3.33pt

Modelos Lineares Generalizados - Regressão Logística

Erica Castilho Rodrigues

01 de Julho de 2016

3.33pt

AIC

Introdução

- Vamos ver um critério para comparação de modelos.
- É muito utilizado para vários tipos de modelo.
- Mede o quão bem o modelo se ajusta aos dados.
- Porém penaliza pelo número de parâmetros.
- Leva assim à escolha de um modelo mais parcimonioso.

Akaike information criterion (AIC)

- É uma medida de qualidade de um modelo estatístico.
- Lida com o balancemento que devemos ter entre:
  - complexidade e qualidade do ajuste.
- Deve ser usado apenas para comparação entre modelos.
- Não pode ser usado para testar o ajuste de modo absoluto.

- O AIC é definido por

$$AIC = 2p - 2l(\mathbf{y}, \theta)$$

  onde
  - onde p é um número de parâmetros do modelo;
  - $l(\mathbf{y}, \theta)$ é o máximo da log-verossimilhança.
- Se um modelo é bom o AIC deve ser alto ou baixo?

- O AIC é definido por

$$AIC = 2p - 2l(\mathbf{y}, \theta)$$

  onde
  - onde p é um número de parâmetros do modelo;
  - $l(\mathbf{y}, \theta)$ é o máximo da log-verossimilhança.
- Se um modelo é bom o AIC deve ser alto ou baixo? Baixo.
- O melhor modelo será aquele com menor AIC.

- Vamos ver com detalhes os modelos para dados binários.
- Exemplos de resposta:
  - morto/vivo;
  - presente/absente.
- Podemos ter dois tipos de resposta.
- Resposta zero/um:
  - a variável resposta (Y_i) é uma Bernoulli.
- Temos então que

$$Y_i = \begin{cases} 1 & \text{se ocorre um sucesso} \\ 0 & \text{se ocorre um fracasso.} \end{cases}$$

- A função de densidade conjunta então é dada por

$$f(\mathbf{y}, \pi) =$$

- Vamos ver com detalhes os modelos para dados binários.
- Exemplos de resposta:
    - morto/vivo;
    - presente/absente.
- Podemos ter dois tipos de resposta.
- Resposta zero/um:
    - a variável resposta (Y_i) é uma Bernoulli.
- Temos então que

$$Y_i = \begin{cases} 1 & \text{se ocorre um sucesso} \\ 0 & \text{se ocorre um fracasso.} \end{cases}$$

- A função de densidade conjunta então é dada por

$$f(\mathbf{y}, \boldsymbol{\pi}) = \prod_i \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

- Número de sucessos:
  - a variável resposta (Y_i) é uma Binomial.
- A função de densidade conjunta então é dada por

$$f(\mathbf{y}, \pi) =$$

- Número de sucessos:
  - a variável resposta (Y_i) é uma Binomial.
- A função de densidade conjunta então é dada por

$$f(\mathbf{y}, \boldsymbol{\pi}) = \prod_i \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

- Queremos descrever a proporção de sucessos $P_i = \frac{Y_i}{n_i}$ em cada subgrupo.
- Essa proporção deve ser explicada por covariáveis do problema.

- Sabemos que

$$E(Y_i) =$$

- Sabemos que

$$E(Y_i) = n_i \pi_i \quad \Rightarrow \quad E\left(\frac{Y_i}{n_i}\right) = \pi_i \; .$$

- Vamos modelar $E(P_i) = \pi_i$ da seguinte maneira

$$g(\pi_i) = \mathbf{x}_i^t \beta$$

onde $g(.)$ é

- Sabemos que

$$E(Y_i) = n_i\pi_i \quad \Rightarrow \quad E\left(\frac{Y_i}{n_i}\right) = \pi_i \ .$$

- Vamos modelar $E(P_i) = \pi_i$ da seguinte maneira

$$g(\pi_i) = \mathbf{x}_i^t\beta$$

onde $g(.)$ é função de ligação.
- Três tipos de funções de ligação mais usadas:
  - logística,
  - complemento log-log,
  - probit.

- A ligação logística é definida da seguinte maneira

- A ligação logística é definida da seguinte maneira

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \eta$$

- A ligação logística é definida da seguinte maneira

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \eta$$

$$\left(\frac{\pi_i}{1-\theta_i}\right) = e^{\eta}$$

$$\pi_i = e^{\eta} - \theta_i e^{\eta} \quad \therefore \quad \theta_i(1+e^{\eta}) = e^{\eta}$$

$$\pi_i = \frac{e^{\eta}}{1+e^{\eta}}$$

- Dividindo o numerador e o denominador por e^η ficamos com

$$\theta_i = \frac{e^\eta(1/e^\eta)}{(1+e^\eta)(1/e^\eta)} = \frac{1}{1+e^{-\eta}}$$

- Ou seja, a função de ligação é dada por

$$E(Y_i) = \theta_i = \frac{1}{1+e^{-x_i^T\boldsymbol{\beta}}}$$

- Esse tipo de ligação é chamada ligação logísitica.
- Essa ligação pode levar a valores não válidos de θ_i?

- Dividindo o numerador e o denominador por e^{η} ficamos com
$$\theta_i = \frac{e^{\eta}(1/e^{\eta})}{(1+e^{\eta})(1/e^{\eta})} = \frac{1}{1+e^{-\eta}}$$
- Ou seja, a função de ligação é dada por
$$E(Y_i) = \theta_i = \frac{1}{1+e^{-x_i^T\boldsymbol{\beta}}}$$
- Esse tipo de ligação é chamada ligação logísitica.
- Essa ligação pode levar a valores não válidos de θ_i? Não.
- Em que intervalo θ_i está definido?

- Dividindo o numerador e o denominador por e^{η} ficamos com
$$\theta_i = \frac{e^{\eta}(1/e^{\eta})}{(1+e^{\eta})(1/e^{\eta})} = \frac{1}{1+e^{-\eta}}$$
- Ou seja, a função de ligação é dada por
$$E(Y_i) = \theta_i = \frac{1}{1+e^{-x_i^T \boldsymbol{\beta}}}$$
- Esse tipo de ligação é chamada ligação logísitica.
- Essa ligação pode levar a valores não válidos de θ_i? Não.
- Em que intervalo θ_i está definido? $[0, 1]$.
- Observe que a função
$$\frac{1}{1+e^{-\eta}}$$
está sempre dentro do intervalo [0,1].

- A ligação Probit é definida por

$$\Phi^{-1}(\theta_i) = x_i^T \boldsymbol{\beta} \Rightarrow \theta_i = \Phi(x_i^T \boldsymbol{\beta})$$

  onde $\Phi(.)$ é a função de distribuição acumulada da $N(0, 1)$.
- Essa função também está definida no intervalo [0,1].
- A ligação Complemento log-log é dada por

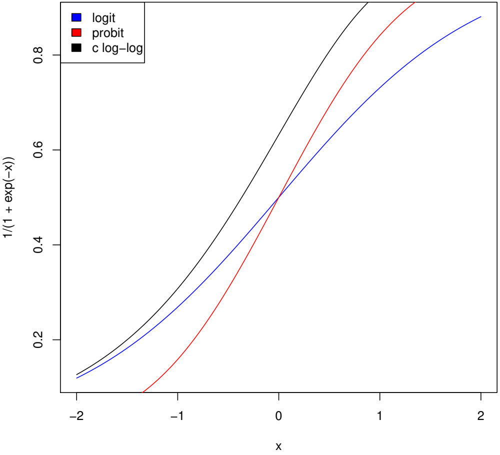
$$\log\left(-\log\left(1 - \theta_i\right)\right) = x_i^T \boldsymbol{\beta}$$

$$-\log(1 - \theta_i) = e^{x_i^T \boldsymbol{\beta}} \quad \therefore \quad (1 - \theta_i) = \exp\left(-e^{x_i^T \boldsymbol{\beta}}\right)$$

$$\theta_i = 1 - \exp\left(-e^{x_i^T \boldsymbol{\beta}}\right)$$

- Também está definida no intervalo [0,1].
- Para valores pequenos de θ_i:
  - fica bem próxima da ligação logística.

O gráfico a seguir compara as três funções de ligação.



- Para esses modelos temos duas formas de resíduos mais usados:
  - componente da Deviance;
  - componente da Estatística de Pearson generalizada.
- O **Resíduo de Pearson** é definido por

$$X_i = \frac{(y_i - n_i\hat{\pi}_i)}{\sqrt{n_i\hat{\pi}_i(1-\hat{\pi}_i)}}$$

- A soma dos seus quadrados resulta na Estatística de Pearson Generalizda

$$X^2 = \sum_i X_i^2 \,.$$

- Podemos a verão padronizada do Resíduo de Pearson

$$X_i' = \frac{X_i}{\sqrt{1 - h_{ii}}}$$

onde $h_i i$ é do termo da diagonal da matriz **H**.

- Vimos que a Deviance desse modelo é dada por

$$D = 2\left[\sum_i y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i)\log\left(\frac{n_i - \hat{y}_i}{n_i - y_i}\right)\right] .$$

- A **Deviance Residual** é então definida por

$$d_i = sign(y_i - \hat{y}_i)\left(y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i)\log\left(\frac{n_i - \hat{y}_i}{n_i - y_i}\right)\right)$$

a Deviande Residual Padronizada é definida por

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

onde $h_i i$ é o i-ésimo termo da diagonal da matriz **H**.

Exemplo:

- Deseja-se verificar a presença de sintomas de senilidade em idosos.
- Um amostra é selecionado e aplica-se um exame psiquiátrico.
- Para cada indivíduo a variável resposta é dada por

$$S_i = \begin{cases} 1 & \text{se os sintomas estão presentes,} \\ 0 & \text{se os sintomas não estão presentes.} \end{cases}$$

- Mediu-se ainda um escore que mede a inteliência de adultos e adolescentes:
  - Wechsler Adult Intelligence Scale (WAIS) .

Exemplo: (continuação)

- A tabela a seguir apresenta os dados do problema

x	s	x	s	x	s	x	s	x	s
9	1	7	1	7	0	17	0	13	0
13	1	5	1	16	0	14	0	13	0
6	1	14	1	9	0	19	0	9	0
8	1	13	0	9	0	9	0	15	0
10	1	16	0	11	0	11	0	10	0
4	1	10	0	13	0	14	0	11	0
14	1	12	0	15	0	10	0	12	0
8	1	11	0	13	0	16	0	4	0
11	1	14	0	10	0	10	0	14	0
7	1	15	0	11	0	16	0	20	0
9	1	18	0	6	0	14	0		

Exemplo: (continuação)

- Algumas pessoas possuem o mesmo valor para o score WAIS.
- Podemos colocar essas pessoas em um mesmo grupo.
- E contamos o número de sucessos nesse grupo.
- Exemplo:
  - existem 3 pessoas com escore 7;
  - dessas 3, 2 tem os sintomas e uma não;
  - podemos agrupá-lase em uma com n_i=3;
  - número de sucessos $y_i = 2$.

Exemplo: (continuação)

- Vamos usar a seguinte notação:
  - Y_i número de casos no i-ésimo grupo;
  - n_i número total de pessoas no gurpo.
- Vamos ajustar um modelo Binomial com ligação logística, ou seja

Exemplo: (continuação)

- Vamos usar a seguinte notação:
  - Y_i número de casos no i-ésimo grupo;
  - n_i número total de pessoas no gurpo.
- Vamos ajustar um modelo Binomial com ligação logística, ou seja

$$Y_i \sim Bin(n_i, \pi_i)$$

Exemplo: (continuação)

- Vamos usar a seguinte notação:
  - Y_i número de casos no i-ésimo grupo;
  - n_i número total de pessoas no gurpo.
- Vamos ajustar um modelo Binomial com ligação logística, ou seja

$$Y_i \sim Bin(n_i, \pi_i) \quad \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_i \ .$$

- O modelo foi ajustado e os valores estimados são dados por

$$\log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = 2,404 - 0,3235 x_i \ .$$

Exemplo: **(continuação)**

- Qual interpretação do coeficiente de x_i?
- Temos que

$$e^{-0,3235} = 0.7237$$

isso singifica que

Exemplo: (continuação)

- Qual interpretação do coeficiente de x_i?
- Temos que
$$e^{-0,3235} = 0.7237$$
isso singifica que para cada aumento em uma unidade no escore WAIS espera-se uma redução dem 28% na chance de ter os sintomas.
- A Deviance e Estatística de Pearson são dados por
$$D = 8,083 \quad X^2 = 9,419\,.$$
- Devemos comparar com a distribuição

Exemplo: (continuação)

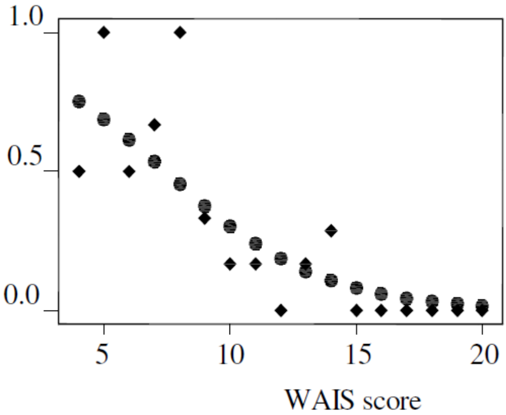
- Qual interpretação do coeficiente de x_i?
- Temos que
$$e^{-0,3235} = 0.7237$$
isso singifica que para cada aumento em uma unidade no escore WAIS espera-se uma redução dem 28% na chance de ter os sintomas.
- A Deviance e Estatística de Pearson são dados por
$$D = 8,083 \quad X^2 = 9,419\,.$$
- Devemos comparar com a distribuição $X^2_{n-p} = X^2_{17-2}$.
- O valor crítico é dado por 24,95.
- Conclusão:

Exemplo: (continuação)

- Qual interpretação do coeficiente de x_i?
- Temos que
$$e^{-0,3235} = 0.7237$$
isso singifica que para cada aumento em uma unidade no escore WAIS espera-se uma redução dem 28% na chance de ter os sintomas.
- A Deviance e Estatística de Pearson são dados por
$$D = 8,083 \quad X^2 = 9,419 \,.$$
- Devemos comparar com a distribuição $X^2_{n-p} = X^2_{17-2}$.
- O valor crítico é dado por 24,95.
- Conclusão: o modelo parece estar bem ajustado.

Exemplo: **(continuação)**

- No grafico a seguir os círculos são as propoções estimadas pelo modelo e os losangos são as proporções observadas.



- O modelo parece se ajustar bem aos dados.

Exemplo: (continuação)

- A tabela a seguir mostra os valores ajustados e os resíduos de Pearson (X) e Deviance residual (d).

x	y	n	$\widehat{\pi}$	X	d
4	1	2	0.751	-0.826	-0.766
5	0	1	0.687	0.675	0.866
6	1	2	0.614	-0.330	-0.326
7	1	3	0.535	0.458	0.464
8	0	2	0.454	1.551	1.777
9	4	6	0.376	-0.214	-0.216
10	5	6	0.303	-0.728	-0.771
11	5	6	0.240	-0.419	-0.436
12	2	2	0.186	-0.675	-0.906
13	5	6	0.142	0.176	0.172
14	5	7	0.107	1.535	1.306
15	3	3	0.080	-0.509	-0.705
16	4	4	0.059	-0.500	-0.696
17	1	1	0.043	-0.213	-0.297
18	1	1	0.032	-0.181	-0.254
19	1	1	0.023	-0.154	-0.216
20	1	1	0.017	-0.131	-0.184

Exemplo: (continuação)

- Os resíduos não apresentam nenhum valor muito discrepante.

- O modelo binomial retorna uma estimativa das probabilidade de sucesso.
- Em muitos casos queremos classificar os elementos:
    - e-mail $\Rightarrow$ Spam ou não;
    - cliente $\Rightarrow$ inadinplente ou não;
    - cliente $\Rightarrow$ alto ou baixo risco.
- Precisamos definir um ponto de corte para as probabilidades estimadas.
- Veremos uma ferramenta para nos auxiliar nessa escolha.

- Quando estamos classificando itens em sucesso/fracasso podemos cometer dois erros.

- Quando estamos classificando itens em sucesso/fracasso podemos cometer dois erros.
- Dizer que é sucesso quando é fracasso (falso positivo).
- Dizer que é fracasso quando é sucesso (falso negativo).
- Esses dois tipos de erros levan duas medidas de qualidade do modelo:
  - **Sensibilidade:** 1-proporção de falsos positivos;
  - **Especificidade:** 1-proporção de falsos negativos.

- Em ensaios clínicos esse tipo de nomenclatura é muito comum.
- A **Sensibilidade** e **Especificidade** são usadas para medir a confiabilidade do teste para diagnóstico de doenças.
- Suponha que queremos verificar se o paciente tem uma doença observando seus sintomas.
- Para verificar a qualidade do teste precisamos de dois grupos de pacientes:
  - um grupo saudável, temos certeza que não tem a doença;
  - um grupo que tem a doença com certeza.

- Podemos construir uma tabela de contigência.
- Ela contém o número de vezes que o resultado do teste condiz com a realidade.

Doença	Teste		
	Poisitivo	Negativo	Total
Presente	a	b	$n_1 = a + b$
Ausente	c	d	$n_2 = c + d$
Total	a+c	b+d	

- Dentre os n_1 pacientes com a doença o teste detectou a.
- Dentre os n_2 pacientes sem a doença o teste foi correto d vezes.

Sensibilidade

- Probabilidade do teste acertar em detectar a doença

Sensibilidade

- Probabilidade do teste acertar em detectar a doença

$$P(\text{positivo}|\text{paciente tem a doença})\ .$$

- Quão sensível o teste é para detectar a doença.
- A partir da tabela pode ser estimada por

$$\text{Sensibilidade} =$$

Sensibilidade

- Probabilidade do teste acertar em detectar a doença

$$P(\text{positivo}|\text{paciente tem a doença})\ .$$

- Quão sensível o teste é para detectar a doença.
- A partir da tabela pode ser estimada por

$$\text{Sensibilidade} = \frac{a}{n_1}\ .$$

Especificidade

- Probabilidade do teste diagnosticar corretamente um paciente saudável

Especificidade

- Probabilidade do teste diagnosticar corretamente um paciente saudável

$$P(\text{negativo}|\text{paciente não tem a doença})\ .$$

- Quão específico o teste é para aquela doença? Não sei detectando o que não deveria.
- A partir da tabela pode ser estimada por

$$\text{Especificidade} =$$

Especificidade

- Probabilidade do teste diagnosticar corretamente um paciente saudável

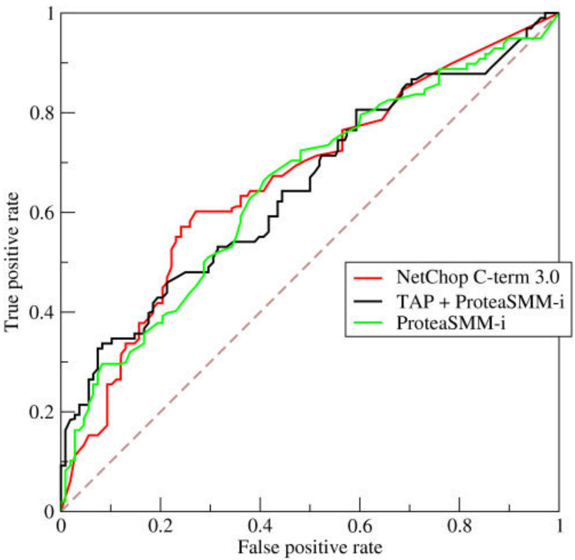
$$P(\text{negativo}|\text{paciente não tem a doença}) .$$

- Quão específico o teste é para aquela doença? Não sei detectando o que não deveria.
- A partir da tabela pode ser estimada por

$$\text{Especificidade} = \frac{d}{n_2} .$$

Curva ROC (Receiver operating characteristic)

- É um gráfico que descreve a qualidade de um operador binário.
- É uma representação gráfica das seguintes quantidades:
  - sensibilidade vs 1-especificidade.
- Ou seja, consiste em fazer o gráfico de:
  - verdadeiros positivos vs falsos positivos
- Tais quantidades variam de acordo com o limiar de probabilidade que é considerado na classificação.
- Pode ser usada para determinar qual o limiar ideal.


1
0.8
0.6
0.4
0.2
0
True positive rate
0
0.2
0.4
0.6
0.8
1
False positive rate
NetChop C-term 3.0
TAP + ProteaSMM-i
ProteaSMM-i

- O nome “Receiver operating characteristic” vem do fato de:
  - comparar duas caraterísticas do teste.
- Se colocamos um limiar de probabilidade baixo:

- O nome “Receiver operating characteristic” vem do fato de:
  - comparar duas caraterísticas do teste.
- Se colocamos um limiar de probabilidade baixo:
  - todo mundo é classificado como

- O nome “Receiver operating characteristic” vem do fato de:
  - comparar duas caraterísticas do teste.
- Se colocamos um limiar de probabilidade baixo:
  - todo mundo é classificado como sucesso;
  - a taxa de verdadeiros positivos é

- O nome “Receiver operating characteristic” vem do fato de:
  - comparar duas caraterísticas do teste.
- Se colocamos um limiar de probabilidade baixo:
  - todo mundo é classificado como sucesso;
  - a taxa de verdadeiros positivos é alta;
  - a taxa de verdadeiros negativos é

- O nome “Receiver operating characteristic” vem do fato de:
  - comparar duas caraterísticas do teste.
- Se colocamos um limiar de probabilidade baixo:
  - todo mundo é classificado como sucesso;
  - a taxa de verdadeiros positivos é alta;
  - a taxa de verdadeiros negativos é baixa;
  - a taxa de falso posititivos é

- O nome “Receiver operating characteristic” vem do fato de:
  - comparar duas caraterísticas do teste.
- Se colocamos um limiar de probabilidade baixo:
  - todo mundo é classificado como sucesso;
  - a taxa de verdadeiros positivos é alta;
  - a taxa de verdadeiros negativos é baixa;
  - a taxa de falso posititivos é alta;
  - sensibilidade

- O nome “Receiver operating characteristic” vem do fato de:
  - comparar duas caraterísticas do teste.
- Se colocamos um limiar de probabilidade baixo:
  - todo mundo é classificado como sucesso;
  - a taxa de verdadeiros positivos é alta;
  - a taxa de verdadeiros negativos é baixa;
  - a taxa de falso posititivos é alta;
  - sensibilidade alta;
  - especificidade

- O nome “Receiver operating characteristic” vem do fato de:
  - comparar duas caraterísticas do teste.
- Se colocamos um limiar de probabilidade baixo:
  - todo mundo é classificado como sucesso;
  - a taxa de verdadeiros positivos é alta;
  - a taxa de verdadeiros negativos é baixa;
  - a taxa de falso posititivos é alta;
  - sensibilidade alta;
  - especificidade baixa.

- Se colocamos um limiar de probabilidade alto:

- Se colocamos um limiar de probabilidade alto:
  - todo mundo é classificado como

- Se colocamos um limiar de probabilidade alto:
  - todo mundo é classificado como fracasso;
  - a taxa de verdadeiros positivos é

- Se colocamos um limiar de probabilidade alto:
  - todo mundo é classificado como fracasso;
  - a taxa de verdadeiros positivos é baixa;
  - a taxa de verdadeiros negativos é

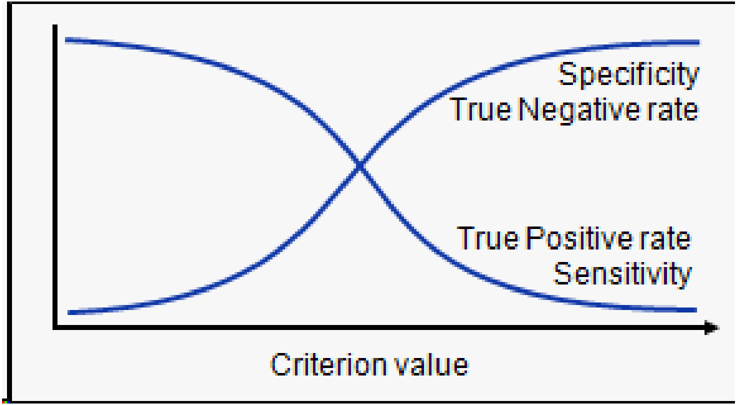
- Se colocamos um limiar de probabilidade alto:
  - todo mundo é classificado como fracasso;
  - a taxa de verdadeiros positivos é baixa;
  - a taxa de verdadeiros negativos é alta;
  - a taxa de falso posititivos é

- Se colocamos um limiar de probabilidade alto:
  - todo mundo é classificado como fracasso;
  - a taxa de verdadeiros positivos é baixa;
  - a taxa de verdadeiros negativos é alta;
  - a taxa de falso posititivos é baixa;
  - sensibilidade

- Se colocamos um limiar de probabilidade alto:
  - todo mundo é classificado como fracasso;
  - a taxa de verdadeiros positivos é baixa;
  - a taxa de verdadeiros negativos é alta;
  - a taxa de falso posititivos é baixa;
  - sensibilidade baixa;
  - especificidade

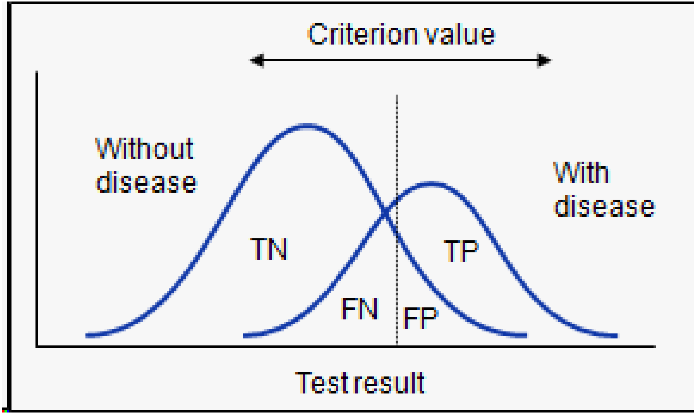
- Se colocamos um limiar de probabilidade alto:
    - todo mundo é classificado como fracasso;
    - a taxa de verdadeiros positivos é baixa;
    - a taxa de verdadeiros negativos é alta;
    - a taxa de falso posititivos é baixa;
    - sensibilidade baixa;
    - especificidade alta.
- Conclusão:
    - a especificidade diminui com o aumento da sensibilidade;
    - existe um “trade-off”, como no erro do tipo I e tipo II;
    - 1-especificadade aumenta com o aumento da sensibilidade.

- A figura abaixo ilustra como essas duas quantidades variam.

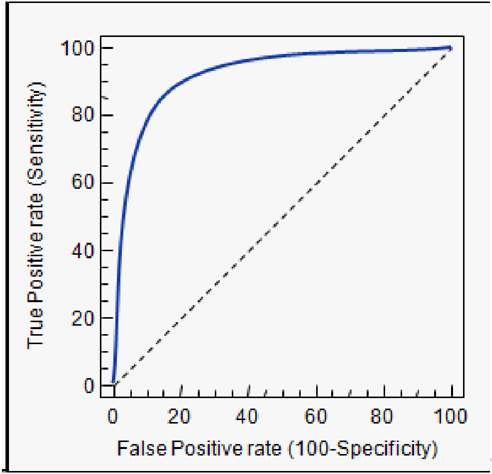


- A curva ROC é uma representação gráfica da:
  - sensibilidade vs 1-especificidade;
  - verdadeiros positivos (TPR) vs falsos positivos (FPR).
- É uma comparação entre duas caraterísticas de operação de um classificador:
  - TPR e FPR.
- Mede quão bem um parâmetro consegue distinguir dois grupos.

- Os indivíduos com e sem a doença têm distribuições distintas.
- Definimos um ponto de corte.
- Podemos errar nas interceções.
- Semelhante ao teste de hipótese para comparar populações.



- As quantidades são plotadas para diferentes limiares de probabilidade.
- Cada ponto corresponde ao TPR e FPR para um limiar específico.



- Qual ponto ideal?

- Qual ponto ideal?
- O ponto onde:
  - sensibilidade=TPR=

- Qual ponto ideal?
- O ponto onde:
  - sensibilidade=TPR=100%;
  - 1-especificidade=FPR=

- Qual ponto ideal?
- O ponto onde:
  - sensibilidade=TPR=100%;
  - 1-especificidade=FPR=0%.
- Uma curva ideal passa pela quina à esquerda (como aquela mostrada no exemplo).
- O que acontece com uma curva que passa na reta?

- Qual ponto ideal?
- O ponto onde:
  - sensibilidade=TPR=100%;
  - 1-especificidade=FPR=0%.
- Uma curva ideal passa pela quina à esquerda (como aquela mostrada no exemplo).
- O que acontece com uma curva que passa na reta?
- TPR=FPR
- Erra o mesmo tanto que acerta.
- Esse resultado pode ser obtido sem modelo nenhum.
- Basta jogar uma moeda para cima e classificar como sucesso ou fracasso.

- Considere um elemento que é sucesso.
- Nesse caso temos que:
  - probabilidade de ser detectado=probabilidade de não ser detectado.
- O modelo não consegue distinguir os dois grupos:
  - não consegue fazer com que probabilidade de detecção seja maior do que a probabilidade de não detecção para um objeto que é sucesso.
- Chega um paciente no consultório.
- O médico não tem uma probabilidade maior de dizer que está doente do que dizer que não está.

- Uma medida para verificar se o classificador é bom:
  - área abaixo da curva (AUC - Area Under Curve).
- Mede o quanto a curva se afasta da reta.
- Considere um elemento que é positivo.
- É a probabilidade do classificador ter maior chance de dizer que ele é positivo do que negativo.
- Quanto **maior** a AUC $\Rightarrow$ **maior** poder de classificação.
- Pode ser usada para escolher qual o melhor modelo.

Exemplo:

- Considere os seguintes dados e o modelo ajustado.

```
x1<-c(29, 30, 31, 31, 32, 29, 30, 31, 32, 33)
x2<-c(62, 83, 74, 88, 68, 41, 74, 21, 50, 33)
Y<-c(0, 0, 0, 0, 0, 1, 1, 1, 1, 1)

modelo=glm(Y~x1+x2,family="binomial")
```

Exemplo: (continuaçã)

- O resultado do ajuste do modelo é mostrado a seguir.

```
> summary(modelo)
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  7.941953  24.469012   0.325    0.746
x1           0.009833   0.770999   0.013    0.990
x2          -0.132726   0.081249  -1.634    0.102

Residual deviance:  6.7789  on 7  degrees of freedo
AIC: 12.779
```

Exemplo: (continuação)

- Queremos encontrar um ponto de corte para as probabilidades estimadas.
- A tabela a seguir mostra os resultados para alguns valores testados.

$Pr\{Y = 1\} > 0.5$

		$\hat{Y}$ 0	$\hat{Y}$ 1
Y	0	5	0
	1	1	4

$Pr\{Y = 1\} > 0.7$

		$\hat{Y}$ 0	$\hat{Y}$ 1
Y	0	5	0
	1	1	4

$Pr\{Y = 1\} > 0.4$

		$\hat{Y}$ 0	$\hat{Y}$ 1
Y	0	4	1
	1	1	4

$Pr\{Y = 1\} > 0.3$

		$\hat{Y}$ 0	$\hat{Y}$ 1
Y	0	3	2
	1	1	4

Exemplo: (continuação)

- Quais pontos de corte parecem adequados?

Exemplo: (continuação)

- Quais pontos de corte parecem adequados?
  - 0,5 ou 0,7.
- Vamos calcular a sensibilidade e especificidades do teste.
- Os comandos a seguir fazem esse cálculo para $p = 0,5$

```
p=.5
> predicted=ifelse(modelo$fitted>p,1,0)
>
> (sensibilidade=sum(predicted==modelo$y
& predicted==1)/sum(Y==1))
[1] 0.8
> (especificidade=sum(predicted==modelo$y
& predicted==0)/sum(Y==0))
[1] 1
```

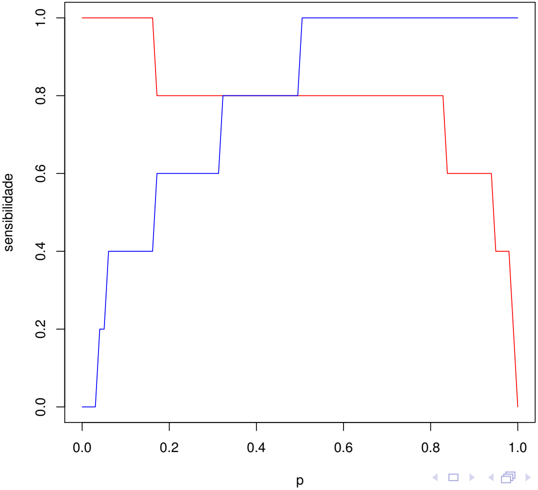
Exemplo: (continuação)

- Os comandos a seguir fazem o gráfico das duas quantidades para vários valores de p.

```
p=seq(0,1,length=100)

sensibilidade=NULL
especificidade=NULL
for( i in 1:length(p)){
predicted=ifelse(modelo$fitted>p[i],1,0)
sensibilidade[i]=sum(predicted==modelo$y
 & predicted==1)/sum(Y==1)
especificidade[i]=sum(predicted==modelo$y
 & predicted==0)/sum(Y==0)

}

plot(p,sensibilidade,type='l',col='red')
lines(p,especificidade,col='blue')
```

Exemplo: (continuação)

- O gráfico obtido é mostrado a seguir.



Exemplo: (continuação)

- Para p grande, classica todo mundo como

Exemplo: (continuação)

- Para p grande, classica todo mundo como fracasso.
- A chance de errar ao classificar como fracasso é

Exemplo: **(continuação)**

- Para p grande, classica todo mundo como fracasso.
- A chance de errar ao classificar como fracasso é alta.
- A especificidade é

Exemplo: (continuação)

- Para p grande, classica todo mundo como fracasso.
- A chance de errar ao classificar como fracasso é alta.
- A especificidade é baixa.
- Classifica poucos como sucesso.
- A chance de errar ao classificar como sucesso é

Exemplo: (continuação)

- Para p grande, classica todo mundo como fracasso.
- A chance de errar ao classificar como fracasso é alta.
- A especificidade é baixa.
- Classifica poucos como sucesso.
- A chance de errar ao classificar como sucesso é baixa.
- A sensibilidade é

Exemplo: (continuação)

- Para p grande, classica todo mundo como fracasso.
- A chance de errar ao classificar como fracasso é alta.
- A especificidade é baixa.
- Classifica poucos como sucesso.
- A chance de errar ao classificar como sucesso é baixa.
- A sensibilidade é alta.

Exemplo: (continuação)

- Para p pequeno, classica todo mundo como

Exemplo: (continuação)

- Para p pequeno, classica todo mundo como sucesso.
- A chance de errar ao classificar como sucesso é

Exemplo: (continuação)

- Para p pequeno, classica todo mundo como sucesso.
- A chance de errar ao classificar como sucesso é alta.
- A sensibilidade é

Exemplo: (continuação)

- Para p pequeno, classica todo mundo como sucesso.
- A chance de errar ao classificar como sucesso é alta.
- A sensibilidade é baixa.
- Classifica poucos como fracasso.
- A chance de errar ao classificar como fracasso é

Exemplo: **(continuação)**

- Para p pequeno, classica todo mundo como sucesso.
- A chance de errar ao classificar como sucesso é alta.
- A sensibilidade é baixa.
- Classifica poucos como fracasso.
- A chance de errar ao classificar como fracasso é baixa.
- A especificidade é

Exemplo: (continuação)

- Para p pequeno, classica todo mundo como sucesso.
- A chance de errar ao classificar como sucesso é alta.
- A sensibilidade é baixa.
- Classifica poucos como fracasso.
- A chance de errar ao classificar como fracasso é baixa.
- A especificidade é alta.
- Podemos escolher como corte o ponto em que

$$\text{sensibilidade} = \text{especificidade} \approx 0,5\,.$$