CE-003: Estatística II - Turma K/O
Avaliações Periódicas (1^o semestre 2018)

Avaliação 01

1. Em cada um dos itens a seguir forneça o espaço amostral e também sua caracterização quanto a ser (i) finito ou infinito, (ii) enumerável ou não enumerável, (iii) equiprovável ou não equiprovável, justificando as respostas.
   (a) O número de solicitações de internamento em UTI de um determinado hospital em um dia.
   (b) O tempo de navegação de um acesso a determinado *site* da internet.
   (c) O número de pontos que um time fará no seu próximo jogo no campeonato brasileiro.
   (d) O número de tentativas sem sucesso de telemarketing até que se consiga a concretização de um negócio (sucesso).
   (e) O número de processos de usuários rodando em uma servidora de processamento em um dado instante.

   Solucao:

---

2. Dois dados são lançados. Calcule a probabilidade de:
   (a) sairem dois números consecutivos,
   (b) a soma ser um número par e maior que nove,
   (c) a soma ser um número par ou maior que nove,
   (d) a soma dos valores ser maior ou igual a oito, sabendo-se que um dos dados era cinco,
   (e) a soma ser maior que sete sabendo que sairam dois números iguais.

   Solucao:

Avaliação 02

1. Um algorítmo de classificação deve tentar resolver corretamente dois problemas, A e B. A probabilidade resolver A corretamente é de 0,6. Caso resolva A corretamente, a probabilidade de resolver B corretamente é de 0,85; caso contrário, essa probabilidade é de 0,25.
   (a) Qual a probabilidade de ele:
      i. resolver corretamente os dois problemas?
      ii. resolver corretamente apenas um dos problemas?
      iii. não resolver nenhum corretamente?
   (b) os eventos "resolver corretamente A" e "resolver corretamente B",
      i. são independentes? (justifique)
      ii. são mutuamente exclusivos? (justifique)

   Solução:

$$A: \text{ resolver corretamente o problema A}$$
$$B: \text{ resolver corretamente o problema B}$$
$$P[A] = 0,6 \quad ; \quad P[B|A] = 0,85 \quad ; \quad P[B|\overline{A}] = 0,25$$
$$P[\overline{A}] = 0,4 \quad ; \quad P[\overline{B}|A] = 0,15 \quad ; \quad P[\overline{B}|\overline{A}] = 0,75$$

   (a) i. $P[A \cap B] = P[A] \cdot P[B|A] = (0,6) \cdot (0,85) = 0.51$
      ii. $P[A \cap \overline{B}] + P[\overline{A} \cap B] = P[A] \cdot P[\overline{B}|A] + P[\overline{A}] \cdot P[B|\overline{A}] = (0,6) \cdot (0,15) + (0,4) \cdot (0,25) = 0.19$
      iii. $P[\overline{A} \cap \overline{B}] = P[\overline{A}] \cdot P[\overline{B}|\overline{A}] = (0,4) \cdot (0,75) = 0.3$

(b) i. Não, pois $P[A \cap B] = 0.51 \neq P[A] \cdot P[B] = 0.61$,
em que $P[B] = P[B \cap A] + P[B \cap \overline{A}] = (0,6)(0,85) + (0,4)(0,25) = 0.61$

ii. Não, pois $P[A \cap B] \neq 0$

---

2. Para selecionar seus funcionários, uma empresa oferece aos candidatos um curso de treinamento durante uma semana. No final do curso, eles são submetidos a uma prova e 25% são classificados como bons, 50% como médios e os restantes como fracos. Para facilitar a seleção a empresa pretende substituir o treinamento por um teste. Para isso, gostaria de conhecer qual a probabilidade de um indivíduo aprovado no teste ser considerado fraco caso fizesse o treinamento. Neste ano, antes do início do curso foi aplicado o teste. Sabe-se que a probabilidade do candidato ser aprovado no teste dado que ele é considerado bom no treinamento é 0,8, de ser aprovado no teste dado que ele é considerado médio no treinamento é 0,5 e de ser aprovado no teste dado que ele é considerado fraco no treinamento é 0,2.

Solucao:

Avaliação 03

1. Assume-se que o tempo entre acessos a um blog tem uma distribuição com média de 1,5 segundos. Assumindo a distribuição exponencial, responda os itens a seguir.
   (a) Qual a probabilidade de haver duas conexões com intervalo inferior a 1,5 segundos?
   (b) Qual a probabilidade de se passarem 5 segundos sem conexão alguma?
   (c) Tendo havido uma conexão, qual a probabilidade da próxima conexão ocorrer entre 0,5 e 2,5 segundos?
   (d) Se já se passou 1 segundo sem conexão, qual a probabilidade de se passar mais 0,5 segundos adicionais sem conexão?
   (e) Qual a probabilidade do intervalo entre conexões não superar 3,5 segundos se já se passaram 2 segundos sem conexão?

Solução:
A distribuição exponencial é uma possivel escolha considerando: (i) que devem ser valores positivos, (ii) pela possibilidade de cálculos com as informações fornecidas.

$$X : \text{intervalo de tempo entre conexões (segundos)}$$
$$X \sim \text{Exp}(\lambda = 1/1,5 = 2/3)$$
$$f(x) = \frac{2}{3}\mathrm{e}^{-2x/3}\ \mathrm{I}_{(0,\infty)}(x) \qquad F(x) = 1 - \mathrm{e}^{-2x/3}$$

(a) $P[X < 1,5] = \int_0^{1,5} f(x)\mathrm{d}x = F(1,5) = 0.63$
(b) $P[X > 5] = \int_5^{\infty} f(x)\mathrm{d}x = 1 - F(5) = 0.036$
(c) $P[X < 0,5] = \int_{0,5}^{2,5} f(x)\mathrm{d}x = F(2,5) - F(0,5) = 0.53$
(d) $P[X > 1,5|X > 1] = \frac{\int_{1,5}^{\infty} f(x)\mathrm{d}x}{\int_1^{\infty} f(x)\mathrm{d}x} = {}^1P[X > 0,5] = 1 - F(0,5) = 0.72$
(e) $P[X < 3,5|X > 2] = \frac{\int_2^{3,5} f(x)\mathrm{d}x}{\int_2^{3,5} f(x)\mathrm{d}x} = \frac{F(3,5)-F(2)}{1-F(2)} = {}^1P[X < 1,5] = F(1,5) = 0.63$

Solução computacional com o programa R:

```
> (pa <- pexp(1.5, rate=2/3))
[1] 0.6321
> (pb <- pexp(5, rate=2/3, lower=F))
[1] 0.03567
> (pc <- diff(pexp(c(0.5,2.5), rate=2/3)))
[1] 0.5277
> (pd <- pexp(0.5, rate=2/3, lower=F))
[1] 0.7165
> (pe <- pexp(1.5, rate=2/3))
[1] 0.6321
```

[1] propriedade de falta de memória da exponencial

2. Um indivíduo vai participar de uma competição que consiste em responder questões que são lhe são apresentadas sequencialmente. Com o nível de conhecimento que possui, a chance de acertar uma questão escolhida ao acaso é de 75% . Neste contexto, para cada diferente situação apresentada a seguir, defina a variável aleatória, sua distribuição de probabilidades e calcule a probabilidade solicitada. Se preciso, faça suposições necessárias e adequadas em cada caso.
   (a) Se for responder até errar uma pergunta, qual a probabilidade de conseguir acertar quatro ou mais questões?
   (b) Se for responder cinco perguntas, qual a probabilidade de acertar quatro ou mais?
   (c) Se for responder até acertar a terceira pergunta, qual a probabilidade de errar apenas uma?
   (d) Se o candidato selecionar aleatoriamente seis questões de um banco de 40 questões das quais o candidato sabe a resposta de 30 delas (75%), qual a probabilidade de acertar ao menos cinco delas.

Solução:

(a)

$$\begin{aligned} &X : \text{Número de acertos até o primeiro erro} \\ &X \sim \mathrm{G}(0,25) \\ &P[X \geq 4] = 1 - P[X \leq 3] = 1 - \sum_{i=0}^{3} (1-0,25)^i (0,25) = 0.316 \end{aligned}$$

(b)

$$\begin{aligned} &X : \text{Número de acertos em cinco perguntas} \\ &X \sim \mathrm{B}(\mathrm{n}=5, \mathrm{p}=0,75) \\ &P[X \geq 4] = P[X=4] + P[X=5] = \sum_{i=4}^{5} \binom{5}{i} 0,75^i (1-0,75)^{5-i} = 0.633 \end{aligned}$$

(c)

$$\begin{aligned} &X : \text{Número de erros até o terceiro acerto} \\ &X \sim \mathrm{BN}(\mathrm{r}=3, \mathrm{p}=0,75) \\ &P[X=1] = \binom{3+1-1}{3-1} 0,75^3 (1-0,75)^1 = 0.316 \end{aligned}$$

(d)

$$\begin{aligned} &X : \text{Número de acertos nas seis questões selecionadas} \\ &X \sim \mathrm{HG}(30,10,6) \\ &P[X \geq 5] = P[X=5] + P[X=6] = \sum_{i=5}^{6} \frac{\binom{30}{i}\binom{10}{6-i}}{\binom{40}{6}} = 0.526 \end{aligned}$$

Solução computacional com o programa *R*:

```
> (pa <- pgeom(3, prob=0.25, lower=F))
[1] 0.3164
> (pb <- pbinom(3, size=5, prob=0.75, lower=F))
[1] 0.6328
> (pc <- dnbinom(1, size=3, prob=0.75))
[1] 0.3164
> (pd <- phyper(4, m=30, n=10, k=6, lower=F))
[1] 0.526
```

Avaliação 04

1. A duração do ”tonner”da uma impressora pode ser modelada pela distribuição normal com média 10.000 cópias e desvio padrão de 1.200 cópias. A duração do ”tonner”será anotada e pergunta-se a probabilidade de ser:
   (a) inferior a 9.000 cópias;
   (b) inferior a 11.000 cópias;

(c) superior a 10.500 cópias;
(d) entre 9.200 e 10.000 cópias;
(e) entre 8.200 e 11.800 cópias.
(f) Qual a probabilidade de não de ser desviar da média mais que dois desvios padrões?
(g) Qual o número de cópias que espera-se que ao menos 80% dos tonners consiga imprimir?
(h) Se for definido como um limite máximo o número de cópias que apenas 2% dos tonners consegue atingir. qual será este limite máximo de cópias?
(i) Mantendo-se o desvio padrão de 1.200, qual deveria ser a média do número de impressões para que 90% dos tonners conseguisse imprimir ao menos 9.000 cópias.
(j) E se a média não puder ser alterada de 10.000, qual deveria ser o desvio padrão para garantir a condição acima?

Solução:

$$X : \text{duração do tonner (em número de cópias)}$$
$$X \sim N(\mu = 10.000, \sigma^2 = 1.200^2)$$

(a) $P[X < 9.000] = P[Z < \frac{9.000-10.000}{1.200}] = P[Z < -0.8333] = 0.2023$
(b) $P[X < 11.000] = P[Z < \frac{11.000-10.000}{1.200}] = P[Z < 0.8333] = 0.7977$
(c) $P[X > 10.500] = P[Z > \frac{10.500-10.000}{1200}] = P[Z > 0.4167] = 0.3385$
(d) $P[9.200 < X < 10.000] = P[\frac{9.200-10.000}{1.200} < Z < \frac{10.000-10.000}{1.200}] = P[-0.6667 < Z < 0] = 0.2475$
(e) $P[8.200 < X < 11.800] = P[\frac{8.200-11.800}{1.200} < Z < \frac{11.800-10.000}{1.200}] = P[-1.5 < Z < 1.5] = 0.8664$
(f) $P[-2\sigma < X < 2\sigma] = P[-2 < Z < 2] = 0.9545$
(g) $P[X > x_g] = 0,80 \rightarrow P[Z > -0.8416] = 0,80 \rightarrow -0.8416 = \frac{x_g-10.000}{1.200} \rightarrow x_g = 8990$
(h) $P[X > x_h] = 0,02 \rightarrow P[Z > 2.054] = 0,02 \rightarrow 2.054 = \frac{x_h-10.000}{1.200} \rightarrow x_h = 12464$
(i) $P[X > 9.000] = 0,90 \rightarrow P[Z > -1.282] = 0,90 \rightarrow -1.282 = \frac{9.000-\mu}{1.200} \rightarrow \mu = 10538$
(j) $P[X > 9.000] = 0,90 \rightarrow P[Z > -1.282] = 0,90 \rightarrow -1.282 = \frac{9.000-10.000}{\sigma} \rightarrow \sigma = 780.3$

Solução computacional com o sistema **R**:

```
> (qa <- pnorm(9000, m=10000, sd=1200))
[1] 0.2023
> (qb <- pnorm(11000, m=10000, sd=1200))
[1] 0.7977
> (qc <- pnorm(10500, m=10000, sd=1200, low=F))
[1] 0.3385
> (qd <- diff(pnorm(c(9200, 10000), m=10000, sd=1200)))
[1] 0.2475
> (qe <- diff(pnorm(c(8200, 11800), m=10000, sd=1200)))
[1] 0.8664
> (qf <- diff(pnorm(c(-2, 2))))
[1] 0.9545
> (qg <- qnorm(0.20, m=10000, sd=1200))
[1] 8990
> (qh <- qnorm(0.98, m=10000, sd=1200))
[1] 12464
> (qi <- 9000 - qnorm(0.1)*1200)
[1] 10538
> (qj <- (9000 - 10000)/qnorm(0.1))
[1] 780.3
```

Um conjunto de imagens (1 a 10) foi submetido a dois algoritmos (A e B) de tratamento (filtragem, correção e classificação) e foram registrados os tempos de processamento. Alguns resumos dos dados encontram-se a seguir.

$$\overline{x}_A = 36.19 \quad \overline{x}_B = 22.98$$
$$S_A = 17.62 \quad S_B = 17.14$$

Responda as questões a seguir baseando-se nos resumos dados e justificando as respostas.

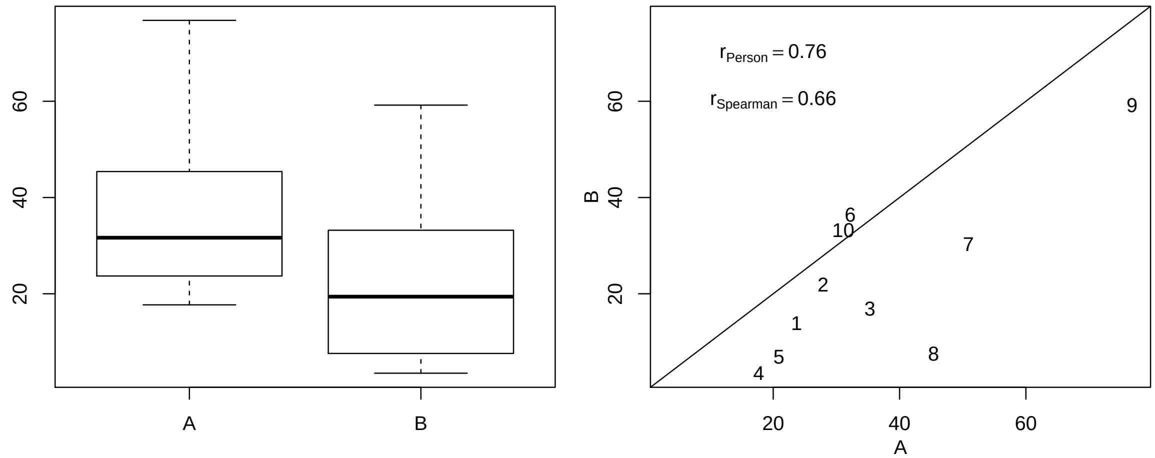

Figura 1: Box-plot e diagrama de dispersão dos tempos de processamento de dois algorítmos aplicados a um mesmo conjunto de problemas

1. Descreva o comportamento cada um dos algorítmos individualmente e compare os seus desempenhos.
2. Existem observações discrepantes (atípicas)? Dê respostas baseando-se em cada um dos gráficos.
3. Como voce descreveria a relação e correlação entre o desempenho dos algorítmos?
4. Os algorítmos possuem variabilidades relativas, medida pelo coeficiente de variação, semelhantes?
5. Os algorítmos possuem variabilidades, medida pela amplitude interquartílica, semelhantes?

Avaliação 06

Em uma avaliação de um novo algoritmo de classificação foi analisada uma amostra de 1200 cenários dentre os quais 780 foram classificados corretamente.

1. Identifique no contexto do problema: a população ("informal" e estatística), a amostra, o parâmetro, o estimador, a distribuição amostral e a estimativa (pontual).
2. Obtenha a estimativa pontual e intervalar (95% de confiança) para a proporção de classificações corretas.
3. Qual deveria ser o tamanho da amostra para que a margem de erro fosse de 1,5% (com 95% de confiança)?
4. Um algoritmo atualmente utilizado possui um percentual de acerto de 62%. Há evidências baseadas no estudo de que o novo algoritmo tem desempenho diferente do utilizado atualmente? Justifique sua resposta.
5. Quais as suposições relevantes para os cálculos feitos nos item anteriores?

Solução:

$$X : \text{resultado da classificação (correto/incorreto)}$$
$$X \sim \mathrm{B}(p) \quad \mathrm{E}[X] = p \quad \mathrm{Var}[X] = p(1-p)$$
$$\hat{p} \sim \mathrm{N}(p, p(1-p)/n)$$

1.

2.

$$\hat{p} = \frac{780}{1200} = 0.65$$

I.C. assintótico:

$$IC_{95\%} : \hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \longrightarrow (0.623 \;\; ; \;\; 0.677)$$

I.C.conservador :

$$IC_{95\%} : \hat{p} \pm z\sqrt{\frac{1}{4n}} \longrightarrow (0.622 \;\; ; \;\; 0.678)$$

3. (a) Utilizando $p = \hat{p}$

$$\begin{aligned}
\text{ME} &= z_{95\%}\sqrt{\frac{p(1-p)}{n}} \\
0.015 &= 1.96\sqrt{\frac{0.65(1-0.65)}{n}} \\
n &= \lceil \frac{1.96^2}{0,015^2} 0.65(1-0.65) \\
n &= 3885
\end{aligned}$$

(b) Utilizando $p = 0,5$

$$\begin{aligned}
\text{ME} &= z_{95\%}\sqrt{\frac{0,5(1-0,5)}{n}} = z_{95\%}\sqrt{\frac{1}{4n}} \\
0.015 &= 1.96\sqrt{\frac{0.5(1-0.5)}{n}} \\
n &= \lceil \frac{1.96^2}{0,015^2} \frac{1}{4} \\
n &= 4269
\end{aligned}$$

4. Resposta e justificativa baseada no valor estar ou não contido no I.C..

5.
