



**Universidade Federal de Uberlândia**  
**Faculdade de Matemática - Prof.: José Waldemar**  
**Disciplina: Est. Comp. - FAMAT31307 - Lista 01 - R e Python**  
**Curso: Estatística - 2024.1**

**1ª Questão**

Considere os dados a seguir, os quais constam no livro “Noções de Probabilidade e Estatística” (MAGALHÃES e LIMA, 2004) e também estão disponibilizado no Teams.

**Tabela.** Informações de um questionário aplicado a estudantes de duas turmas do primeiro ano de uma escola.

Id	Turma	Sexo	Idade	Alt	Peso	Filhos	Fuma	Toler	Exerc	Cine	OpCine	TV	OpTV
1	A	F	17	1,60	60,50	2	NAO	P	0	1	B	16	R
2	A	F	18	1,69	55,00	1	NAO	M	0	1	B	7	R
3	A	M	18	1,85	72,80	2	NAO	P	5	2	M	15	R
4	A	M	25	1,85	80,90	2	NAO	P	5	2	B	20	R
5	A	F	19	1,58	55,00	1	NAO	M	2	2	B	5	R
6	A	M	19	1,76	60,00	3	NAO	M	2	1	B	2	R
7	A	F	20	1,60	58,00	1	NAO	P	3	1	B	7	R
8	A	F	18	1,64	47,00	1	SIM	I	2	2	M	10	R
9	A	F	18	1,62	57,80	3	NAO	M	3	3	M	12	R
10	A	F	17	1,64	58,00	2	NAO	M	2	2	M	10	R
11	A	F	18	1,72	70,00	1	SIM	I	10	2	B	8	N
12	A	F	18	1,66	54,00	3	NAO	M	0	2	B	0	R
13	A	F	21	1,70	58,00	2	NAO	M	6	1	M	30	R
14	A	M	19	1,78	68,50	1	SIM	I	5	1	M	2	N
15	A	F	18	1,65	63,50	1	NAO	I	4	1	B	10	R
16	A	F	19	1,63	47,40	3	NAO	P	0	1	B	18	R
17	A	F	17	1,82	66,00	1	NAO	P	3	1	B	10	N
18	A	M	18	1,80	85,20	2	NAO	P	3	4	B	10	R
19	A	F	20	1,60	54,50	1	NAO	P	3	2	B	5	R
20	A	F	18	1,68	52,50	3	NAO	M	7	2	B	14	M
21	A	F	21	1,70	60,00	2	NAO	P	8	2	B	5	R
22	A	F	18	1,65	58,50	1	NAO	M	0	3	B	5	R
23	A	F	18	1,57	49,20	1	SIM	I	5	4	B	10	R
24	A	F	20	1,55	48,00	1	SIM	I	0	1	M	28	R
25	A	F	20	1,69	51,60	2	NAO	P	8	5	M	4	N
26	A	F	19	1,54	57,00	2	NAO	I	6	2	B	5	R
27	B	F	23	1,62	63,00	2	NAO	M	8	2	M	5	R
28	B	F	18	1,62	52,00	1	NAO	P	1	1	M	10	R
29	B	F	18	1,57	49,00	2	NAO	P	3	1	B	12	R
30	B	F	25	1,65	59,00	4	NAO	M	1	2	M	2	R
31	B	F	18	1,61	52,00	1	NAO	P	2	2	M	6	N
32	B	M	17	1,71	73,00	1	NAO	P	1	1	B	20	R
33	B	F	17	1,65	56,00	3	NAO	M	2	1	B	14	R
34	B	F	17	1,67	58,00	1	NAO	M	4	2	B	10	R
35	B	M	18	1,73	87,00	1	NAO	M	7	1	B	25	B
36	B	F	18	1,60	47,00	1	NAO	P	5	1	M	14	R
37	B	M	17	1,70	95,00	1	NAO	P	10	2	M	12	N
38	B	M	21	1,85	84,00	1	SIM	I	6	4	B	10	R
39	B	F	18	1,70	60,00	1	NAO	P	5	2	B	12	R
40	B	M	18	1,73	73,00	1	NAO	M	4	1	B	2	R
41	B	F	17	1,70	55,00	1	NAO	I	5	4	B	10	B
42	B	F	23	1,45	44,00	2	NAO	M	2	2	B	25	R
43	B	M	24	1,76	75,00	2	NAO	I	7	0	M	14	N
44	B	F	18	1,68	55,00	1	NAO	P	5	1	B	8	R
45	B	F	18	1,55	49,00	1	NAO	M	0	1	M	10	R
46	B	F	19	1,70	50,00	7	NAO	M	0	1	B	8	R
47	B	F	19	1,55	54,50	2	NAO	M	4	3	B	3	R
48	B	F	18	1,60	50,00	1	NAO	P	2	1	B	5	R
49	B	M	17	1,80	71,00	1	NAO	P	7	0	M	14	R
50	B	M	18	1,83	86,00	1	NAO	P	7	0	M	20	B

**Codificação e respectivos significados das variáveis:**

Id: Identificação do estudante	Toler: Indiferente (I), Incomoda Pouco (P) e Incomoda Muito (M)
Turma: Turma a que o aluno pertence	Exerc: Horas de exercício físico por semana
Sexo: F se feminino e M se masculino	Cine: Números de vezes que vai ao cinema por semana.
Idade: Idade em anos	OpCine: Opinião a respeito do cinema da cidade: se regular a boa B e se muito boa M
Alt: Altura em metros	TV: Horas assistindo TV por semana
Peso: Peso em kg	OapTV: opinião a respeito da qualidade da programa da TV: R-ruim, M-Média, B-Boa e N-não sabe
Filhos: Número de filhos na família	
Fuma: Se fuma SIM se não, NAO	

- a) Mostre que não existem dados ausentes.
- b) Tanto no R quanto no Python, exclua a variável Id do banco de dados.
- c) Apresente a média, mediana, valor mínimo e valor máximo, variância, desvio padrão e coeficiente de variação para todas as variáveis quantitativas.
- d) Apresente a média, mediana, valor mínimo e valor máximo, variância, desvio padrão e coeficiente de variação para todas as variáveis quantitativas, por turma. As turmas parecem similar quanto à média em todas as variáveis? E quanto à dispersão? Justifique.
- e) Você utilizaria o histograma para descrever graficamente todas as variáveis numéricas deste banco de dados? Justifique.
- f) Descreva todas as variáveis por meio de gráficos observando o tipo de gráfico mais adequado para cada uma.
- g) Descreva todas as variáveis por meio de gráficos observando o tipo de gráfico mais adequado para cada uma, por turma. O padrão das distribuições parece similar nas duas turmas? Justifique.

**Obs.: Obtenha os resultados tanto no R quanto no Python.**

### 2ª Questão

Sabe-se que quanto maior a amostra menor será o erro de estimativa. Para ilustrar este fato, gere amostras de uma distribuição normal com média 30 e desvio padrão 3 ( $\mu = 30$  e  $\sigma = 3$ ), começando com uma amostra de tamanho um ( $n = 1$ ) e aumentando de uma em uma unidade, iterativamente, até que o erro de estimativa seja menor que 0,001. Lembre-se que o estimador natural de  $\mu$  é a média amostral ( $\bar{x}$ ) e portanto as amostras devem aumentar a cada iteração até que  $|\bar{x} - 30| < 0,001$ . Utilize a estrutura *while* no R.

### 3ª Questão

O Teorema do Limite Central garante que se cada amostra for grande o suficiente ( $n$  tendendo a infinito), a distribuição da média amostral é aproximadamente normal. Se a população onde são realizadas as amostras for normal então, a distribuição da média amostral será exatamente normal. Para ilustrar este fato, retire amostras de uma distribuição gama

$$f(x : \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\text{Gamma}(\alpha)} \quad \text{com } x > 0, \alpha > 0 \text{ e } \beta > 0 \quad (1)$$

mais especificamente uma distribuição gama com parâmetros  $\alpha = 2$  e  $\beta = 10$  utilizando o comando *rgamma(n,2,10)* do R. Realize amostras repetidamente (muitas vezes) de tamanho 5 ( $n=5$ ), o suficiente para ilustrar a distribuição da média empiricamente (por meio de um histograma). Repita este mesmo procedimento para  $n = 10, 15, 20, 30, 40, 50, 70, 100, 200$  e observe o formato do histograma em cada caso. A partir de qual tamanho amostral o histograma parece indicar que a distribuição da variável média amostral é normal? Teoricamente tem-se normalidade com  $n \rightarrow \infty$ ) mas, na prática, qual a partir de qual tamanho de amostral testado você diria que a hipótese de normalidade seria aceita? Justifique.

#### 4ª Questão

A distribuição beta é uma distribuição de probabilidade contínua definida no intervalo  $[0; 1]$  parametrizada por dois parâmetros positivos, denotados por  $\alpha$  e  $\beta$ , ou,  $a$  e  $b$  no R, que aparecem como expoentes da variável aleatória e controlam o formato da distribuição. A expressão desta distribuição pode ser vista no help do R, acionado com o comando `?rbeta` e, dentro do help, localizando o tópico **Details**.

A percentagem de impurezas por lote, em um determinado produto químico, é uma variável aleatória com distribuição beta de parâmetros  $\alpha = 4$  e  $\beta = 2$ . Um lote com mais de 40% ( $x = 0,4$ ) de impurezas não pode ser vendido.

a) Qual é a probabilidade de que um lote selecionado ao acaso não possa ser vendido por causa do excesso de impurezas? E qual a probabilidade de que ele possa ser vendido?

b) Se a distribuição beta fosse descrita com os parâmetros  $\alpha = 2$  e  $\beta = 4$ , quais seriam as respostas às questões apresentadas no item a.

c) Ilustre em um mesmo gráfico as duas curvas, isto é, para a distribuição beta com parâmetros  $\alpha = 4$  e  $\beta = 2$  e com parâmetros  $\alpha = 2$  e  $\beta = 4$ .