ACH0021 – Tratamento e Análise de Dados/Informações (1/2012)

Primeira Prova – Abril/2012

Nome: ______________________________ Nº USP: ____________

Turma/Horário: ____________________ Curso: ____________________

Observação 1: Duração da prova: **90** (**noventa**) minutos.
Observação 2: O uso de calculadora é opcional, e seu empréstimo durante a prova é proibido.

Formulário (conjunto de n elementos $\{w_i\}$)

Média: $\overline{w} = \frac{1}{n}\sum_{k=1}^{n} w_k$ Variância: $\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}(w_k - \overline{w})^2$ Desvio padrão: σ

1) Um estudante resolveu medir o tempo que ele necessita para chegar à EACH a partir de seu lar, registrando o tempo de percurso por 80 dias. Os dados de sua pesquisa estão organizados na tabela abaixo.

Tempo de percurso (minutos)	Frequência absoluta
$10q \vdash 14q$	10
$14q \vdash 16q$	26
$16q \vdash 18q$	20
$18q \vdash 22q$	24
TOTAL	80

a) [2,0 pontos] Estimar a média $\bar{t}$ do tempo de percurso e o desvio padrão σ.

A fim de decidir se um dado é típico dentre os registrados, adotou-se o seguinte critério: "Um dado será considerado típico se estiver até a uma distância de $m\sigma$ da média $\bar{t}$, onde $m \in \mathbb{R}$."

b) [1,5 pontos] No octogésimo primeiro dia (que não está contemplado na tabela), o aluno gastou $17,5q$ minutos para chegar à EACH. Estimar m de sorte que este dia não seja um dado típico quando comparado com os oitenta já registrados (que estão na tabela acima).

Prova A: $q = 1$ Prova B: $q = 2$ Prova C: $q = 3$ Prova D: $q = 4$

2a) Assumindo uma distribuição uniforme dos dados em cada intervalo da variável e tomando o ponto médio t_i de cada um destes como sendo o respectivo representante, tem-se

Tempo de percurso (min)	t_i (min)	Frequência absoluta (n_i)	Frequência relativa (f_i)	Amplitude (Δ_i)	Densidade (d_i)
$10q \vdash 14q$	$12q$	10	$10/80 = 0{,}125$	$4q$	$1/(32q)$
$14q \vdash 16q$	$15q$	26	$26/80 = 0{,}325$	$2q$	$13/(80q)$
$16q \vdash 18q$	$17q$	20	$20/80 = 0{,}250$	$2q$	$1/(8q)$
$18q \vdash 22q$	$20q$	24	$24/80 = 0{,}300$	$4q$	$3/(40q)$
TOTAL	-	80	1,000	-	-

Estimativa da média dos $n = 80$ dados:

$$\bar{t} = \frac{1}{n}\sum_i n_i t_i = \frac{1}{80}\left[10 \cdot 12q + 26 \cdot 15q + 20 \cdot 17q + 24 \cdot 20q\right] = \frac{133q}{8} = 16{,}625q \ (\text{min})\,.$$

Estimativa da variância:

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\sum_i n_i\left(t_i - \bar{t}\right)^2 = \frac{1}{80}\Big[10\,(12q - 16{,}625q)^2 + 26\,(15q - 16{,}625q)^2 + \\ &\quad +20\,(17q - 16{,}625q)^2 + 24\,(20q - 16{,}625q)^2\Big] = \frac{447q^2}{64}\,,\end{aligned}$$

que implica um desvio padrão de $\sigma = \sqrt{\frac{447}{64}}q$ (min). Em suma, tem-se

$$\begin{cases} \text{Prova A } (q=1): & \text{Estimativas: } \bar{t} = 16,625 \text{ (minutos) e } \sigma \approx 2,643 \text{ (minutos)} \\ \text{Prova B } (q=2): & \text{Estimativas: } \bar{t} = 33,250 \text{ (minutos) e } \sigma \approx 5,286 \text{ (minutos)} \\ \text{Prova C } (q=3): & \text{Estimativas: } \bar{t} = 49,875 \text{ (minutos) e } \sigma \approx 7,928 \text{ (minutos)} \\ \text{Prova D } (q=4): & \text{Estimativas: } \bar{t} = 66,500 \text{ (minutos) e } \sigma \approx 10,571 \text{ (minutos)} \end{cases} .$$

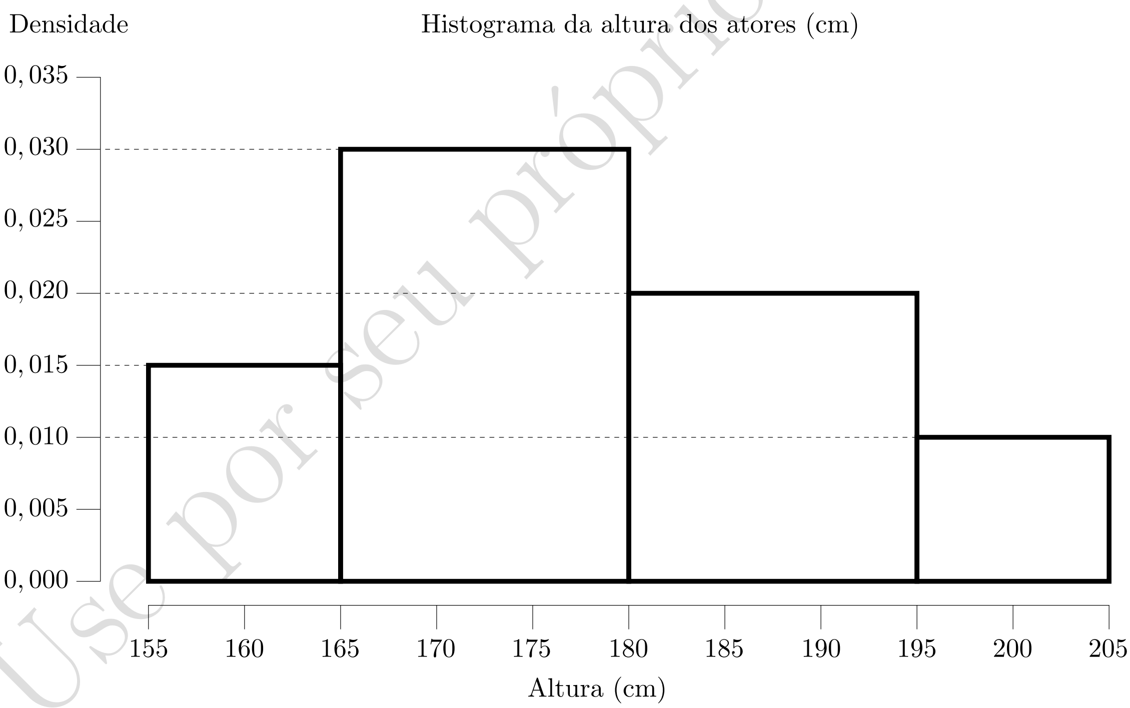
2b) Seguindo o critério de tipicidade, para que o octogésimo primeiro tempo não seja típico, é necessário e suficiente que esse não pertença ao intervalo $(\bar{t} - m\sigma, \bar{t} + m\sigma)$, tomando $m \geq 0$. Como o tempo de $17,5q$ minutos é superior à média $\bar{t}$, é suficiente analisar a situação $\bar{t} + m\sigma < 17,5q$, que implica

$$(0 \leq)\; m < \frac{17,5q - \bar{t}}{\sigma} = \frac{17,5q - 16,625q}{\sqrt{\frac{447}{64}}q} = \frac{7}{\sqrt{447}} \approx 0,331 .$$

Logo, se $m \in [0, \frac{7}{\sqrt{447}}) \approx [0; 0,331)$, o tempo de $17,5q$ é excluído do intervalo $(\bar{t} - m\sigma, \bar{t} + m\sigma)$.
Nota: Reparar que m independe do valor de q, sendo o mesmo para as quatro provas.

---

2) Para fazer o papel de herói em uma peça de teatro, N atores enviaram seus currículos ao diretor, e o histograma referente à altura deles é exibido abaixo.



a) [2,0 pontos] Para contracenar de forma satisfatória com a atriz principal, o diretor julgou que os atores deveriam ter uma altura entre a cm e b cm. Estimar o <u>número</u> de candidatos situados nessa faixa.
b) [2,0 pontos] Nesta mesma peça, há um papel secundário de "árvore", que seria melhor desempenhado por atores com uma dada altura mínima. A fim de selecionar os candidatos que satisfaçam esta condição, o diretor requisitou separar p dos candidatos mais altos. Estimar a altura mínima m destes atores – isto é, determinar a altura m acima da qual encontram-se exatamente p de todos os candidatos.

Prova A:	$N = 330$	$a = 168$	$b = 183$	$p = 12\%$
Prova B:	$N = 270$	$a = 171$	$b = 186$	$p = 14\%$
Prova C:	$N = 180$	$a = 174$	$b = 189$	$p = 16\%$
Prova D:	$N = 120$	$a = 177$	$b = 192$	$p = 18\%$

2a) Tabela de frequência a partir do histograma:

Altura (cm)	Densidade	Amplitude (cm)	Frequência Relativa	Frequência Absoluta
$155 \vdash 165$	0,015	10	$0,015 \cdot 10 = 0,15 = 15\%$	$0,15N$
$165 \vdash 180$	0,030	15	$0,030 \cdot 15 = 0,45 = 45\%$	$0,45N$
$180 \vdash 195$	0,020	15	$0,020 \cdot 15 = 0,30 = 30\%$	$0,30N$
$195 \vdash 205$	0,010	10	$0,010 \cdot 10 = 0,10 = 10\%$	$0,10N$
TOTAL	-	-	$1,00 = 100\%$	N

Assumindo uma distribuição uniforme dos dados nas barras do histograma, e notando que $a \in [165, 180)$ (em cm) e $b \in [180, 195)$ (em cm), tem-se

$$d_{165\vdash 180} = d_{a\vdash 180} \Rightarrow 0,030 = \frac{f_{a\vdash 180}}{180 - a} \Rightarrow f_{a\vdash 180} = \frac{3}{100}(180 - a)$$

e

$$d_{180\vdash 195} = d_{180\vdash b} \Rightarrow 0,020 = \frac{f_{180\vdash b}}{b - 180} \Rightarrow f_{180\vdash b} = \frac{1}{50}(b - 180) \ .$$

A fração de atores situadas na faixa estabelecida é, pois,

$$f_{a\vdash b} = f_{a\vdash 180} + f_{180\vdash b} = \frac{3}{100}(180 - a) + \frac{1}{50}(b - 180) = \frac{180 - 3a + 2b}{100} \ ,$$

ou

$$\begin{cases} \text{Prova A:} & f_{168\vdash 183} = 0,42 = 42\% \quad \Rightarrow \quad 0,42 \cdot 330 = 138,6 \approx 139 \text{ (atores)} \\ \text{Prova B:} & f_{171\vdash 186} = 0,39 = 39\% \quad \Rightarrow \quad 0,39 \cdot 270 = 105,3 \approx 105 \text{ (atores)} \\ \text{Prova C:} & f_{174\vdash 189} = 0,36 = 36\% \quad \Rightarrow \quad 0,36 \cdot 180 = 64,8 \approx 65 \text{ (atores)} \\ \text{Prova D:} & f_{177\vdash 192} = 0,33 = 33\% \quad \Rightarrow \quad 0,33 \cdot 120 = 39,6 \approx 40 \text{ (atores)} \end{cases} .$$

2b) Sendo $f_{195\vdash 205} = 10\%$ e $f_{180\vdash 195} = 30\%$, nota-se que o ponto m (altura requisitada) situa-se no intervalo $180 \vdash 195$ (em cm); ademais, é imediato que $f_{m\vdash 195} = f_{m\vdash 205} - f_{195\vdash 205} = p - 0,10$. Assumindo uma distribuição uniforme dos dados nessa faixa, tem-se

$$d_{180\vdash 195} = d_{m\vdash 195} \Rightarrow 0,020 = \frac{\overbrace{f_{m\vdash 195}}^{p-0,10}}{195 - m} \Rightarrow m = 200 - 50p \, .$$

Logo,

$$\begin{cases} \text{Prova A } (p = 12\%): & \text{Estimativa da altura mínima requisitada: } 200 - 50 \cdot 0,12 = 194,0 \text{ (cm)} \\ \text{Prova B } (p = 14\%): & \text{Estimativa da altura mínima requisitada: } 200 - 50 \cdot 0,14 = 193,0 \text{ (cm)} \\ \text{Prova C } (p = 16\%): & \text{Estimativa da altura mínima requisitada: } 200 - 50 \cdot 0,16 = 192,0 \text{ (cm)} \\ \text{Prova D } (p = 18\%): & \text{Estimativa da altura mínima requisitada: } 200 - 50 \cdot 0,18 = 191,0 \text{ (cm)} \end{cases} .$$

---

3) O desempenho de um jogador de basquete em cobranças de lances livres foi registrado. De um total de n lançamentos ($n > 1$), cada acerto foi indicado por 1 ponto e cada erro, por 0 ponto.
[1,5 pontos] a) Determinar o desvio padrão caso o jogador tenha sido bem sucedido em $1/4$ de suas cobranças (suponha, em caso de desconforto, que n seja múltiplo de 4).
[1,0 ponto] b) Determinar o desvio padrão máximo.

3a) Seja n_1 o número de acertos (e, por conseguinte, o número de erros é $n_0 = n - n_1$). A média de pontos $\overline{x}$ é dada por

$$\overline{x} = \frac{1}{n}\big[n_0 \cdot 0 + n_1 \cdot 1\big] = \frac{n_1}{n}\,,$$

que implica uma variância σ^2 de

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\big[n_0\,(0-\overline{x})^2 + n_1\,(1-\overline{x})^2\,\big] = \frac{1}{n}\left[(\overbrace{n-n_1}^{n_0})\left(0-\frac{n_1}{n}\right)^2 + n_1\left(1-\frac{n_1}{n}\right)^2\right] \\ &= \frac{n_1}{n}\left(1-\frac{n_1}{n}\right)\,, \end{aligned} \tag{1}$$

donde se chega ao desvio padrão

$$\sigma = \sqrt{\frac{n_1}{n}\left(1-\frac{n_1}{n}\right)}\,. \tag{2}$$

Caso o jogador acerte 1/4 de suas cobranças ($n_1 = n/4$), tem-se

$$\sigma = \sqrt{\frac{n/4}{n}\left(1-\frac{n/4}{n}\right)} = \frac{\sqrt{3}}{4}\,, \tag{3}$$

conforme requisitado.

3b) Da equação (2), é imediato que o valor máximo do desvio padrão coincide com o valor máximo da função $f(x) = x\,(1-x)$, onde $x := n_1/n$ (quando então f coincide com a variância). Como f é uma parábola côncava, o ponto máximo realiza-se no ponto médio de suas raízes, que é $x = 1/2$. Como $f(1/2) = 1/4$, o desvio padrão máximo é 1/2.