

Atividade 3: Lista de exercícios

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

Prazo de entrega: 30/09/2024 (deve ser feito via Google Classroom)

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

Para realização dos exercícios abaixo, 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 $\rightarrow a_1 = 3$; $a_2 = 7$; $a_3 = 8$; $a_4 = 9$; $a_5 = 1$; $a_6 = 7$; $a_1a_3 = 38$; $2a_2 = 27$; $a_4a_2a_3 = 978$; $0,a_4 = 0,9$

1. Um estudo foi planejado para investigar a eficácia de uma nova vacina para o tratamento de uma grave doença em diferentes faixas etárias. Os desfechos delineados no estudo foram os seguintes: i) paciente não apresentou doença; ii) paciente apresentou doença com sintomas leves; iii) paciente apresentou doença com sintomas graves; iv) paciente veio a óbito. A Tabela 1 apresenta o conjunto de dados obtidos no estudo, considerando unicamente o grupo de controle, ou seja, as pessoas que receberam a vacina.

Tabela 1. Dados levantados no estudo para eficácia da vacina.

Faixa etária (anos)	Desfechos para o grupo de pessoas que receberam a vacina (controle)			
	Sem infecção	Sintomas leves	Sintomas graves	Óbitos
0-19	10 a_4 1	20	2	0
20-44	70 a_4	242	10	0
45-54	48 a_5	44 a_6	27	9
55-64	3 a_6 1	475	45	1 a_5
> 64	1 a_4 2	5 a_3 1	60	30

- (a) Qual a probabilidade de uma pessoa escolhida ao acaso neste grupo apresentar sintomas graves?
- (b) Qual a probabilidade de uma pessoa escolhida ao acaso neste grupo estar na faixa etária 45-54 anos e vir a óbito?
- (c) Qual a probabilidade de uma pessoa escolhida ao acaso neste grupo estar na faixa etária 20-44 anos ou apresentar sintomas leves (ou ambos)?
- (d) Dado que uma pessoa veio a óbito, qual é a probabilidade de ela estar na faixa etária > 64 anos?

- (e) Os eventos “Sintomas graves” e “faixa etária > 64 anos” são estatisticamente independentes? Justifique e comente possíveis implicações para o entendimento do estudo.
- (f) Dado que uma pessoa está na faixa 20-44, qual é a probabilidade de ela não ser infectada após receber a vacina?
- (g) Dado que uma pessoa não foi infectada, qual é a probabilidade de ela estar nas faixas “faixa etária > 64 anos” ou “faixa etária 20-44 anos”?

2. Um fabricante de peças deseja controlar a qualidade de seus produtos descartando lotes que apresentem um alto valor de peças defeituosas. Se em um lote de $9a_5$ peças, ao menos $2a_6$ peças defeituosas forem encontradas, então o lote deverá ser descartado. Diante disso calcule:

- (a) A probabilidade de um lote ser rejeitado no caso em que uma peça defeituosa ocorre com probabilidade de 5%.
- (b) A probabilidade de um lote ser rejeitado no caso em que uma peça defeituosa ocorre com probabilidade de $1a_4\%$.
- (c) A probabilidade de um lote ser rejeitado no caso em que uma peça defeituosa ocorre com probabilidade de $2a_3\%$.
- (d) A probabilidade de um lote ser rejeitado no caso em que uma peça defeituosa ocorre com probabilidade de $3a_5\%$.
- (e) Trace a função distribuição de probabilidade para cada um dos casos expressos nos itens (a), (b), (c) e (d).

3. A lei de Benford se originou a partir de uma interessante observação do físico Frank Benford nos anos 1930. Ele observou que, em muitas aplicações, a distribuição de probabilidade do primeiro dígito significativo dos números observados não é uniforme, ou seja, há dígitos mais prováveis que outros. Mais precisamente, a lei de Benford indica que o primeiro dígito de um número pode ser modelado por uma variável aleatória discreta D , de modo que a função distribuição de probabilidade de D é dada por:

$$f_D(d) = \log_{10} \left(1 + \frac{1}{d} \right), \quad d = 1, 2, 3, \dots, 9.$$

- (a) Esboce no Python a função distribuição de probabilidade associada à lei de Benford
- (b) Mostre que $f_D(d)$ satisfaz as condições para ser uma variável aleatória.
- (c) Considerando a lei de Benford. Calcule a probabilidade do primeiro dígito ser um número par e a probabilidade do primeiro dígito ser um número ímpar.
- (d) Calcule a esperança $E\{D\}$ e a variância σ_D^2 .
- (e) Verifique no Python que a média e a variância calculadas para números gerados a partir de D convergem para os valores de esperança e variância calculados no item (d).