

Prova 3 - Conceitual

Pontuação: 10

Prazo: 26/11/2024 - A ser marcado em aula (mesma data que a prova prática)

Nome:	Matrícula:
Nome:	Matrícula:
Nome:	Matrícula:
Nome:	Matrícula:

**Observação:** Somente assinale, associe ou responda quando tiver certeza, pois, em cada questão, 1 (uma) resposta equivocada, cancela 1 (uma) uma correta.

## 1 Questão (1.0)

Considerando o assunto **Introdução à Estatística**, assinale verdadeiro (V) ou falso (F) as afirmativas, conceitos e definições:

- ( ) A palavra estatística significa, originalmente, uma coleção de informações de interesse para o Estado sobre a população e economia.
- ( ) O campo da análise estatística é relacionado à coleção, organização e interpretação de dados de acordo com procedimentos bem definidos.
- ( ) A análise exploratória de dados tem como objetivo a redução de grandes conjuntos de dados, via apresentações tabulares, gráficas e medidas estatísticas que resumem ou sumarizam as observações, sem o sacrifício de informações críticas.
- ( ) Um dos grandes objetivos da estatística é a tomada de decisão, em um processo particular qualquer, a respeito de uma amostra, em geral desconhecida, realizada a partir de populações representativas.
- ( ) A inferência estatística possibilita a tomada de decisão sobre os grandes grupos de observações (populações) com base na observação de apenas uma parte ou fração desses grandes grupos (amostras).
- ( ) A análise multivariada é o ramo da análise estatística que considera simultaneamente duas ou mais características mensuradas em um conjunto de objetos.
- ( ) Quando se está interessado em uma característica isolada de um conjunto de objetos, desconsiderando as outras características, o domínio é o da análise univariada.
- ( ) População: conjunto, finito ou infinito, de indivíduos, objetos ou medidas que apresentam pelo menos uma característica observável em comum.
- ( ) Amostra: consiste de uma parte (subconjunto) dos indivíduos, objetos ou medidas, selecionados a partir da população.
- ( ) Parâmetro: qualquer quantidade numérica medindo algum aspecto de uma amostra.
- ( ) Estimador do parâmetro: qualquer quantidade numérica medindo algum aspecto de uma amostra obtido, ou calculado, a partir de uma população.
- ( ) Escala é o esquema usado para a representação dos possíveis valores de uma variável.
- ( ) Escala intervalar: diferenças iguais entre valores representam quantidades iguais, mas proporções não tem significado devido à arbitrariedade do zero (ex: temperatura).
- ( ) Escala nominal: números agem como rótulos apenas indicando diferenças no tipo (ex: identificação).
- ( ) Escala ordinal: números representam ordem e as diferenças entre eles não tem significado (ex: localização).
- ( ) Escala proporcional: diferenças iguais entre valores representam quantidades iguais. Proporções iguais dos valores também são equivalentes devido a posição genuína do zero (ex: peso).

## 2 Questão (1.0)

Considerando o assunto **Noções de Amostragem**, associe as afirmativas aos conceitos e definições:

### Afirmativas

- ( ) A população é dividida em grupos (estratos) que são mutuamente exclusivos de acordo com alguma(s) característica(s) relevante(s). Posteriormente uma amostra aleatória simples é retirada de cada estrato.
- ( ) Baseia-se no sorteio de residências ou pessoas com base no mapa da região a ser pesquisada. A população regional é transformada numa população de áreas para a qual uma lista, em forma de mapa, existe.
- ( ) Consiste na escolha dos elementos da amostra por um especialista no assunto, que seleciona os elementos que julga os mais apropriados e representativos para o estudo em questão.
- ( ) De posse de uma listagem dos elementos da população, resulta da escolha sistemática, a partir de um número inicial qualquer, onde os demais elementos são selecionados de forma intervalar.
- ( ) O princípio se opõe ao da estratificação, pois o que se busca é a heterogeneidade: quanto maior a variabilidade, maior a precisão.
- ( ) Os elementos da amostra são escolhidos por serem os mais acessíveis ou fáceis de serem avaliados.
- ( ) É o método de selecionar, sem reposição, n elementos de uma população de tamanho N, conhecido e finito, onde cada elemento tem a mesma chance de ser selecionado.

### Conceitos e definições de amostragem

1. Acidental ou de conveniência
2. Aleatória simples
3. Conglomerado
4. Estratificada
5. Por julgamento
6. Por áreas
7. Sistemática

## 3 Questão (1.0)

Considerando o assunto **Análise Exploratória de Dados**, e as abreviações abaixo, associe as afirmativas às medidas estatísticas:

### Abreviações das Medidas Estatísticas - ME:

- Associação - MEA
- Dispersão - MED
- Posição - MEP
- Tendência Central - METC

### Afirmativas

- ( ) MEA adimensional, útil para comparar a associação em situações em que foram usadas diferentes escalas para mensurar as variáveis aleatórias em estudo.
- ( ) MEA influenciada pelas unidades de medida das variáveis aleatórias em estudo.
- ( ) MED adimensional, útil para comparar dispersão em estudos em que foram usadas diferentes escalas.
- ( ) MED com a mesma unidade de expressão da variável em estudo, porém, elevada ao quadrado.
- ( ) MED com a mesma unidade de expressão da variável em estudo.
- ( ) MEP que divide a série em 4 partes iguais.
- ( ) MEP que divide a série em 10 partes iguais.
- ( ) MEP que divide a série em 100 partes iguais.
- ( ) METC que não é afetada por valores extremos (outliers).
- ( ) METC que utiliza todos os valores da série, independente da posição.

### Medidas estatísticas

1. Coeficiente de variação
2. Correlação
3. Covariância
4. Decil
5. Desvio padrão
6. Mediana
7. Média aritmética
8. Percentil
9. Quartil
10. Variância

## 4 Questão (2.0)

O diagrama de caixa ("boxplot") é uma das apresentações gráficas mais importantes na análise exploratória de dados univariados.

### 4.1 Identificação (1.0)

Identifique na Figura ?? suas características (são oito no total) indicando a localização dos pontos e todas as medidas estatísticas associadas.

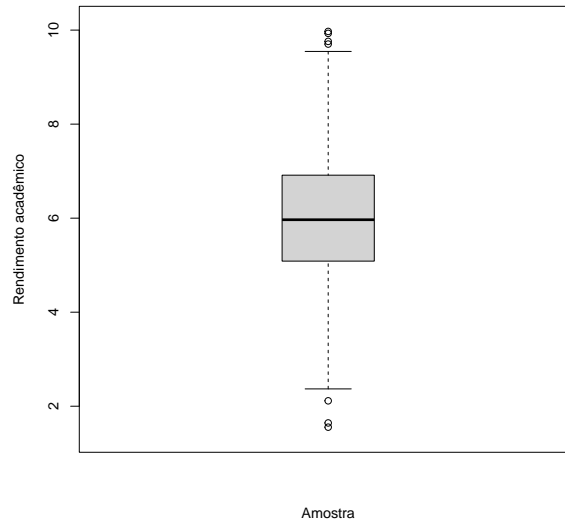


Figura 1 – Diagrama de caixa do rendimento acadêmico da disciplina Probabilidade e Estatística de uma amostra dos estudantes da UESC - 2024.

### 4.2 Entendimento da apresentação gráfica (1.0)

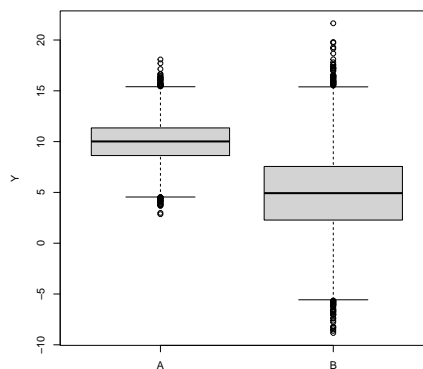


Figura 2 – Diagrama de caixa da variável Y dos cursos A e B da disciplina Probabilidade e Estatística de uma amostra dos estudantes da UESC - 2024.

Considerando a Figura ?? assinale verdadeiro (V) ou falso (F) as afirmativas abaixo:

- ( ) A medida de tendência central de A é maior que B.
- ( ) A amplitude interquartil de A é claramente maior que em B.
- ( ) A distribuição dos dados interquartil tanto de A quanto B é claramente simétrica.
- ( ) Os pontos acima do quartil 4 de A e B são outliers.
- ( ) O traço horizontal em negrito no interior das caixas de A e B representa as respectivas médias aritméticas.
- ( ) Os dados mais representativos de A são os que estão abaixo de 4.5 e acima de 15.4.
- ( ) Os dados mais representativos de B são os que estão abaixo de -5.6 e acima de 15.5.
- ( ) Não é possível afirmar nada sobre os dois grupos baseado apenas no boxplot.

## 5 Questão (2.0)

Considerando os assuntos **correlação linear (CL)** e **regressão linear (RL)**, assinale verdadeiro (V) ou falso (F) as afirmativas abaixo:

- ( ) A análise de CL é indicada para estudar o grau de associação linear entre duas variáveis quaisquer, aleatórias ou não.
- ( ) Enquanto a análise de RL simples nos mostra como duas variáveis aleatórias se relacionam linearmente, ( $\hat{Y}_1 = f(Y_2)$ ), a análise CL vai nos mostrar apenas o grau,  $r_{(Y_1, Y_2)}$ , desse mesmo relacionamento.
- ( ) O coeficiente de correlação,  $r_{(Y_1, Y_2)}$ , é uma medida do grau de associação linear entre duas variáveis aleatórias ( $Y_1$  e  $Y_2$ ). O coeficiente de determinação,  $r^2$ , é uma medida da adequação de um modelo de regressão. Ou seja, mede a quantidade da variação de uma variável aleatória (Y) que é explicada pela variação de uma outra (X: aleatória ou fixa), segundo um modelo ajustado ( $\hat{Y} = f(X)$ ).
- ( ) A análise de RL é uma técnica menos poderosa que a análise de CL.
- ( ) Se um modelo linear ajustado ( $\hat{Y} = f(X)$ ) contemplar apenas uma variável independente (X: fixa ou aleatória), trata-se de RL simples; se mais de uma ( $\hat{Y} = f(X_1 + X_2 + \dots + X_n)$ ), trata-se de RL múltipla.
- ( ) Se uma RL ( $\hat{Y} = f(X)$ ) for feita entre uma variável aleatória (Y) e uma variável fixa (X) a relação  $r = \sqrt{r^2}$  é verdadeira e tem significado estatístico.
- ( ) Se uma RL ( $\hat{Y} = f(X)$ ) for feita entre uma variável aleatória (Y) e uma variável aleatória (X) a relação  $r = \sqrt{r^2}$  é verdadeira e tem significado estatístico.
- ( ) Em um modelo linear  $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ ,  $\hat{\alpha}$  significa o intercepto; ou seja, o ponto onde a reta tangencia o eixo das ordenadas (Y) quando  $X = 0$ .  $\hat{\beta}$  é o coeficiente angular, ou seja, mede a variação prevista em Y decorrente da variação unitária em X.
- ( ) O método de ajustamento mais utilizado para estimar os parâmetros de um modelo de regressão ( $\hat{Y} = f(X)$ ,  $\hat{Y} = f(X_1 + X_2 + \dots + X_n), \dots$ ), é o que minimiza a soma de quadrados dos erros. Neste método erro é considerado a diferença entre o valor observado Y e o valor estimado  $\hat{Y}$ :  $\sum_{i=1}^n (Y_i - \hat{Y})^2$ .

## 6 Questão (2.0)

Com a finalidade de avaliar a vida útil de um grande lote de placas de circuito integrado, uma amostra é submetida a condições elevadas de salinidade e umidade relativa em condições laboratoriais.

Considere as seguintes variáveis de resposta:

- $Y_1$ : vida útil da placa (dias)
- $Y_2$ : número de pontos de oxidação da placa ( $n.cm^2$ )

e as seguintes variáveis preditoras

- $X_1$ : tempo (pré-estabelecido) de exposição da placa: 5, 10, 15, 20, 25 e 30 horas
- $X_2$ : dose de sais corrosivos adicionados a solução salina: 10, 20, 30, 40 e 50  $mg.l^{-1}$
- $X_3$ : teor de cloro nas soluções salinas detectado por um método analítico B ( $mg.l^{-1}$ ).

No estudo quantitativo de variáveis (**correlação linear (C)** e **regressão linear (R)**), com fundamentação estatística, pode-se estudar os seguintes pares ( $Y_i, X_i$ ) ou ( $Y_i, Y_i$ ) de variáveis usando ( $\rightarrow$  C/R). Assinale verdadeiro (V) ou falso (F) as afirmativas abaixo.

- ( )  $Y_1, Y_2 \rightarrow C$
- ( )  $Y_1, X_1 \rightarrow C$
- ( )  $Y_1, X_2 \rightarrow C$
- ( )  $Y_1, X_3 \rightarrow C$
- ( )  $Y_2, X_1 \rightarrow C$
- ( )  $Y_2, X_2 \rightarrow C$
- ( )  $Y_2, X_3 \rightarrow C$
- ( )  $Y_1, Y_2 \rightarrow R$
- ( )  $Y_1, X_1 \rightarrow R$
- ( )  $Y_1, X_2 \rightarrow R$
- ( )  $Y_1, X_3 \rightarrow R$
- ( )  $Y_2, X_1 \rightarrow R$
- ( )  $Y_2, X_2 \rightarrow R$
- ( )  $Y_2, X_3 \rightarrow R$

## 7 Questão (1.0)

De forma sucinta e objetiva, discorra sobre análise exploratória de dados abordando os principais objetivos e utilidade.