

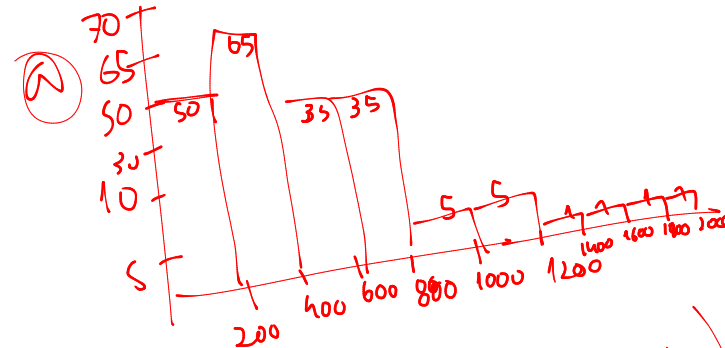
CE-003: Estatística II - Turma: K/O, 2ª Prova (13/03/2013)

GRR: _____ Nome: _____ Turma: _____

1. O número diário de solicitações em um serviço de atendimento online foi registrado por um período de 200 dias e os resultados foram resumidos na tabela a seguir.

Número de Solicitações diárias	Número dias
[0, 200)	50
[200, 400)	65
[400, 800)	70
[800, 1200)	10
[1200, 2000]	5
Total	200

- (a) Faça um histograma para representar estes dados
 (b) Obtenha o número médio de solicitações
 (c) Obtenha o número mediano de solicitações
 (d) Obtenha o coeficiente de variação do número de solicitações



Solução:

```
> xm <- c(100, 300, 600, 1000, 1600)
> fAbs <- c(50, 65, 70, 10, 5)
> (media <- (sum(xm * fAbs)/sum(fAbs)))
```

```
[1] 422.5
```

```
> xI <- c(0, 200, 400, 800, 1200)
> xS <- c(200, 400, 800, 1200, 2000)
> freq <- c(50, 65, 70, 10, 5)
> (freqAc <- cumsum(freq)/sum(freq))
```

```
[1] 0.250 0.575 0.925 0.975 1.000
```

```
> (ind50 <- min(which(freqAc > 0.5)))
```

```
[1] 2
```

```
> (xI[ind50] + ((0.5 - freqAc[ind50 - 1])/diff(freqAc[(ind50 - 1):ind50])) *
+ (xS - xI)[ind50])
```

```
[1] 353.8
```

```
> S2 <- sum(((xm - media)^2) * fAbs)/(sum(fAbs) - 1)
> (CV <- 100 * sqrt(S2)/media)
```

$$\textcircled{b} \left(\frac{100 \cdot 50}{200} + \dots + \frac{1000 \cdot 10}{200} \right)$$

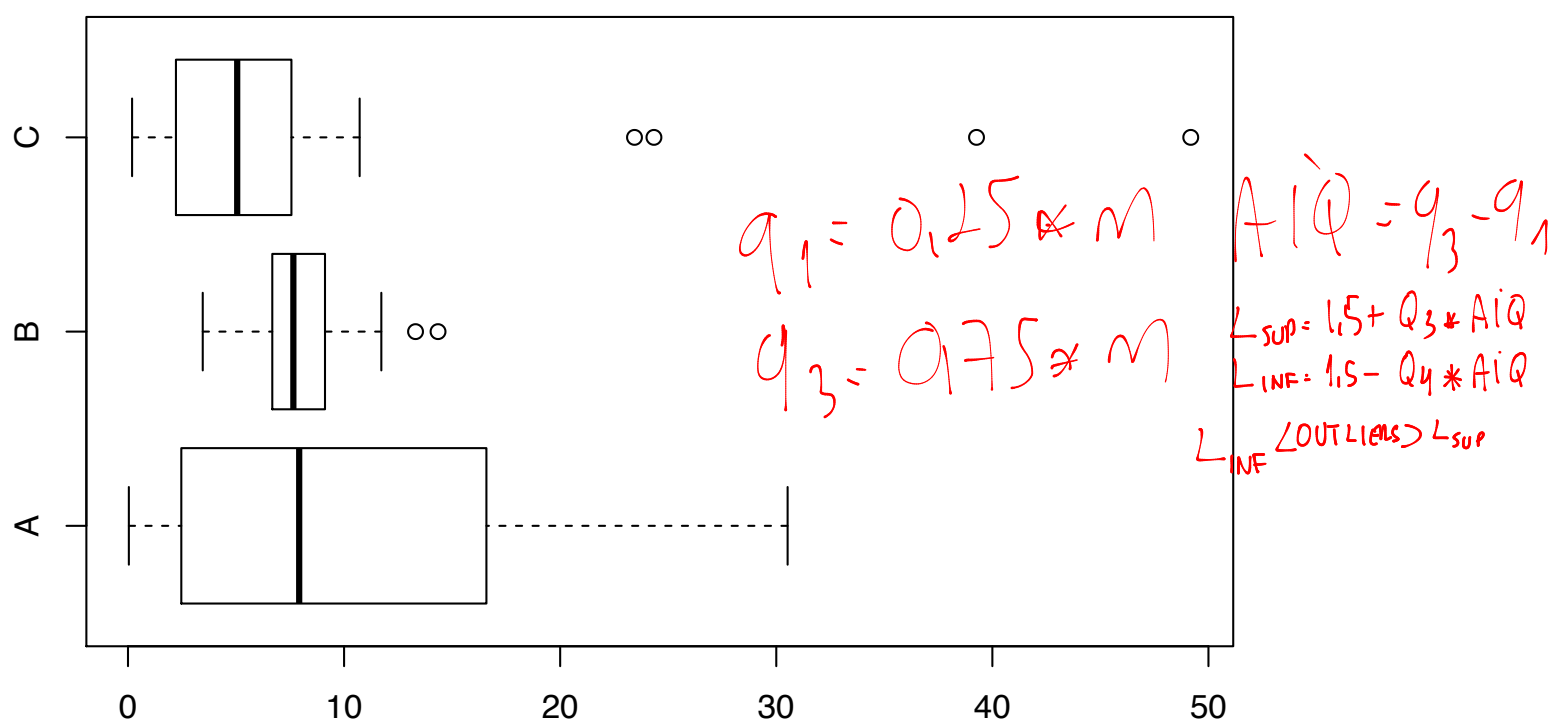
$$\textcircled{c} \sim 300$$

$$\textcircled{d} CV = 100 * \left(\frac{6}{\bar{x}} \right)$$

$$S^2 = \sum (x - \bar{x})^2$$

$$6 = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

2. Os tempos de atendimento e solução de problemas foram medidos em três *call-centers* distintos de uma mesma empresa e os dados foram representados no gráfico a seguir. Baseando-se no gráfico, avalie cada uma das afirmações a seguir, dizendo se está certa ou errada, e, corrigindo se errada.



- (a) Os valores no local C possuem uma distribuição simétrica. *QUASE*
NÃO HA *NÃO ACEITA* *MÉDIA = MEDIANA*
- (b) Os dados discrepantes do local A afetam (aumentam) a mediana do local. *BOX-LOT FORNECE A MEDIANA*
- (c) Os locais B e C possuem médias e desvios padrão semelhantes. *DESVIO B < DESVIO A*
DADOS CONTRÁRIOS
- (d) O local B possui o menor coeficiente de variação. *DEVIDO AO MENOR DESVIO*
- (e) As médias dos três locais devem ser semelhantes. *✓*

3. Considere um estudo na qual deseja-se estimar a proporção de solicitações atendidas e resolvidas de uma central do usuário através de uma amostra aleatória simples.

- (a) Qual deveria ser o tamanho da amostra para se obter a estimativa com uma margem de erro de 2,5% com 95% de confiança?
- (b) E para uma margem de erro de 3% com 99% de confiança?

Handwritten red notes and diagrams:

- $ME = Z_{1-\alpha} \cdot \sqrt{\frac{p(1-p)}{m}}$; *M = O AMOSTRA* ; $p(1-p) \equiv \frac{pq}{4} \equiv \text{DESVIO}$; $p = 0.5$; *SEM PRE*
- $p(1-p) = -p^2 + p \Rightarrow (pq)_{max} = 0.25$
- $0.025 = Z_{0.95} \sqrt{\frac{0.25}{m}}$
- DECORAR*
- $Z_{0.95} = 1.96$
- $Z_{0.99} = 2.57$
- $Z_{0.99} = 2.57$

Diagram showing a normal distribution curve with the area under the curve shaded blue, representing the confidence interval. The x-axis is labeled with $1-\alpha/2$, $1-\alpha$, and $1+\alpha/2$. The y-axis is labeled ME and $CONFIANÇA$.

Solução:

$$X \sim B(p)$$

$$\text{Teo 2: } \hat{p} = \bar{X} \approx N\left(p, \frac{p(1-p)}{n}\right)$$

$$\hat{p} \pm z_{1-\alpha} \sqrt{\frac{p(1-p)}{n}}$$

$$\hat{p} \pm \text{ME}$$

$$\text{ME} = z_{1-\alpha} \sqrt{\frac{p(1-p)}{n}}$$

$$n = \frac{(z_{1-\alpha})^2}{\text{ME}^2} p(1-p)$$

$p(1-p)$ é limitado superiormente para $p = 0,5$

$$n = \frac{(z_{1-\alpha})^2}{\text{ME}^2} 0,25$$

$$(a) \ n = \frac{(1.96)^2}{(0,025)^2} 0,25 = 1537$$

$$(b) \ n = \frac{(2.576)^2}{(0,03)^2} 0,25 = 1844$$

4. O consumo de combustível de uma frota de ônibus é uma v.a. medida pelo rendimento em km/l que tem distribuição normal com média de 8,2 e desvio padrão de 0,7. Seleccionando-se um veículo ao acaso qual a probabilidade de

(a) ter rendimento inferior a 7,0 km/l , $P(X < 7) = P\left(Z < \frac{7-8,2}{0,7}\right)$

(b) ter rendimento acima de 9,0 km/l , $P(X > 9)$

(c) ter rendimento entre 8,0 e 8,5 km/l . $P(8 < X < 8,5)$

Seleccionando-se uma amostra de 5 veículos, qual a probabilidade de que

(d) nenhum deles tenha rendimento inferior a 7,0 km/l $B(5, 0,5)$ $P(X > 7)^5$

(e) ao menos 4 tenham rendimento entre 8,0 e 8,5 km/l . $B(5, 0,5)$ $P(8 < X < 8,5)^4 \cdot (1 - P(8 < X < 8,5))^{5-4}$

Considerando-se ainda a amostra de 5 veículos, que a probabilidade de que

(f) o rendimento médio esteja entre 8,0 e 8,5 km/l ,

(g) o rendimento médio esteja abaixo de 7,5 km/l .

Solução:

X : consumo de combustível de cada veículo

$$X \sim N(\mu = 8,2; \sigma^2 = 0,7^2)$$

```
> (prA <- pnorm(7, m=8.2, sd=0.7))
```

```
[1] 0.04324
```

① $\bar{X} \sim N\left(8,2, \left(\frac{0,7}{\sqrt{5}}\right)^2\right)$
 $P(8 < \bar{X} < 8,5)$
 $P\left(\frac{8-8,2}{\frac{0,7}{\sqrt{5}}} < Z < \frac{8,5-8,2}{\frac{0,7}{\sqrt{5}}}\right)$
 ② idem ① $P(X < 7,5)$

```
> (prB <- pnorm(9, m=8.2, sd=0.7, lower=FALSE))
```

```
[1] 0.1265
```

```
> (prC <- diff(pnorm(c(8, 8.5), m=8.2, sd=0.7)))
```

```
[1] 0.2783
```

Y_1 : número de veículos com rendimento inferior a 7,0 km/l

$Y_1 \sim B(n = 5, p = prA)$

```
> (prD <- dbinom(0, size=5, prob=prA))
```

```
[1] 0.8017
```

Y_2 : número de veículos com rendimento entre 8,0 e 8,5 km/l

$Y_2 \sim B(n = 5, p = prC)$

```
> (prE <- pbinom(3, size=5, prob=prC, lower=FALSE))
```

```
[1] 0.02333
```

$\bar{X}$: rendimento médio de cinco veículos

$X \sim N(\mu = 8, 2; \sigma^2 = 0, 7^2/5)$

```
> (prF <- diff(pnorm(c(8, 8.5), m=8.2, sd=0.7/sqrt(5))))
```

```
[1] 0.5696
```

```
> (prG <- pnorm(7.5, m=8.2, sd=0.7/sqrt(5)))
```

```
[1] 0.01267
```

-
5. Com o objetivo de se dimensionar um serviço de atendimento ao cliente, foram tomados os tempos de atendimento em minutos de 50 consultas escolhidas ao acaso. Na análise dos dados adotou-se a distribuição exponencial e, o valor calculado do tempo médio de atendimento nas 50 consultas, foi de 6,5 minutos. Neste problema identifique e descreva: qual a população, qual a amostra, qual o parâmetro de interesse, qual o estimador e qual é a estimativa. Além disto discuta qual é a distribuição amostral relevante para inferências neste problema, justificando sua resposta.

→ A REGRA

$$\frac{\sum_{i=1}^n x_i}{n}$$

→ TEMPO
ATENDIMENTO

→ 6,5

DIST APROX.
DA
NORMAL

+ AMOSTRAS

O ESTIMADOR
É UMA MÉDIA

DIST AMOSTRAL

Solução:

$$\begin{aligned} X &: \text{ tempo de atendimento em consulta} \\ X &\sim \text{Exp}(\lambda) \\ f(x) &= \frac{1}{\lambda} \exp\left\{-\frac{x}{\lambda}\right\} \end{aligned}$$

- população: tempo de atendimento de cliente no SAC
 - amostra: os 50 tempos selecionados
 - parâmetro: o valor de λ na expressão de $f(x)$
 - estimador: a regra para se estimar o parâmetro, no caso $\hat{\lambda} = \bar{X} = \frac{\sum_i X_i}{n}$
 - estimativa: valor do estimador para uma particular amostra, $\hat{\lambda} = \bar{x} = 6,5$
 - distribuição amostral: distribuição do estimador. No caso, como o estimador é uma média, pelo *teorema do limite central* tem-se que esta distribuição aproxima-se da normal na medida que cresce o tamanho da amostra.
-