

Estudante: _____ Turma: _____ 04/04/2025

Atividade Avaliativa - **Análise exploratória de dados utilizando o R**

Orientação geral:

- ✓ Desenvolva as análises no R, reportando os resultados no word de modo organizado, na mesma sequência em que as questões são apresentadas. Ao final gere um pdf sendo nome do arquivo o seu nome. Exemplo **FULANO_DA_SILVA**
- ✓ Incluir as **sintaxes** que você utilizou no **Apêndice (fim do arquivo)**;
- ✓ Data de entrega: **25/04/2025**. Não é necessário imprimir. Entregar via **e-mail** jeovani.schmitt@ifc.edu.br No assunto do e-mail informar **ATIVIDADE DE ESTATÍSTICA**

1. Com o objetivo de conhecer o perfil de 110 clientes, uma seguradora solicitou uma análise estatística. Para tanto disponibilizou a base **seguro.csv**, conforme abaixo:

A	B	C	D	E	F	G
ID	sinistro	idade	estado	tempo	renda	instrucao
001	não	22	solteiro	3	2,1	fundamental
002	sim	24	casado	4	2,2	fundamental
003	não	45	casado	20	4	médio
004	não	58	solteiro	25	2,3	médio
005	não	27	casado	7	4,5	superior
006	sim	31	solteiro	10	4,5	superior

Dados dos 5
primeiros clientes
da amostra

Dicionário de variáveis:

sinistro	não quando não tem sinistro ou sim quando tem
idade	idade do segurado em anos completos
estado	estado civil do segurado, sendo solteiro e casado
tempo	tempo em anos que possui habilitação
renda	renda em milhares de reais
instrucao	escolaridade do segurado: fundamental, médio ou superior
ID	indicador sequencial de 001 a 110 (não é variável)

Utilizando a base dados fornecida, realize uma análise exploratória de dados:

- Construa uma distribuição de frequências (uma tabela simples) para a variável **sinistro**; (incluir a frequência absoluta e relativa)
- Construa um gráfico de colunas para a variável **instrucao**;
- Construa uma tabela cruzada para as variáveis **sinistro x estado**;
- Construa um gráfico de colunas empilhadas para as variáveis **sinistro e instrucao**;

- e) Para a variável **renda**, construir um Histograma e um Box plot. Há presença de valores discrepantes? Classifique a variável em relação à assimetria.
- f) Calcular para a **renda** as seguintes estatísticas: média, mediana, desvio padrão, coeficiente de variação e quartis (1º. e 3º.).
- g) Construir uma tabela para a variável **renda** categorizando-a em 4 faixas:
 Até 1,9
 2.0 a 3.9
 4.0 a 6.9
 7.0 ou mais
- h) Construir um Box plot e compare a variável **renda por instrucao** (um box para cada **instrução na mesma figura**)
- i) **Para a variável renda**, calcular média, mediana, quartis, desvio padrão e coeficiente de variação **por instrução**.
 (estatísticas para fundamental, médio e superior)

2. Considere que a base **seguros.csv** se refere a uma população de segurados (N = 110). Determinar o tamanho da amostra para estimar a **idade** média dos segurados dessa população, com margem de erro de **6** anos e nível de confiança de 95%, considere o desvio padrão igual a **14 anos**.

$$n_0 = \frac{z^2 \sigma^2}{d^2}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

$$\begin{cases} z = \\ \sigma = \\ d = \end{cases}$$

3. Com o tamanho da amostra calculado na questão anterior, faça o sorteio usando o método **ASA**, estime a **idade** média dos segurados (\bar{x}) e construa o intervalo de precisão com a margem de erro de **6** anos. O intervalo construído contém a verdadeira idade média populacional μ ?