

## Prova P1

FT043 (Fundamentos da Ciência de Dados) – 2S2024 – Prof. Leonardo Tomazeli Duarte – FCA/UNICAMP

Prazo de entrega: 4/10/2024 (deve ser feito via Google Classroom)

**Para realização da prova, você deverá considerar a seguinte correspondência:**

$a_1$  = Primeiro dígito do seu RA ;  $a_2$  = Segundo dígito do RA ;  $a_3$  = Terceiro dígito do RA ...  $a_6$  = Último dígito do seu RA.

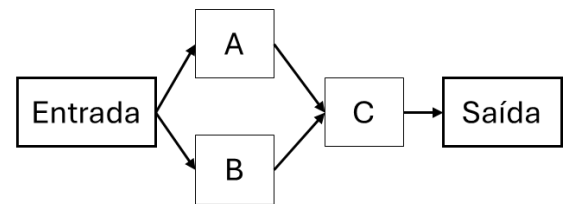
Exemplos: RA=378917  $\rightarrow a_1 = 3$ ;  $a_2 = 7$ ;  $a_5 = 1$  ;  $a_6 = 7$ ;  $a_1a_3 = 38$ ;  $a_4a_2 = 97$

**A resolução da parte teórica deve ser feita em folha de papel e redigida à caneta ou lápis. Enviar foto da resolução, preferencialmente num arquivo único em pdf.**

**A resolução da parte computacional deve ser feita por meio do Google Colab ou outra ferramenta de Notebook. Em todo caso, anexar códigos e respostas**

### PARTE TEÓRICA (4 pontos)

1. Para que o sistema elétrico da Figura 1 funcione corretamente, é necessário haver, pelo menos, um caminho fechado entre a entrada e a saída. As probabilidades de funcionamento dos componentes que integram esse sistema são dadas por  $P(A) = 9a_4\%$ ,  $P(B) = 85\%$  e  $P(C) = \alpha$ , e as falhas ocorrem de modo independente. Diante disso:



**Figura 1.** Sistema elétrico.

**(a)(1 ponto)** Determine  $\alpha$  de modo que a probabilidade do circuito funcionar seja superior a 80%.

**(b) (1 ponto)** Para um valor de  $\alpha = 90\%$ , obtenha a probabilidade de B funcionar, dado que o sistema funciona.

2. Deseja-se estimar a média de gastos anual de todos os clientes de um aplicativo de entrega de comidas. A variabilidade desses gastos é conhecida, de modo que o desvio padrão é dado por 100 Reais. Para realizar a estimação desejada, foram coletados (de maneira aleatória) os gastos anuais de 121 clientes, de modo que a média amostral obtida foi de 645 Reais por ano. Diante disso:

**(a) (1 ponto)** Obtenha um intervalo de confiança, de nível de confiança de  $9a_4\%$ , para a estimação do gasto médio anual dos clientes do aplicativo. Determine a respectiva margem de erro.

**(b) (1 ponto)** Seja ME a margem de erro obtida no item (a) e considerando o mesmo nível de confiança do item (a), qual seria o número mínimo de amostras necessárias para se obter uma margem de erro inferior a  $ME/4$ ?

### PARTE COMPUTACIONAL (6 pontos)

3. Considere os dados dos arquivos *banco\_dados\_salarios.xlsx*. Tais dados representam os salários mensais (em kR\$) de 20000 pessoas de uma região de uma grande cidade. Cada entrada da tabela, que corresponde a uma pessoa, também apresenta o gênero da pessoa. Diante disso, realize o seguinte procedimento.

**(a)** Calcule a média salarial desta população.

**(b)** Obtenha os histogramas e os box-plots para cada um dos grupos do banco de dados.

**(c)** Obtenha um intervalo de confiança de  $9a_5\%$  considerando  $n = 200$  amostras dessa população. Considere a situação na qual o desvio populacional é conhecido. Tal desvio pode ser calculado a partir considerando todas as amostras da população (todas as entradas da planilha fornecida).

**(d)** Realize 5000 realizações (com sorteios diferentes) do item (c), ou seja, obtenha 5000 intervalos de confiança com as especificações do item (c). Discuta o resultado obtido à luz do valor da média populacional obtida no item (a). Em particular, calcule a proporção de vezes (nessas 5000 realizações) em que o intervalo obtido conteve a média populacional.

**(e)** Realize o mesmo procedimento do item (d), porém apenas para as mulheres. Discuta o resultado obtido à luz do valor da média populacional obtida no item (a). Em particular, calcule a proporção de vezes (nessas 5000 realizações) em que o intervalo obtido conteve a média populacional. Discuta possíveis diferenças com relação ao item (d).

**(f)** Considere que uma pesquisa estatística para estimação da média populacional (dada pelo banco de dados fornecido) teve acesso a 100 amostras de mulheres e 50 amostras de homens. Proponha uma estratégia para se obter um intervalo de confiança sem nenhum tipo de viés. Valide a sua ideia por meio de um experimento computacional.