DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA

Probabilidades e Estatística

LEE, LEGI, LEMat, LERC/LETI, LMAC, MEAer, MEAmb, MEBiol, MEBiom, MEEC, MEFT, MEQ

2º semestre – 2016/2017
16/06/2017 – **11h:00**

Duração: 90 minutos **2º teste**

Justifique convenientemente todas as respostas

Grupo I	10 valores

1. Amostras de solo de certa região têm acidez descrita por uma variável aleatória X cuja função de densidade de probabilidade é dada por

$$f_X(x) = \begin{cases} \frac{2x}{\theta}\exp\left(-\frac{x^2}{\theta}\right), & x \geq 0 \\ 0, & x < 0, \end{cases}$$

onde θ é um parâmetro positivo desconhecido.

(a) Deduza o estimador de máxima verosimilhança de θ com base numa amostra aleatória $(X_1, \ldots, X_n)$ proveniente da população X. (3.0)

- **V.a. de interesse**
  X = acidez do solo em certa região
- **F.d.p. de** X
  $$f_X(x) = \begin{cases} \frac{2x}{\theta}\exp\left(-\frac{x^2}{\theta}\right), & x \geq 0 \\ 0, & x < 0 \end{cases}$$
- **Parâmetro desconhecido**
  $\theta, \quad \theta > 0$
- **Amostra**
  $\underline{x} = (x_1, \ldots, x_n)$ amostra de dimensão n proveniente da população X [com $x_i > 0$, $i = 1, \ldots, n$].
- **Obtenção do estimador de MV de** θ

  Passo 1 — Função de verosimilhança

$$\begin{aligned} L(\theta|\underline{x}) &= f_{\underline{X}}(\underline{x}) \\ &\stackrel{X_i\ indep}{=} \prod_{i=1}^n f_{X_i}(x_i) \\ &\stackrel{X_i \sim X}{=} \prod_{i=1}^n f_X(x_i) \\ &= \prod_{i=1}^n \left[\frac{2x_i}{\theta}\exp\left(-\frac{x_i^2}{\theta}\right)\right] \\ &= 2^n \times \theta^{-n} \times \left(\prod_{i=1}^n x_i\right) \times \exp\left(-\frac{1}{\theta}\sum_{i=1}^n x_i^2\right), \quad \theta > 0 \end{aligned}$$

  Passo 2 — Função de log-verosimilhança

$$\ln L(\theta|\underline{x}) = n\ln(2) - n\ln(\theta) + \sum_{i=1}^n \ln(x_i) - \frac{1}{\theta}\sum_{i=1}^n x_i^2$$

  Passo 3 — Maximização
  A estimativa de MV de θ é doravante representada por $\hat{\theta}$ e

$$\hat{\theta} \quad : \quad \begin{cases} \left.\frac{d\ln L(\theta|\underline{x})}{d\theta}\right|_{\theta=\hat{\theta}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2\ln L(\theta|\underline{x})}{d\theta^2}\right|_{\theta=\hat{\theta}} < 0 & \text{(ponto de máximo)} \end{cases}$$

$$\hat{\theta} \quad : \quad \begin{cases} -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n x_i^2 = 0 \\ \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^n x_i^2 < 0 \end{cases}$$

$$\begin{cases} \hat{\theta} = \frac{1}{n}\sum_{i=1}^n x_i^2 \\ \frac{n}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)^2} - \frac{2}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)^3}\sum_{i=1}^n x_i^2 = -\frac{n^3}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)^2} < 0 \qquad \text{(proposição verdadeira).} \end{cases}$$

Passo 4 — Estimador de MV de θ

$$EMV(\theta) = \frac{1}{n}\sum_{i=1}^n X_i^2$$

(b) Tendo-se recolhido a amostra $(x_1, \ldots, x_5) = (2.9, 8.3, 3.6, 8.9, 3.9)$ para a qual $\sum_{i=1}^5 x_i^2 = 184.68$, obtenha a estimativa de máxima verosimilhança da probabilidade de a acidez de uma amostra de solo da região pertencer ao intervalo $[5.5, 6.5]$. (2.0)

Nota: A função de distribuição de X é dada por $F_X(x) = 1 - \exp\left(-\frac{x^2}{\theta}\right)$, para $x \geq 0$.

- **Estimativa de MV de θ**

$$\begin{aligned} \hat{\theta} &= \frac{1}{n}\sum_{i=1}^n x_i^2 \\ &= \frac{184.68}{5} \\ &\simeq 36.936 \end{aligned}$$

- **Outro parâmetro desconhecido**

$$\begin{aligned} h(\theta) &= P(5.5 \leq X \leq 6.5) \\ &= F_X(6.5) - F_X(5.5) \\ &= \left[1 - \exp\left(-\frac{6.5^2}{\theta}\right)\right] - \left[1 - \exp\left(-\frac{5.5^2}{\theta}\right)\right] \\ &= \exp\left(-\frac{5.5^2}{\theta}\right) - \exp\left(-\frac{6.5^2}{\theta}\right) \end{aligned}$$

- **Estimativa de MV de $h(\theta)$**

  Invocando a propriedade de invariância dos estimadores de máxima verosimilhança, pode concluir-se que a estimativa de MV de $h(\theta)$ é

$$\begin{aligned} \widehat{h(\theta)} &= h(\hat{\theta}) \\ &= \exp\left(-\frac{5.5^2}{\hat{\theta}}\right) - \exp\left(-\frac{6.5^2}{\hat{\theta}}\right) \\ &= \exp\left(-\frac{5.5^2}{36.936}\right) - \exp\left(-\frac{6.5^2}{36.936}\right) \\ &\simeq 0.122296. \end{aligned}$$

2. Um inquérito a 1000 lisboetas revelou que, entre eles, 285 são favoráveis à aplicação de uma taxa à circulação automóvel no centro histórico da cidade.

(a) Com base nestes dados, construa um intervalo de confiança a aproximadamente 95% para a proporção de lisboetas que são favoráveis à proposta. (2.5)

- **V.a. de interesse**

  X = resposta de um lisboeta (escolhido ao acaso) ao inquérito

- **Situação**

  $X \sim \text{Bernoulli}(p)$
  $p = P(\text{resposta favorável à proposta})$ DESCONHECIDA
  $n = 1000 >> 30$ (suficientemente grande).

- **Obtenção de IC aproximado para p**

  Passo 1 — Selecção da v.a. fulcral para p

  [Uma vez que nos foi solicitada a determinação de um IC aproximado para uma probabilidade e a dimensão da amostra é suficientemente grande para justificar o recurso à seguinte v.a. fulcral para p com distribuição aproximada:]

  $$Z = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \overset{a}{\sim} \text{normal}(0,1)$$

  Passo 2 — Obtenção dos quantis de probabilidade

  Os quantis a utilizar são

  $$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = -\Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(0.975) \overset{tabela}{=} -1.9600 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.975) = 1.9600. \end{cases}$$

  [Estes enquadram a v.a. fulcral para p com probabilidade aproximadamente igual a $(1-\alpha) = 0.95$.]

  Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$

  $$P(a_\alpha \leq Z \leq b_\alpha) \simeq 1 - \alpha$$

  $$P\left[a_\alpha \leq \frac{\bar{X}-p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \leq b_\alpha\right] \simeq 1-\alpha$$

  $$P\left[\bar{X} - b_\alpha \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} - a_\alpha \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] \simeq 1-\alpha$$

  $$P\left[\bar{X} - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] \simeq 1-\alpha.$$

  Passo 4 — Concretização

  Ao ter-se em consideração que

  - $n = 1000$
  - $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{285}{1000} = 0.285$ [≡ proporção observada de respostas favoráveis]
  - $\Phi^{-1}(1-\alpha/2) = 1.9600$,

  conclui-se que o intervalo de confiança aproximadamente igual a 95% para p é dado por

  $$\left[\bar{x} - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}},\quad \bar{x} + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

  $$= \left[0.285 - 1.9600 \times \sqrt{\frac{0.285\times(1-0.285)}{1000}},\quad 0.285 + 1.9600 \times \sqrt{\frac{0.285\times(1-0.285)}{1000}}\right]$$

  $$= [0.257021, 0.312979].$$

(b) Uma engenheira afirmou que "um quarto dos lisboetas é favorável à proposta". Avalie se os dados recolhidos contrariam esta afirmação. Decida com base no valor-p. (2.5)

- **Hipóteses**

  $H_0 : p = p_0 = 0.25$

  $H_1 : p \neq p_0$

- **Estatística de teste**

  [Sabe-se que o estimador de MV de p é $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, onde $X_i \sim_{i.i.d.} X$. Para além disso, $E(\bar{X}) = E(X) = p$ e $V(\bar{X}) = \frac{1}{n}V(X) = \frac{p(1-p)}{n} < +\infty$. Então pelo TLC pode afirmar-se que $\frac{\bar{X}-E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{a}{\sim} \text{normal}(0,1)$, pelo que a estatística de teste é]

$$T = \frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \stackrel{a}{\sim}_{H_0} \text{normal}(0,1).$$

- **Região de rejeição de H_0** (para valores de T)
  Tratando-se de um teste bilateral ($H_1 : p \neq p_0$), a região de rejeição de H_0, escrita para valores da estatística de teste, é do tipo $W = (-\infty, -c) \cup (c, +\infty)$.

- **Decisão (com base no valor-p)**
  O valor observado da estatística de teste é
  $$\begin{aligned} t &= \frac{\bar{x} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ &= \frac{0.285 - 0.25}{\sqrt{\frac{0.25\,(1-0.25)}{1000}}} \\ &\simeq 2.56. \end{aligned}$$
  Dado que a região de rejeição deste teste é a reunião de dois intervalos simétricos, temos:
  $$\begin{aligned} valor-p &= 2 \times P(T > |t| \mid H_0) \\ &\simeq 2 \times [1 - \Phi(|t|)] \\ &\simeq 2 \times [1 - \Phi(2.56)] \\ &\stackrel{calc/tabela}{=} 2 \times (1 - 0.9948) \\ &= 0.0104. \end{aligned}$$
  Consequentemente, é suposto:
  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 1.04\%$, pelo que $H_0 : p = p_0 = 0.25$ não é contrariada pelos dados ao n.u.s. de 1%;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > 1.04\%$, nomeadamente aos n.u.s. de 5% e 10%.

Grupo II — 10 valores

1. Um arquitecto conjectura que o primeiro algarismo (X) da altura de uma estrutura artificial, escolhida ao acaso entre estruturas com pelo menos 100 metros, possui função de probabilidade

$$P(X = i) = \log_{10}\left(1 + \tfrac{1}{i}\right), \quad \text{para} \quad i = 1, 2, \ldots, 9.$$

Uma amostra relativa a 50 estruturas artificiais nas condições referidas, conduziu aos seguintes dados:

Primeiro algarismo da altura	1	2	3	4	$\{5,\ldots,9\}$
Frequência absoluta observada	26	4	11	6	3
Frequência absoluta esperada	E_1	8.8	6.2	4.8	E_5

(a) Calcule os valores das frequências absolutas esperadas E_1 e E_5 (aproximando-os às décimas). (0.5)

- **V.a. de interesse**
  X = primeiro algarismo da altura de uma estrutura artificial com pelo menos 100 metros

- **F.p. conjecturada**
  $\log_{10}\left(1 + \frac{1}{i}\right)$, $i = 1, 2, \ldots, 9$

- **Frequências absolutas esperadas omissas**
  Atendendo à dimensão da amostra $n = 50$ e à f.p. conjecturada, segue-se:
  $$\begin{aligned} E_1 &= 50 \times \log_{10}\left(1 + \frac{1}{1}\right) \\ &\simeq 15.1; \end{aligned}$$

$$
\begin{aligned}
E_5 &= n - \sum_{i=1}^{4} E_i \\
&\simeq 50 - (15.1 + 8.8 + 6.2 + 4.8) \\
&= 15.1.
\end{aligned}
$$

(b) Teste a hipótese conjeturada pelo arquiteto, ao nível de significância de 1%. (3.0)

- **Hipóteses**
  $H_0 : P(X = i) = \log_{10}\left(1 + \frac{1}{i}\right), \quad \text{para} \quad i = 1, 2, \ldots, 9$
  $H_1 : \neg H_0$

- **Nível de significância**
  $\alpha_0 = 1\%$

- **Estatística de Teste**
  $$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim}_{H_0} \chi^2_{(k-\beta-1)},$$
  onde:
  k = No. de classes = 5
  O_i = Frequência absoluta observável da classe i
  E_i = Frequência absoluta esperada, sob H_0, da classe i
  β = No. de parâmetros a estimar = 0 [dado que a distribuição conjecturada em H_0 está completamente especificada, i.e., H_0 é uma hipótese simples.]

- **Frequências absolutas esperadas sob H_0**
  De acordo com a tabela facultada e a alínea (a), os valores das frequências absolutas esperadas sob H_0 aproximados às décimas são: $E_1 \simeq 15.1$; $E_2 \simeq 8.8$; $E_3 \simeq 6.2$; $E_4 \simeq 4.8$; $E_5 \simeq 15.1$.

  [De notar que não é necessário fazer qualquer agrupamento de classes uma vez que em pelo menos 80% das classes se verifica $E_i \geq 5$ e que $E_i \geq 1$ para todo o i. Caso fosse preciso efectuar agrupamento de classes, os valores de k e c teriam que ser recalculados...]

- **Região de rejeição de H_0** (para valores de T)
  Tratando-se de um teste de ajustamento, a região de rejeição de H_0 escrita para valores de T é o intervalo à direita $W = (c, +\infty)$, onde
  $$
  \begin{aligned}
  c &= F^{-1}_{\chi^2_{(k-\beta-1)}}(1 - \alpha_0) \\
  &= F^{-1}_{\chi^2_{(5-0-1)}}(1 - 0.01) \\
  &= F^{-1}_{\chi^2_{(4)}}(0.99) \\
  &\overset{tabela/calc}{=} 13.28.
  \end{aligned}
  $$

- **Decisão**
  No cálculo do valor observado da estatística de teste convém recorrer à seguinte tabela auxiliar:

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esper. sob H_0 E_i	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$\{1\}$	26	15.1	$\frac{(26-15.1)^2}{15.1} \simeq 7.868$
2	$\{2\}$	4	8.8	2.618
3	$\{3\}$	11	6.2	3.716
4	$\{4\}$	6	4.8	0.300
5	$\{5, \ldots, 9\}$	3	15.1	9.696
		$\sum_{i=1}^{k} o_i = n = 50$	$\sum_{i=1}^{k} E_i = n = 50$	$t = \sum_{i=1}^{k} \frac{(o_i-E_i)^2}{E_i} \simeq 24.198$

Como $t \simeq 24.199 \in W = (13.28, +\infty)$, devemos rejeitar H_0 ao n.s. de $\alpha_0 = 1\%$ [ou qualquer outro n.s. superior a α_0].

2. A perda percentual de massa (Y) de uma certa substância metálica (quando exposta a oxigénio seco a $500^o C$) depende do período de exposição (x, em hora). Cinco medições conduziram a:

$$\sum_{i=1}^5 x_i = 12, \quad \sum_{i=1}^5 x_i^2 = 32.5, \quad \sum_{i=1}^5 y_i = 0.177, \quad \sum_{i=1}^5 y_i^2 = 0.006789, \quad \sum_{i=1}^5 x_i y_i = 0.4685,$$

onde $[\min_{i=1,\dots,5} x_i, \max_{i=1,\dots,5} x_i] = [1.0, 3.5]$.

(a) Calcule as estimativas de mínimos quadrados dos parâmetros da reta de regressão linear simples de Y em x. (1.5)

- **Estimativas de MQ de β_0 e β_1**

  Dado que

  $n = 5$

  $\sum_{i=1}^n x_i = 12$
  $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{12}{5} = 2.4$
  $\sum_{i=1}^n x_i^2 = 32.5$
  $\sum_{i=1}^n x_i^2 - n(\bar{x})^2 = 32.5 - 5 \times 2.4^2 = 3.7$

  $\sum_{i=1}^n y_i = 0.177$
  $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{0.177}{5} = 0.0354$
  $\sum_{i=1}^n y_i^2 = 0.006789$
  $\sum_{i=1}^n y_i^2 - n(\bar{y})^2 = 0.006789 - 5 \times 0.0354^2 \simeq 0.000523$

  $\sum_{i=1}^n x_i y_i = 0.4685$
  $\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} = 0.4685 - 5 \times 2.4 \times 0.0354 = 0.0437,$

  as estimativas de MQ de β_1 e β_0 são, para este modelo de RLS, iguais a:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n(\bar{x})^2} \\ &= \frac{0.0437}{3.7} \\ &\simeq 0.011811 \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\ &\simeq 0.0354 - 0.011811 \times 2.4 \\ &= 0.007054 \end{aligned}$$

(b) Obtenha a estimativa de mínimos quadrados do valor esperado da perda percentual de massa quando a substância metálica é exposta por um período de 3 horas. (1.0)

- **Estimativa de MQ para $E(Y \mid x_0) = \beta_0 + \beta_1 x_0$**

$$\begin{aligned} \hat{E}(Y \mid x_0) &= \hat{\beta}_0 + \hat{\beta}_1 x_0 \\ &\simeq 0.007054 + 0.011811 \times 3 \\ &\simeq 0.042487. \end{aligned}$$

  [Não estamos a cometer qualquer erro de extrapolação ao estimar pontualmente $E(Y \mid x = 3) = \beta_0 + \beta_1 \times 3$ dado que $3 \in [\min_{i=1,\dots,5} x_i, \max_{i=1,\dots,5} x_i] = [1.0, 3.5]$.]

(c) Teste a significância do modelo de regressão linear simples, ao nível de significância de 5%. Enuncie as hipóteses de trabalho que assumir. (3.0)

- **Hipóteses de trabalho**
  No modelo de RLS, $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, consideraremos $\epsilon_i \overset{i.i.d.}{\sim} \text{Normal}(0, \sigma^2)$, $i = 1, \ldots, n$.

- **[Obs.**
  Pretende confrontar-se
  - $H_0 : \beta_1 = \beta_{1,0} = 0$ (regressão linear não é significativa, i.e., o valor esperado da variável resposta Y, $E(Y) = \beta_0 + \beta_1\, x$, não depende linearmente da variável explicativa x)
  - $H_1 : \beta_1 \neq \beta_{1,0}$ (regressão linear é significativa).]

- **Hipóteses**
  $H_0 : \beta_1 = \beta_{1,0} = 0$
  $H_1 : \beta_1 \neq \beta_{1,0}$

- **Nível de significância**
  $\alpha_0 = 5\%$

- **Estatística de teste**
  $$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)}$$

- **Região de rejeição de** H_0 (para valores de T)
  Estamos a lidar com um teste bilateral ($H_1 : \beta_1 \neq \beta_{1,0}$), pelo que a região de rejeição de H_0 é $W = (-\infty, -c) \cup (c, +\infty)$, onde
  $$\begin{aligned} c \quad : \quad & P(\text{Rejeitar } H_0 | H_0) = \alpha_0 \\ & c = F^{-1}_{t_{(n-2)}}\left(1 - \frac{\alpha_0}{2}\right) \\ & c = F^{-1}_{t_{(3)}}(0.975) \\ & c \overset{tabela/calc}{=} 3.182 \end{aligned}$$

- **Decisão**
  Tendo em conta que
  $$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2 \left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\ &= \frac{1}{5-2}\left(0.000523 - 0.011811^2 \times 3.7\right) \\ &\simeq 2.28368 \times 10^{-6}, \end{aligned}$$
  o valor observado da estatística de teste é dado por
  $$\begin{aligned} t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \\ &= \frac{0.011811 - 0}{\sqrt{\frac{2.28368 \times 10^{-6}}{3.7}}} \\ &\simeq 15.033848. \end{aligned}$$
  Como $t \simeq 15.033848 \in W = (-\infty, -3.182) \cup (3.182, +\infty)$, devemos rejeitar H_0 [hipótese de inexistência de relação de tipo linear entre o valor esperado da variável resposta Y e a variável explicativa x], ao nível de significância $\alpha_0 = 5\%$ [ou a qualquer outro n.s. superior 5%].

(d) Calcule e interprete o valor do coeficiente de determinação do modelo ajustado. (1.0)

- **Cálculo do coeficiente de determinação**

$$
\begin{aligned}
r^2 &= \frac{\left(\sum_{i=1}^{n} x_i\, y_i - n\,\bar{x}\,\bar{y}\right)^2}{\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right) \times \left(\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2\right)} \\
&= \frac{0.0437^2}{3.7 \times 0.000523} \\
&\simeq 0.987.
\end{aligned}
$$

- **Interpretação coeficiente de determinação**
  Cerca de 98.7% da variação total da perda percentual de massa é explicada pelo tempo de exposição, através do modelo de regressão linear simples considerado. Assim, podemos adiantar que a recta estimada parece ajustar-se muito bem ao conjunto de dados.