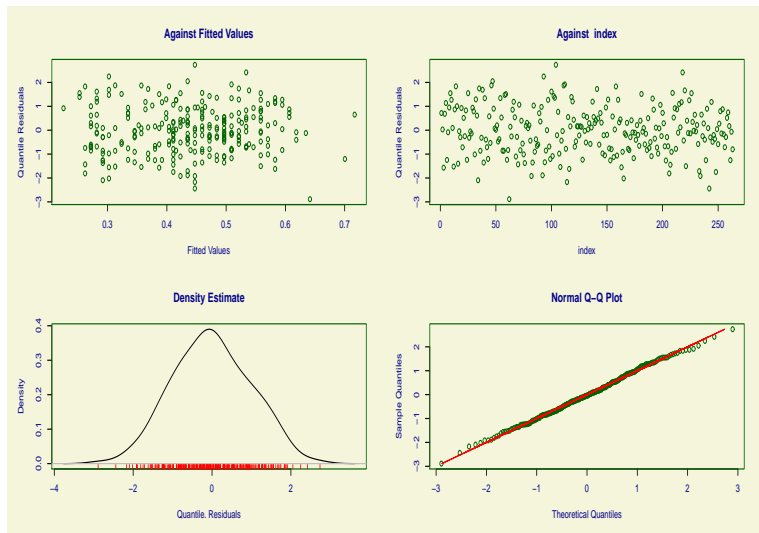
O teste de ausência de interação entre idade e estado civil forneceu o seguinte valor para a estatística da razão de verossimilhanças  $\xi_{RV} = 0,81$  com nível descritivo  $P=0,368$ . Portanto não rejeitamos a ausência de interação entre idade e estado civil.

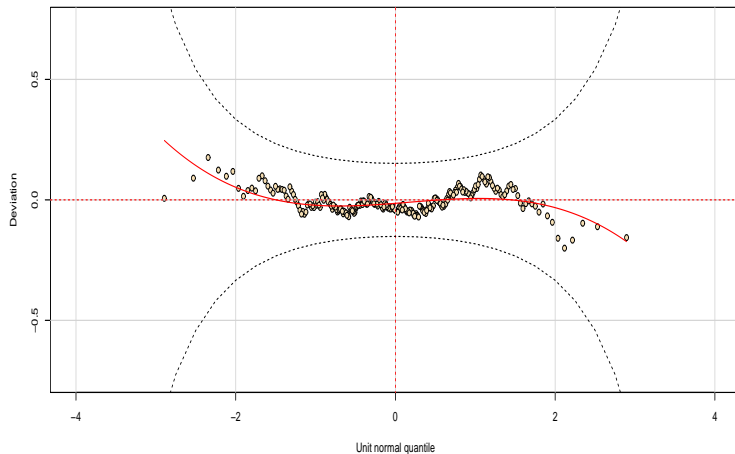
# Resíduos Modelo Final



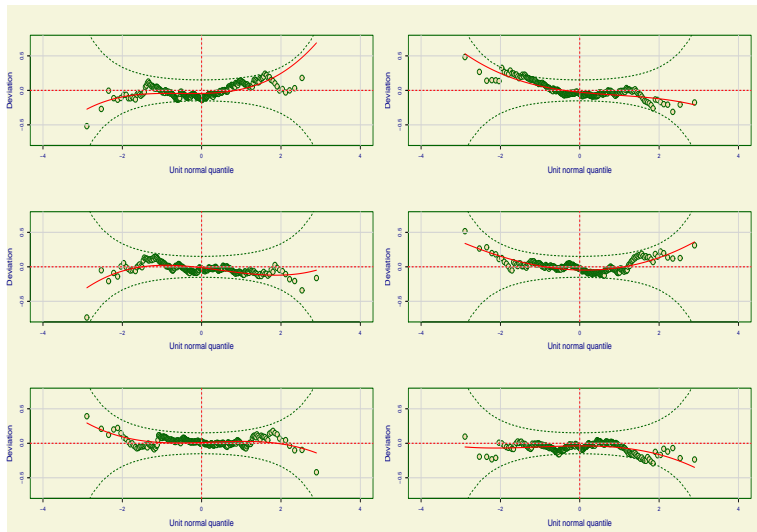
# Distância de Cook Modelo Final











## Sem observação #99

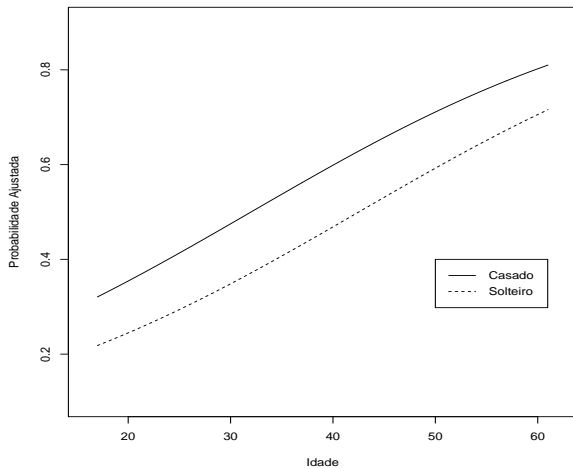
Efeito	Estimativa	z-valor	Variação
Constante	-1,942	-2,65	-21,4%
Idade	0,060	2,65	22,4%
ECivils	-0,474	-1,72	9,9%

Este comprador tem **60 anos, comprou carro japonês e é solteiro**. Sua eliminação aumenta a significância da idade e diminuiu a significância do estado civil.

### Preferência por Automóvel Americano

A probabilidade ajustada de preferência por automóvel americano fica expressa na forma

$$\hat{\pi} = \frac{\exp(-1,600 + 0,050 \times \text{Idade} - 0,526 \times \text{ECivil})}{1 + \exp(-1,600 + 0,050 \times \text{Idade} - 0,526 \times \text{ECivil})}$$



## Efeito Idade

A probabilidade de preferência por automóvel americano aumenta com a idade do comprador.

## Razão de Chances

A razão de chances entre um comprador casado e um comprador solteiro preferir automóvel americano é estimada por

$$\begin{aligned}\hat{\psi} &= \exp(0,526) \\ &= 1,69.\end{aligned}$$

Ou seja, um comprador casado tem uma chance 69% maior de preferir automóvel americano em relação a um comprador solteiro.

## Estimativa Intervalar

Estimativa intervalar de 90% para a razão de chances entre um comprador casado e um comprador solteiro preferir automóvel americano

$$\begin{aligned} e^{0,526 \pm 1,65 \times 0,272} &= e^{0,526 \pm 0,449} \\ &= [1,080; 2,651] [8,0\%; 165,1\%]. \end{aligned}$$

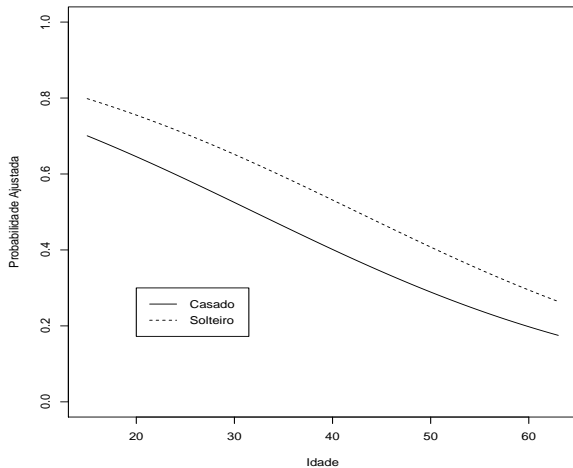
Portanto, um comprador casado tem uma chance entre 8% e 165,1% maior de preferir automóvel americano em relação a um comprador solteiro.

## Preferência por Automóvel Japonês

A probabilidade ajustada de preferência por automóvel japonês fica expressa na forma

$$1 - \hat{\pi} = \frac{\exp(1,600 - 0,050 \times \text{Idade} + 0,526 \times \text{ECivil})}{1 + \exp(1,600 - 0,050 \times \text{Idade} + 0,526 \times \text{ECivil})}$$

# Probabilidade Ajustada





## Efeito Idade

A probabilidade de preferência por automóvel japonês diminui com a idade do comprador.

## Razão de Chances

A razão de chances entre um comprador solteiro e um comprador casado preferir automóvel japonês é estimada por

$$\begin{aligned}\hat{\psi} &= \exp(0,526) \\ &= 1,69.\end{aligned}$$

Ou seja, um comprador solteiro tem uma chance 69% maior de preferir automóvel japonês em relação a um comprador casado.

## Estimativa Intervalar

Estimativa intervalar de 90% para a razão de chances entre um comprador solteiro e um comprador casado preferir automóvel japonês

$$\begin{aligned} e^{0,526 \pm 1,65 \times 0,272} &= e^{0,526 \pm 0,449} \\ &= [1,080; 2,651] [8,0\%; 165,1\%]. \end{aligned}$$

Portanto, um comprador solteiro tem uma chance entre 8% e 165,1% maior de preferir automóvel japonês em relação a um comprador casado.

- 1 Preferência Automóveis
- 2 Análise de Dados Preliminar
- 3 Modelo Binomial
- 4 Resultados Modelo Ajustado
- 5 Curva ROC**
- 6 Conclusões
- 7 Referências

## Métricas de Desempenho da Predição

Seja  $\hat{\pi}$  a probabilidade ajustada de sucesso de um modelo binário e queremos classificar um novo indivíduo como sucesso ou fracasso. Portanto, é razoável classificar esse indivíduo como sendo sucesso à medida que  $\hat{\pi}$  se aproxima de 1 e como fracasso à medida que  $\hat{\pi}$  se aproxima de 0. Assim, definindo um ponto de corte podemos construir para os dados da amostra a tabela abaixo.

Classificação pelo Modelo	Classificação Correta		Total
	Sucesso	Fracasso	
Sucesso	VP	FP	VP+FP
Fracasso	FN	VN	FN+VN
Total	VP+FN	FP+VN	n

## Métricas de Desempenho da Predição

Principais métricas:

- Acurácia: proporção de predições corretas

$$ACC = \frac{VP+VN}{n}$$

- Sensibilidade: proporção de verdadeiros positivos

$$SENS = \frac{VP}{VP+FN}$$

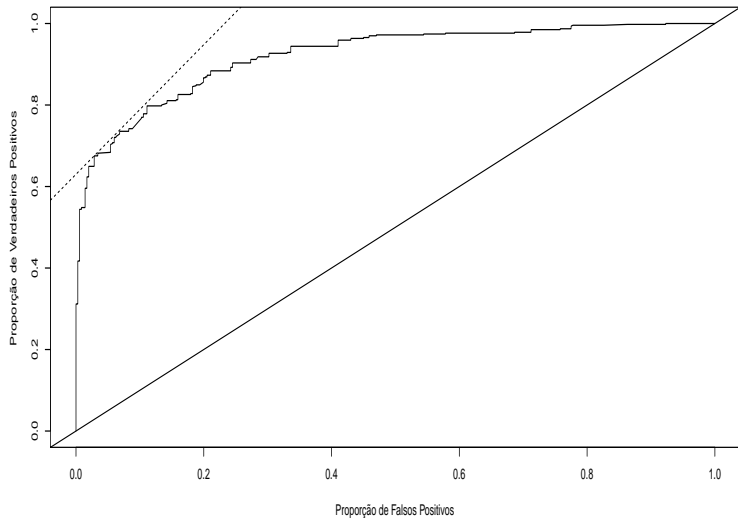
1 - SENS: proporção de falsos negativos

- Especificidade: proporção de verdadeiros negativos

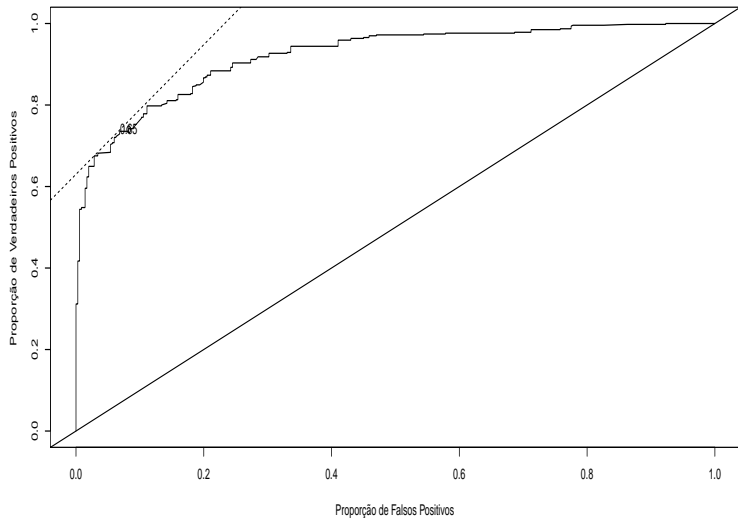
$$ESPEC = \frac{VN}{FP+VN}$$

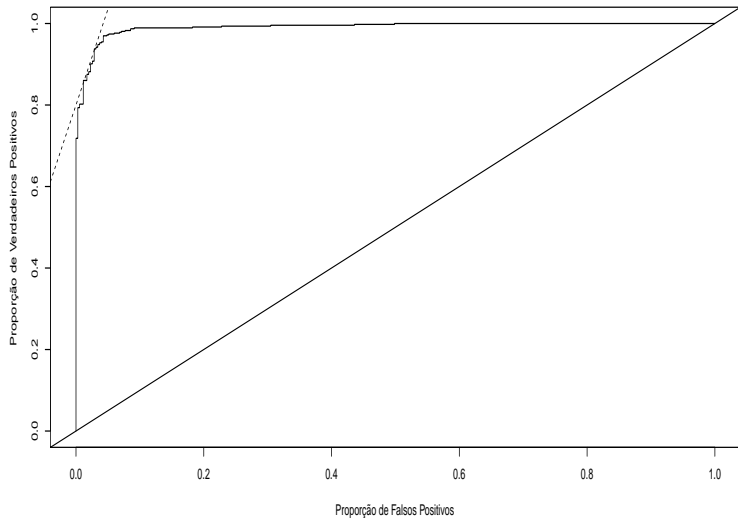
1 - ESPEC: proporção de falsos positivos

# Curva ROC

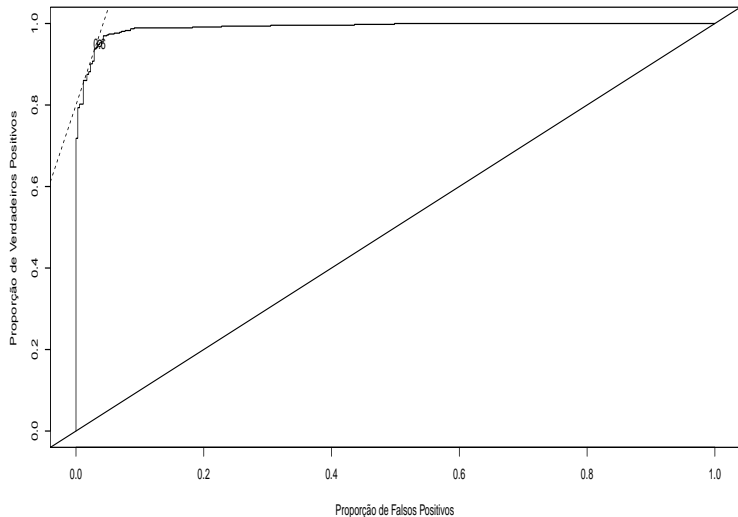


# Curva ROC









## Preferência Automóveis

Seja  $\hat{\pi}$  a probabilidade ajustada de preferência por automóvel americano. Vamos classificar como comprador de automóvel americano se  $\hat{\pi} \geq 0,44$  e como comprador de automóvel japonês se  $\hat{\pi} < 0,44$  ( $0,44 \cong 115/263$  proporção amostral de compradores de automóvel americano).

Classificação pelo Modelo	Classificação Correta		Total
	Americano	Japonês	
Americano	68	62	130
Japonês	47	86	133
Total	115	148	263

## Preferência Automóveis

- Acurácia: proporção de predições corretas

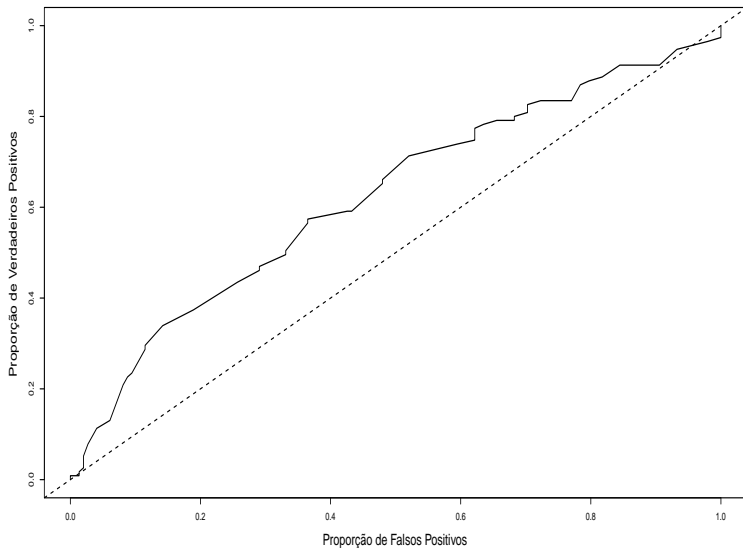
$$ACC = \frac{68+86}{263} \cong 0,586(58,6\%)$$

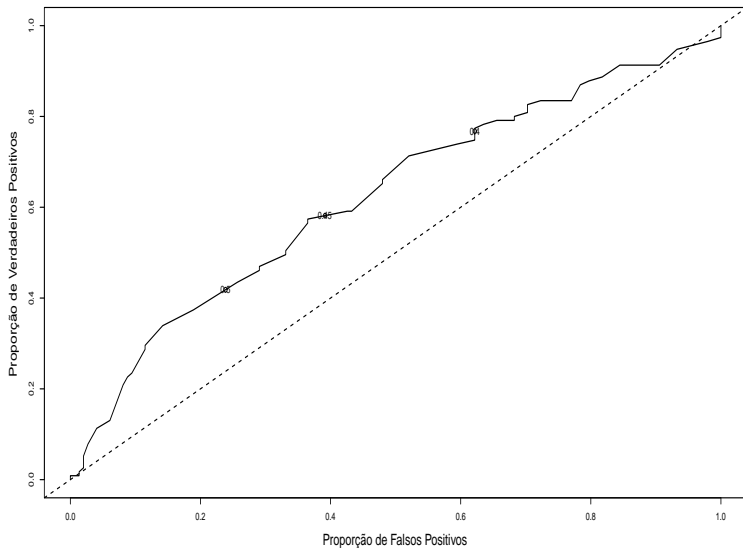
- Sensibilidade: proporção de verdadeiros positivos

$$SENS = \frac{68}{115} \cong 0,591(59,1\%)$$

- Especificidade: proporção de verdadeiros negativos

$$ESPEC = \frac{86}{148} \cong 0,581(58,1\%)$$





- 1 Preferência Automóveis
- 2 Análise de Dados Preliminar
- 3 Modelo Binomial
- 4 Resultados Modelo Ajustado
- 5 Curva ROC
- 6 Conclusões**
- 7 Referências

## Considerações Finais

- Neste exemplo em que ajustamos a probabilidade de um comprador preferir automóvel de marca americana em relação a marca japonesa, notamos que a idade e o estado civil do comprador são variáveis importantes.
- O ponto influente referente a um comprador com perfil atípico, embora mude de forma desproporcional as estimativas não muda a inferência ao nível de 10%.
- Não há indícios pelos gráficos de resíduos que a distribuição das respostas não seja Bernoulli.

- 1 Preferência Automóveis
- 2 Análise de Dados Preliminar
- 3 Modelo Binomial
- 4 Resultados Modelo Ajustado
- 5 Curva ROC
- 6 Conclusões
- 7 Referências**



## Referência

- Foster, D. P.; Stine, R. A. e Waterman, R. P. (1998). *Business Analysis using Regression*. New York: Springer.