ACH0021 – Tratamento e Análise de Dados/Informações (2015.1)

Primeira Prova – Abril/2015

Nome: ______________________________ Nº USP: ____________

Turma/Horário: ____________________ Curso: ____________________

Observação 1: Duração da prova: **75** (**setenta e cinco**) minutos.
Observação 2: O uso de calculadora é **proibido**.
Observação 3: É **proibido** fazer perguntas **durante** a prova.

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

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

1) Uma empresa de transportes mediu o tempo que um ônibus necessita para ir do ponto X ao ponto Y, registrando o tempo de percurso 10 vezes, e constatou-se que este varia bastante conforme as condições do clima e da estrada. Os dados de sua pesquisa estão organizados na tabela abaixo.

Tempo de percurso (minutos)	Frequência absoluta
$00q \vdash 04q$	02
$04q \vdash 06q$	04
$06q \vdash 12q$	04
TOTAL	10

a) [2,0 pontos] Estimar a média $\bar{t}$ do tempo de percurso e o desvio padrão σ (**não** é necessário calcular explicitamente a raíz quadrada; exemplos: a resposta pode estar na forma $\sqrt{19}$, $\sqrt{4,5}$, $\sqrt{7/3}$, *et cætera*).

b) [3,0 pontos] Estimar o **número de viagens** cujo tempo de percurso seja superior a $5,15q$ minutos.

Nota: Explicitar/justificar o raciocínio na resolução.

Prova A: $q = 2$ Prova B: $q = 4$ Prova C: $q = 8$ Prova D: $q = 10$

1a) 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)
$00q \vdash 04q$	$02q$	02	$02/10 = 0{,}20$	$4q$	$1/(20q)$
$04q \vdash 06q$	$05q$	04	$04/10 = 0{,}40$	$2q$	$1/(5q)$
$06q \vdash 12q$	$09q$	04	$04/10 = 0{,}40$	$4q$	$1/(10q)$
TOTAL	-	10	1,00	-	-

Estimativa da média dos $n = 10$ dados:

$$\bar{t} = \frac{1}{n}\sum_i n_i t_i = \frac{1}{10}\left[2 \cdot 2q + 4 \cdot 5q + 4 \cdot 9q\right] = 6q \text{ (min)}\,.$$

Estimativa da variância:

$$\sigma^2 \approx \frac{1}{n}\sum_i n_i \left(t_i - \bar{t}\right)^2 = \frac{1}{10}\left[2\,(2q-6q)^2 + 4\,(5q-6q)^2 + 4\,(9q-6q)^2\right] = \frac{36q^2}{5} \text{ (min}^2\text{)}\,,$$

que implica um desvio padrão de $\sigma = \frac{6}{\sqrt{5}}q$ (min). Em suma, tem-se

$$\begin{cases} \text{Prova A } (q=2): & \text{Estimativas: } \bar{t} = 12 \text{ (minutos) e } \sigma = \frac{12}{\sqrt{5}} \approx 5,37 \text{ (minutos)} \\ \text{Prova B } (q=4): & \text{Estimativas: } \bar{t} = 24 \text{ (minutos) e } \sigma = \frac{24}{\sqrt{5}} \approx 10,73 \text{ (minutos)} \\ \text{Prova C } (q=8): & \text{Estimativas: } \bar{t} = 48 \text{ (minutos) e } \sigma = \frac{48}{\sqrt{5}} \approx 21,47 \text{ (minutos)} \\ \text{Prova D } (q=10): & \text{Estimativas: } \bar{t} = 60 \text{ (minutos) e } \sigma = \frac{60}{\sqrt{5}} \approx 26,83 \text{ (minutos)} \end{cases} .$$

1b) O ponto $5,15q$ situa-se na faixa $4q \vdash 6q$, e deseja-se saber $f_{5,15q \vdash 6q}$ que, juntamente com $f_{6q \vdash 12q} = 0,40$ (vide tabela do exercício (1a)), fornece $f_{5,15q \vdash 12q} = f_{5,15q \vdash 6q} + f_{6q \vdash 12q}$, que é a fração das viagens com duração superior a $5,15q$ minutos. Assumindo uma distribuição uniforme dos dados em cada intervalo da variável, e da tabela do exercício (1a), tem-se:

$$d_{5,15q \vdash 6q} = d_{4q \vdash 6q} \Leftrightarrow \frac{f_{5,15q \vdash 6q}}{\underbrace{\Delta_{5,15q \vdash 6q}}_{=0,85q}} = \frac{1}{5q} ,$$

donde é imediato que $f_{5,15q \vdash 6q} = 0,17$. Logo, estima-se que $f_{5,15q \vdash 12q} = f_{5,15q \vdash 6q} + f_{6q \vdash 12q} = 0,17 + 0,40 = 0,57$ do total de dados foi superior a $5,15q$ minutos, e isto corresponde a $10 \times 0,57 \approx 6$ viagens.

Nota: Notar que o resultado do exercício (1b) independe do valor de q, e a resposta é a mesma para as quatro provas.

---

2) Ainda em relação ao exercício anterior,
a) [3,0 pontos] Estimar o tempo t abaixo do qual estariam p de **todos** os percursos mais rápidos. Dar a resposta com ao menos **duas** casas decimais.

b) [2,0 pontos] Um estudante notou que a variância amostral (vide formulário) na pluviometria diária numa dada região em cinco dias (logo, são cinco dados) foi de $1,0\text{mm}^2$. Na semana posterior, os dados referentes ao segundo dia foram perdidos; contudo, sabe-se que, nos outros dias, o registro foi de uma pluviometria de exatamente $1,0$mm em cada. Determinar a média dos cinco dados e a medida da pluviometria do segundo dia.

Nota 1: Explicitar/justificar o raciocínio na resolução.

Nota 2: A questão (2b) é **independente** de (2a).

Prova A: $p = 10,4\%$	Prova B: $p = 11,1\%$	Prova C: $p = 15,1\%$	Prova D: $p = 9,43\%$

2a) Os p dos tempos mais rápidos encontram-se no intervalo $00q \vdash 04q$, visto que $f_{00q \vdash 04q} = 20\% > p$. Assumindo uma distribuição uniforme dos dados em cada intervalo[1], tem-se

$$d_{0q \vdash t} = d_{0q \vdash 4q} \Rightarrow \frac{f_{0q \vdash t}}{\Delta_{0q \vdash t}} = \frac{1}{20q} \Rightarrow t = 20qp \text{ (min)} .$$

Desta forma,

$$\begin{cases} \text{Prova A } (q = 2 \text{ e } p = 0,104): & t = 4,16 \text{ min} \\ \text{Prova B } (q = 4 \text{ e } p = 0,111): & t = 8,88 \text{ min} \\ \text{Prova C } (q = 8 \text{ e } p = 0,151): & t = 24,16 \text{ min} \\ \text{Prova D } (q = 10 \text{ e } p = 0,0943): & t = 18,86 \text{ min} \end{cases} .$$

[1] Neste exercício, é suficiente considerar esta hipótese somente no intervalo $00q \vdash 04q$.

2b) Denote por m a medida da pluviometria do segundo dia e por $\overline{x}$ e σ^2 a média e variância amostral dos dados, respectivamente. Desta forma, tem-se

$$\left\{\begin{array}{rcl} \overline{x} & = & \dfrac{4 \times 1,0 + m}{5} \\ \\ \sigma^2 & = & \dfrac{4\,(1,0-\overline{x})^2 + (m-\overline{x})^2}{5} \end{array}\right. ,$$

donde é imediato que

$$\left\{\begin{array}{rcl} \overline{x} & = & 1 \pm \dfrac{\sigma}{2} \\ \\ m & = & 1 \pm \dfrac{5\sigma}{2} \end{array}\right. .$$

De $\sigma^2 = 1$, os resultados realísticos indicam que a medida do segundo dia foi de $\frac{7}{2} = 3,5$(mm), o que conduz a uma média dos cinco dias de $\frac{3}{2} = 1,5$(mm). Naturalmente, a solução $m = -\frac{3}{2}$ com $\overline{x} = \frac{1}{2}$ não é aceitável.