

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

Prova de Recuperação – Julho/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 $\{w_i\}$)

$$\text{Média: } \bar{w} = \frac{1}{n} \sum_{k=1}^n w_k$$

$$\text{Variância: } \sigma^2 = \frac{1}{n} \sum_{k=1}^n (w_k - \bar{w})^2$$

$$\text{Desvio padrão: } \sigma$$

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(-a \leq Z \leq a) = \gamma, \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

1) [3,0 pontos] Deseja-se estimar o tempo que os alunos ingressantes da EACH estudam por dia. Com base em estatísticas dos anos anteriores, este tempo pode ser aproximado por uma distribuição normal e a variância pode ser considerada como sendo de 1,21 horas². Estimar o número de alunos para a pesquisa de forma que a amplitude do intervalo de 95% de confiança seja de, no máximo, 0,5 hora.

1) [3,0 pontos] Seja \bar{X} a média amostral com n dados; sejam μ e σ , respectivamente, a média e o desvio padrão das baterias ($\sigma^2 = 1,21$ horas², o que implica $\sigma = \sqrt{1,21} = 1,1$ horas, e μ desconhecido). Pelo teorema central do limite (assumindo as variáveis envolvidas como sendo independentes e identicamente distribuídas), a variável

$$Z := \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

segue uma distribuição normal de média 0 e variância 1. Logo, pela tabela de valores para a distribuição normal, sabe-se que

$$P(-a \leq Z \leq a) = \gamma = 0,95$$

realiza-se para $a \approx 1,96$, visto que $P(0 \leq Z \leq a) = \frac{1}{2}P(-a \leq Z \leq a) = \frac{0,925}{2} = 0,475$. Como $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, então

$$P\left(-a \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq a\right) = 0,95 \quad (\text{com } a \approx 1,96),$$

donde se tem

$$P\left(\bar{X} - \frac{a\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{a\sigma}{\sqrt{n}}\right) = 0,95 \quad (\text{com } a \approx 1,96).$$

A amplitude Δ do intervalo de 85% de confiança no problema é, segundo a equação acima,

$$\Delta = \left(\bar{X} + \frac{a\sigma}{\sqrt{n}}\right) - \left(\bar{X} - \frac{a\sigma}{\sqrt{n}}\right) = \frac{2a\sigma}{\sqrt{n}} \text{ (horas)}.$$

Para que esta amplitude seja de no máximo 0,5 hora (*id est*, $\Delta \leq 0,5$ hora), deve-se ter

$$\begin{aligned} 0,5 &\geq \Delta = \frac{2a\sigma}{\sqrt{n}} \\ &\approx \frac{2 \cdot 1,96 \cdot 1,1}{\sqrt{n}}, \end{aligned}$$

donde é imediato que $n \geq 75$, que é o tamanho estimado da amostra para as condições impostas.

2) [3,0 pontos] Um motorista X, que costuma estar embriagado em 20% das vezes que dirige à noite, é parado durante seu passeio noturno de carro e é convidado a fazer o teste do bafômetro. O aparelho em questão acusa um condutor indevidamente alcolizado com probabilidade $\frac{19}{20}$, mas também acusa, injustamente, um condutor sóbrio com chance de 5,0%. Se o teste do bafômetro indicou que o motorista X estava com excesso de álcool em seu sangue, determinar a probabilidade desta pessoa de fato estar embriagado.

2) Definição dos eventos:

A: Evento “motorista X com excesso de álcool no sangue”.

B: Evento “bafômetro indicar excesso de álcool no sangue”.

Do enunciado, tem-se $P(A) = 0,20$ (e, por conseguinte, $P(A^c) = 1 - P(A) = 0,80$), $P(B|A) = 0,95$ e $P(B|A^c) = 0,05$. Como os eventos A e A^c formam uma partição, pode-se escrever $B = (B \cap A) \cup (B \cap A^c)$, donde é imediato que

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap A^c) \\ &= P(B|A)P(A) + P(B|A^c)P(A^c) = 0,95 \cdot 0,20 + 0,05 \cdot 0,80 = 0,23, \end{aligned}$$

que é a probabilidade do bafômetro indicar irregularidade. A probabilidade do motorista X de fato estar indevidamente alcolizado, dado que o bafômetro indicou o fato, é dada por

$$P(A|B) = \frac{P(B \cap A)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{0,95 \cdot 0,20}{0,23} = \frac{19}{23} \approx 0,826 \text{ (ou aproximadamente 82,6\%)}. \quad \cdot$$

3) Em um curso hipotético, a distribuição de notas de uma turma é dada na tabela ao lado.

a) [1,0 ponto] Estimar a média e o desvio padrão das notas.

b) [1,5 pontos] Estimar a fração de alunos aprovados (nota maior ou igual a 5,0).

c) [1,5 pontos] O professor sorteará um prêmio aos melhores alunos. Estimar a nota acima da qual encontram-se os 40% dos melhores alunos do total.

Nota	Frequência
0,0 \vdash 2,0	07
2,0 \vdash 4,0	11
4,0 \vdash 6,0	29
6,0 \vdash 10,0	13
TOTAL	60

3a) Assumindo uma distribuição uniforme em cada barra do histograma e tomando o ponto médio g_i do i -ésimo intervalo como representante do mesmo, tem-se

Nota	g_i	Frequência absoluta	Frequência relativa	Amplitude	Densidade
0,0 \vdash 2,0	1,0	07	$07/60 \approx 0,1167$	2,0	$7/120 \approx 0,058$
2,0 \vdash 4,0	3,0	11	$11/60 \approx 0,1833$	2,0	$11/120 \approx 0,092$
4,0 \vdash 6,0	5,0	29	$29/60 \approx 0,4833$	2,0	$29/120 \approx 0,242$
6,0 \vdash 10,0	8,0	13	$13/60 \approx 0,2167$	4,0	$13/240 \approx 0,054$
TOTAL	-	60	1	-	-

Tabela 1: Tabela de frequências, amplitudes e densidades do problema.

A média \bar{n} dos $n = 60$ dados é estimada como sendo

$$\bar{n} = \frac{1}{n} \sum_i n_i g_i = \frac{1}{60} [7 \cdot 1,0 + 11 \cdot 3,0 + 29 \cdot 5,0 + 13 \cdot 8,0] = \frac{289}{60} \approx 4,817.$$

Por outro lado, a variância σ^2 é estimada como

$$\begin{aligned} \sigma^2 &= \frac{1}{n} \sum_i n_i (g_i - \bar{n})^2 \\ &= \frac{1}{60} \left[7 \left(1,0 - \frac{289}{60} \right)^2 + 11 \left(3,0 - \frac{289}{60} \right)^2 + 29 \left(5,0 - \frac{289}{60} \right)^2 + 13 \left(8,0 - \frac{289}{60} \right)^2 \right] = \frac{16259}{3600}, \end{aligned}$$

que conduz ao desvio padrão $\sigma = \frac{\sqrt{16259}}{60} \approx 2,125$.

3b) Assumindo uma distribuição uniforme dos dados no intervalo $4,0 \vdash 6,0$, tem-se

$$d_{4,0 \vdash 6,0} = d_{5,0 \vdash 6,0} \Rightarrow \frac{29}{120} = \frac{f_{5,0 \vdash 6,0}}{6,0 - 5,0} \Rightarrow f_{5,0 \vdash 6,0} = \frac{29}{120}.$$

A fração de alunos estimada com nota acima de $5,0$ é, pois,

$$f_{5,0 \vdash 6,0} + f_{6,0 \vdash 10,0} = \frac{29}{120} + \frac{13}{60} = \frac{11}{24} \approx 0,458 \text{ (ou aproximadamente } 45,8\%).$$

3c) Como $f_{6,0 \vdash 10,0} = \frac{13}{60} \approx 0,2167$ e $f_{4,0 \vdash 10,0} = f_{4,0 \vdash 6,0} + f_{6,0 \vdash 10,0} = \frac{29}{60} + \frac{13}{60} = \frac{7}{10} > 40\%$, a nota x acima da qual encontram-se os 40% dos melhores alunos situa-se no intervalo $4,0 \vdash 6,0$. Assumindo uma distribuição uniforme dos dados neste intervalo, tem-se

$$d_{4,0 \vdash 6,0} = d_{x \vdash 6,0} \Rightarrow \frac{29}{120} = \frac{f_{x \vdash 6,0}}{6,0 - x} = \frac{\overbrace{f_{x \vdash 10,0} - f_{6,0 \vdash 10,0}}^{0,40}}{6,0 - x} \Rightarrow x = \frac{152}{29} \approx 5,24.$$

Distribuição Normal/Gaussiana: Valores de q , onde $P(0 \leq Z \leq z_\alpha) = q$

Segunda decimal de z_α

	0	1	2	3	4	5	6	7	8	9	
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359	0,0
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753	0,1
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141	0,2
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517	0,3
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879	0,4
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224	0,5
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549	0,6
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852	0,7
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133	0,8
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389	0,9
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621	1,0
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830	1,1
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015	1,2
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177	1,3
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319	1,4
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441	1,5
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545	1,6
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633	1,7
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706	1,8
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767	1,9
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817	2,0
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857	2,1
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890	2,2
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916	2,3
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936	2,4
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952	2,5
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964	2,6
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974	2,7
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981	2,8
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986	2,9
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990	3,0
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993	3,1
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995	3,2
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997	3,3
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998	3,4
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	3,5
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	3,6
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	3,7
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	3,8
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	3,9

Parte inteira e primeira decimal de z_α