

Universidade Federal de Lavras
Departamento de Estatística
Prof. Daniel Furtado Ferreira

8ª Lista de Exercícios **Teoria da Estimação: médias de uma população qualquer**

1) Marcar como verdadeira ou falsa as seguintes afirmativas:

a) () O intervalo de confiança para média de uma normal, dado por:

$$IC_{1-\alpha}(\mu) : \left[\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right],$$

é pouco útil, embora seja válido, pois exige o conhecimento do desvio padrão populacional σ , o que em situações reais é muito difícil de acontecer.

b) () O intervalo de confiança para μ dado por

$$IC_{1-\alpha}(\mu) : \bar{X} \pm t_{\alpha/2; \nu=n-1} \frac{S}{\sqrt{n}} = \bar{X} \pm e, \quad (1)$$

é válido sob normalidade da amostra aleatória, independência das observações amostrais entre si, garantida pela casualização ou sorteio e distribuições idênticas entre os diferentes X_j 's, $j = 1, 2, \dots, n$.

c) () Quanto maior for o coeficiente de confiança $1 - \alpha$, menor é o comprimento do intervalo de confiança, para uma mesma amostra.

d) () Quanto menor for o valor de α , maior é comprimento do intervalo de confiança, para uma mesma amostra.

e) () Uma alternativa para se obter um intervalo de confiança mais estreito (menor comprimento), quando se pretende aumentar a confiança é aumentar o tamanho da amostra n .

f) () A violação do pressuposto de normalidade no intervalo de confiança (1) provoca um aumento da probabilidade de cobertura real em relação ao valor nominal de $1 - \alpha$ fixado *a priori*.

g) () Se a população não for normal, a aplicação do intervalo (1) causa uma redução do valor da probabilidade de cobertura real em relação ao valor nominal de $1 - \alpha$.

h) () Não existem métodos exatos para se obter intervalos de confiança exatos para médias das distribuições exponenciais e Poisson.

i) () Em populações discretas, como a Binomial e Poisson, quando dizemos que os intervalos de confiança para suas médias são exatos, significa que a cobertura real dos intervalos é de **pelo menos**. $100(1 - \alpha)\%$.

j) () O método *bootstrap* para estimar por intervalo médias de um determinada população pode ser aplicado para qualquer distribuição.

k) () O método *bootstrap* é um procedimento simples em que a amostra original é reamostrada um grande número de vezes e cada reamostragem tem o tamanho que é o dobro do tamanho n da amostra original.

l) () Um dos procedimentos mais simples do método *bootstrap* é reamostrar a amostra original formando amostras do mesmo tamanho um grande número de vezes (milhares de vezes), ordenar os valores obtidos e tomar os quantis $100\alpha/2\%$ e $100(1 - \alpha/2)\%$ para serem os limites inferior (LI) e superior (LS), respectivamente, do intervalo de $100(1 - \alpha)\%$ de confiança.

m) () Outro método *bootstrap* para obter intervalos de confiança chamado método *bootstrap* com correção de viés é idêntico ao processo descrito no item 1l, exceto pelo fato de os quantis inferior e superior para determinar os LI e LS são corrigidos para contemplar assimetrias na distribuição *bootstrap* do estimador.

2) Uma amostra de tamanho $n = 20$ mourões de uma população de eucalipto citriodora (*Corymbia citriodora* (Hook.) K.D. Hill & L.A.S. Johnson) tratada com arseniato de cobre cromatado. Os resultados estão apresentados a seguir e referem-se ao tempo em anos de durabilidade de cada mourão. Existe uma garantia da empresa que comercializa os mourões tratados de uma durabilidade de 15 anos. As indústrias químicas fabricantes do produto preservativo, afirmam que a durabilidade média é de 20 a 30 anos. Faça o intervalo de confiança para a média do tempo de durabilidade dos mourões assumindo que o modelo apropriado é o exponencial

$$f_X(x) = \lambda e^{-\lambda x}$$

e identifique quem está com a razão, a empresa que comercializa os mourões tratados ou os fabricantes dos produtos químicos.

a) Aplique o intervalo de confiança exato com 95% de confiança. Este intervalo é dado por:

$$IC_{1-\alpha}(\mu) : \left[\frac{2n\bar{X}}{\chi_{\alpha/2, \nu=2n}^2}, \frac{2n\bar{X}}{\chi_{1-\alpha/2, \nu=2n}^2} \right].$$

b) Aplique o intervalo dado pela aproximação normal a seguir e compare com o resultado do intervalo exato.

$$IC_{1-\alpha}(\mu) : \left[\frac{\bar{X}\sqrt{n}}{\sqrt{n} + Z_{\alpha/2}}; \frac{\bar{X}\sqrt{n}}{\sqrt{n} - Z_{\alpha/2}} \right].$$

c) Determine quem tem razão na afirmativa feita. Estaria a firma que trata os mourões para comercialização sendo muito conservadora?

3,1	68,4	7,4	40,3	45,6	3,9	39,9	9,1	19,7	2,7
49,6	20,1	30,2	45,4	9,5	11,4	48,5	30,9	47,7	9,5

3) Uma amostra de $n = 30$ unidades amostrais de 1 ha cada foi realizada para determinar a abundância de espécies da família *Fabaceae*. Os resultados, supondo que seguem uma distribuição Poisson estão apresentados a seguir. O pesquisador afirma que a abundância das espécies da família é de 5 plantas/ha. Com base nos intervalos de 95% de confiança é possível afirmar que o pesquisador está certo? Utilize para isso o intervalo exato

$$IC_{1-\alpha}(\lambda) : \left[\frac{\chi^2_{1-\alpha/2; \nu=2k}}{2n}, \frac{\chi^2_{\alpha/2; \nu=2k+2}}{2n} \right]$$

e as aproximações normais

$$IC_{1-\alpha}(\mu) : \left[\bar{X} - Z_{\alpha/2}\sqrt{\frac{\bar{X}}{n}}, \bar{X} + Z_{\alpha/2}\sqrt{\frac{\bar{X}}{n}} \right]$$

e

$$IC_{1-\alpha}(\mu) : \left[\frac{1}{4} \left(\frac{Z_{\alpha/2}}{\sqrt{n}} - \sqrt{\frac{Z_{\alpha/2}^2}{n} + 4\bar{X}} \right)^2, \frac{1}{4} \left(\frac{Z_{\alpha/2}}{\sqrt{n}} + \sqrt{\frac{Z_{\alpha/2}^2}{n} + 4\bar{X}} \right)^2 \right].$$

Compare os resultados entre si. Avalie se o pesquisador tem ou não razão com base no intervalo exato. Os resultados obtidos são:

11	8	14	10	12	7	7	8	3	8	16	9	12	13	10
11	9	11	3	6	9	10	9	12	13	17	14	13	10	9

4) Utilizar os métodos *bootstrap* e obter os intervalos de 95% para a durabilidade média dos mourões. O programa necessário para aplicar este método usando o R é:

```
# necessita do pacote boot - instale-o
library(boot)
# função para média
samplemean <- function(x, d) {
  return(mean(x[d]))
}
# Amostra
x<- c(3.1, 68.4, 7.4, 40.3, 45.6, 3.9, 39.9, 9.1, 19.7, 2.7,
      49.6, 20.1, 30.2, 45.4, 9.5, 11.4, 48.5, 30.9, 47.7, 9.5)
#Média
ICboot <- boot(x, samplemean, R = 19999)
boot.ci(ICboot, conf = 0.95, type = c("norm","basic", "perc", "bca"))
print(sd(ICboot$t[,1]))
ICboot
```

5) Quais são os problemas de uma apresentação de resultados da média amostral como este: $1,70 \pm 0,20$?

Resolução

1) Marcar como verdadeira ou falsa as seguintes afirmativas:

a) (V) O intervalo de confiança para média de uma normal, dado por:

$$IC_{1-\alpha}(\mu) : \left[\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right],$$

é pouco útil, embora seja válido, pois exige o conhecimento do desvio padrão populacional σ , o que em situações reais é muito difícil de acontecer.

b) (V) O intervalo de confiança para μ dado por

$$IC_{1-\alpha}(\mu) : \bar{X} \pm t_{\alpha/2; \nu=n-1} \frac{S}{\sqrt{n}} = \bar{X} \pm e, \quad (2)$$

é válido sob normalidade da amostra aleatória, independência das observações amostrais entre si, garantida pela casualização ou sorteio e distribuições idênticas entre os diferentes X_j 's, $j = 1, 2, \dots, n$.

- c) (F) Quanto maior for o coeficiente de confiança $1 - \alpha$, menor é o comprimento do intervalo de confiança, para uma mesma amostra.
- d) (V) Quanto menor for o valor de α , maior é comprimento do intervalo de confiança, para uma mesma amostra.
- e) (V) Uma alternativa para se obter um intervalo de confiança mais estreito (menor comprimento), quando se pretende aumentar a confiança é aumentar o tamanho da amostra n .
- f) (F) A violação do pressuposto de normalidade no intervalo de confiança (1) provoca um aumento da probabilidade de cobertura real em relação ao valor nominal de $1 - \alpha$ fixado *a priori*.
- g) (V) Se a população não for normal, a aplicação do intervalo (1) causa uma redução do valor da probabilidade de cobertura real em relação ao valor nominal de $1 - \alpha$.
- h) (F) Não existem métodos exatos para se obter intervalos de confiança exatos para médias das distribuições exponenciais e Poisson.
- i) (V) Em populações discretas, como a Binomial e Poisson, quando dizemos que os intervalos de confiança para suas médias são exatos, significa que a cobertura real dos intervalos é de **pelo menos**. $100(1 - \alpha)\%$.
- j) (V) O método *bootstrap* para estimar por intervalo médias de um determinada população pode ser aplicado para qualquer distribuição.
- k) (F) O método *bootstrap* é um procedimento simples em que a amostra original é reamostrada um grande número de vezes e cada reamostragem tem o tamanho que é o dobro do tamanho n da amostra original.
- l) (V) Um dos procedimentos mais simples do método *bootstrap* é reamostrar a amostra original formando amostras do mesmo tamanho um grande número de vezes (milhares de vezes), ordenar os valores obtidos e tomar os quantis $100\alpha/2\%$ e $100(1 - \alpha/2)\%$ para serem os limites inferior (LI) e superior (LS), respectivamente, do intervalo de $100(1 - \alpha)\%$ de confiança.
- m) (V) Outro método *bootstrap* para obter intervalos de confiança chamado método *bootstrap* com correção de viés é idêntico ao processo descrito no item ll, exceto pelo fato de os quantis inferior e superior para determinar os LI e LS são corrigidos para contemplar assimetrias na distribuição *bootstrap* do estimador.

2) A média amostral é $\bar{X} = 27,138$. Os intervalos exato e aproximado são dados na sequência.

a) O intervalo exato de 95% de confiança, sabendo que os quantis superiores 2,5% e 97,5% com $\nu = 2n = 40\%$ graus de liberdade da distribuição qui-quadrado são $\chi_{0,025; \nu=40} = 59,342$ e $\chi_{0,975; \nu=40} = 24,433$, é dado por:

$$\begin{aligned} IC_{1-\alpha}(\mu) : \left[\frac{2n\bar{X}}{\chi_{\alpha/2, \nu=2n}^2}, \frac{2n\bar{X}}{\chi_{1-\alpha/2, \nu=2n}^2} \right] &= \left[\frac{40 \times 27,138}{59,342}, \frac{40 \times 27,138}{24,443} \right] \\ &= [18,29, 44,43]. \end{aligned}$$

A verdadeira média da durabilidade dos mourões é um valor entre 18,29 e 44,43 anos com 95% de confiança.

b) Sabendo que $Z_{0,025} = 1,96$, o intervalo usando a aproximação normal é:

$$\begin{aligned} IC_{1-\alpha}(\mu) : \left[\frac{\bar{X}\sqrt{n}}{\sqrt{n} + Z_{\alpha/2}}, \frac{\bar{X}\sqrt{n}}{\sqrt{n} - Z_{\alpha/2}} \right] &= \left[\frac{27,138 \times \sqrt{20}}{\sqrt{20} + 1,96}, \frac{27,138 \times \sqrt{20}}{\sqrt{20} - 1,96} \right] \\ &= [18,87, 48,31]. \end{aligned}$$

A verdadeira média da durabilidade dos mourões é um valor entre 18,87 e 48,31 anos com aproximadamente 95% de confiança. O intervalo aproximado forneceu uma boa estimativa do limite inferior, mas uma estimativa pior do limite superior do intervalo.

- c) Ambos os intervalos nos permitem concluir que a empresa que comercializa os mourões tratados está errada em relação a informação das empresas que fabricam o produto. Neste caso, o erro é no sentido de estarem sendo conservativas. A verdadeira média é um valor maior que 18 anos com 95% de confiança.
- 3) A média amostral é $\bar{X} = 10,1333$. Os intervalos estão apresentados a seguir. Neste caso $k = \sum_{j=1}^{30} x_j = 304$.
- a) Os quantis superiores $\chi_{0,975;\nu=608}^2$ e $\chi_{0,025;\nu=610}^2$ são obtidos com os comandos R dados por `qchisq(0.025,608)` e `qchisq(0.975,610)`, respectivamente, sendo iguais a 541,567 e 680,332. Assim, o intervalo exato é:

$$IC_{1-\alpha}(\lambda) : \left[\frac{\chi_{1-\alpha/2;\nu=2k}^2}{2n}, \frac{\chi_{\alpha/2;\nu=2k+2}^2}{2n} \right] = \left[\frac{541,567}{60}, \frac{680,332}{60} \right] \\ = [9,03, 11,34].$$

A verdadeira média da abundância com **pelo menos** 95% de confiança é um valor entre 9,03 e 11,34 plantas/ha.

- b) Utilizando as aproximações normais temos, para o primeiro caso:

$$IC_{1-\alpha}(\mu) : \left[\bar{X} - Z_{\alpha/2} \sqrt{\frac{\bar{X}}{n}}, \bar{X} + Z_{\alpha/2} \sqrt{\frac{\bar{X}}{n}} \right] = \left[10,1333 - 1,96 \sqrt{\frac{10,1333}{30}}, 10,1333 + 1,96 \sqrt{\frac{10,1333}{30}} \right] \\ = 10,1333 \pm 1,1391 \\ = [8,9942, 11,2724]$$

e, no segundo:

$$IC_{1-\alpha}(\mu) : \left[\frac{1}{4} \left(\frac{Z_{\alpha/2}}{\sqrt{n}} - \sqrt{\frac{Z_{\alpha/2}^2}{n} + 4\bar{X}} \right)^2, \frac{1}{4} \left(\frac{Z_{\alpha/2}}{\sqrt{n}} + \sqrt{\frac{Z_{\alpha/2}^2}{n} + 4\bar{X}} \right)^2 \right] \\ = \left[\frac{1}{4} \left(\frac{1,96}{\sqrt{30}} - \sqrt{\frac{1,96^2}{30} + 4 \times 10,1333} \right)^2, \frac{1}{4} \left(\frac{1,96}{\sqrt{30}} + \sqrt{\frac{1,96^2}{30} + 4 \times 10,1333} \right)^2 \right] \\ = [9,3806, 11,7007].$$

Concluimos que a verdadeira média da abundância com aproximadamente 95% de confiança é um valor entre 8,99 e 11,27 plantas/ha, no primeiro caso, e entre 9,38 e 11,70 plantas/ha, no segundo. As duas aproximações foram razoáveis. Sabemos por simulação Monte Carlo que a segunda aproximação tem desempenho muito superior a primeira.

- c) Com base no intervalo exato, a afirmativa do pesquisador não procede, ou seja, a abundância das plantas da família *Fabaceae*, com 95% de confiança, é superior a 5 plantas/ha.
- 4) Os resultados dos intervalos *bootstrap* são:

```
> boot.ci(ICboot, conf = 0.95, type = c("norm","basic", "perc", "bca"))
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 19999 bootstrap replicates

CALL :
boot.ci(boot.out = ICboot, conf = 0.95, type = c("norm", "basic",
"perc", "bca"))

Intervals :
Level      Normal                Basic
95%      (18.67, 35.62 )      (18.59, 35.61 )

Level      Percentile                BCa
95%      (18.68, 35.70 )      (18.89, 35.88 )

Calculations and Intervals on Original Scale

> print(sd(ICboot$t[,1]))
[1] 4.322
> ICboot
```

ORDINARY NONPARAMETRIC BOOTSTRAP

```
Call:
boot(data = x, statistic = samplemean, R = 19999)
```

```
Bootstrap Statistics :
      original    bias    std. error
t1*    27.14 0.0009505      4.322
```

Os quatro métodos apresentaram resultados bem próximos entre si. Comparativamente aos resultados do intervalo exato, observamos uma boa proximidade dos limites inferiores os métodos *bootstrap* com o limite inferior do método exato. Eles tiveram grande divergência para o limite superior do intervalo.

- 5) Podemos dizer que são essencialmente dois problemas: a) o uso do “ \pm ” sem a devida identificação do significado do termo que vem logo após este sinal. Não sabemos se o valor 0,20 significa o erro padrão da média, o desvio padrão amostral ou a margem de erro da pesquisa. b) não sabemos qual é a unidade que foi utilizada para expressar tanto o valor da estimativa quanto o valor do termo após o sinal \pm . c) não foi informado o valor do tamanho da amostra utilizado.

O ideal seria apresentar o resultado por: em uma amostra de tamanho $n = 50$ foi estimada a média e margem de erro em uma pesquisa sobre a altura média dos estudantes da Ufla por 1,70 m (0,20) m. Neste caso informou-se a unidade (em metros), o tamanho da amostra n e que o valor entre parênteses é a margem de erro da pesquisa, que é a semiamplitude do intervalo de confiança.