CE-003: Estatística II - Turma: K/O, 2^a Prova (22/06/2016)

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

1. Calcule as medidas descritivas pedidas para o conjunto de dados, supondo que eles representam uma amostra.
`83, 92, 100, 57, 85, 88, 84, 82, 94, 93, 91, 95`

   (a) a média, a mediana e os quartis,

   (b) medidas de dispersão: amplitude total e interquartílica, desvio médio, desvio padrão e coeficiente de variação,

   (c) produza um gráfico *box-plot* dos dados.

   Quais os valores da mediana, média, variância, desvio padrão e o coeficiente de variação quando:

   (d) cada observação é multiplicada por 2;

   (e) soma-se 10 a cada observação;

   (f) de cada observação subtrai-se 3 e multiplica-se por 0,25;

   (g) subtrai-se a média de cada observação;

   (h) de cada observação subtrai-se a média e divide- se pelo desvio padrão.

Solução:

(a) Medidas de posição

	media	mediana	Q1	Q2	Q3
1	87,00	89,50	83,50	89,50	93,50

(b) Medidas de dispersão

	A	AI	DM	DP	CV	Var
1	43,00	10,00	7,33	10,93	12,56	119,45

(c) *box-plot*

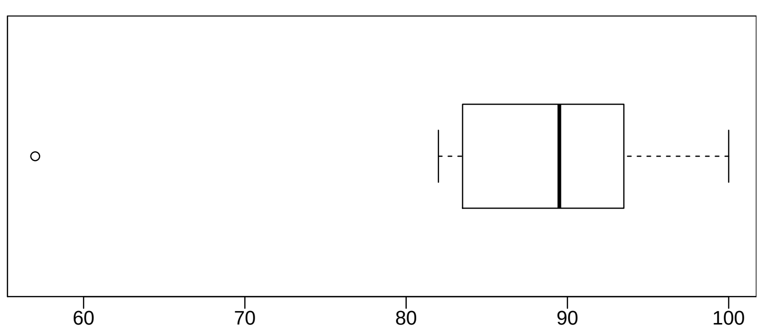


Figura 1: Diagrama box-plot dos dados fornecidos.

(d) cada observação é multiplicada por 2;

mediana fica multiplicada por 2
média fica multiplicada por 2
variância fica multiplicada por 4
desvio padrão fica multiplicado por 2
CV não se altera

(e) soma-se 10 a cada observação;

mediana fica acrescida de 10

média fica acrescida de 10
variância não se altera
desvio padrão não se altera
CV diminui

(f) de cada observação subtrai-se 3 e multiplica-se por 0,25;

mediana é 1/4 da mediana original diminuída de três unidades
média é 1/4 da media original diminuída de três unidades
variância é 1/16 da variância original
desvio padrão é 1/4 do desvio padrão original
CV aumenta

(g) subtrai-se a média de cada observação;

mediana fica subtraída do valor da média
média é igual a zero
variância não se altera
desvio padrão não se altera
CV indeterminado (Infinito)

(h) de cada observação subtrai-se a média e divide-se pelo desvio padrão.

mediana fica subtraída da média e dividida pelo desvio padrão
média fica igual a zero (0)
variância fica igual a um (1)
desvio padrão fica igual a um (1)
CV indeterminado (Infinito)

	mediana	media	var	DP	CV
original	89,50	87,00	119,45	10,93	12,56
d	179,00	174,00	477,82	21,86	12,56
e	99,50	97,00	119,45	10,93	11,27
f	21,62	21,00	7,47	2,73	13,01
g	2,50	0,00	119,45	10,93	Inf
h	0,23	0,00	1,00	1,00	Inf

Soluções computacionais (linguagem R):

```
> x1 <- c(83,  92,  100, 57,  85,  88,  84,  82,  94,  93,  91,  95)
> a1 <- mean(x1)
> a2 <- median(x1)
> a3 <- fivenum(x1)[2:4]
> b1 <- diff(range(x1))
> b2 <- diff(fivenum(x1)[c(2,4)])
> b3 <- mean(abs(x1-mean(x1)))
> b4 <- sd(x1)
> b5 <- 100*sd(x1)/mean(x1)
> b6 <- var(x1)
> boxplot(x1, horizontal=TRUE)
```

---

2. Seja X a variável tempo de serviço dos funcionários de determinada localidade. Se X tem distribuição normal com média 15 anos desvio padrão 10 anos:

(a) Qual a probabilidade de um funcionário, aleatoriamente escolhido, ter pelo menos 10 anos de tempo de serviço?

(b) Tomando-se uma amostra de 16 funcionários, qual a probabilidade do tempo médio de serviço estar entre 13 e 16 anos?

Solução:

$$\begin{aligned}
X &: \text{tempo de serviço de um funcionário} \\
X &\sim \mathrm{N}(\mu = 15, \sigma^2 = 10^2) \\
\overline{X} &: \text{tempo médio de serviço de um grupo de 16 funcionários} \\
\overline{X}_{16} &\sim \mathrm{N}(\mu = 15, \sigma^2 = 10^2/16)
\end{aligned}$$

(a) $P[X \geq 10] = P[Z \geq \frac{10-15}{10}] = P[Z \geq -0.5] = 0,6915$

(b) $P[13 \leq \overline{X}_{16} \leq 16] = P[\frac{13-15}{10/\sqrt{16}} \leq Z \leq \frac{16-15}{10/\sqrt{16}}] = P[-0,8 \leq Z \leq 0,4] = 0,4436$

Soluções computacionais (linguagem R):

```
> pa <- pnorm(10, m=15, sd=10, lower=FALSE)
> pb <- diff(pnorm(c(13,16), m=15, sd=10/4))
```

---

3. Uma companhia de cigarros anuncia que o índice médio de nicotina dos seus cigarros apresenta-se abaixo de 26 mg. Um laboratório realiza 6 análises, obtendo as seguintes quantidades de nicotina: 27, 24, 21, 25, 26, 22. Sabe-se que o índice de nicotina se distribui normalmente. Pode-se aceitar, ao nível de 10% de significância, a afirmação do fabricante? E ao nível de 5%?

Solução:

$$H_0 : \mu \geq 26 \text{ vs } H_a : \mu < 26$$
$$\alpha = 0,10 \text{ ou } \alpha = 0,05$$
$$t_c = \frac{\overline{x} - \mu_0}{S/\sqrt{n}} = \frac{24,17 - 26}{2,317/\sqrt{6}} = -1,938$$

(a) Para $\alpha = 0,10$

$$RC : \{t < -1,48\}$$
$$Rejeita - seH0...$$

(b) Para $\alpha = 0,05$

$$RC : \{t < -2,02\}$$
$$NaoRejeita - seH0...$$

Soluções computacionais (linguagem R):

```
> y <- c(27,  24,  21,  25,  26,  22)
> (m.y <- mean(y))

[1] 24,17

> (v.y <- var(y))

[1] 5,367

> (RC0.10 <- qt(0.10, df=length(y)-1))

[1] -1,476

> (RC0.05 <- qt(0.05, df=length(y)-1))

[1] -2,015

> (tc <- (mean(y)-26)/sqrt(var(y)/length(y)))

[1] -1,938

> ## solução "direta" usando função apropriada do R
> (T1 <- t.test(y, alternative="less", mu=26, conf=0.90))

One Sample t-test

data:  y
t = -1,9, df = 5, p-value = 0,06
alternative hypothesis: true mean is less than 26
90 percent confidence interval:
  -Inf 25,56
sample estimates:
mean of x
    24,17
```

```
> (T2 <- t.test(y, alternative="less", mu=26, conf=0.95))

One Sample t-test

data:  y
t = -1,9, df = 5, p-value = 0,06
alternative hypothesis: true mean is less than 26
95 percent confidence interval:
  -Inf 26,07
sample estimates:
mean of x
    24,17
```

---

4. Uma pesquisa mostrou que 30 dos 120 alunos entrevistados de uma universidade eram chefes de família.
   (a) Calcule um intervalo com 95% de confiança para a proporção de estudantes chefes de família.
   (b) Quantos alunos deveriam ser entrevistados para que o intervalo de confiança tivesse a metade da amplitude?
   (c) Uma segunda universidade fez a mesma pesquisa e encontrou 50 chefes de família entre os 150 entrevistados. Pode-se afirmar que há uma diferença significativa entre as proporções de chefes de família nas duas universidades?

Solução:

(a) IC (95%)

$$
\begin{aligned}
\hat{p} &= \frac{30}{120} = 0,25 \\
\text{Assintótico}: & \\
& \hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
& 0,25 \pm 1,96\sqrt{\frac{0,25(1-0,25)}{120}} \\
& 0,25 \pm 0,0775 \\
& (0,173 \;\; ; \;\; 0,327) \\
\text{Conservador}: & \\
& \hat{p} \pm z\sqrt{\frac{1}{4n}} \\
& 0,25 \pm 1,96\sqrt{\frac{1}{4 \cdot 120}} \\
& 0,25 \pm 0,0895 \\
& (0,161 \;\; ; \;\; 0,339)
\end{aligned}
$$

(b) n

$$
\begin{aligned}
\text{baseado no IC assintótico}: & \\
1,96\sqrt{\frac{0,25(1-0,25)}{n}} &= \frac{0,0775}{2} \\
n &= \left\lceil \frac{1,96^2 \cdot 0,25 \cdot (1-0,25)}{0,0387^2} \right. \\
n &= 480 \\
\text{baseado no IC conservador}: & \\
1,96\sqrt{\frac{1}{4n}} &= \frac{0,0895}{2} \\
n &= \left\lceil \frac{1,96^2}{4 \cdot 0,0387^2} \right. \\
n &= 640
\end{aligned}
$$

(c) Comparação de duas proporções

$$
\begin{aligned}
H_0: & p_1 = p_2 (p_1 - p_2 = 0) \text{ vs } H_a: p_1 \neq p_2 (p_1 - p_2 \neq 0) \\
& \alpha = 0,05 \\
& z_c = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} = \frac{(0,25 - 0,33) - 0}{\sqrt{\frac{0,25(1-0,25)}{120} + \frac{0,33(1-0,33)}{150}}} = -1,43 \\
RC: & \{z < 1,96 \text{ ou } z > 1,96\} \\
& z_c \notin RC
\end{aligned}
$$

Não rejeita-se H_0, não há evidência suficiente para rejeitar a hipótese de que as proporções sejam iguais.

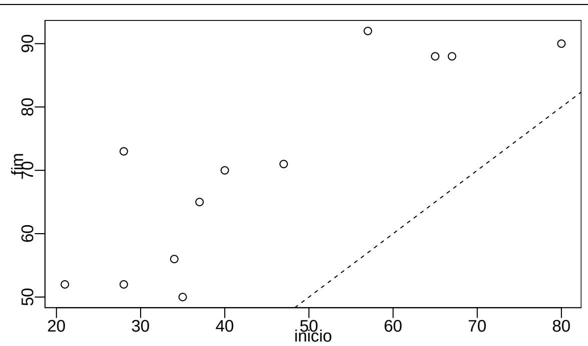
Soluções computacionais (linguagem R):

```
> ## resultado "direto" usando funções do R
> (a <- prop.test(30, 120, conf=0.95)$conf.int)

[1] 0,1775 0,3389
attr(,"conf.level")
[1] 0,95

> (TAB <- matrix(c(30,50,(120-30),(150-50)), nrow=2, dimnames=list(c("UniA","UniB"), c("Chefe","NaoChefe")))

     Chefe NaoChefe
UniA    30       90
UniB    50      100

> (b <- prop.test(TAB, conf=0.95))

2-sample test for equality of proportions with continuity correction

data:  TAB
X-squared = 1,8, df = 1, p-value = 0,2
alternative hypothesis: two.sided
95 percent confidence interval:
 -0,1990  0,0323
sample estimates:
prop 1 prop 2
0,2500 0,3333
```

---

5. Em um procedimento de avaliação um mesmo exame foi aplicado aos mesmos alunos quando estavam em períodos inicial e final de um curso. Os resultados das notas obtidas são apresentados a seguir. Interprete, faça análises estatísticas adequadas e discuta os resultados.

	media	desvio	CV	Amplitude	mediana	q1	q3	AIQ
inicio	44,92	18,39	40,95	59,00	38,50	31,00	61,00	30,00
fim	70,58	15,97	22,63	42,00	70,50	54,00	88,00	34,00



Solução:

(a) Será verificada a comparação e discussão das medidas descritivas da tabela e o gráfico.

(b) Teste de hipótese para duas amostras pareadas
Diferenças das observações:

```
[1] 28 35 22 30 31 45 15 10 23 24 24 21
```

$$H_0 : \mu_{fim} = \mu_{inicio} \ \ (d = 0) \text{ vs } \mu_{fim} = \mu_{inicio} \ \ (d > 0)$$
$$\alpha = 0,05$$
$$\overline{d} = 25,7 \text{ e } S_d^2 = 83,7$$
$$t_c = \frac{25,7 - 0}{\sqrt{83,7/12}} = 9,72$$
$$RC : \{t_c > 1,8\}$$
$$t_c \in RC$$

Rejeita-se H_0, ou seja, rejeita-se que a diferença das notas seja nula econsidera-se portanto que houve um aumento nas notas.

Soluções computacionais (linguagem R):

```
> t.test(fim, inicio, paired=TRUE, conf=0.95)

Paired t-test

data:  fim and inicio
t = 9,7, df = 11, p-value = 1e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 19,85 31,48
sample estimates:
mean of the differences
                  25,67
```