FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA
Departamento de Matemática

IPEIO - Probabilidades e Estatística
Ano Lectivo 2017/18
2º Teste - 9 de maio de 2018
Duração: 0h45

Resolução abreviada do 2º Teste

Versão 1

1. Considere a amostra aleatória $(X_1, X_2, \ldots, X_n)$ de uma população com distribuição $N(2\mu; 1)$.

(0.4) (a) O estimador $\hat{\mu} = \frac{\overline{X}}{2}$ é centrado para μ.

(0.4) (b) Dada a amostra (0.5 , 3.2 , 1.8 , 1.7 , 2.8), uma estimativa pontual de μ resultante de $\hat{\mu} = \frac{\overline{X}}{2}$, é?

(A) 1 (B) 2 (C) 3 (D) 4 (E) Nenhuma das anteriores

$\boxed{\text{V}}$ Considere-se a população X de onde é proveniente a amostra aleatória. Como $X \sim N(2\mu; 1)$ então sabe-se que $E[X] = 2\mu$.
O estimador $\hat{\mu}$ será estimador centrado para μ se $E[\hat{\mu}] = \mu$. Como,

$$E[\hat{\mu}] = E\left[\frac{\bar{X}}{2}\right] = \frac{E[\bar{X}]}{2} = \frac{E[X]}{2} = \frac{2\mu}{2} = \mu,$$

uma vez que é conhecido que $E[\bar{X}] = E[X]$ (ver Exemplo 5.11, pag. 39 do texto de apoio), pelo que o estimador $\hat{\mu}$ é centrado.

$\boxed{\text{A}}$ Para a amostra considerada, teremos que $\bar{x} = \dfrac{0.5 + 3.2 + 1.8 + 1.7 + 2.8}{5} = 2$, logo obteremos como estimativa de μ, o valor

$$\hat{\mu} = \frac{\bar{x}}{2} = \frac{2}{2} = 1.$$

2. Considere-se uma população com distribuição normal de variância 36. Recolhida uma amostra de dimensão $n = 25$ dessa população, obteve-se $\bar{x} = 50$.

(0.4) (a) O intervalo de 95% de confiança para o valor médio da população é (com valores arredondados a 3 casas decimais):

(A) $]48.026\,;\,51.974[$ (B) $]47.648\,;\,52.352[$ (C) $]47.947\,;\,52.053[$ (D) $]47.523\,;\,52.477[$

(0.4) (b) Qual deve ser a dimensão da amostra para que a amplitude do intervalo de 95% de confiança para o valor médio da população seja inferior a 2:

(A) $n = 54$ (B) $n = 98$ (C) $n = 106$ (D) $n = 139$ (E) Nenhuma das anteriores

$\boxed{\text{B}}$ Represente-se por X a população.
Informação populacional: $X \sim N\left(\mu, \sigma^2\right)$, $\mu \equiv E\left(X\right) = ?$, $\sigma^2 \equiv V\left(X\right) = 36$
Informação amostral: $n = 25$, $\bar{x} = 50$

- Estatística pivot: $Z = \sqrt{n}\dfrac{\overline{X} - \mu}{\sigma} \sim N(0, 1)$
- Determinação da constante $\underline{c}$ que garante que $P\left(-c < Z < c\right) = 0.95$
  $c = z_{0.05/2} = z_{0.025} = 1.96$
- $-1.96 < \sqrt{n}\dfrac{\overline{X} - \mu}{\sigma} < 1.96 \Leftrightarrow \overline{X} - 1.96\dfrac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\dfrac{\sigma}{\sqrt{n}}$
- $IC_{95\%}\left(\mu\right) \equiv \left]\overline{X} - 1.96\dfrac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\dfrac{\sigma}{\sqrt{n}}\right[$
- Estimativa por intervalo de 95% de confiança para μ
  $IC_{95\%}\left(\mu\right) = \left]50 - 1.96\dfrac{6}{5}, 50 + 1.96\dfrac{6}{5}\right[=]47.648; 52.352[$

$\boxed{\text{D}}$ Como a amplitude do intervalo de confiança a 95% é dada por $2 \times 1.96 \times \frac{6}{\sqrt{n}}$ então queremos determinar n tal que:

$$2 \times 1.96 \times \frac{6}{\sqrt{n}} < 2 \Leftrightarrow 1.96 \times 6 < \sqrt{n} \Rightarrow n > (1.96 \times 6)^2 \Rightarrow n > 138.2976 \Rightarrow n = 139$$

3. As classificações do 2ºteste de IPEIO têm distribuição normal de valor médio desconhecido. Recolhida uma amostra de dimensão $n = 20$, obteve-se $s = 1.5$.

(0.4) (a) Para o teste de hipóteses $H_0 : \sigma \geq 2$ vs $H_1 : \sigma < 2$, a região de rejeição para um nível de 5% de significância é:

(A) $R_{0.05} =]30.143; +\infty[$ (B) $R_{0.05} =]0; 30.143[$ (C) $R_{0.05} =]0; 10.117[$ (D) $R_{0.05} =]10.117; +\infty[$

(0.4) (b) Se num determinado teste de hipóteses a decisão é de rejeitar a hipótese nula para um nível de significância $\alpha = 5\%$, então também se rejeita a hipótese nula para um nível de significância $\alpha = 10\%$.

$\boxed{\text{C}}$ A estatística de teste é:

$$X^2 = \frac{(n-1)S^2}{2^2} \underset{sob\ H_0}{\sim} \chi^2_{n-1}.$$

Como o teste é unilateral esquerdo, e $n = 20$ a região de rejeição é $R_{0.05} =]0, x^2_{19;0.95}[$ com $x^2_{19;0.95} = 10.117$.

$\boxed{\text{V}}$ Se a hipótese nula é rejeitada para um nível de significância de $\alpha = 5\%$, então o valor observado da estatística encontra-se nessa região de rejeição ($R_{0.05}$). Como, mantendo todas as outras condições, teremos que $R_{0.05} \subseteq R_{0.10}$ então o valor observado da estatística encontra-se também na região de rejeição de 10%, pelo que se rejeita a hipótese nula para um nível de significância $\alpha = 10\%$.

4. Num determinado curso de àgua, pretende-se modelar a concentração Y de um certo poluente (em gr/m^3), em função da distância x à fonte poluidora, em Km. Para tal, registaram-se os dados relativos a 15 localizações.

Distância, x	1	2	3	4	6	8	10	12	14	16	18	20	22	25	30
Concentração, Y	53.4	46.2	48.6	43.5	44.8	42.0	41.4	37.6	39.2	33.2	34.4	29.0	26.1	24.2	15.2

Resolva as questões com base nos resultados do R:

Coefficients:

	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	51.5511	1.0028	51.41	2e-16	***
x	-1.1229	0.0651	-17.25	2.44e-10	***

Residual standard error: 2.185 on 13 degrees of freedom
Multiple R-squared: 0.9581, Adjusted R-squared: 0.9549

Assumindo que existe uma relação linear entre as variáveis x e Y:

(0.4) (a) Escreva a expressão da reta de regressão linear estimada e comente a qualidade do ajustamento.

(0.4) (b) Qual o valor estimado da variância dos erros do modelo de regressão linear simples?

(0.4) (c) Qual prevê que seja a concentração de poluente a uma distância da fonte de poluição de $15Km$? E a uma distância de $40Km$?

(0.4) (d) Teste para um nível de significância de 5%, a hipótese de o verdadeiro declive da recta de regressão ser nulo, indicando:

- Hipóteses:
- Decisão(justifique):

(a) A reta estimada é:

$$\hat{Y} = 51.5511 - 1.1229x$$

e como o coeficiente de determinação vale

$$R^2 = 0.9581 \geq 0.8,$$

a qualidade do ajustamento é "‘razoável”’.

(b) O valor estimado da variância dos erros do modelo de regressão linear, $\hat{\sigma}^2 = 2.185^2 = 4.774225$

(c) Para $x = 15Km$ temos o valor previsto de concentração de poluente:

$$\hat{Y}(15) = 51.5511 - 1.1229 \times 15 = 34.7076.$$

Para $x = 40Km$ não se pode fazer previsão pois $40 \notin [\min(x_i), \max(x_i)] \equiv [1, 30]$ e nada sabemos sobre o ajustamento do modelo fora deste intervalo.

(d)

- Como o verdadeiro declive da reta de regressão é β_1, temos as hipóteses: $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$.
- Utilizando o $valor - p$, que para este teste vale $valor - p = 2.44 \times 10^{-10}$ e ao ser menor que o nível de significância de 5%, permite-nos rejeitar H_0 ao nível de significância de 5%.

Observação: Tomamos a mesma decisão se realizarmos o teste seguindo os passos tradicionais.

- Região de rejeição para $\alpha = 0.05$: $t_{13:\,0.025} = 2.160$ e $R_{0.05} =]-\infty, -2.160[\cup]2.160, +\infty[$
- Valor observado da estatística de teste: $t_{obs} = -17.25$, dado pelo *output* do R
- Decisão: $t_{obs} \in R_{0.05}$ logo rejeitamos H_0 ao nível de significância de 5%.

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA
Departamento de Matemática

IPEIO - Probabilidades e Estatística
Ano Lectivo 2017/18
2º Teste - 9 de maio de 2018
Duração: 0h45

Resolução abreviada do 2º Teste

Versão 2

1. Considere a amostra aleatória $(X_1, X_2, \ldots, X_n)$ de uma população com distribuição $N(4\mu; 16)$.

(0.4) (a) O estimador $\hat{\mu} = \frac{\overline{X}}{4}$ é centrado para μ.

(0.4) (b) Dada a amostra (2.0 , 12.8 , 7.2 , 6.8 , 11.2), uma estimativa pontual de μ resultante de $\hat{\mu} = \frac{\overline{X}}{4}$, é?

(A) 1 (B) 2 (C) 3 (D) 4 (E) Nenhuma das anteriores

$\boxed{\text{V}}$ Considere-se a população X de onde é proveniente a amostra aleatória. Como $X \sim N(4\mu; 1)$ então sabe-se que $E[X] = 4\mu$.
O estimador $\hat{\mu}$ será estimador centrado para μ se $E[\hat{\mu}] = \mu$. Como,

$$E[\hat{\mu}] = E\left[\frac{\bar{X}}{4}\right] = \frac{E[\bar{X}]}{4} = \frac{E[X]}{4} = \frac{4\mu}{4} = \mu,$$

uma vez que é conhecido que $E[\bar{X}] = E[X]$ (ver Exemplo 5.11, pag. 39 do texto de apoio), pelo que o estimador $\hat{\mu}$ é centrado.

$\boxed{\text{B}}$ Para a amostra considerada, teremos que $\bar{x} = \dfrac{2.0 + 12.8 + 7.2 + 6.8 + 11.2}{5} = 8$, logo obteremos como estimativa de μ, o valor

$$\hat{\mu} = \frac{\bar{x}}{4} = \frac{8}{4} = 2.$$

2. Considere-se uma população com distribuição normal de variância 25. Recolhida uma amostra de dimensão $n = 25$ dessa população, obteve-se $\bar{x} = 50$.

(0.4) (a) O intervalo de 95% de confiança para o valor médio da população é (com valores arredondados a 3 casas decimais):

(A) $]47.936\,;\,52.064[$ (B) $]48.289\,;\,51.711[$ (C) $]48.355\,;\,51.645[$ (D) $]48.040\,;\,51.960[$

(0.4) (b) Qual deve ser a dimensão da amostra para que a amplitude do intervalo de 95% de confiança para o valor médio da população seja inferior a 2:

(A) $n = 54$ (B) $n = 68$ (C) $n = 74$ (D) $n = 97$ (E) Nenhuma das anteriores

$\boxed{\text{D}}$ Represente-se por X a população.
Informação populacional: $X \sim N\left(\mu, \sigma^2\right)$, $\mu \equiv E\left(X\right) = ?$, $\sigma^2 \equiv V\left(X\right) = 25$
Informação amostral: $n = 25$, $\bar{x} = 50$

- Estatística pivot: $Z = \sqrt{n}\dfrac{\overline{X} - \mu}{\sigma} \sim N(0, 1)$
- Determinação da constante $\underline{c}$ que garante que $P\left(-c < Z < c\right) = 0.95$
  $c = z_{0.05/2} = z_{0.025} = 1.96$
- $-1.96 < \sqrt{n}\dfrac{\overline{X} - \mu}{\sigma} < 1.96 \Leftrightarrow \overline{X} - 1.96\dfrac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\dfrac{\sigma}{\sqrt{n}}$
- $IC_{95\%}\left(\mu\right) \equiv \left]\overline{X} - 1.96\dfrac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\dfrac{\sigma}{\sqrt{n}}\right[$
- Estimativa por intervalo de 95% de confiança para μ
  $IC_{95\%}\left(\mu\right) = \left]50 - 1.96\dfrac{5}{5}, 50 + 1.96\dfrac{5}{5}\right[=]48.040; 51.960[$

$\boxed{\text{D}}$ Como a amplitude do intervalo de confiança a 95% é dada por $2 \times 1.96 \times \frac{5}{\sqrt{n}}$ então queremos determinar n tal que:

$$2 \times 1.96 \times \frac{5}{\sqrt{n}} < 2 \Leftrightarrow 1.96 \times 5 < \sqrt{n} \Rightarrow n > (1.96 \times 5)^2 \Rightarrow n > 96.04 \Rightarrow n = 97$$

3. As classificações do 2ºteste de IPEIO têm distribuição normal de valor médio desconhecido. Recolhida uma amostra de dimensão $n = 20$, obteve-se $s = 3.5$.

(0.4) (a) Para o teste de hipóteses $H_0 : \sigma \leq 3 \;\; vs \;\; H_1 : \sigma > 3$, a região de rejeição para um nível de 5% de significância é:

(A) $R_{0.05} =]30.143; +\infty[$ (B) $R_{0.05} =]0; 30.143[$ (C) $R_{0.05} =]0; 10.117[$ (D) $R_{0.05} =]10.117; +\infty[$

(0.4) (b) Se num determinado teste de hipóteses a decisão é de não rejeitar a hipótese nula para um nível de significância $\alpha = 5\%$, então também não se rejeita a hipótese nula para um nível de significância $\alpha = 1\%$.

$\boxed{\text{A}}$ A estatística de teste é:

$$X^2 = \frac{(n-1)S^2}{3^2} \underset{sob\ H_0}{\sim} \chi^2_{n-1}.$$

Como o teste é unilateral direito, e $n = 20$ a região de rejeição é $R_{0.05} =]x^2_{19;0.05}, +\infty[$ com $x^2_{19;0.05} = 30.143$.

$\boxed{\text{V}}$ Se a hipótese nula não é rejeitada para um nível de significância de $\alpha = 5\%$, então o valor observado da estatística encontra-se fora da região de rejeição ($R_{0.05}$). Como, mantendo todas as outras condições, teremos que $R_{0.01} \subseteq R_{0.05}$ então o valor observado da estatística encontra-se também fora da região de rejeição de 1%, pelo que não se rejeita a hipótese nula para um nível de significância $\alpha = 1\%$.

4. Num determinado curso de àgua, pretende-se modelar a concentração Y de um certo poluente (em gr/m^3), em função da distância x à fonte poluidora, em Km. Para tal, registaram-se os dados relativos a 15 localizações.

Distância, x	1	2	3	4	6	8	10	12	14	16	18	20	22	25	30
Concentração, Y	53.4	46.2	48.6	43.5	44.8	42.0	41.4	37.6	39.2	33.2	34.4	29.0	26.1	24.2	15.2

Resolva as questões com base nos resultados do R:

Coefficients:

	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	51.5511	1.0028	51.41	2e-16	***
x	-1.1229	0.0651	-17.25	2.44e-10	***

Residual standard error: 2.185 on 13 degrees of freedom
Multiple R-squared: 0.9581, Adjusted R-squared: 0.9549

Assumindo que existe uma relação linear entre as variáveis x e Y:

(0.4) (a) Escreva a expressão da reta de regressão linear estimada e comente a qualidade do ajustamento.

(0.4) (b) Qual o valor estimado da variância dos erros do modelo de regressão linear simples?

(0.4) (c) Qual prevê que seja a concentração de poluente a uma distância da fonte de poluição de $15Km$? E a uma distância de $40Km$?

(0.4) (d) Teste para um nível de significância de 5%, a hipótese de o verdadeiro declive da recta de regressão ser nulo, indicando:

- Hipóteses:
- Decisão(justifique):

(a) A reta estimada é:

$$\hat{Y} = 51.5511 - 1.1229x$$

e como o coeficiente de determinação vale

$$R^2 = 0.9581 \geq 0.8,$$

a qualidade do ajustamento é "'razoável"'.

(b) O valor estimado da variância dos erros do modelo de regressão linear, $\hat{\sigma}^2 = 2.185^2 = 4.774225$

(c) Para $x = 15Km$ temos o valor previsto de concentração de poluente:

$$\hat{Y}(15) = 51.5511 - 1.1229 \times 15 = 34.7076.$$

Para $x = 40Km$ não se pode fazer previsão pois $40 \notin [\min(x_i), \max(x_i)] \equiv [1, 30]$ e nada sabemos sobre o ajustamento do modelo fora deste intervalo.

(d)
- Como o verdadeiro declive da reta de regressão é β_1, temos as hipóteses: $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$.
- Utilizando o $valor - p$, que para este teste vale $valor - p = 2.44 \times 10^{-10}$ e ao ser menor que o nível de significância de 5%, permite-nos rejeitar H_0 ao nível de significância de 5%.

Observação: Tomamos a mesma decisão se realizarmos o teste seguindo os passos tradicionais.
- Região de rejeição para $\alpha = 0.05$: $t_{13:\,0.025} = 2.160$ e $R_{0.05} =]-\infty, -2.160[\cup]2.160, +\infty[$
- Valor observado da estatística de teste: $t_{obs} = -17.25$, dado pelo *output* do R
- Decisão: $t_{obs} \in R_{0.05}$ logo rejeitamos H_0 ao nível de significância de 5%.