Probabilidades e Estatística

TODOS OS CURSOS

2º semestre – 2016/2017
05/07/2017 – **15:00**

Duração: 90 minutos **2º Teste C**

Justifique convenientemente todas as respostas

Grupo I	10 valores

1. Admita que a proporção de um ingrediente em determinado produto alimentar é representada pela variável aleatória X com função de densidade de probabilidade

$$f_X(x) = \begin{cases} \theta\, x^{\theta-1}, & 0<x<1 \\ 0, & \text{caso contrário,} \end{cases}$$

onde θ é um parâmetro positivo desconhecido.

(a) Mostre que o estimador de máxima verosimilhança de θ, com base numa amostra aleatória $(X_1,\ldots,X_n)$ proveniente da população X, é dado por $-\frac{n}{\sum_{i=1}^n \ln(X_i)}$. (3.0)

- **V.a. de interesse**
  X = proporção de um ingrediente em determinado produto alimentar
- **F.d.p. de** X
  $$f_X(x) = \begin{cases} \theta\, x^{\theta-1}, & 0<x<1 \\ 0, & \text{caso contrário} \end{cases}$$
- **Parâmetro desconhecido**
  $\theta, \quad \theta>0$
- **Amostra**
  $\underline{x} = (x_1,\ldots,x_n)$ amostra de dimensão n proveniente da população X [onde $0<x_i<1$, $i=1,\ldots,n$].
- **Obtenção do estimador de MV de** θ

  Passo 1 — Função de verosimilhança
  $$\begin{aligned} L(\theta|\underline{x}) &= f_{\underline{X}}(\underline{x}) \\ &\stackrel{X_i\,indep}{=} \prod_{i=1}^n f_{X_i}(x_i) \\ &\stackrel{X_i\sim X}{=} \prod_{i=1}^n f_X(x_i) \\ &= \prod_{i=1}^n \left(\theta\, x_i^{\theta-1}\right) \\ &= \theta^n \left(\prod_{i=1}^n x_i\right)^{\theta-1}, \quad \theta>0 \end{aligned}$$

  Passo 2 — Função de log-verosimilhança
  $$\ln L(\theta|\underline{x}) = n\ln(\theta) + (\theta-1)\sum_{i=1}^n \ln(x_i)$$

  Passo 3 — Maximização
  A estimativa de MV de θ passa a ser representada por $\hat\theta$ e
  $$\hat\theta \;:\; \begin{cases} \left.\frac{d\ln L(\theta|\underline{x})}{d\theta}\right|_{\theta=\hat\theta} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2\ln L(\theta|\underline{x})}{d\theta^2}\right|_{\theta=\hat\theta} < 0 & \text{(ponto de máximo)} \end{cases}$$

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

$$\begin{cases} \hat\theta = -\frac{n}{\sum_{i=1}^n \ln(x_i)} & \text{[uma vez que } 0 < x_i < 1\ (i = 1,\dots,n) \text{ tem-se } \sum_{i=1}^n \ln(x_i) < 0, \\ & \text{logo } \hat\theta \text{ está bem definido]} \\ -\frac{\left[\sum_{i=1}^n \ln(x_i)\right]^2}{n} < 0 & \text{(proposição verdadeira porque } \sum_{i=1}^n \ln(x_i) < 0). \end{cases}$$

Passo 4 — Estimador de MV de θ

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

(b) Determine a estimativa de máxima verosimilhança de $E(X) = \frac{\theta}{\theta+1}$ baseada na amostra $(x_1, x_2, x_3, x_4, x_5) = (0.42, 0.31, 0.53, 0.05, 0.63)$ para a qual $\sum_{i=1}^5 \ln(x_i) \simeq -6.131$. (1.5)

- **Estimativa de MV de θ**

$$\begin{aligned} \hat\theta &= -\frac{n}{\sum_{i=1}^n \ln(x_i)} \\ &\simeq -\frac{5}{-6.131} \\ &\simeq 0.815528 \end{aligned}$$

- **Outro parâmetro desconhecido**

$$\begin{aligned} h(\theta) &= E(X) \\ &= \frac{\theta}{\theta+1} \end{aligned}$$

- **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 dada por

$$\begin{aligned} \widehat{h(\theta)} &= h(\hat\theta) \\ &= \frac{\hat\theta}{\hat\theta+1} \\ &= \frac{0.815528}{0.815528+1} \\ &\simeq 0.449196. \end{aligned}$$

2. A massa (X, em grama) de um dado tipo de peça possui distribuição normal, sendo os respectivos valor esperado e variância desconhecidos. Sabendo que a concretização $(x_1, \dots, x_{10})$ de uma amostra aleatória proveniente da população X conduziu a $\sum_{i=1}^{10} x_i = 293$ e $\sum_{i=1}^{10} x_i^2 = 8\,745$:

(a) Construa um intervalo de confiança a 90% para o desvio padrão da massa de uma peça do tipo referido. (2.5)

- **V.a. de interesse**
  X = massa da peça

- **Situação**
  $X \sim \text{normal}(\mu, \sigma^2)$
  μ desconhecido
  σ^2 DESCONHECIDO

- **Obtenção do IC para σ**

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

$$Z = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

[uma vez que é suposto determinar um IC para a variância de uma população normal, com valor esperado desconhecido.]

Passo 2 — Obtenção dos quantis de probabilidade
Ao ter-se em consideração que $n = 10$ e $(1-\alpha)\times 100\% = 90\%$, far-se-á uso dos quantis

$$(a_\alpha, b_\alpha) \quad : \quad \begin{cases} P(a_\alpha \le Z \le b_\alpha) = 1-\alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2. \end{cases}$$

$$\begin{cases} a_\alpha = F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) = F^{-1}_{\chi^2_{(9)}}(0.05) \stackrel{tabela/calc.}{=} 3.325 \\ b_\alpha = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2) = F^{-1}_{\chi^2_{(9)}}(0.95) \stackrel{tabela/calc.}{=} 16.92. \end{cases}$$

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

$$P(a_\alpha \le Z \le b_\alpha) = 1-\alpha$$

$$P\left[a_\alpha \le \tfrac{(n-1)S^2}{\sigma^2} \le b_\alpha\right] = 1-\alpha$$

$$P\left[\tfrac{1}{b_\alpha} \le \tfrac{\sigma^2}{(n-1)S^2} \le \tfrac{1}{a_\alpha}\right] = 1-\alpha$$

$$P\left[\sqrt{\tfrac{(n-1)S^2}{b_\alpha}} \le \sigma \le \sqrt{\tfrac{(n-1)S^2}{a_\alpha}}\right] = 1-\alpha$$

Passo 4 — Concretização
Tendo em conta o par de quantis acima e o facto de

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left[\sum_{i=1}^{10} x_i^2 - n\,(\bar{x})^2\right] \\ &= \frac{1}{10-1}\left(8745 - 10\times 29.3^2\right) \\ &= 17.7(8) \end{aligned}$$

$$IC_{(1-\alpha)\times 100\%}(\sigma) = \left[\sqrt{\frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2)}}, \sqrt{\frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(\alpha/2)}}\right]$$

temos:

$$\begin{aligned} IC_{90\%}(\sigma) &= \left[\sqrt{\frac{(10-1)\times 17.7(8)}{16.92}}, \sqrt{\frac{(10-1)\times 17.7(8)}{3.325}}\right] \\ &\simeq [\sqrt{9.4627}, \sqrt{48.1487}] \\ &\simeq [3.0762, 6.9389]. \end{aligned}$$

(b) Avalie se os dados apoiam a hipótese de que o valor esperado da massa desse tipo de peças é igual a 27 g. Decida com base no valor-p. (3.0)

- **Hipóteses**
  $H_0 : \mu = \mu_0 = 27$
  $H_1 : \mu \neq \mu_0$

- **Estatística de teste**

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim_{H_0} t_{(n-1)}$$

  [pois pretendemos efectuar um teste sobre o valor esperado de uma população normal, com variância desconhecida.]

- **Região de rejeição de H_0** (para valores de T)
  Tratando-se de um teste biilateral ($H_1 : \mu \neq \mu_0$), a região de rejeição de H_0 (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 dado por
$$\begin{aligned} t &= \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} \\ &= \frac{29.3-27}{\sqrt{\frac{17.7(8)}{10}}} \\ &\simeq 1.724461. \end{aligned}$$
  Dado que a região de rejeição deste teste é uma reunião de intervalos simétricos, temos:
$$\begin{aligned} valor-p &= P(|T|>|t| \mid H_0) \\ &= 2\times[1-F_{t_{(n-1)}}(|t|)] \\ &= 2\times[1-F_{t_{(9)}}(1.724461)] \\ &\stackrel{calc.}{=} 0.118712. \end{aligned}$$
  Consequentemente, é suposto:
  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 11.8712\%$, pelo que $H_0 : \mu = \mu_0 = 27$ é consistente a qualquer dos n.u.s. $(1\%, 5\%, 10\%)$;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > 11.8712\%$.

  [**Decisão (com base em intervalo para o valor-p)**
  Recorrendo às tabelas de quantis da distribuição de t-student podemos adiantar um intervalo para o valor-p:
$$\begin{aligned} F^{-1}_{t_{(9)}}(0.925) = 1.574 \quad &< \quad 1.724461 \quad < \quad 1.833 = F^{-1}_{t_{(9)}}(0.95) \\ 0.10 = 2\times(1-0.95) \quad &< \quad valor-p = 2\times[1-F_{t_{(9)}}(1.724461)] \quad < \quad 2\times(1-0.925) = 0.15. \end{aligned}$$
  Consequentemente:
  - não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 10\%$, nomeadamente aos n.u.s. $(1\%, 5\%, 10\%)$;
  - devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 15\%$.]

Grupo II — 10 valores

1. Numa experiência química, registou-se o índice de ferro numa determinada solução em 100 unidades experimentais escolhidas aleatoriamente, tendo-se obtido os seguintes dados:

Classe	≤ 0.7	$]0.7, 0.9]$	$]0.9, 1.1]$	$]1.1, 1.3]$	> 1.3
Frequência absoluta observada	4	31	35	26	4
Frequência absoluta esperada	6.68	24.17	38.30	E_4	E_5

Um engenheiro químico conjetura que o índice de ferro na solução segue uma distribuição normal com valor esperado e desvio padrão iguais a 1.0 e 0.2 (respectivamente).

(a) Calcule os valores das frequências absolutas esperadas E_4 e E_5 (aproximando-os às centésimas). (1.0)

- **V.a. de interesse**
  X = índice de ferro na solução

- **Dist. conjecturada**
  $\text{Normal}(1, 0.2^2)$

- **Frequências absolutas esperadas omissas**
  Atendendo à dimensão da amostra $n = 100$ e à dist. conjecturada, segue-se:
$$\begin{aligned} E_4 &= 100\times\left[\Phi\left(\frac{1.3-1}{0.2}\right)-\Phi\left(\frac{1.1-1}{0.2}\right)\right] \\ &= 100\times[\Phi(1.5)-\Phi(0.5)] \\ &\stackrel{tabela/calc.}{=} 100\times(0.9332-0.6915) \\ &= 24.17 \end{aligned}$$

$$
\begin{aligned}
E_5 &= n - \sum_{i=1}^{4} E_i \\
&= 100-(6.68+24.17+38.30+24.17) \\
&= 6.68.
\end{aligned}
$$

(b) Teste a hipótese do engenheiro químico, ao nível de significância de 5%. (3.0)

- **Hipóteses**
  $H_0 : X \sim \text{Normal}(1, 0.2^2)$
  $H_1 : X \not\sim \text{Normal}(1, 0.2^2)$
- **Nível de significância**
  $\alpha_0 = 5\%$
- **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 = ESTIMADOR da frequência absoluta esperada, sob H_0, da classe i
  β = No. de parâmetros a estimar = 0 [dado que em H_0 se conjectura uma distribuição específica.]
- **Frequências absolutas esperadas sob H_0**
  De acordo com a tabela facultada e (a), os valores das freq. absolutas esperadas sob H_0 aproximados às centésimas são: $E_1 \simeq 6.68$; $E_2 \simeq 24.17$; $E_3 \simeq 38.30$; $E_4 \simeq 24.17$; $E_5 \simeq 6.68$.

  [Constata-se 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 $c = F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0)$ 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
  $$
  \begin{aligned}
  c &= F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) \\
  &= F^{-1}_{\chi^2_{(5-0-1)}}(1-0.05) \\
  &= F^{-1}_{\chi^2_{(4)}}(0.95) \\
  &\overset{tabela/calc.}{=} 9.488.
  \end{aligned}
  $$
- **Decisão**
  No cálculo do valor obs. da estat. de teste convém recorrer à seguinte tabela auxiliar:

i	Classe i	Freq. abs. obs. o_i	Estim. freq. abs. esp. sob H_0 $e_i = n \times \hat{p}_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-e_i)^2}{e_i}$
1	$]-\infty, 0.7]$	4	6.68	$\frac{(4-6.68)^2}{6.68} \simeq 1.075$
2	$]0.7, 0.9]$	31	24.17	1.930
3	$]0.9, 1.1]$	35	38.30	0.284
4	$]1.1, 1.3]$	26	24.17	0.139
5	$]1.3, +\infty[$	4	6.68	1.075
		$\sum_{i=1}^{k} o_i = n$ $= 100$	$\sum_{i=1}^{k} e_i = n$ $= 100$	$t = \sum_{i=1}^{k} \frac{(o_i-e_i)^2}{e_i}$ $\simeq 4.503$

  Como $t \simeq 4.503 \not\in W = (9.488, +\infty)$, não devemos rejeitar H_0 ao n.s. de $\alpha_0 = 5\%$ [nem a qualquer outro n.s. inferior a α_0].

2. Por forma a analisar a relação entre a percentagem de humidade relativa no local de armazenagem (x) e o teor de humidade de fibra sintética aí armazenada (Y), foram obtidos os seguintes dados em 10 localizações distintas:

$$\sum_{i=1}^{10} x_i = 4.50\,, \quad \sum_{i=1}^{10} x_i^2 = 2.1062\,, \quad \sum_{i=1}^{10} y_i = 1.18\,, \quad \sum_{i=1}^{10} y_i^2 = 0.1470\,, \quad \sum_{i=1}^{10} x_i y_i = 0.5551\,,$$

onde $[\min_{i=1,\dots,10} x_i, \max_{i=1,\dots,10} x_i] = [0.29, 0.61]$.

Nota: Recorra a pelo menos seis casas decimais em todos os cálculos que efetuar nesta questão.

(a) Após ter enunciado as hipóteses de trabalho que entender convenientes, calcule as estimativas de máxima verosimilhança dos parâmetros da reta de regressão linear simples de Y em x. (2.0)

- **Hipóteses de trabalho**
  No modelo de RLS, $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, consideraremos $\epsilon_i \overset{i.i.d.}{\sim} \text{Normal}(0, \sigma^2)$, $i = 1, \dots, n$.
- **Estimativas de MV de β_0 e β_1**
  Dado que

  $n = 10$

  $\sum_{i=1}^n x_i = 4.50$
  $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{4.50}{10} = 0.45$
  $\sum_{i=1}^n x_i^2 = 2.1062$
  $\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2 = 2.1062 - 10 \times 0.45^2 = 0.0812$

  $\sum_{i=1}^n y_i = 1.18$
  $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{1.18}{10} = 0.118$
  $\sum_{i=1}^n y_i^2 = 0.1470$
  $\sum_{i=1}^n y_i^2 - n\,(\bar{y})^2 = 0.1470 - 10 \times 0.118^2 = 0.007760$

  $\sum_{i=1}^n x_i y_i = 17931.50$
  $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 0.5551 - 10 \times 0.45 \times 0.118 = 0.0241,$

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

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\,(\bar{x})^2} \\
&= \frac{0.0241}{0.0812} \\
&\simeq 0.296798 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&\simeq 0.118 - 0.296798 \times 0.45 \\
&\simeq -0.015559.
\end{aligned}$$

(b) Obtenha a estimativa de máxima verosimilhança do valor esperado do teor de humidade de fibra sintética quando armazenada num local com humidade relativa de 0.35. (1.0)

- **Estimativa de MV para $E(Y \mid x_0) = \beta_0 + \beta_1 x_0$**
  [Invocando a propriedade de invariância dos estimadores de MV, tem-se:]

$$\begin{aligned}
\hat{E}(Y \mid x_0) &= \hat{\beta}_0 + \hat{\beta}_1 x_0 \\
&\simeq -0.015559 + 0.296798 \times 0.35 \\
&\simeq 0.088320.
\end{aligned}$$

  [Não estamos a cometer qualquer erro de extrapolação ao estimar pontualmente $E(Y \mid x = 0.35) = \beta_0 + \beta_1 \times 0.35$ dado que $0.35 \in [\min_{i=1,\dots,10} x_i, \max_{i=1,\dots,10} x_i] = [0.29, 0.61]$.]

(c) Deduza um intervalo de confiança a 95% para o declive da reta de regressão linear simples de Y em x. O que conclui sobre a significância do modelo de regressão linear simples de Y em x, ao nível de (3.0) significância de 5% ?

- **Hipóteses de trabalho**

  $\epsilon_i \overset{i.i.d.}{\sim} \text{Normal}(0,\sigma^2),\ i=1,\dots,n$

  β_0, β_1 e σ^2 DESCONHECIDOS

- **Obtenção de IC para β_1**

  Passo 1 — Selecção da v.a. fulcral para β_1

  $$Z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}} \sim t_{(n-2)}$$

  Passo 2 — Obtenção dos quantis de probabilidade

  Neste caso $n = 10$ e $(1-\alpha)\times 100\% = 95\%$, logo usaremos os quantis de probabilidade

  $$(a_\alpha, b_\alpha) \quad : \quad \begin{cases} P(a_\alpha \le Z \le b_\alpha) = 1-\alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2. \end{cases}$$

  $$\begin{cases} a_\alpha = -F^{-1}_{t_{(n-2)}}(1-\alpha/2) = -F^{-1}_{t_{(8)}}(0.975) \overset{tabela/calc.}{=} -2.306 \\ b_\alpha = F^{-1}_{t_{(n-2)}}(1-\alpha/2) = F^{-1}_{t_{(8)}}(0.975) \overset{tabela/calc.}{=} 2.306. \end{cases}$$

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

  $P\,(a_\alpha \le Z \le b_\alpha) = 1-\alpha$

  $\dots$

  $$P\left[\hat{\beta}_1 - F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}} \le \beta_1 \le \hat{\beta}_1 + F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}\right] = 1-\alpha.$$

  Passo 4 — Concretização

  Atente-se que

  $$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\bar{y}^2\right) - \left(\hat{\beta}_1\right)^2\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right)\right] \\ &= \frac{1}{10-2}\left[0.007760 - 0.296798^2 \times 0.0812\right] \\ &= 0.000076 \end{aligned}$$

  $$IC_{(1-\alpha)\times 100\%}(\beta_1) = \left[\hat{\beta}_1 \pm F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}\right].$$

  Logo

  $$\begin{aligned} IC_{95\%}(\beta_1) &\simeq \left[0.296798 \pm 2.306\times\sqrt{\frac{0.000076}{0.0812}}\right] \\ &\simeq [0.296798 \pm 0.070549]] \\ &= [0.226249, 0.367347]. \end{aligned}$$

- **Teste de significância do modelo de RLS**

  Hipóteses

  $H_0 : \beta_1 = \beta_{1,0} = 0$

  $H_1 : \beta_1 \neq \beta_{1,0}$

  N.s.

  $\alpha_0 = 0.05$

  Decisão

  Invocando a relação entre intervalos de confiança e testes de hipóteses, podemos afirmar que

  - o valor conjecturado para β_1 em H_0 é

    $$\begin{aligned} \beta_{1,0} &= 0 \\ &\not\in IC_{95\%}(\beta_1) = [0.226249, 0.367347], \end{aligned}$$

  - logo a hipótese $H_0 : \beta_1 = \beta_{1,0} = 0$ deve ser rejeitada ao nível de significância $\alpha = 5\%$ [ou a qualquer n.s. $\alpha_0 > 5\%$].