



<b>Iniciado em</b>	sexta-feira, 19 abr. 2024, 10:52
<b>Estado</b>	Finalizada
<b>Concluída em</b>	sexta-feira, 19 abr. 2024, 11:31
<b>Tempo empregado</b>	39 minutos 3 segundos

**Questão 1**  
Completo  
Vale 2,00 ponto(s).

Sobre a visualização e exploração dos dados, assinale V ou F (verdadeiro ou falso)

I - A análise exploratória deve idealmente ser realizada de forma personalizada para cada conjunto de dados. Ferramentas de visualização como gráficos de barras e de setores, comumente conhecidos como gráficos de pizza, são usados para descrever dados qualitativos, enquanto que histogramas e boxplots são usados para descrever dados quantitativos.

II - Os boxplots ou gráficos de caixa estão entre as representações mais informativas para variáveis quantitativas, pois apresentam medidas-resumo de localização, dispersão, assimetria e presença de pontos atípicos.

III - A associação entre variáveis qualitativas pode ser analisada por meio de tabelas ou gráficos e deve ser confirmada com análises inferenciais, como testes de hipóteses.

IV - Os boxplots auxiliam na análise de associação entre variáveis qualitativas e quantitativas, e dispensam o uso de testes estatísticos para avaliar a significância estatística das diferenças encontradas graficamente.

A alternativa mais adequada para avaliar as afirmações é:

- ☒ a. V - V - V - F
- ☐ b. F - F - F - V
- ☐ c. F - V - V - F
- ☐ d. V - F - V - V
- ☐ e. V - V - V - V

**Questão 2**  
Completo  
Vale 2.00 ponto(s).

Sobre as medidas-resumo usadas para caracterizar a distribuição de dados quantitativos, são feitas as afirmações abaixo

I - A média amostral tem boas propriedades estatísticas mas pode ser influenciada por observações atípicas, especialmente em conjuntos de dados pequenos, enquanto que a mediana é uma medida mais resistente, no sentido de que é menos afetada do que a média por dados atípicos.

II - Média, moda, mediana, quartis e coeficiente de correlação linear são medidas de posição usadas para descrever variáveis quantitativas.

III - O coeficiente de variação descreve a razão entre o desvio-padrão e a média e é uma medida adimensional.

IV - Desvio-padrão e amplitude são medidas de dispersão usadas para descrever dados quantitativos.

Estão corretos os itens:

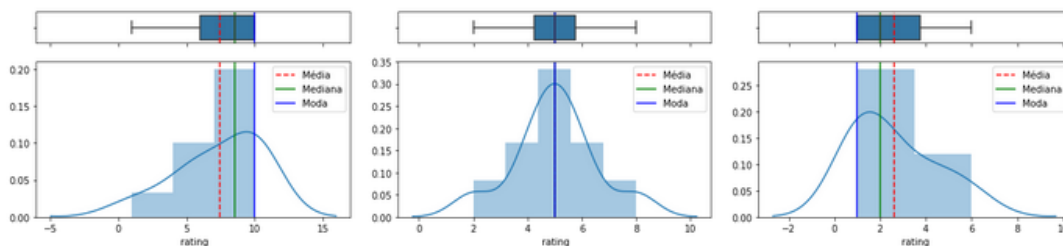
- ☒ a. I, III e IV apenas
- ☐ b. II, III e IV apenas
- ☐ c. III apenas
- ☐ d. I e II apenas
- ☐ e. I, II, III apenas

### Questão 3

Completo

Vale 2,00 ponto(s).

As medidas de posição média, mediana e moda são frequentemente utilizadas para descrever a distribuição de dados quantitativos. Considere para esta questão, distribuições bem comportadas e unimodais, isto é, que possuem uma única moda, apresentadas a seguir na ordem: distribuição assimétrica à esquerda, distribuição simétrica e distribuição assimétrica à direita. Assinale a alternativa **incorreta**:



- ☐ a. Em dados assimétricos à direita, temos que moda < mediana < média.
- ☐ b. Em dados assimétricos à esquerda, temos que moda < mediana < média.
- ☐ c. Em boxplots, é possível observar a assimetria comparando as distâncias entre mediana e primeiro quartil e entre mediana e

terceiro quartil.

- ☐ d. Nos gráficos, são representados histogramas, densidade alisada e boxplots, que são representações adequadas para dados quantitativos.
- ☒ e. Em dados simétricos, moda, média e mediana coincidem.

#### Questão 4

Completo

Vale 2,00 ponto(s).

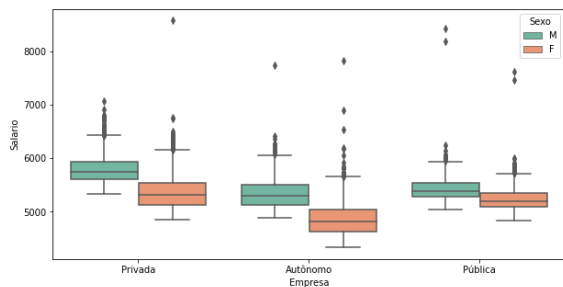
Considere as afirmações abaixo, contendo códigos em Python

I - `tab.plot.pie(y='count')` constrói um gráfico de setores, popularmente conhecido como gráfico de pizza, a partir de uma tabela de frequências `tab`

II - `displot(dados['Salario'], bins=10)` é um exemplo do uso da função do pacote `seaborn` que mostra um histograma dos dados com 10 classes.

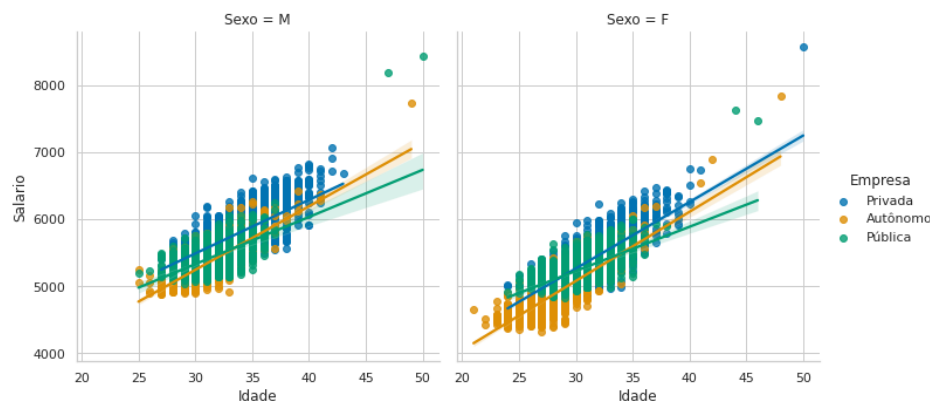
III - `dados.describe()` fornece medidas de posição, dispersão, assimetria, curtose e correlações para as variáveis quantitativas do dataframe `dados`.

IV - `boxplot(x='Empresa', y='Salario', hue='Sexo', data=dados, palette='Set2')`, que produz o gráfico abaixo



é um exemplo de uso da função do pacote `seaborn` que produz boxplots de salário para os dados agrupados por tipo de empresa e sexo.

V - `lplot(x='Idade', y='Salario', hue='Empresa', col='Sexo', data=dados)`, que produz o gráfico abaixo



é um exemplo de uso de função do pacote

`seaborn` para representar gráficos de dispersão entre salário e idade, separadamente por tipo de empresa e sexo dos clientes, com retas de regressão diferentes para cada grupo

Quais itens estão corretos?

- ☒ a. todos
- ☐ b. I, IV e V somente
- ☐ c. III, IV e V somente
- ☐ d. I, II, IV e V somente
- ☐ e. I, II e III somente

### Questão 5

Completo

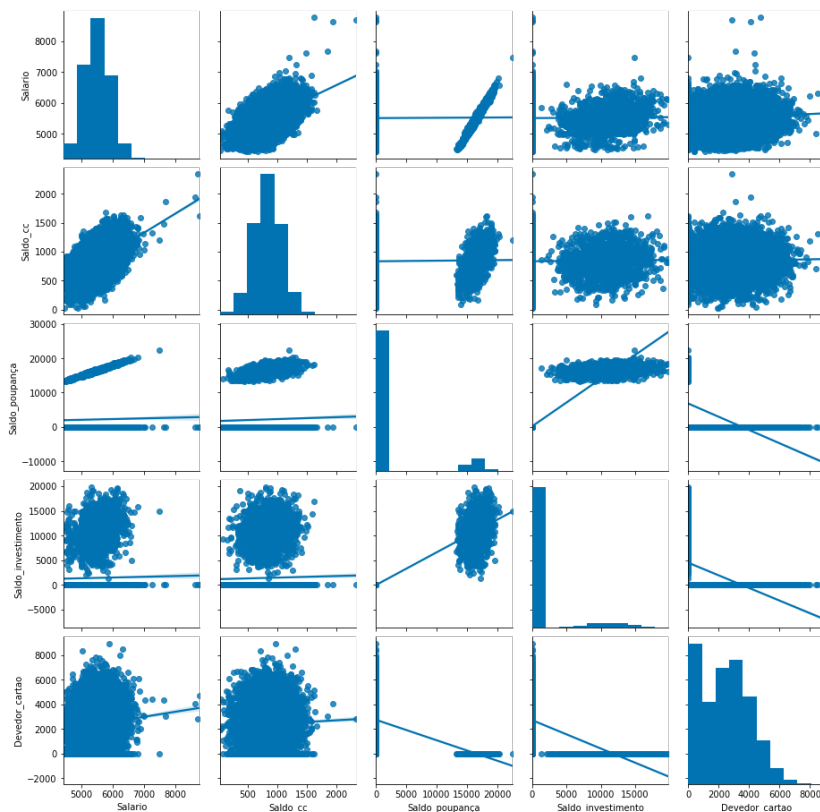
Vale 2,00 ponto(s).

Sobre a visualização e exploração de dados, considere as afirmações abaixo

I - A análise exploratória dos dados muitas vezes é negligenciada, sendo pouco valorizada em favor de técnicas de inferência e modelagem de dados. Mas ela traz insights importantes para entender a estrutura e comportamento dos dados.

II - Há pacotes prontos em Python que produzem relatórios de análises exploratórias suficientes para substituir, sem perda alguma, o trabalho do pesquisador em explorar os dados.

III - As retas de regressão apresentadas no gráfico a seguir não são adequadas para avaliar a associação entre alguns pares de variáveis, já que são afetadas pela grande quantidade de zeros existente nas colunas “Saldo em poupança” e “Saldo em investimento”.



IV - Em pacotes como sweetviz e pandas-profiling, é possível produzir relatórios de análises exploratórias gerais e que podem auxiliar o pesquisador, que devem ser usados com cautela e não substituem o trabalho de detetive do pesquisador nas análises iniciais de dados.

São corretos os itens:

- ☐ a. II somente
- ☒ b. I, III e IV somente
- ☐ c. I somente
- ☐ d. I e IV somente
- ☐ e. III somente

[◀ ECD-01-Prática 1-Gabarito](#)

Seguir para...

[TSS - Tarefa Substitutiva 1 ▶](#)