

Nome(legível): \_\_\_\_\_

**ATENÇÃO:** Os valores de cada questão estão entre colchetes no início de cada item. A questão 4 é obrigatória. Para os itens envolvendo cálculo, o código em R correspondente deve ser indicado. Caso você tenha feito o trabalho, faça no máximo o total de 6.0 pontos na prova, indicados através de círculos nos itens correspondentes. Itens sem círculo não serão corrigidos. Cada item feito a mais na prova será considerado para a subtração da nota final.

1. Na tabela abaixo temos dados sobre hábitos de fumo de uma amostra de moradores de uma pequena cidade (dados fictícios).

Hábitos de Fumo	Idade			Total
	< 20	[20, 30)	$\geq 30$	
Fumante	143	171	40	354
Ex-fumante	11	152	140	303
Nunca fumou	66	57	20	143

- (a) [0.5] Defina claramente as variáveis envolvidas, estabelecendo o tipo de cada uma.
- (b) [0.5] É possível estabelecer uma relação dependente/independente entre elas? Em caso afirmativo, qual é a variável independente e qual é a variável dependente?
- (c) [1.0] Construa o gráfico apropriado para representar esses dados.
2. Considere os dados apresentados na tabela a seguir, que contém as populações, em 1000 habitantes, dos estados brasileiros.

RR	325	MS	2079	PB	3444
PR	9564	AP	478	MT	2505
GO	5004	RS	10188	AC	558
RN	2777	SC	5357	BA	13071
TO	1158	AM	2813	MA	5652
RJ	14392	RO	1380	AL	2823
PA	6193	MG	17892	SE	1785
PI	2844	CE	7431	SP	37033
DF	2052	ES	3098	PE	7919

- (a) [0.2] Construa o rol das observações.
- (b) [0.4] Calcule todos os quartis.
- (c) [0.4] Encontre os *outliers*.
- (d) [1.0] Construa o boxplot.

3. Considere os dados das tabelas a seguir, referentes a um levantamento feito por professores da Universidade Federal de Santa Catarina (UFSC), no qual o principal objetivo era avaliar os efeitos políticos dos programas de alimentação popular.

Os dados são referentes a 40 famílias residentes em três locais: Conjunto Residencial Monte Verde, Conjunto Residencial Parque da Figueira e na Encosta do Morro. As variáveis apresentadas são:

- PAP: variável indicadora de uso de programas de alimentação (1 = Sim, 0 = Não),
- GI: grau de instrução do chefe da casa (1 = nenhum grau oficialmente completo, 2 = primeiro grau completo, 3 = segundo grau completo),
- RES: número de pessoas residentes na casa,
- RENDA: renda familiar mensal, em salários mínimos.

Ident.	PAP	GI	RES	RENDA	Ident.	PAP	GI	RES	RENDA
1	0	3	4	10,3	21	1	3	5	5,8
2	0	3	4	15,4	22	1	3	5	12,9
3	1	2	4	9,6	23	0	3	5	7,7
4	0	2	5	5,5	24	0	2	4	1,1
5	1	3	4	9,0	25	0	2	8	7,5
6	1	1	1	2,4	26	1	3	4	5,8
7	0	3	2	4,1	27	1	1	5	7,2
8	1	3	3	8,4	28	0	3	3	8,6
9	1	3	6	10,3	29	1	2	4	5,1
10	1	2	4	4,6	30	0	3	5	2,6
11	0	2	6	18,6	31	1	3	5	7,7
12	1	1	4	7,1	32	1	2	2	2,4
13	0	2	4	12,9	33	1	3	5	4,8
14	0	2	6	8,4	34	1	1	2	2,1
15	0	3	3	19,3	35	1	1	6	4,0
16	0	2	5	10,4	36	1	1	8	12,5
17	1	3	3	8,9	37	1	3	3	6,8
18	0	3	4	12,9	38	1	3	5	3,9
19	0	3	4	5,1	39	1	3	5	9,0
20	1	3	4	12,2	40	1	3	3	10,9

Tabela 1: Conjunto residencial Monte Verde

- [0.4] Classifique as variáveis da pesquisa de acordo com o seu tipo.
- [1.2] Para as variáveis qualitativas e quantitativa discreta, construa tabelas de frequência sem perda de informação, considerando as três localidades em conjunto.
- [0.8] Para a variável quantitativa contínua, construa uma tabela de frequências trabalhando com quatro classes de mesmo tamanho.
- [0.6] Trabalhar com classes de mesmo tamanho não é recomendável em análises de distribuição de renda, pois geralmente há uma grande concentração em poucas classes. Defina novas classes para que a distribuição das classes fique mais detalhada.

4. Considere a tabela a seguir com o peso e a altura de cinco homens e cinco mulheres.

Peso ( $P$ )	Altura Homens ( $H$ )	Altura Mulheres ( $M$ )
80	1,80	—
85	1,83	—
50	—	1,65
70	—	1,90
55	—	1,60
77	1,80	—
85	—	1,78
93	1,86	—
65	1,70	—
60	—	1,65

Tabela 2: Peso e a altura de cinco homens e cinco mulheres

- [0.4] Construa um diagrama de dispersão para esta tabela.
- [0.6] Calcule o coeficiente de correlação linear para as variáveis  $P \times H$ ,  $P \times M$  e  $P \times (H + M)$ .
- [1.2] Encontre a reta de regressão linear para os três casos.
- [0.4] Combine o diagrama de dispersão e as três retas encontradas.
- [0.4] Faça uma análise crítica dos resultados obtidos.

### Fórmulas

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = n \left( \sum_{i=1}^n x_i^2 \right) - \left( \sum_{i=1}^n x_i \right)^2$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = n \left( \sum_{i=1}^n y_i^2 \right) - \left( \sum_{i=1}^n y_i \right)^2$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = n \left( \sum_{i=1}^n x_i \cdot y_i \right) - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

$$y = \alpha + \beta x \quad \beta = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2} \quad \alpha = \frac{\sum y_i - \beta \sum x_i}{n}$$