

PROVA OBJETIVA P₃ -- CONHECIMENTOS ESPECÍFICOS II --

ANÁLISE DE DADOS

Questão 91

Dados estruturados são tipos de dados que

- Ⓐ não podem ser indexados, mesmo com organização tabular, devendo ser percorridos integralmente em buscas sequenciais.
- Ⓑ dependem de metadados embutidos em arquivos binários para permitirem qualquer forma de busca ou filtragem.
- Ⓒ podem ser armazenados somente em sistemas de gerenciamento de banco de dados relacionais.
- Ⓓ não podem conter valores nulos, pois a não atribuição de valores descaracterizaria a consistência estrutural exigida.
- Ⓔ têm formato e organização que, definidos por esquemas rígidos que estabelecem tipos e restrições para cada campo, permitem consultas por meio de linguagens como SQL.

Questão 92

Assinale a opção correta a respeito de dados abertos.

- Ⓐ Os dados abertos são aqueles cujo acesso exige cadastro prévio e aceite de termos de uso restritivos que proíbem a modificação das informações.
- Ⓑ Os dados abertos são aqueles que somente podem ser compartilhados com órgãos governamentais mediante acordos de confidencialidade para se evitar mau uso por terceiros.
- Ⓒ Dados abertos são disponibilizados de forma gratuita, em formatos abertos e legíveis por máquina, podendo ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa, desde que citada a fonte.
- Ⓓ Dados abertos podem ser publicados em qualquer formato eletrônico, independentemente de estarem protegidos por direitos autorais ou licenças restritivas.
- Ⓔ Para que mantenham o *status* de abertos, os dados devem ser obrigatoriamente atualizados em tempo real.

Questão 93

No que diz respeito aos desafios e às técnicas de integração de dados, assinale a opção correta.

- Ⓐ Em ambientes com dados semiestruturados e não estruturados, a integração de dados limita-se à replicação dos dados originais, a fim de se evitarem perdas de informação.
- Ⓑ A integração de dados pode ser realizada sem políticas de qualidade e validação, pois qualquer inconsistência é detectada e corrigida automaticamente pelos sistemas de origem.
- Ⓒ O uso de *middleware* para integração de dados elimina a necessidade de mapeamento entre os esquemas heterogêneos dos diferentes sistemas.
- Ⓓ A utilização de um modelo unificado de dados e de um dicionário de dados corporativo é essencial para a resolução de conflitos semânticos e a garantia da governança da integração.
- Ⓔ A técnica de extração, transformação e carga (ETL) tradicional é insuficiente para integrar dados em tempo real, sendo necessária a adoção de processos ELT (*extract, load, transform*) em todos os casos.

Questão 94

Os bancos de dados relacionais

- Ⓐ têm estrutura lógica, com dados organizados em tabelas interligadas por chaves primárias e estrangeiras, e suporte à integridade referencial definida no próprio sistema gerenciador de banco de dados (SGBD).
- Ⓑ mantêm todos os registros em estruturas hierárquicas que exigem navegação física entre os níveis para que as informações sejam recuperadas.
- Ⓒ garantem a integridade e a consistência dos dados por meio de controle realizado apenas na aplicação cliente, e não no próprio banco.
- Ⓓ não permitem que operações de junção (JOIN) sejam utilizadas para relacionar dados, pois cada tabela deve ser consultada de forma isolada.
- Ⓔ dependem de armazenamento exclusivamente local, não podendo ser implementados em ambientes distribuídos ou na nuvem.

Questão 95

Em um banco de dados relacional, os metadados são

- Ⓐ *logs* de transações que registram todas as operações de inserção, atualização e exclusão.
- Ⓑ instruções de código-fonte que controlam a execução de consultas SQL em tempo real.
- Ⓒ sinônimos de *backups* de dados, utilizados para restaurar o banco em caso de falha.
- Ⓓ dados armazenados em tabelas que representam diretamente os registros da aplicação.
- Ⓔ dados que descrevem outros dados, podendo incluir informações sobre conteúdo, contexto, qualidade, formato e restrições de uso.

Questão 96

Considerando as características e funções de uma tabela em um banco de dados relacional, assinale a opção correta.

- Ⓐ Uma tabela deve ser indexada automaticamente para todas as consultas, a fim de se garantir *performance* sem necessidade de intervenção do administrador do banco de dados.
- Ⓑ Para existir no banco de dados relacional, uma tabela deve possuir no mínimo duas colunas, sendo uma delas definida como chave primária e a outra como chave candidata.
- Ⓒ As tabelas são equivalentes a arquivos físicos de armazenamento, sendo independentes de abstrações lógicas.
- Ⓓ Em um banco de dados relacional, a tabela, formada por tuplas e atributos, é a materialização de uma relação, em que a ordem das linhas é irrelevante, mas a ordem das colunas deve ser preservada.
- Ⓔ Todos os atributos de uma tabela devem ser numéricos para que seja possível garantir a consistência matemática do modelo relacional.

Questão 97

Em bancos de dados relacionais, uma *view*

- Ⓐ impede a filtragem e a agregação de dados provenientes das tabelas-base.
- Ⓑ é um objeto que redefine a estrutura física das tabelas-base, alterando o modo como os dados são armazenados em disco.
- Ⓒ pode ser utilizada para restrição de acesso a determinadas colunas de tabelas, funcionando como um mecanismo de segurança adicional no banco de dados.
- Ⓓ armazena fisicamente os dados em disco, duplicando todas as informações das tabelas originais.
- Ⓔ define regras de integridade referencial automaticamente para todas as tabelas-base incluídas.

Questão 98

Com base nos conceitos relacionados aos índices em bancos de dados relacionais, assinale a opção correta.

- Ⓐ A presença de índices acelera qualquer operação no banco de dados, inclusive inserções e atualizações, sem causar impactos negativos de desempenho.
- Ⓑ Índices compostos permitem a indexação de mais de uma coluna, sendo particularmente úteis quando as consultas filtram dados por múltiplos campos, mas a ordem das colunas no índice influencia seu uso pelo otimizador de consultas.
- Ⓒ Por sua baixa necessidade de manutenção, os índices *bitmap* são indicados para sistemas transacionais (OLTP) com alta taxa de atualizações e inserções.
- Ⓓ Em um banco de dados relacional, podem ser criados até quatro índices clusterizados em uma mesma tabela, desde que esta possua menos que 128 TB de tamanho.
- Ⓔ Um índice não clusterizado mantém uma cópia ordenada dos dados da tabela junto aos dados originais, duplicando o armazenamento e permitindo acesso direto sem pesquisa adicional.

Questão 99

Uma chave estrangeira (*foreign key*) de um banco de dados relacional

- Ⓐ representa um ou mais campos que, embora possam ser utilizados como chave primária, não foram escolhidos como tal.
- Ⓑ é uma chave que combina dois ou mais campos para formar um identificador único para registros dentro da mesma tabela.
- Ⓒ é uma chave especial usada apenas em tabelas temporárias com vistas à melhoria da *performance* das consultas.
- Ⓓ é um campo ou conjunto de campos que, em uma tabela, corresponde à *primary key* em outra tabela, e visa garantir a integridade referencial.
- Ⓔ garante que cada registro em uma tabela tenha um identificador único, de forma que não haja valores nulos ou duplicados.

Questão 100

Acerca do conceito e da aplicação da modelagem dimensional, assinale a opção correta.

- Ⓐ O objetivo da modelagem dimensional é substituir o uso de chaves primárias e estrangeiras, de modo que apenas identificadores textuais sejam utilizados para relacionar tabelas.
- Ⓑ A modelagem dimensional é um método de organização de dados em tabelas normalizadas utilizado para se evitarem redundâncias e minimizar a duplicidade de registros, sendo mais adequado para sistemas transacionais (OLTP).
- Ⓒ Na modelagem dimensional, cada dimensão deve ser armazenada em formato de documento JSON para maior flexibilidade no esquema.
- Ⓓ A modelagem dimensional organiza os dados em estruturas multidimensionais, geralmente compostas por tabelas-fato e tabelas-dimensão, otimizando consultas analíticas e sumarizações históricas.
- Ⓔ A modelagem dimensional elimina a necessidade de índices no banco de dados, pois o desempenho de consulta é garantido pela estrutura de tabelas-estrela.

Questão 101

Alguns algoritmos de mineração de dados requerem que os atributos sejam categóricos, o que impossibilita o uso direto de dados numéricos. Para contornar essa limitação, pode-se aplicar a discretização. Acerca desse processo, assinale a opção correta.

- Ⓐ A discretização não é recomendada para atributos contínuos, pois reduz a capacidade dos algoritmos de aprender padrões úteis.
- Ⓑ A discretização baseada em histograma utiliza os intervalos de frequência do histograma para segmentar os dados, e os valores são substituídos conforme a faixa em que se encontram.
- Ⓒ A discretização baseada em entropia é uma função linear por meio da qual se busca maximizar o ganho de informação ao se dividir o conjunto de dados, obtendo-se intervalos com maior relevância para classificação.
- Ⓓ A discretização por agrupamento atribui aleatoriamente os valores do atributo numérico a grupos fixos, sem levar em conta critérios ou similaridades entre os dados.
- Ⓔ A técnica de encaixotamento (*binning*) substitui os valores de um atributo por limites máximos e mínimos arbitrários, desconsideradas a média e a mediana.

Questão 102

No que diz respeito ao fenômeno de sobreajuste (*overfitting*), assinale a opção correta.

- Ⓐ O sobreajuste ocorre quando um modelo preditivo é avaliado por seu desempenho em exemplares não utilizados no treinamento, o que resulta em avaliações positivas.
- Ⓑ Modelos que se sobreajustam não representam os dados de treinamento com total fidelidade, o que resulta em avaliações ruins para todos os tipos de exemplares.
- Ⓒ O sobreajuste é um fenômeno benéfico, pois garante que o modelo preditivo seja altamente preciso tanto para dados de treinamento quanto para novos exemplares.
- Ⓓ O sobreajuste ocorre apenas quando o conjunto de dados de treinamento é muito pequeno, podendo ser solucionado com a sintonização manual dos parâmetros do modelo.
- Ⓔ Modelos que se sobreajustam representam os exemplares de treinamento com alta fidelidade, mas perdem desempenho quando aplicados a novos exemplares, o que resulta em avaliações ruins.

Questão 103

Considerando o contexto da mineração de dados, assinale a opção correta a respeito do funcionamento e das características das árvores de decisão.

- Ⓐ A impureza de uma partição em uma árvore de decisão será considerada maior quanto mais homogênea ela for em relação às classes dos exemplares.
- Ⓑ Por não serem adequadas a atributos numéricos, as árvores de decisão são utilizadas de forma exclusiva para atributos categóricos.
- Ⓒ Cada caminho de uma árvore de decisão — do nó-raiz até uma folha — representa um conjunto de decisões aleatórias aplicadas ao conjunto de treinamento.
- Ⓓ A interpretação de uma árvore de decisão não é recomendada para justificar decisões, pois o modelo resultante tende a ser opaco e de difícil compreensão.
- Ⓔ A árvore de decisão pode ser percorrida por um exemplar de teste desde a raiz até uma folha; as decisões são tomadas com base nos atributos, gerando-se uma classificação.

Questão 104

Considerando a aplicação do algoritmo Apriori sobre um conjunto de dados transacionais, assinale a opção correta no que concerne aos conceitos envolvidos na geração de regras de associação.

- Ⓐ O total de itens frequentes depende exclusivamente da quantidade de regras geradas.
- Ⓑ A confiança de uma regra mede a proporção de transações que contêm o antecedente e também o consequente em relação ao total de transações que contêm apenas o antecedente.
- Ⓒ O total de regras geradas é sempre igual ao número de itens frequentes multiplicado pela confiança mínima.
- Ⓓ O suporte de uma regra é calculado como a razão entre o número de regras geradas e o total de itens frequentes.
- Ⓔ O algoritmo Apriori não leva em consideração o valor mínimo de suporte ao gerar os conjuntos frequentes, mas sim a confiança das regras.

Questão 105

Assinale a opção em que é corretamente apresentada uma limitação do particionamento final produzido pelo algoritmo *k-means*.

- Ⓐ inaplicabilidade para grupos com formato aproximadamente esférico e de tamanhos variados
- Ⓑ exigência de que os dados de entrada estejam previamente rotulados
- Ⓒ imutabilidade do resultado final, independentemente da inicialização dos centroides
- Ⓓ não utilização de medidas de distância, como a euclidiana, para o cálculo da similaridade entre pontos e centroides
- Ⓔ necessidade de que o número de grupos (*k*) seja definido antes da execução do algoritmo

Questão 106

O uso de *big data* no setor jurídico tem-se mostrado promissor na identificação de padrões em jurisprudências e na otimização de processos. Considerando os 5 Vs do *Big Data* e a aplicação dessa tecnologia ao setor jurídico, assinale a opção correta.

- Ⓐ A correlação entre volume e veracidade é o desafio mais comum e mais crítico enfrentado por analistas jurídicos que utilizam o *Big Data*, pois a grande quantidade de informações não garante sua precisão, e a falta de padronização em documentos oficiais exige um esforço significativo para se garantir a confiabilidade dos dados.
- Ⓑ A variedade de documentos digitais e o alto volume de processos judiciais são facilmente superados por sistemas de *big data*, o que torna a análise de dados jurídicos um processo direto e automatizado.
- Ⓒ A velocidade na geração de novas leis e decisões e o ritmo com que novas informações jurídicas são disponibilizadas podem tornar rapidamente desatualizadas as ferramentas de análise de *Big Data*, o que exige atualizações constantes, sendo o principal desafio para a aplicação dessa tecnologia no meio jurídico.
- Ⓓ Uma vez garantidas a veracidade e a autenticidade das fontes oficiais, os demais Vs são desafios tecnológicos secundários no âmbito da aplicação do *big data* ao setor jurídico.
- Ⓔ O valor é o V menos importante no setor jurídico, pois as análises de *Big Data* sempre geram *insights* valiosos que justificam o alto custo de aquisição e manutenção dos sistemas.

Questão 107

município	arrecadação (R\$ milhões)
A	12,4
B	15,2
C	18,0
D	14,7
E	20,3

Considerando os dados apresentados na tabela precedente, relativos à arrecadação de ISS de cinco municípios em determinado mês, assinale a opção em que é apresentada uma interpretação correta das medidas descritivas da amostra (média, mediana e desvio padrão).

- Ⓐ A mediana da amostra é menor que a sua média, o que indica que a maioria dos municípios arrecada valores abaixo da média e sugere uma possível assimetria positiva (distribuição assimétrica à direita).
- Ⓑ A média e a mediana da amostra são iguais, o que indica uma distribuição perfeitamente simétrica da arrecadação entre os municípios.
- Ⓒ O desvio padrão baixo indica que os dados são muito dispersos, o que dificulta a definição de políticas fiscais homogêneas.
- Ⓓ A média da amostra é de aproximadamente R\$ 16,1 milhões, o que indica que os municípios têm arrecadação próxima e pouca variação.
- Ⓔ O desvio padrão elevado sugere que a arrecadação é muito homogênea entre os municípios.

Questão 108

Caso se pretendesse representar, em um relatório técnico, a relação entre duas variáveis quantitativas de determinado estado — o volume de arrecadação tributária e o número de fiscalizações realizadas —, seria mais adequado o uso de um gráfico de

- Ⓐ Gantt, que ilustraria as etapas e a duração de um processo de fiscalização.
- Ⓑ *pizza*, que mostraria a proporção de arrecadação por tipo de tributo (ICMS, IPVA, entre outros impostos) em determinado ano.
- Ⓒ linhas, que demonstraria a evolução da arrecadação de determinado tributo ao longo de cinco anos.
- Ⓓ dispersão, que exibiria a correlação entre o volume de arrecadação tributária e o número de fiscalizações realizadas.
- Ⓔ barras, que compararia o volume de arrecadação tributária entre diferentes regiões do estado.

Questão 109

Um analista de dados da Secretaria de Fazenda do estado X recebeu a incumbência de consolidar dados de repasses intergovernamentais no Microsoft Excel. Com a finalidade de assegurar a consistência e a qualidade dos dados, o analista deve aplicar técnicas de limpeza e transformação.

Nessa situação hipotética,

- Ⓐ para verificar se o nome de um município em uma tabela de referência corresponde a um código específico, o analista pode usar a função `PROCV` combinada com `SE`, que retorna o nome do município ou uma mensagem de erro caso o código não seja encontrado.
- Ⓑ para converter uma coluna de datas armazenadas como texto (por exemplo, 2023-01-15) em formato de data reconhecido pelo Excel, o analista só poderá usar a função `DATA.VALOR` se o texto estiver no formato DD/MM/AAAA.
- Ⓒ para identificar valores únicos em uma coluna de repasses, o analista pode criar uma tabela dinâmica, que só exibirá valores únicos se os dados forem previamente ordenados por ordem alfabética.
- Ⓓ para identificar registros duplicados em uma coluna de códigos de municípios, o analista pode usar a função `CONT.SE` combinada com formatação condicional, mas essa técnica só funciona se a coluna contiver apenas valores numéricos.
- Ⓔ para combinar valores de duas colunas, como um município e seu respectivo estado, em uma única célula no formato município – estado, o analista pode usar a função `CONCATENAR`, que, entretanto, não permite incluir um delimitador personalizado, como o hífen.

Questão 110

```
CREATE TABLE empresas (id_empresa INT PRIMARY KEY,
                        nome_empresa VARCHAR(100),
                        cnpj VARCHAR(18));

CREATE TABLE auditorias (id_auditoria INT PRIMARY KEY,
                          id_empresa INT,
                          data_auditoria DATE,
                          FOREIGN KEY (id_empresa) REFERENCES
empresas(id_empresa));
```

A partir das tabelas precedentes, referentes a empresas em que foram realizadas auditorias, assinale a opção em que a consulta SQL apresentada retorna corretamente a lista de empresas que foram objeto de nenhuma ou de apenas uma auditoria no ano de 2025.

- Ⓐ

```
SELECT e.id_empresa, e.nome_empresa
FROM empresas e LEFT JOIN auditorias a
ON e.id_empresa = a.id_empresa
WHERE a.data_auditoria >= '2025-01-01'
AND a.data_auditoria <= '2025-12-31'
GROUP BY e.id_empresa, e.nome_empresa
HAVING COUNT(*) <= 1
```
- Ⓑ

```
SELECT e.id_empresa, e.nome_empresa
FROM empresas e LEFT JOIN auditorias a
ON e.id_empresa = a.id_empresa
AND a.data_auditoria BETWEEN '2025-01-01' AND
'2025-12-31'
GROUP BY 1, 2
HAVING COUNT(a.id_auditoria) <= 1
```
- Ⓒ

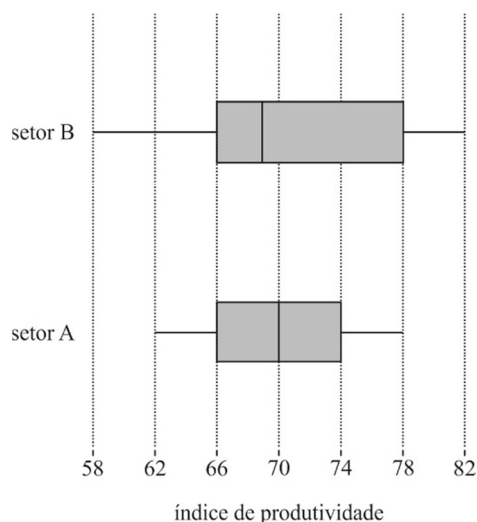
```
SELECT e.id_empresa, e.nome_empresa
FROM empresas e INNER JOIN auditorias a
ON e.id_empresa = a.id_empresa
AND a.data_auditoria BETWEEN '2025-01-01' AND
'2025-12-31'
GROUP BY e.id_empresa, e.nome_empresa
HAVING COUNT(*) <= 1
```
- Ⓓ

```
SELECT e.id_empresa, e.nome_empresa
FROM empresas e LEFT JOIN auditorias a
ON e.id_empresa = a.id_empresa
GROUP BY e.id_empresa, e.nome_empresa
HAVING COUNT(*) <= 1
```
- Ⓔ

```
SELECT e.id_empresa, e.nome_empresa
FROM empresas e LEFT JOIN auditorias a ON
e.id_empresa = a.id_empresa
WHERE a.data_auditoria BETWEEN '2025-01-01'
AND '2025-12-31'
GROUP BY e.id_empresa, e.nome_empresa
HAVING COUNT(*) <= 1
```

ESTATÍSTICA

Texto 4A8



Os diagramas de *box-plot* precedentes correspondem às distribuições das produtividades das equipes de dois setores de determinada secretaria da fazenda.

Questão 111

Considerando os diagramas de *box-plot* apresentados no texto 4A8, assinale a opção correta.

- Ⓐ Em relação ao setor B, $1/6$ da distribuição das produtividades é maior ou igual a 78.
- Ⓑ Em ambas as distribuições, todos os valores observados estão dentro dos limites adjacentes dos respectivos diagramas de *box-plot*.
- Ⓒ No diagrama de *box-plot* do setor A, são considerados pontos exteriores os valores que estão fora da caixa, isto é, os valores que não pertencem ao intervalo $[66, 74]$.
- Ⓓ O primeiro quartil da distribuição da produtividade do setor A é igual a 74.
- Ⓔ A distribuição da produtividade referente ao setor B apresenta assimetria negativa.

Questão 112

Supondo-se, com base nos diagramas de *box-plot* apresentados no texto 4A8, que IQ_A e IQ_B denotem os intervalos interquartil das respectivas distribuições das produtividades dos setores A e B, é correto afirmar que a diferença $IQ_A - IQ_B$ é igual a

- Ⓐ -8.
- Ⓑ -4.
- Ⓒ 0.
- Ⓓ +4.
- Ⓔ +8.

Questão 113

A quantidade diária X de processos gerados por um setor de determinada secretaria da fazenda segue uma distribuição de Poisson com taxa igual a 10 processos por dia. Diariamente, 20% dos processos são gerados por pessoas físicas e 80%, por pessoas jurídicas, independentemente do total de processos gerados no dia.

Tendo como base a situação proposta e considerando que Y denote a quantidade de processos diários gerados por pessoas físicas, assinale a opção correta.

- Ⓐ A diferença $X - Y$ segue uma distribuição de Poisson com taxa igual a 8 processos por dia.
- Ⓑ As variáveis aleatórias Y e $X - Y$ são dependentes.
- Ⓒ A variável aleatória Y segue uma distribuição binomial.
- Ⓓ Se, em determinado dia, forem gerados 10 processos, então a distribuição condicional $Y|X = 10$ seguirá uma distribuição de Poisson.
- Ⓔ Se, em determinado dia, forem gerados 5 processos por pessoas físicas, então a distribuição condicional $X|Y = 5$ seguirá uma distribuição de Poisson.

Questão 114

$$f_V(v) = \begin{cases} \frac{1}{2}, & \text{se } v \in [0, 2] \\ 0, & \text{se } v \notin [0, 2] \end{cases}$$

Se a função de densidade de probabilidade para uma variável aleatória V for dada pela expressão precedente, o valor da probabilidade $P(V = 1,5)$ será igual a

- Ⓐ 0.
- Ⓑ $1/4$.
- Ⓒ $1/2$.
- Ⓓ $3/4$.
- Ⓔ 1.

Questão 115

Considere que $Z_1, \dots, Z_n$ represente uma amostra aleatória simples, de tamanho n , retirada de uma distribuição normal padrão. Nesse caso, à medida que n aumenta, a média $\frac{1}{n} \sum_{i=1}^n Z_i^3$ converge em probabilidade para

- Ⓐ 0.
- Ⓑ 1.
- Ⓒ 2.
- Ⓓ 3.
- Ⓔ 4.

Questão 116

$$f_X(x) = \begin{cases} \frac{1}{a}, & \text{se } x \in [0, a] \\ 0, & \text{se } x \notin [0, a] \end{cases}$$

Considerando uma amostra aleatória simples X_1, X_2, X_3 retirada de uma população cuja função de densidade seja dada pela expressão precedente, assinale a opção em que é apresentada uma estatística suficiente para a estimação do parâmetro desconhecido a .

- Ⓐ $(X_1 + X_2 + X_3)/3$
- Ⓑ $(X_1 + X_2 + X_3)^2$
- Ⓒ $X_1^2 + X_2^2 + X_3^2$
- Ⓓ $X_1^3 + X_2^3 + X_3^3$
- Ⓔ $\text{máximo}(X_1, X_2, X_3)$

Questão 117

Para se estimar o tempo médio de tramitação de determinado tipo de processo, avalia-se uma amostra aleatória simples de 25 processos, obtendo-se média amostral $\bar{x} = 12$ dias e desvio padrão amostral $s = 4$ dias.

A partir dessa situação hipotética, assinale a opção correta, considerando que os tempos de tramitação desse tipo de processo tenham distribuição normal.

- A** O intervalo de confiança para a média populacional deve ser tipicamente construído com base na distribuição normal padrão.
- B** Um intervalo de 95% de confiança para a média populacional pode ser representado como 12 ± 4 .
- C** No que diz respeito a intervalos de confiança simétricos para a média populacional, quanto maior for o nível de confiança, maior será sua margem de erro.
- D** Se $12 \pm \epsilon$ representar o intervalo de 90% de confiança para a média populacional μ , então a probabilidade de ocorrer o evento $\mu \in [12 - \epsilon, 12 + \epsilon]$ será igual a 0,90.
- E** A margem de erro do intervalo de 95% de confiança simétrico para a média populacional é de 0,8 dia.

Questão 118

A fim de se efetuar o teste de hipóteses para certa média populacional especificado como $H_0: \mu = 10$ versus $H_1: \mu > 10$, tomou-se uma amostra aleatória simples de tamanho $n = 16$ de uma população normal. A média amostral foi igual a 11 e o desvio padrão amostral, igual a 2.

A partir dessa situação hipotética, assinale a opção correta.

- A** O nível de significância do teste de hipóteses representa a probabilidade de se rejeitar indevidamente a hipótese alternativa (H_1).
- B** A hipótese nula (H_0) deverá ser rejeitada quando a média amostral for superior a 10, de acordo com a hipótese alternativa (H_1).
- C** O valor da estatística do teste t de Student é igual a -2 .
- D** O nível descrito do teste é superior a 50%.
- E** A função poder do teste em questão tende a aumentar à medida que a média populacional aumenta.

Questão 119

Um estudo está sendo realizado para investigar a relação entre o tempo de tramitação de processos (y) e o número de documentos fiscais por processo (x). A correlação linear entre ambas as variáveis é positiva, sendo o desvio padrão amostral de cada variável mostrado no quadro a seguir. Nesse estudo, utiliza-se o modelo de regressão linear simples na forma $y = \beta_0 + \beta_1 x + \epsilon$, em que ϵ denota o erro aleatório com média zero e variância constante. O coeficiente de determinação desse modelo é igual a 64%.

variável	desvio padrão amostral
y	10
x	8

Com base nessas informações, é correto concluir que a estimativa do coeficiente β_1 obtida pelo método de mínimos quadrados ordinários é igual a

- A** 0,512.
- B** 0,640.
- C** 0,800.
- D** 0,810.
- E** 1,000.

Questão 120

Deseja-se realizar uma auditoria sobre o recolhimento de tributos em microempresas localizadas em diferentes municípios de determinado estado da Federação. Para se garantir representatividade territorial, uma amostra de 15 microempresas por amostragem aleatória simples será retirada de cada município. Dessa forma, havendo 75 municípios, o tamanho total da amostra será de 1.125 microempresas.

O plano amostral descrito nessa situação hipotética é mais apropriadamente caracterizado como uma amostragem

- A** aleatória simples.
- B** estratificada com alocação uniforme.
- C** estratificada com alocação ótima de Neyman.
- D** sistemática.
- E** por conglomerados.

Espaço livre