CE-003: Estatística II - Turma: K/O, 2^a Prova (09/12/2013)

GRR: ________________ **Nome:** ____ **Turma:** ________

1. (adap M & L) O consumo de combustível é uma variável aleatória com parâmetros dependendo do tipo de veículo. Suponha que, para um certo automóvel, o desvio padrão do consumo seja conhecido e igual a 2 km/l Porém, precisamos informações sobre o consumo médio. Para tal, coletamos uma amostra de 40 automóveis desse modelo e observamos seu consumo.
   (a) Identifique no contexto: a população, a amostra, o parâmetro de interesse, o estimador e a distribuição amostral.
   (b) Se a amostra forneceu um consumo médio de 9,3 km/l, construa um intervalo de confiança (95%) para a média de consumo deste modelo de carro.
   (c) Se a amplitude de um intervalo de confiança, construído a partir dessa amostra, é de 1,5 unidades, qual teria sido o nível de confiança?
   (d) Com a amostra já obtida de média $9,3$, teste a hipótese de que o consumo de veículos do modelo está acima de 8,7 km/l
   (e) Uma empresa possui 8 veículos do modelo considerado. Baseando-se nas informações disponíveis, qual a probabilidade de que o consumo médio dos 8 veículos fique acima de 9,5 km/l?
   (f) Uma nova amostra vai ser tomada para outro modelo e supondo o mesmo desvio padrão. Qual deve ser o tamanho da amostra para que o intervalo de confiança (95%) tenha amplitude de 1,2 unidades.

Solução:

(a) A população são os veículos da marca considerada e representada pela v.a.:

$$X : \text{consumo de combustível dos veículos da marca considerada}$$

que possui distribuição com $\mathrm{E}[X] = \mu$ (deconhecida) e $\mathrm{Var}[X] = 2^2$ (conhecida). A amostra é dada pelos consumos $(x_1, x_2, \ldots, x_n)$ dos veículos observados. O parâmetro de interesse é a média populacional $\mathrm{E}[X] = \mu$, o estimador é a média amostral $\overline{X}$ e a distribuição amostral é

$$\overline{X} \sim \mathrm{N}(\mu, \sigma^2/n = 2^2/40 = 0,1)$$

(b) I.C. (95%)

$$\overline{x} \pm z_(0,975)\frac{\sigma}{\sqrt{n}}$$
$$9,3 \pm 1,96\frac{2}{\sqrt{4}}$$
$$(8,68 \ \ ; \ \ 9,92)$$

(c)

$$1,5 = 2 \cdot (z_(1-\alpha/2)\frac{\sigma}{\sqrt{n}})$$
$$z_(1-\alpha/2 = \frac{1,5\sqrt{40}}{2 \cdot 2}$$
$$z_(1-\alpha/2 = 2,37$$
$$1-\alpha = 0,982$$

(d)

$$H_0 : \mu \leq 8,7 \quad vs \quad H_a : \mu > 8,7$$
$$\ldots$$

(e)

$$\overline{X}_8 \approx \mathrm{N}(\mu = 9,3, \sigma^2/n = 2^2/8 = 0,5)$$
$$P[\overline{X}_8 > 9,5] = 1,88e-41$$

(f)

$$1,2 = 2 \cdot (z_(0,975)\frac{2}{\sqrt{n}})$$
$$n = \lceil \frac{2^2 \cdot 1.96^2 \cdot 2^2}{1,2^2}$$
$$n = 43$$

---

2. Em uma avaliação de um novo algoritmo de classificação foi analisada uma amostra de 1200 cenários destre os quais 780 foram classificados corretamente.

   (a) Obtenha a estimativa pontual e intervalar (95% de confiança) para a proporção de classificações corretas.

   (b) Um algoritmo atualmente utilizado possui um percentual de acerto de 62%. Há evidências baseadas no estudo de que o novo algoritmo é superior aos utilizado atualmente? Justifique sua resposta.

Solução:

$$X : \text{resultado da classificação (correto/incorreto)}$$
$$X \sim \mathrm{B}(p) \qquad \mathrm{E}[X] = p \qquad \mathrm{Var}[X] = p(1-p)$$
$$\hat{p} \sim \mathrm{N}(p, p(1-p)/n)$$

(a)

$$\hat{p} = \frac{780}{1200} = 0,63$$
$$\text{I.C.assinttico :}$$
$$IC_{95\%} : \hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \longrightarrow (0,603 \;\; ; \;\; 0,657)$$
$$\text{I.C.conservador :}$$
$$IC_{95\%} : \hat{p} \pm z\sqrt{\frac{1}{4n}} \longrightarrow (0,602 \;\; ; \;\; 0,658)$$

(b) Resposta e justificativa baseada no valor estar ou não contido no I.C..

---

3. Quinze homens com idades entre 35 e 50 anos participaram em um estudo para avaliar o efeito de uma dieta e exercícios no nível de colesterol. O colesterol total foi medido em cada indivíduo inicialmente e depois novamente medido após 3 meses após participação em um programa de exercícios aeróbicos combinado com uma dieta de baixa caloria. Os dados estão a seguir.

antes	265	240	258	295	251	245	287	314	260	279	283	240	238	225	247
depois	229	231	227	240	238	241	234	256	247	239	246	218	219	226	233

Tabela 1: Medidas de colesterol de 15 homens antes de depois de dieta combinada com exercícios.

   (a) Calcule a média e mediana para as medidas alteração do colesterol.

   (b) Calcule desvio padrão e amplitude interquartílica para alteração do coleterol.

   (c) Construa um gráfico *boxplot* para as medidas de alteração do colesterol.

   (d) Obtenha um intervalo de confiança (95%) para o nível médio de coleterol antes da dieta/exercícios.

   (e) Use algum procedimento estatístico adequado para discutir se a dieta combinada com exercícios foi eficiente para reduzir o nível de colesterol total.

```
> antes <- c(265, 240, 258, 295, 251, 245, 287, 314, 260, 279, 283, 240, 238, 225, 247)
> depois <- c(229, 231, 227, 240, 238, 241, 234, 256, 247, 239, 246, 218, 219, 226, 233)
> (ad <- depois - antes)
```

```
 [1] -36  -9 -31 -55 -13  -4 -53 -58 -13 -40 -37 -22 -19   1 -14

> # a)
> c(media= mean(ad), mediana = median(ad))

  media mediana
 -26,87  -22,00

> # b)
> c(desvioP= sd(ad), AI = diff(fivenum(ad)[c(2,4)]))

desvioP      AI
  19,04   25,50

> # c)
> boxplot(ad)
> # d)
> mean(antes) + qt(c(0.025, 0.975), df=length(antes)-1) * sd(antes)/sqrt(length(antes))

[1] 248,0 275,6

> t.test(antes)$conf

[1] 248,0 275,6
attr(,"conf.level")
[1] 0,95

> # e)
```