



COMISSÃO DE VALORES MOBILIÁRIOS

TARDE

ANALISTA CVM - PERFIL 7 - CIÊNCIA DE DADOS

PROVA OBJETIVA DE CONHECIMENTOS ESPECÍFICOS – NÍVEL SUPERIOR
TIPO 2 – VERDE



SUA PROVA

Além deste caderno de provas contendo setenta questões objetivas, você receberá do fiscal de sala:

- uma folha para a marcação das respostas das questões objetivas



TEMPO

- **4 horas** é o período disponível para a realização da prova, já incluído o tempo para a marcação da folha de respostas da prova objetiva
- **3 horas** após o início da prova é possível retirar-se da sala, sem levar o caderno de provas
- **30 minutos** antes do término do período de prova é possível retirar-se da sala **levando o caderno de provas**



NÃO SERÁ PERMITIDO

- Qualquer tipo de comunicação entre os candidatos durante a aplicação da prova
- Usar o sanitário ao término da prova, após deixar a sala



INFORMAÇÕES GERAIS

- As questões objetivas têm cinco alternativas de resposta (A, B, C, D, E) e somente uma delas está correta
- Verifique se este caderno de provas está completo, sem repetição de questões ou falhas. Caso contrário, notifique imediatamente o fiscal da sala, para que sejam tomadas as devidas providências
- Na folha de respostas da prova objetiva, confira seus dados pessoais, especialmente nome, número de inscrição e documento de identidade, e leia atentamente as instruções para preenchimento
- Use somente caneta esferográfica, fabricada em material transparente, com tinta preta ou azul
- Assine seu nome apenas no(s) espaço(s) reservado(s)
- Confira o cargo, a cor e o tipo do seu caderno de provas. Caso tenha recebido caderno de cargo, cor ou tipo diferente do impresso em sua folha de respostas, o fiscal deve ser **obrigatoriamente** informado para o devido registro na ata da sala
- O preenchimento das respostas da prova objetiva é de sua responsabilidade e não será permitida a troca da folha de respostas em caso de erro
- Para fins de avaliação, serão levadas em consideração apenas as marcações realizadas na folha de respostas da prova objetiva, não sendo permitido anotar informações relativas às respostas em qualquer outro meio que não seja o caderno de provas
- Os candidatos serão submetidos ao sistema de detecção de metais quando do ingresso e da saída de sanitários durante a realização das provas

Boa sorte!

PROVA OBJETIVA

CONHECIMENTOS ESPECÍFICOS

Ciência de Dados

1

Visando a maximizar a eficiência de uma equipe de auditores fiscais, um sistema de classificação de documentação foi encomendado à equipe de ciência de dados, com o objetivo de decidir, com base nos documentos obtidos durante uma fiscalização, se um exame detalhado de documentação é ou não necessário.

Idealmente, o sistema permitiria aos auditores direcionar mais tempo às auditorias complexas e agilizar a análise dos casos mais simples, otimizando o custo de pessoal e equipamento especializado. Contudo, não examinar detalhadamente um caso complexo pode custar muito caro ao governo, a ponto de anular quaisquer ganhos obtidos usando o sistema com um pequeno número de erros.

Considerando esse cenário, e o fato de o sistema de classificação responder apenas “sim” ou “não” quanto à necessidade de exame detalhado, a métrica de classificação a ser maximizada pela equipe que irá implementar o sistema é:

- (A) acurácia;
- (B) revocação;
- (C) F1-score;
- (D) precisão;
- (E) AUC.

2

O cientista de dados Miguel decidiu buscar um serviço de nuvem que forneça recursos de computação sob demanda, tais como servidores, rede, armazenamento e outros, para construir seu ambiente de análise e exploração de dados, podendo incorporar sistemas operacionais e aplicativos.

Para isso, Miguel deve contratar o serviço de nuvem:

- (A) DaaS;
- (B) FaaS;
- (C) IaaS;
- (D) PaaS;
- (E) SaaS.

3

Flávia, responsável pelo setor de análise de dados de uma rede de concessionárias de carros, está realizando o pré-processamento dos dados dos clientes da rede. Entre os atributos do conjunto de dados, estão os CPFs dos clientes, o seu sexo e a quantidade de carros que eles já compraram na rede.

Esses três atributos podem ser classificados, respectivamente, como:

- (A) qualitativo nominal, qualitativo nominal e quantitativo discreto;
- (B) qualitativo nominal, qualitativo ordinal e quantitativo discreto;
- (C) qualitativo nominal, qualitativo nominal e quantitativo contínuo;
- (D) qualitativo ordinal, qualitativo nominal e quantitativo discreto;
- (E) quantitativo discreto, qualitativo nominal e quantitativo contínuo.

4

Observe os conjuntos de dados a seguir.

Conjunto de Dados Sensíveis A:

ID	Nome	Profissão	CEP	Situação
10	Luis Alberto Nascimento	Agente Autonomo	20321-560	Em Análise
11	Juliana Lopes Quinta	Agente Fiduciário	70125-120	Suspensão
12	Décio Roberto Cunha	Analista de Sistemas	15080-121	Cancelado
13	Edson Nilo do Nascimento	Engenheiro de Software	32800-090	Normal

Conjunto de Dados B:

ID	Nome	Profissão	CEP	Situação
2	Luis Alberto Souza	Agente	20321	Sem acesso
4	Juliana Lima Quinta	Agente	70125	Sem acesso
6	Décio Rui Cunha	Profissional de TI	15080	Sem acesso
8	Edson Nilo da Silva	Profissional de TI	32800	Com acesso

Para desidentificar o Conjunto de Dados Sensíveis A e gerar o Conjunto de Dados B, a técnica de anonimização que deve ser aplicada é o(a):

- (A) adição de ruído no atributo ID;
- (B) truncagem no atributo Nome;
- (C) encobrimento no atributo Profissão;
- (D) supressão no atributo CEP;
- (E) blurring no atributo Situação.

5

Uma certa organização busca melhorar a qualidade e agilidade do seu atendimento eletrônico. Para isso um projeto foi criado para agrupar os e-mails recebidos de acordo com o tipo de problema a ser resolvido e assim repassá-los para o setor mais apropriado.

A equipe responsável pela implementação do projeto resolveu utilizar um modelo de linguagem recente para representar o máximo possível de informação contida num e-mail em um vetor de dimensão 768. Entretanto, depararam-se com o seguinte problema: as distâncias entre os vetores se mostraram muito pequenas, tornando o agrupamento por diversos algoritmos muito pouco significativo.

Com esse último problema em mente, a sequência mais apropriada de algoritmos a ser aplicada sobre os vetores, de forma a obter um agrupamento significativo dos e-mails, é:

- (A) PCA → t-SNE → KNN;
- (B) UMAP → KNN;
- (C) t-SNE → HDBSCAN → K-Means;
- (D) UMAP → HDBSCAN;
- (E) K-Means → t-SNE.

6

Considere o código python a seguir.

```
import torch
from torch import nn, Tensor

class CVMNet(nn.Module):
    def __init__(self,
                 dim_in: int,
                 dim_hidden: int,
                 n_classes: int):
        super().__init__()
        self.i_layer = nn.Linear(dim_in, dim_hidden)
        self.h_layer = nn.Sequential(
            nn.Linear(dim_hidden, dim_hidden // 2),
            nn.Tanh()
        )
        self.o_layer = nn.Sequential(
            nn.Dropout(0.2),
            nn.Linear(dim_hidden // 2, n_classes),
            nn.Softmax(dim=-1)
        )

    def forward(self, x: Tensor):
        return self.o_layer(self.h_layer(self.i_layer(x)))

model = CVMNet(400, 100, 3)
pred = model(torch.eye(20).flatten()).detach()
```

Sobre a variável pred, é correto afirmar que:

- (A) seus gradientes foram preparados para a retropropagação;
- (B) seus valores, quando somados, resultam em 1.0;
- (C) sua dimensão é 50;
- (D) seus valores serão os mesmos para qualquer execução do código;
- (E) sua dimensão é 1.

7

Considere o código python a seguir.

```
import spacy

nlp = spacy.load("pt_core_news_lg")
doc = nlp("O rato roeu a roupa do rei de Roma")

print(doc[2].pos_, doc[2].dep_)
```

Os valores exibidos pela última linha são:

- (A) NOUN, NSUBJ;
- (B) VERB, ROOT;
- (C) NMOD, DET;
- (D) VERB, CASE;
- (E) NSUBJ, NOUN.

Texto 1

Aline, cientista de dados da CVM, foi designada para aferir a reação à prova da CVM entre os usuários de uma rede social de textos curtos usando técnicas de análise de sentimentos. Para isso, ela realiza um processo de KDD. Nesse processo, Aline opta por representar os textos obtidos da rede social no formato de vetores reais de baixa dimensionalidade, calculados a partir das representações das palavras obtidas de um modelo de linguagem pré-treinado utilizando a técnica word2vec.

8

Considerando o texto 1, a representação das palavras que será utilizada por Aline é chamada de:

- (A) TF-IDF;
- (B) tokens;
- (C) n-gramas;
- (D) bag-of-words;
- (E) word embeddings.

9

Considerando o texto 1, a fase do KDD em que Aline gera os vetores a partir dos textos é chamada de:

- (A) seleção dos dados;
- (B) limpeza dos dados;
- (C) mineração dos dados;
- (D) transformação dos dados;
- (E) enriquecimento dos dados.

10

Uma equipe de analistas de dados preparou um modelo preditivo cuja entrada consiste em planilhas contendo uma matriz de valores reais entre 1 e 10. Tais planilhas são obtidas de um sistema externo à equipe. O modelo foi treinado com um conjunto de planilhas que foi coletado pelos analistas, de forma a obter uma amostra representativa dos dados a serem utilizados. A média e o desvio padrão de duas colunas importantes foram calculados do conjunto de treinamento, como uma forma simples de verificar a consistência da distribuição dos dados, sendo seus valores 4,89 e 3,08, respectivamente. O modelo obteve bons resultados durante sua etapa de testes, com uma precisão de 94%.

Ao iniciar a operação do modelo com planilhas atuais, entretanto, os analistas observaram que o modelo teve um desempenho muito inferior, com precisão de apenas 72%. Investigando as planilhas recebidas, obtiveram a média e o desvio padrão para as duas colunas importantes com valores 5,34 e 3,68, respectivamente.

A explicação mais adequada à situação descrita é:

- (A) as planilhas atuais contêm ruído: valores ausentes nas colunas importantes;
- (B) a distribuição dos valores nas planilhas atuais não corresponde mais aos dados usados no treinamento, pois foram coletadas em momentos diferentes (drifting);
- (C) a distribuição dos valores nas planilhas atuais não corresponde mais aos dados usados no treinamento devido a uma mudança na posição das colunas;
- (D) a distribuição dos valores nas planilhas de treinamento não corresponde aos dados atuais, pois a amostra obtida pelos analistas não foi representativa;
- (E) as planilhas atuais contêm ruído: valores fora do intervalo definido para o modelo.

11

Alexandre recebe a tarefa de treinar um sistema de detecção de fraudes no banco em que trabalha. Para isso, ele testa cinco modelos, M1, M2, M3, M4 e M5, que possuem, respectivamente, 2, 2, 2, 3 e 3 parâmetros. Alexandre realiza uma seleção bayesiana dos modelos, usando o critério de informação bayesiano.

Sabendo que o tamanho da amostra é 200 e que os valores maximizados das funções de verossimilhança dos modelos são 0,3; 0,4; 0,5; 0,3 e 0,5, respectivamente, Alexandre seleciona o modelo:

(se necessário, use $\ln(2) = 0,7$; $\ln(3) = 1,1$ e $\ln(5) = 1,6$)

- (A) M1;
- (B) M2;
- (C) M3;
- (D) M4;
- (E) M5.

12

O sistema CVMWeb armazena os dados pessoais dos consultores de valores mobiliários, como nome, telefone e e-mail. Uma entidade financeira privada solicita, ao setor de recursos humanos da CVM, os dados de contato dos consultores cadastrados para oferecer linhas de crédito e financiamento imobiliário.

Nesse contexto, é correto afirmar que o pedido foi:

- (A) negado, pois há incompatibilidade entre a finalidade original da coleta e a finalidade proposta para o compartilhamento, além de não haver enquadramento em uma das exceções previstas (Art. 26º, § 1, da LGPD);
- (B) analisado pela autoridade competente, que decidiu que os dados dos consultores inativos poderão ser compartilhados atendendo ao princípio da não discriminação (Art. 6º, IX, da LGPD);
- (C) enviado para a Autoridade Nacional de Privacidade de Dados, que deverá emitir um parecer autorizando o compartilhamento (Art. 30º, da LGPD);
- (D) atendido, com base no Art. 7º, II, da LGPD, segundo o qual o compartilhamento de dados pessoais pelo Poder Público poderá ser realizado “para o cumprimento de obrigação legal ou regulatória pelo controlador”;
- (E) legitimado pelo Art. 26º § 1º da LGPD: “É autorizado ao Poder Público transferir a entidades privadas dados pessoais constantes de bases de dados a que tenha acesso desde que previamente tratados.”

13

Ao receber um conjunto de dados para elaborar um modelo preditivo, uma equipe de analistas de dados percebeu que havia uma quantidade significativa de dados faltantes em certos atributos. Foi então debatido o uso de duas técnicas para lidar com esse problema: (1) remoção de observações contendo dados ausentes e (2) “imputação” multivariável, sendo que apenas uma das duas seria aplicada.

Dois características do conjunto de dados que devem ser prioritariamente consideradas na escolha entre as duas técnicas são:

- (A) o intervalo dos atributos com dados faltantes e a fração de dados faltantes;
- (B) o número de observações e a fração de dados faltantes;
- (C) a fração de atributos com dados faltantes e o intervalo dos atributos com dados faltantes ;
- (D) uma medida de dependência estatística entre os atributos e a fração de atributos com dados faltantes;
- (E) o número de observações e uma medida de dependência estatística entre os atributos.

14

As informações são a base de toda tomada de decisão e gestão de empresas, sendo um diferencial importante o uso de grandes volumes de dados de diversas fontes.

Nesse contexto, as soluções de Big Data para análise de dados devem ter a capacidade de:

- (A) processar dados heterogêneos, de alto volume e alta velocidade, utilizando estruturas computacionais aprimoradas para a automação de processos e tomadas de decisão;
- (B) operar dados homogêneos, de alto volume e limitada velocidade, gerando valor agregado para o negócio da empresa;
- (C) manusear volumes de dados armazenados localmente, heterogêneos e com atividades semiestruturadas, agregando valor ao processo decisório da empresa;
- (D) processar grandes volumes de dados homogêneos, cujo objetivo principal é a implementação de técnicas avançadas de criptografia;
- (E) auxiliar no processo de gestão empresarial, processando dados heterogêneos, de alto volume e alta velocidade, descartando domínios de dados e transações.

15

A abordagem que garante uma exploração completa das combinações de hiperparâmetros na identificação da configuração ideal para maximizar o desempenho do modelo de aprendizado de máquina é a:

- (A) realização de uma Grid Search, explorando sistematicamente os valores de hiperparâmetros predefinidos para identificar a combinação ideal para maximizar o desempenho do modelo;
- (B) utilização de uma Busca Aleatória, amostrando aleatoriamente o espaço de hiperparâmetros, oferecendo uma abordagem mais eficiente para o ajuste de hiperparâmetros em comparação com os métodos exaustivos;
- (C) implementação de uma Validação Cruzada K-fold, particionando o conjunto de dados em K subconjuntos para treinar e avaliar iterativamente o modelo, fornecendo uma avaliação robusta de seu desempenho de generalização;
- (D) aplicação de métodos Ensemble, combinando vários modelos para aproveitar suas diversas forças, visando a melhorar o desempenho geral, particularmente quando os modelos individuais apresentam diferentes tipos de erros;
- (E) utilização de uma Feature Scaling, normalizando a escala das características de entrada para garantir comparações justas entre diferentes variáveis, o que é particularmente benéfico para algoritmos sensíveis a diferenças de escala.

16

Um dos principais fatores que tornam viável a aplicação de modelos grandes de linguagem (LLMs) é o controle do espaço de probabilidade de tokens através da redução de dimensionalidade do vocabulário, sem perda da capacidade de reconstruir qualquer token válido da linguagem sendo modelada.

Considerando esse objetivo, dois algoritmos que podem ser utilizados para esse fim são:

- (A) UMAP, BPE;
- (B) T5, UMAP;
- (C) BPE, WordPiece;
- (D) WordPiece, T5;
- (E) UMAP, GTE.

17

O cientista de dados Pedro trabalha em um projeto que envolve a previsão dos movimentos de um braço robótico em um ambiente complexo. Pedro tem um fluxograma de um algoritmo de aprendizado por reforço que é capaz de se adaptar dinamicamente ao ambiente e ajustar suas ações com base nos resultados de ações anteriores.

O algoritmo representado pelo referido fluxograma que deve ser empregado para a realização da tarefa de Pedro é o:

- (A) Máquina de Vetores de Suporte (SVM);
- (B) Rede Neural Convolucional (CNN);
- (C) *Long Short-Term Memory* (LSTM);
- (D) *Twin Delayed Deep Deterministic Policy Gradient* (TD3);
- (E) K-Vizinhos Mais Próximos (KNN).

18

Uma certa organização gostaria de compartilhar dados com um grupo de pesquisadores de uma universidade para a condução de um estudo sobre problemas ergonômicos nos seus escritórios. Entre os dados coletados, há informações sensíveis sobre seus funcionários; portanto, o responsável pela coleta decidiu anonimizar os dados. Isso foi feito removendo-se nomes e outros campos identificadores e adicionando-se um número identificador próprio a cada funcionário. Dessa forma, a identidade dos funcionários seria preservada. Após a verificação de uma amostra, o pesquisador responsável pelo estudo recomendou medidas que deveriam ser aplicadas antes que os dados pudessem ser aceitos para o estudo.

O problema que mais provavelmente motivou a recomendação do pesquisador e uma medida que pode mitigar esse problema são, respectivamente:

- (A) os dados podem ser reidentificados usando uma combinação dos campos restantes → embaralhamento dos campos;
- (B) o número identificador não é suficiente para separar os dados → hashing usando os campos restantes;
- (C) os dados podem ser reidentificados utilizando dados externos ao conjunto compartilhado → privacidade diferencial;
- (D) os dados podem ser reidentificados através de engenharia reversa do número identificador → randomização não linear dos identificadores;
- (E) os campos restantes podem ser usados para deduzir informação não inclusa no conjunto de dados → k-anonimato.

19

Considere um modelo de rede neural projetado para tarefas de classificação de imagens. A arquitetura da rede consiste em várias camadas, incluindo camadas de entrada, ocultas e de saída. Durante o treinamento, o modelo aprende a atribuir imagens de entrada a classes predefinidas (por exemplo, “gato”, “cachorro”, “pássaro”, etc.).

Sobre o papel da camada de saída nessa rede neural, é correto afirmar que ela:

- (A) garante que a rede neural generalize bem para dados não vistos, prevenindo o sobreajuste;
- (B) ajusta os pesos dos neurônios ocultos para minimizar o erro de classificação durante o treinamento;
- (C) mapeia as características aprendidas para rótulos de classe específicos com base em padrões aprendidos;
- (D) realiza a extração de características aprendendo representações hierárquicas a partir de valores de pixel brutos;
- (E) calcula a soma ponderada das características de entrada e aplica uma função de ativação para produzir probabilidades de classe.

20

Considere uma estrutura específica de rede neural recorrente, conhecida como *Long Short-Term Memory* (LSTM). Essa estrutura é projetada para enfrentar desafios como capturar dependências de longo prazo e mitigar o problema do gradiente que desaparece.

A inovação arquitetônica distintiva da LSTM é(são):

- (A) o recorte de gradiente;
- (B) as células de memória com mecanismos de gating;
- (C) os mecanismos de atenção;
- (D) as conexões de salto;
- (E) as camadas de Max-pooling.

21

Para ser utilizado em um modelo neural de regressão, um conjunto de dados precisa ser tratado de tal forma que todos os atributos de entrada sejam representados como um ou mais valores numéricos no intervalo [0, 1].

Os atributos de uma observação são: idade (inteiro ≥ 18), escolaridade (fundamental, médio, superior, pós-graduação), estado de residência (Acre, Alagoas, ..., Tocantins, incluindo Distrito Federal) e local de trabalho (empresa, home office, misto).

O número mínimo de valores necessários para representar uma observação com os atributos acima descritos para o modelo de regressão, de forma que não ocorra perda de informação ordinal nem inserção de vieses nos dados, é:

- (A) 4;
- (B) 10;
- (C) 16;
- (D) 32;
- (E) 35.

22

Considere-se a aplicação de um modelo grande de linguagem (LLM) com 3 bilhões de parâmetros, distribuído em formato não quantizado, meia-precisão.

A quantidade mínima de memória necessária para carregar os pesos do modelo para inferência (sem gradientes), considerando apenas o espaço ocupado pelos pesos, é:

- (A) 8GB;
- (B) 16GB;
- (C) 32GB;
- (D) 64GB;
- (E) 128GB.

23

Para tomar decisões diárias, o analista João precisa consultar o preço de diversas ações do mercado financeiro, bem como outros dados da CVM. Contudo, ao acessar o ambiente de BigData da CVM, João verificou que os preços das ações desse ambiente demoravam para ser atualizados.

João procurou o arquiteto de BigData da CVM para tratar a dimensão da qualidade de dados denominada:

- (A) validade (*validity*);
- (B) correspondência (*matching*);
- (C) unicidade (*uniqueness*);
- (D) consistência (*consistency*);
- (E) temporalidade (*timeliness*).

24

Um cientista trabalha em um projeto de cibersegurança no qual deve identificar atividades de rede incomuns, que podem indicar possíveis ameaças de segurança. Após coletar dados sobre o tráfego de rede, o cientista percebe que alguns pontos de dados exibem padrões significativamente diferentes da maioria.

O método que deve ser aplicado para detectar as anomalias, identificando outliers isolados nos dados e possíveis ameaças de segurança, é o:

- (A) Isolation Forest;
- (B) Box Plot (IQR);
- (C) *Local Outlier Factor* (LOF);
- (D) do desvio padrão;
- (E) DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*).

25

O analista João desenvolveu diversos dashboards na plataforma Microsoft Power BI e solicitou ao programador web Pedro que disponibilizasse aquelas análises no portal de acesso à informação da CVM.

Para inserir os dashboards desenvolvidos por João na página web da CVM, o recurso do Power BI que Pedro deve usar é o(a):

- (A) Modelo semântico;
- (B) Relatório Paginado;
- (C) Conexão com fontes de dados Web;
- (D) Análise Integrada (*Embedded Analytics*);
- (E) Visualização Narrativa Inteligente (*Smart Narrative*).

Linguagens e Banco de Dados

26

Igor, analista de dados da CVM, escreveu e rodou o código a seguir.

```
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
```

```
texto = "Eu sou um analista de dados da CVM!"
```

```
stop_words =
set(stopwords.words('portuguese'))
tokens = word_tokenize(texto)
```

```
tokens_processados = [w for w in tokens if
not w in stop_words]
```

```
print(tokens_processados)
```

Considerando que o código foi executado sem erros e sabendo que Igor está usando Python 3.10.12 e NLTK 3.8.1, a saída do terminal foi:

- (A) ['analista', 'dados', 'CVM'];
- (B) ['analista', 'dados', 'CVM', '!'];
- (C) ['Eu', 'analista', 'dados', 'CVM'];
- (D) ['Eu', 'analista', 'dados', 'CVM', '!'];
- (E) ['Eu', 'sou', 'analista', 'dados', 'CVM', '!'].

27

O ambiente analítico da CVM armazena dados no formato multidimensional, implementa um cubo de dados e disponibiliza uma ferramenta OLAP para apoiar os analistas no desempenho de suas funções. Diante da grande quantidade de dados disponíveis, eles precisam reduzir o seu domínio de análise.

Para isso, a operação OLAP, que extrai um subcubo da seleção de duas ou mais dimensões de um cubo de dados, é a:

- (A) Slice;
- (B) Dice;
- (C) Pivot;
- (D) Roll up;
- (E) Drill down.

28

Um cientista de dados utiliza a biblioteca scikit-learn para treinar um estimador clf usando um conjunto de treinamento X_1 e seu respectivo conjunto de atributos-alvo y . Posteriormente, o cientista estima os atributos-alvo do conjunto X_2 .

Para realizar o treinamento e a predição, o cientista de dados deve usar, respectivamente, os métodos:

- (A) $fit(X_1, y)$ e $estimate(X_2)$;
- (B) $fit(X_1, y)$ e $predict(X_2)$;
- (C) $train(X_1, y)$ e $estimate(X_2)$;
- (D) $train(X_1, y)$ e $predict(X_2)$;
- (E) $train(X_1, y)$ e $test(X_2)$.

29

O dado no formato JSON – padrão utilizado para armazenar e transportar dados – deve ser de algum tipo.

O dado que representa o tipo de dado JSON Array é o:

- (A) `["nome": "Joca"] {"idade": "30"}`
- (B) `{"idade": 30} {"idade": 20}`
- (C) `{"funcionario": {"nome": "Joca", "idade": 30}}`
- (D) `["Joca", "Ana", "Pedro"]`
- (E) `{"vendido": true} {"sobrenome": null}`

30

Documentos do Jupyter Notebook são salvos com a extensão `.ipynb`, mas internamente eles são documentos do tipo:

- (A) DTD;
- (B) HTML;
- (C) JSON;
- (D) XHTML;
- (E) XML.

31

No método `tensorflow.keras.layers.Dense(...)`, se nenhuma função de ativação é especificada, é utilizada por padrão a função:

- (A) linear;
- (B) ReLU;
- (C) sigmoide;
- (D) softmax;
- (E) softplus.

32

As demandas de dados pelos analistas e cientistas de dados da CVM estão aumentando a cada dia. Para atendê-las com agilidade, é necessário obter dados de diversas fontes heterogêneas no seu formato original para posterior seleção e processamento sob demanda.

Para armazenar dados estruturados, não estruturados e semiestruturados, deve-se implementar um(a):

- (A) Stage Area;
- (B) Data Lake;
- (C) Data Mining;
- (D) Data Warehouse;
- (E) Operational Data Store.

33

Janine é a responsável pela administração dos bancos de dados gerenciados pelo PostgreSQL de uma autarquia federal. Durante a criação de um banco de dados, Janine especificou a criação de um tablespace diferente do tablespace default.

O tablespace criado por Janine:

- (A) é independente dos metadados contidos no diretório de dados principal;
- (B) será tratado pelo PostgreSQL como uma coleção autônoma de arquivos de dados;
- (C) pode ser anexado a outro cluster de banco de dados diferente daquele no qual foi criado;
- (D) pode ter uma rotina de backup individual, independente de outros componentes do banco de dados;
- (E) requer que o superusuário do banco de dados conceda o privilégio de criação no tablespace aos usuários comuns do banco de dados.

34

As transações financeiras da CVM são realizadas sobre ativos nacionais, utilizando a moeda real (R\$), e ativos internacionais, utilizando a moeda dólar americano (US\$).

Para implementar um Data Mart Financeiro da CVM, permitindo análises dos ativos nas duas moedas, deve-se:

- (A) criar um par de colunas para cada moeda na Tabela Fato, uma contendo o valor da moeda utilizada na transação e outra contendo o valor da moeda convertida. Criar uma Dimensão relacionada contendo a moeda utilizada na transação;
- (B) criar uma Minidimensão contendo duas colunas, uma com o valor da transação em real e a outra coluna com o valor em dólar. Correlacionar cada registro da Minidimensão com o respectivo valor da Tabela Fato;
- (C) normalizar a Dimensão que contém as moedas utilizadas em transações financeiras criando um snowflake apenas para as moedas real e dólar americano. Na Tabela Fato, criar um registro para cada transação relacionando-o com o respectivo registro da Dimensão normalizada;
- (D) criar uma Dimensão Auditoria para armazenar a descrição da moeda utilizada na transação armazenada na Tabela Fato, a qual deve armazenar o valor convertido;
- (E) entrevistar o gestor financeiro para levantar a regra de negócio da conversão de câmbio. Criar a tabela Dimensão correspondente e implementar a regra de negócio em um pacote de Extração, Transformação e Carga (ETL) para carregar o valor convertido a cada nova transação.

35

O analista Gabriel fez um levantamento das bases de dados existentes na CVM e percebeu que havia Data Marts distintos, criados para atender a requisitos analíticos específicos de cada Superintendência, como: Relações Institucionais, Auditoria e Registro de Valores Imobiliários. Cada Data Mart foi construído de forma independente, o que dificultava análises integradas para relacionar dados das diferentes Superintendências. Gabriel observou que havia várias dimensões em comum nos Data Marts. Para permitir análises integradas padronizando e compartilhando as dimensões em comum dos Data Marts da CVM, Gabriel implementou um(a):

- (A) Data Mart Híbrido;
- (B) Data Mart Dependente;
- (C) Modelo de Dados Normalizado;
- (D) Repositório de Metadados Mestres Corporativos;
- (E) Arquitetura e Matriz de Barramento de Data Warehouse.

36

Uma das principais atividades de implementação de um ambiente analítico é a limpeza dos conjuntos de dados origem. A biblioteca Pandas do Python é utilizada para analisar e também para limpar conjuntos de dados. Observe o trecho de código Python a seguir, que utiliza a biblioteca Pandas.

```
import pandas as pd
df = pd.read_csv('data.csv')
df.dropna(inplace = True)
```

O resultado da execução do código apresentado é a alteração do conjunto original de dados com o(a):

- (A) eliminação das linhas duplicadas;
- (B) truncamento dos dados do tipo boolean;
- (C) correção dos dados que estão na posição errada;
- (D) remoção das linhas com células vazias;
- (E) alteração do formato errado dos dados.

37

Os Sistemas Gerenciadores de Banco de Dados (SGBD) comerciais implementam internamente técnicas para processar, otimizar e executar consultas de alto nível.

Uma estratégia eficiente utilizada pelo otimizador de consultas do SGBD considera o uso de:

- (A) regras heurísticas de forma a realizar restrições (seleções) antes de projeções para reduzir o número de atributos para as projeções;
- (B) estimativas de custo para identificar restrições (seleções) mais eficientes como entrada para execução das projeções com menos atributos;
- (C) regras heurísticas de forma a realizar restrições (seleções) antes de projeções para reduzir o número de tuplas para as projeções;
- (D) estimativas de custo para consultas compiladas com o intuito de realizar junções e projeções antes das restrições (seleções);
- (E) regras heurísticas associadas a estimativas de custo para consultas interpretadas para eliminar o produto cartesiano.

38

O modelo relacional representa o banco de dados como uma coleção de relações. Considere a relação COLABORADOR apresentada a seguir, cuja chave primária é Matrícula.

COLABORADOR

Nome	Matricula	Cpf	Celular	Endereco
Fernando Azul	1234	422.111.232-78	(21) 9888-9991	Rua das Flores, 8
Carla Vermelha	1357	444.222.333-77	(21) 9777-7474	Rua da Árvore, 4
Vivian Amarela	1458	147.854.111-12	NULL	Rua das Frutas, 2

Na relação COLABORADOR, o(a):

- (A) superchave padrão é representada pelos atributos Matrícula e Cpf;
- (B) estado de relação atual reflete a cardinalidade dos domínios e o produto cartesiano;
- (C) ordenação das tuplas faz parte da definição da relação sendo influenciada pela alteração da ordem das tuplas;
- (D) domínio de cada atributo é um conjunto de valores atômicos, indivisível em se tratando de modelo relacional formal;
- (E) restrição de integridade referencial garante que nenhum valor da chave primária será NULL.

39

As transações em banco de dados possuem propriedades que buscam proteger dados contra perdas ou danos.

A propriedade durabilidade tem relação com:

- (A) a técnica de recuperação, para desfazer quaisquer efeitos de transações não completadas por algum motivo;
- (B) as restrições de integridade especificadas para o esquema, bem como quaisquer outras restrições no banco de dados;
- (C) o subsistema de recuperação do Sistema Gerenciador de Banco de Dados, por meio do log do sistema, para transações confirmadas e modificáveis pela execução de outra transação futura;
- (D) o ocultamento das atualizações por parte de cada transação até que esta esteja confirmada;
- (E) os níveis de durabilidade de uma transação variando do nível 1 ao 3, de forma a evitar atualizações perdidas.

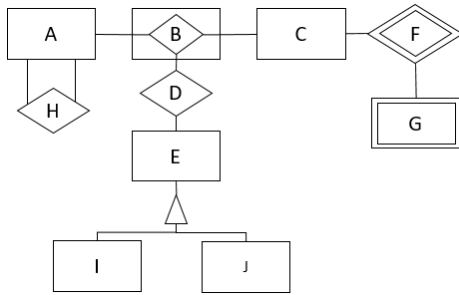
40

Diante de várias reclamações de performance em resposta a consultas a dados por meio de um dos sistemas estruturantes de uma autarquia federal, a equipe de tecnologia identificou que o motivo estava na lentidão para recuperação de registros na base de dados utilizada pelo sistema. Para agilizar a recuperação de registros em resposta a uma pesquisa que utiliza um campo que comporta valores repetidos, a equipe de tecnologia criou índices. Considerando que já existe um índice primário para o conjunto de dados em questão, a equipe criou um índice:

- (A) multinível secundário esparsa de chave candidata;
- (B) de agrupamento (clustering) denso para o campo de ordenação física;
- (C) bitmap para um ou mais campos de pesquisa para os quais há grandes variações de valores;
- (D) de hash para o campo de pesquisa que é utilizado para organização do índice primário;
- (E) secundário não denso para o campo de pesquisa com um nível de indireção extra para múltiplos ponteiros.

41

Observe o Modelo de Entidades e Relacionamentos a seguir.

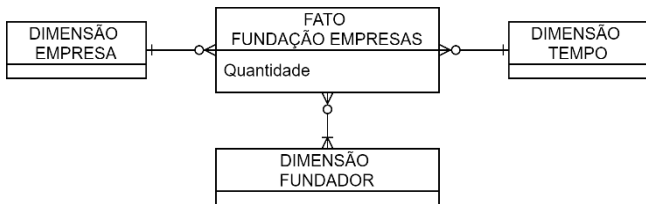


Com base nos relacionamentos apresentados, está explícito que:

- (A) I e J podem se relacionar com B;
- (B) H transforma A em uma entidade-forte;
- (C) B é um relacionamento semântico ternário;
- (D) G tem uma dependência do tipo todo-parte em relação a C;
- (E) I é um tipo de E dependente do relacionamento com B.

42

Para apoiar análises sobre os fundadores de empresas ao longo do tempo, elaborou-se, inicialmente, o seguinte modelo multidimensional de dados, no qual a tabela FATO FUNDAÇÃO EMPRESAS se relaciona com múltiplos valores da tabela DIMENSÃO FUNDADOR.



No caso apresentado, a implementação de uma dimensão multivalorada deve ser realizada por meio da aplicação da técnica de modelagem multidimensional:

- (A) Tabela Ponte (*Bridge Tables*);
- (B) Tabela Fato sem Fatos (*Factless Fact Tables*);
- (C) Dimensão Degenerada (*Degenerate Dimensions*);
- (D) Dimensão Floco de Neve (*Snowflaked Dimensions*);
- (E) Dimensão que Desempenha Múltiplos Papéis (*Role-Playing Dimensions*).

43

Observe o script SQL a seguir.

```
CREATE TABLE Pessoa (Nome varchar(255));
INSERT INTO Pessoa (Nome)
VALUES ('Ana Luca');
INSERT INTO Pessoa (Nome)
VALUES ('Antonio Silva');
INSERT INTO Pessoa (Nome)
VALUES ('Tania Rosendo');
INSERT INTO Pessoa (Nome)
VALUES ('Rosa Flor');
INSERT INTO Pessoa (Nome)
VALUES ('Lara Bel');
SELECT * FROM Pessoa
WHERE (Nome LIKE 'A A%'
OR Nome LIKE 'R%' /*OR Nome LIKE '%BEL'
OR Nome LIKE '%T%*/ OR Nome LIKE '%F');
```

O resultado JSON da execução do script apresentado é:

- (A) [{"Nome": "Ana Luca"}, {"Nome": "Rosa Flor"}]
- (B) [{"Nome": "Rosa Flor"}, {"Nome": "Lara Bel"}]
- (C) [{"Nome": "Ana Luca"}, {"Nome": "Tania Rosendo"}, {"Nome": "Lara Bel"}]
- (D) [{"Nome": "Antonio Silva"}, {"Nome": "Tania Rosendo"}, {"Nome": "Rosa Flor"}]
- (E) [{"Nome": "Ana Luca"}, {"Nome": "Antonio Silva"}, {"Nome": "Tania Rosendo"}, {"Nome": "Lara Bel"}]

44

Durante a preparação de dados para análise, alterar a quantidade de dimensões e/ou elementos de cada dimensão de um dado do tipo Array é uma operação importante na manipulação de dados. Observe o seguinte trecho de código escrito na linguagem de programação Python:

```
import numpy as np
arr = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9,
10, 11, 12])
newarr = arr.reshape(2, 3, 2)
print(newarr)
```

O resultado da execução do código apresentado é:

- (A)

```
[[ 1  2  3]
 [ 4  5  6]
 [ 7  8  9]
 [10 11 12]]
```
- (B)

```
[[ 1  2  3  4]
 [ 5  6  7  8]
 [ 9 10 11 12]]
```
- (C)

```
[[[ 1  2]
 [ 3  4]
 [ 5  6]]
 [[ 7  8]
 [ 9 10]
 [11 12]]]
```
- (D)

```
[[ 1  2  3  4  5  6]
 [ 7  8  9 10 11 12]]
```
- (E)

```
[[[ 1  2  3]
 [ 4  5  6]]
 [[ 7  8  9]
 [10 11 12]]]
```

45

Observe o script SQL a seguir.

```
CREATE TABLE AUDITOR (ID_A INT NOT NULL,
Nome varchar(20), PRIMARY KEY (ID_A));
CREATE TABLE AUDITADO (ID_O INT NOT NULL,
Nome varchar(20), ID_A INT,
PRIMARY KEY (ID_O), FOREIGN KEY (ID_A)
REFERENCES AUDITOR(ID_A));
INSERT INTO AUDITOR (ID_A, Nome)
VALUES (1, 'Maite');
INSERT INTO AUDITOR (ID_A, Nome)
VALUES (2, 'Lucca');
INSERT INTO AUDITOR (ID_A, Nome)
VALUES (3, 'Maria Clara');
INSERT INTO AUDITADO (ID_O, Nome, ID_A)
VALUES (1, 'Felipe', 1);
INSERT INTO AUDITADO (ID_O, Nome, ID_A)
VALUES (2, 'Stella', 2);
INSERT INTO AUDITADO (ID_O, Nome, ID_A)
VALUES (3, 'Patricia', NULL);
```

Para analisar quais auditores estão realizando auditoria em quais auditados, é necessária a execução de uma consulta SQL que apresente o seguinte resultado:

Auditor	Auditado
Maite	Felipe
Lucca	Stella

Para obter o resultado apresentado, deve-se executar a consulta SQL:

- (A)

```
SELECT AUDITOR.Nome as 'Auditor'
FROM AUDITOR
WHERE AUDITOR.ID_A = ALL
(SELECT AUDITADO.Nome as 'Auditado'
FROM AUDITADO
WHERE AUDITADO.ID_A=AUDITOR.ID_A);
```
- (B)

```
SELECT AUDITOR.Nome as 'Auditor',
AUDITADO.Nome as 'Auditado'
FROM AUDITOR, AUDITADO
HAVING AUDITADO.ID_A=AUDITOR.ID_A;
```
- (C)

```
SELECT AUDITOR.Nome as 'Auditor',
AUDITADO.Nome as 'Auditado'
FROM AUDITOR INNER JOIN AUDITADO ON
AUDITADO.ID_A=AUDITOR.ID_A;
```
- (D)

```
SELECT AUDITOR.Nome as 'Auditor'
FROM AUDITOR
WHERE EXISTS (SELECT AUDITADO.Nome as
'Auditado'
FROM AUDITADO
WHERE AUDITADO.ID_A=AUDITOR.ID_A);
```
- (E)

```
SELECT AUDITOR.Nome as 'Auditor'
FROM AUDITOR
UNION SELECT AUDITADO.Nome as 'Auditado'
FROM AUDITADO;
```

Matemática e Estatística

Para a resolução das questões 61 a 70, pode ser necessário utilizar alguns dos resultados a seguir.

- Probabilidades aproximadas da Normal padrão ($Z \sim N(0,1)$):

$P(Z > 0.25) = 0.40$	$P(Z > 0.5) = 0.31$
$P(Z > 0.8) = 0.21$	$P(Z > 1) = 0.16$
$P(Z > 1.2) = 0.12$	$P(Z > 1.28) = 0.1,$
$P(Z > 1.5) = 0.07$	$P(Z > 1.64) = 0.05$
$P(Z > 1,96) = 0.025$	$P(Z > 2) = 0.02$
$P(Z > 2,33) = 0,01$	$P(Z > 2.5) = 0.06;$
$Pr(Z > 2,575) = 0,005$	$P(Z > 3) = 0.013$

- Valores aproximados da função exponencial:

$\exp(-1/40) = 0.97$	$\exp(-1) = 0,368$
$\exp(-2) = 0,135$	$\exp(-4) = 0,018$

- Valores aproximados da função logaritmo natural:

$\ln(2) = 0,7$	$\ln(3) = 1,1$	$\ln(4) = 1,4.$
----------------	----------------	-----------------

Também podem ser úteis os trechos de tabelas das distribuições a seguir.

- Distribuição t de Student:

Graus de Liberdade	Área de extremidade superior				
	10%	5%	2,5%	1%	0,5%
15	1,341	1,753	2.131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845

- Distribuição qui-quadrado:

Graus de Liberdade	Área de extremidade superior				
	2,5%	2%	1%	0,2%	0,1%
6	14,449	15,033	16,812	20,791	22,457
7	16,013	16,622	18,472	22,601	24,322
8	17,534	18,168	20,090	24,352	26,125
9	19,023	19,679	21,666	26,056	27,877
10	20,483	21,161	23,209	27,722	29,588

- Distribuição qui-quadrado:

Graus de Liberdade	Área de extremidade superior				
	99%	98%	97,5%	95%	90%
8	1,646	2,032	2,180	2,733	3,490
9	2,088	2,532	2,700	3,325	4,168
10	2,558	3,059	3,247	3,940	4,865
11	3,053	3,609	3,816	4,575	5,578
12	3,571	4,178	4,404	5,226	6,304

46

Suponha que o tempo T até um que investidor solicite o resgate integral de um fundo, em meses, seja representado por uma variável aleatória contínua com função de densidade

$$f(t) = 0,05e^{-0.05t}, t > 0.$$

De acordo com esse modelo probabilístico, o período até que a metade dos investidores desse fundo venha a solicitar o resgate integral é de, aproximadamente:

- (A) 7 meses;
- (B) 11 meses;
- (C) 14 meses;
- (D) 18 meses;
- (E) 20 meses.

47

Um indicador de desempenho das instituições que atuam no mercado financeiro brasileiro é avaliado com base nas 8 observações de uma amostra aleatória simples, considerando 8 dessas instituições. O desvio padrão amostral do indicador foi igual a 8.

Supondo que a distribuição dos valores do indicador no universo em estudo seja Normal, o limite inferior do intervalo de confiança de 95% para a variância populacional é, aproximadamente (considere probabilidades iguais nas caudas):

- (A) 16;
- (B) 20;
- (C) 24;
- (D) 28;
- (E) 32.

48

Um analista busca evidenciar estatisticamente a conjectura de que a valorização média das cotas dos fundos imobiliários negociados no mercado em 2023 tenha sido superior a 15%. Supõe-se que as valorizações das cotas sigam distribuição Normal, sendo o desvio padrão desconhecido. Com base nas observações de uma amostra aleatória de tamanho 16, ele observa que a valorização média foi de 15,85%, com desvio padrão amostral igual a 2%.

Considerando os três níveis de significância usuais (0,01, 0,05 e 0,1), a conjectura investigada:

- (A) foi evidenciada apenas aos níveis 0,05 e 0,1, o que resulta da comparação do valor da estatística de teste com os valores críticos 1,64 e 1,28;
- (B) foi evidenciada apenas ao nível 0,1, o que se conclui a partir da comparação do valor da estatística de teste com os valores críticos 1,96 e 1,64;
- (C) foi evidenciada apenas ao nível 0,1, o que se conclui a partir da comparação do valor da estatística de teste com os valores críticos 1,64 e 1,28;
- (D) foi evidenciada apenas ao nível 0,1, o que se conclui a partir da comparação do valor da estatística de teste com os valores críticos 1,753 e 1,341;
- (E) não foi evidenciada a nenhum dos três níveis, o que se conclui comparando o valor da estatística de teste com os valores 2,947, 2,131 e 1,753.

49

O número de fraudes anuais detectadas no mercado financeiro, nos últimos 16 anos, foi registrado por um auditor. Ele deseja testar se o resultado fornece evidência de que a média anual de fraudes no mercado é inferior a 4, supondo que esses 16 registros constituam observações de uma amostra aleatória simples obtida a partir de uma população Normal. A variância dessa população é conhecida e igual a 25.

Nessas condições, o auditor obterá evidência estatística de que a média populacional é inferior a 4, ao nível de significância 0,1, se a média na amostra for menor ou igual a:

- (A) 1,6;
- (B) 1,8;
- (C) 2,4;
- (D) 3,6;
- (E) 5,6.

50

Suponha que sejam usados indicadores para avaliar a possibilidade de inadimplência de títulos emitidos no mercado, e seja X um desses indicadores. Se X assume um valor inferior a 4, a probabilidade de que o emissor do título venha a se tornar inadimplente é de apenas 0,2. Por outro lado, se X estiver acima de 7, a probabilidade de inadimplência é de 0,6. Finalmente, se o indicador estiver situado entre 4 e 7 (incluindo os extremos), o título emitido possui probabilidade de inadimplência igual a 0,4. Quando se considera o universo de todos os títulos emitidos neste mercado, os valores de X seguem distribuição Normal com média 6 e variância 4.

Dado que o emissor de um determinado título se tornou inadimplente, a probabilidade de que o valor de X associado a ele estivesse situado entre 4 e 7 é:

- (A) 48/215;
- (B) 62/215;
- (C) 96/215;
- (D) 106/215;
- (E) 158/215.

51

Um analista investiga, mediante um modelo de regressão linear clássico, a relação entre a rentabilidade y de ofertas públicas disponíveis no mercado e um indicador de risco associado ao emissor, representado pela variável explicativa x . Considera-se que o termo de erro do modelo siga distribuição Normal. Foi utilizada uma amostra aleatória simples de 20 pares (x,y) de observações mensais. O modelo estimado está apresentado a seguir (erros padrão entre parênteses).

$$y = 0,528 + 0,627x$$

(0,264) (0,300)

O intervalo de 95% de confiança associado ao impacto de x sobre y é (considere apenas 3 casas decimais):

- (A) [-0,009;1,255], e o impacto não é significativo ao nível 0,05;
- (B) [-0,033;1,257], e o impacto não é significativo ao nível 0,05;
- (C) [0,039;1,215], e o impacto é significativo ao nível 0,05;
- (D) [0,135;1,119], e o impacto é significativo ao nível 0,05;
- (E) [0,327;0,927], e o impacto é significativo ao nível 0,05.

52

Um gestor avalia a expectativa de rentabilidade mensal de um fundo de ações utilizando o modelo de regressão linear clássico $y = \beta_0 + \beta_1x + \epsilon$, em que y é a rentabilidade, x é um indicador econômico, β_0 e β_1 são parâmetros a serem estimados por mínimos quadrados e ϵ é o termo de erro. O modelo satisfaz aos pressupostos para estimação por mínimos quadrados. Com base em uma amostra de 3 meses, na qual os valores observados da variável explicativa x foram $x_1 = 1$, $x_2 = 2$ e $x_3 = 2$, o modelo estimado conduziu aos resíduos $e_1 = 2$, $e_2 = 1$ e $e_3 = 1$.

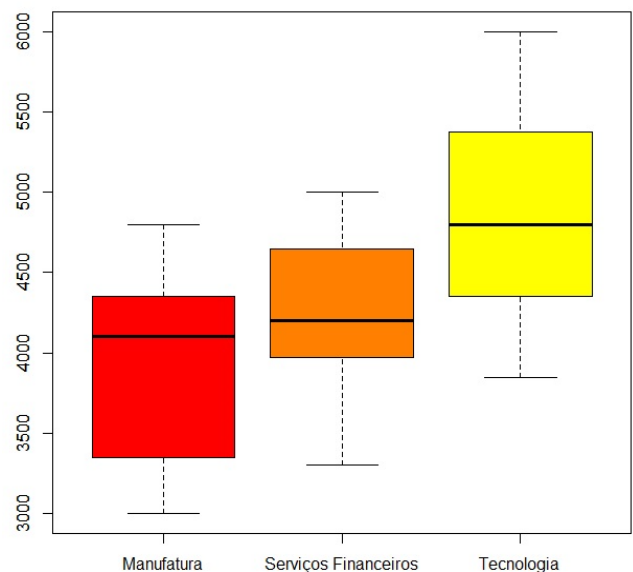
A estimativa, baseada no estimador não viciado, para a covariância entre os estimadores de β_0 e β_1 , é:

- (A) - 4;
- (B) - 5;
- (C) -10;
- (D) -15;
- (E) -20.

53

Um analista estuda discrepâncias salariais entre os seguintes setores: manufatura, serviços financeiros e tecnologia. A figura a seguir apresenta os *box-plots* dos salários desses setores, em reais.

Distribuição dos salários por setor



Dentre as afirmativas a seguir, a única correta é:

- (A) o primeiro quartil de todos os setores é superior a 4.000 reais;
- (B) a amplitude interquartil dos salários do setor de tecnologia vale aproximadamente 2.000 reais;
- (C) um salário de 5.500 seria identificado como uma observação discrepante, ou seja, um *outlier*, para todos os três setores;
- (D) a mediana dos salários do setor de serviços financeiros é menor do que o primeiro quartil do setor de tecnologia;
- (E) o menor salário observado para o setor de tecnologia é maior do que a mediana dos salários do setor de manufatura.

54

Em um concurso, 2.048 candidatos prestam um exame em que são submetidos a 6 questões de múltipla escolha, cada uma com 4 alternativas, das quais apenas uma é correta. Um candidato passa para a segunda fase do concurso caso acerte, pelo menos, 4 questões.

Se todos os candidatos “chutam” as respostas, isto é, sempre escolhem ao acaso uma alternativa, o valor esperado do número de aprovados para a segunda fase é:

- (A) 77;
- (B) 85;
- (C) 98;
- (D) 116;
- (E) 128.

55

O desempenho de uma variável econômica em 2023, em termos das variações percentuais trimestrais, foi armazenado no vetor $(2, -1, 4, -2)'$. Com o intuito de avaliar a variabilidade nesse ano, foram consideradas as quatro métricas a seguir.

1. norma L1 das componentes do vetor;
2. norma L2 das componentes do vetor;
3. soma dos valores absolutos (módulos) dos desvios de cada componente do vetor em relação à mediana de todos;
4. raiz quadrada da soma dos quadrados dos desvios em relação à sua média, isto é, o numerador do desvio padrão.

A respeito da comparação entre os valores assumidos por essas medidas, a única afirmativa correta é:

- (A) 1 e 3 coincidem, e são inferiores a todos os demais valores;
- (B) 3 é inferior a todos os demais, ao passo que 2 e 4 coincidem;
- (C) 1 é inferior a todos os demais, ao passo que 2 e 4 coincidem;
- (D) 4 é inferior a todos os demais, ao passo que 1 e 3 coincidem;
- (E) 2 e 4 coincidem, e são inferiores a todos os demais valores.

56

Uma agência reguladora recebe, em média, uma denúncia a cada 15 minutos.

Se o número de denúncias em um período qualquer segue distribuição de Poisson, a probabilidade de que, no intervalo de 1 hora, cheguem pelo menos 2 denúncias, sabendo-se que pelo menos uma denúncia terá chegado, é de:

- (A) 72/910;
- (B) 18/125;
- (C) 288/982;
- (D) 91/100;
- (E) 455/491.

57

Suponha que o tempo X , em dias, até que uma debênture incentivada aumente seu valor de mercado em 30%, seja uma variável aleatória com função de densidade

$$f(x) = \theta^2 x e^{-\theta x}; x > 0.$$

O tempo médio registrado, com base nas observações de uma amostra aleatória simples, foi de 400 dias.

Com base nessa amostra, a estimativa de máxima verossimilhança do parâmetro θ é:

- (A) 1/40;
- (B) 1/80;
- (C) 1/100;
- (D) 1/200;
- (E) 1/400.

58

Considere duas séries temporais x e y , ambas integradas de ordem 1, ou $I(1)$, representando a evolução de agregados macroeconômicos no tempo. Ao aplicarmos o teste de raiz unitária ADF aos resíduos da regressão linear de y em x (com valores críticos propostos por Engle-Granger para aplicá-lo a resíduos de uma regressão), verifica-se que a hipótese nula não é rejeitada, aos níveis usuais.

É correto concluir que essas séries:

(Obs: os valores críticos propostos por Engle-Granger para esse tipo de teste não são necessários para a resolução da questão)

- (A) são cointegradas, pois tanto as séries quanto os resíduos são estacionários, o que torna a regressão entre elas válida;
- (B) não são cointegradas, pois, apesar de serem estacionárias, os resíduos da regressão entre elas não são estacionários;
- (C) são cointegradas, pois, embora não sejam estacionárias, os resíduos da regressão entre elas são estacionários;
- (D) não são cointegradas, pois não são estacionárias e os resíduos da regressão entre elas não possuem raiz unitária;
- (E) não são cointegradas, pois, embora não sejam estacionárias, os resíduos da regressão entre elas possuem raiz unitária.

59

A proporção de emissões de títulos imobiliários com suspeita de irregularidade em um ano pode ser representada por uma variável aleatória contínua X com função de densidade:

$$f(x) = (\theta + 1)x^\theta, 0 < x < 1$$

Deseja-se conduzir uma análise probabilística dessa proporção em 2024; porém, para isso, é preciso estimar o parâmetro θ . Nos últimos 5 anos, a proporção anual registrada foi: 0,3; 0,2; 0,6; 0,7 e 0,2.

Considerando que esses registros sejam observações de uma amostra aleatória simples da população referenciada por $f(x)$, a estimativa do parâmetro θ a partir dessa amostra, obtida pelo método dos momentos, é:

- (A) -2/3;
- (B) -1/3;
- (C) 1/3;
- (D) 2/5;
- (E) 2/3.

60

Um analista financeiro tenta prever a rentabilidade anual futura de um ativo, em termos reais. Ele considera que a rentabilidade real (em %) siga, ao longo dos anos, um modelo AR(1): $y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$, em que t é o ano, $E(\varepsilon_t) = 0$ e $\text{corr}(\varepsilon_t, \varepsilon_{t-s}) = 0$, para $s = 1, 2, \dots$. Sabe-se que a rentabilidade real prevista pelo modelo para o longuíssimo prazo foi de 4% ao ano. Se a estimativa obtida para o parâmetro ϕ_1 foi 0,8, a estimativa do parâmetro ϕ_0 foi:

- (A) 0,2;
- (B) 0,4;
- (C) 0,5;
- (D) 0,6;
- (E) 0,8.

Gestão de Projetos

61

Na fase de planejamento de projetos, são elaborados os cronogramas. Um cronograma é um modelo para a execução das atividades do projeto, incluindo durações, dependências e outras informações de planejamento. O cronograma do projeto parte da decomposição de seu escopo em atividades específicas. Essas atividades podem estar ordenadas em série, em paralelo e em combinações série/paralelo.

Para tal são usadas as seguintes classes de dependências entre as atividades:

- (A) dependências críticas e não críticas;
- (B) dependências obrigatórias, arbitradas, externas, e internas;
- (C) dependências lógicas e físicas;
- (D) dependências obrigatórias, contratuais, lógicas e internas;
- (E) dependências do caminho crítico, externas e internas.

62

Você é gerente de um projeto cujo prazo total é de oito meses e que está em fase de execução. No último dia do quarto mês do projeto, você solicitou o cálculo do índice de desempenho de prazo do projeto, e o resultado, nessa data, foi de 50%.

Tal índice significa que:

- (A) o projeto está caminhando conforme o previsto, pois está na metade da duração estabelecida;
- (B) o projeto está pronto para ser encerrado, pois as atividades foram realizadas na metade do tempo previsto;
- (C) o projeto teve desembolsos equivalentes a 50% do que havia sido previsto no plano inicial;
- (D) o projeto deve levar o dobro do tempo para ser concluído, caso o desempenho passado permaneça o mesmo a partir de hoje;
- (E) conclusões não podem ser tiradas sem informações das entregas realizadas no período.

63

Em termos de recursos humanos, os projetos possuem gerentes, equipes de gerenciamento e as equipes de projeto, com as pessoas efetivamente alocadas às atividades ao longo do ciclo de vida do projeto.

Independentemente de como as atividades de gerenciamento são estruturadas, existem aspectos comuns de desenvolvimento que são relevantes para a maioria das equipes de projeto.

Tais aspectos incluem:

- (A) atribuição de tarefas; monitoramento e controle; avaliação de desempenho; lições aprendidas;
- (B) liderança; trabalho em equipe; negociação; solução de conflitos;
- (C) visão e objetivos; papéis e responsabilidades; operações da equipe do projeto; orientação; e crescimento;
- (D) visão e objetivos; valor do projeto; benefícios esperados; crescimento pessoal;
- (E) papéis e responsabilidades; atribuições de atividades do projeto; execução das atividades; gestão do conhecimento.

64

Você foi convocado para uma reunião com a cúpula estratégica da sua empresa. Na reunião, foi informado de que um projeto em execução em uma filial localizada em outro estado está enfrentando problemas. Atrasos, gastos em excesso e conflitos técnicos entre membros da equipe têm sido relatados à diretoria. Você foi, então, designado para acompanhar o projeto, coletar e informar os indicadores de desempenho e identificar as causas prováveis dos problemas.

Sabendo que a sua empresa utiliza a metodologia de análise de valor agregado (EVA - *Earned Value Analysis*), você deve iniciar o trabalho com:

- (A) coleta dos indicadores de progresso de entregas, de progresso de cronograma e de progresso de custos;
- (B) coleta dos indicadores de desempenho de prazo e de custo;
- (C) cálculo da variação de prazo e de custo do projeto;
- (D) medição do valor planejado, valor agregado e custo real atuais;
- (E) medição dos valores orçados, dos gastos efetivos e do saldo do projeto.

65

Para aumentar a probabilidade de sucesso dos projetos e assegurar que seus objetivos sejam cumpridos, processos de gerenciamento são utilizados. São cinco os grupos de processos de gerenciamento de projetos.

Você foi designado para gerenciar um projeto que estava em execução no ano passado, mas foi interrompido por falta de recursos e mudança de prioridades na sua empresa. Agora a empresa quer retomar o projeto.

Para obter sucesso, você deverá começar o trabalho a partir do grupo de processos de:

- (A) execução, pois você precisa validar o escopo do trabalho já concluído até a data da interrupção;
- (B) encerramento, para encerrar as atividades realizadas antes da interrupção do projeto;
- (C) planejamento, pois a retomada vai requerer a atualização no plano de gerenciamento do projeto;
- (D) iniciação, pois você deve reiniciar o projeto a partir do ponto em que ele foi interrompido;
- (E) monitoramento e controle, pois, para retomar o projeto, você deve criar nova EAP (Estrutura Analítica de Projeto).

66

Você foi designado como gerente de um novo projeto, que logo terá o respectivo termo de abertura aprovado pela cúpula estratégica da sua organização.

As habilidades do gerente de projetos que são essenciais para você iniciar o trabalho são:

- (A) empreendedorismo; alocação de recursos; análise das partes interessadas;
- (B) construção de equipes; liderança; planejamento;
- (C) iniciação; planejamento; análise de partes interessadas;
- (D) resolução de conflitos; competências técnicas; competências comportamentais;
- (E) detalhamento do escopo; identificação das partes interessadas; organização.

67

Trabalhando em uma empresa de produtos de beleza e cuidados pessoais, você foi designado gerente de um projeto de lançamento de uma nova linha de perfumes no mercado. A data do evento de lançamento já foi definida pela cúpula estratégica. O escopo inclui a campanha de lançamento, definição dos pontos de venda, com instalação de quiosques nos principais shoppings, bem como a definição dos estoques iniciais e sua distribuição. Durante a realização dos processos de gerenciamento do cronograma do projeto, você solicitou que a equipe de gerenciamento fizesse a verificação das atividades, seu sequenciamento e respectivas estimativas de duração. Em uma reunião de planejamento, ao perguntar sobre o caminho crítico do projeto, você recebeu diferentes informações a respeito, uma de cada membro.

A informação correta é:

- (A) “Não há como calcular o caminho crítico do projeto, pois dependemos da fábrica para definição do lote inicial dos produtos”;
- (B) “Sem os quiosques prontos nos shoppings não há como concluir o projeto, portanto o caminho crítico passa pelas atividades de montagem dos quiosques”;
- (C) “De acordo com as estimativas de duração de cada atividade do projeto, bem como de seu respectivo sequenciamento, o caminho crítico do projeto foi calculado e será detalhado na apresentação”;
- (D) “O projeto tem muitas atividades que dependem de terceiros, como aquisição de matérias-primas novas, montagem de estruturas em shopping e contratação de agência de publicidade, portanto todas as atividades são críticas”;
- (E) “A diretoria da empresa não poderia ter definido a data de lançamento sem consultar a equipe; portanto, independentemente do caminho crítico, esse projeto deverá atrasar”.

68

Hoje em dia, todas as organizações realizam projetos, a partir de um direcionamento estratégico e visando, entre outros aspectos, a estarem aptas para atender às exigências crescentes de seus respectivos mercados, em um ambiente de intensa competição e inovação tecnológica. Ainda assim, os diferentes tipos de organizações terão abordagens diferentes na condução de projetos, em função de sua atividade fim.

Nesse sentido, é correto afirmar que organizações:

- (A) baseadas em projetos são aquelas empresas intensivas no desenvolvimento de novas tecnologias;
- (B) suportadas por projetos são aquelas nas quais a principal fonte de receita é oriunda das suas operações de produção de bens e/ou serviços;
- (C) caracterizadas por projetos em redes são unidades de negócio de um mesmo grupo empresarial, que executam projetos em parceria;
- (D) baseadas em processos são aquelas que adotam processos de gerenciamento de projetos alinhados com as áreas de conhecimento e os grupos de processos mencionados no Guia PMBOK (6ª. edição);
- (E) baseadas em portfólios são aquelas que possuem uma gestão integrada de projetos, programas e operações, com um portfólio ativo.

69

O planejamento de um projeto sempre envolve a determinação de estimativas, que se dividem em estimativas de esforço, duração, custos, pessoas e recursos físicos do trabalho. Dentre os aspectos ligados às estimativas, incluem-se as dimensões de intervalo, exatidão, precisão e confiança.

Dentre as caracterizações de cada uma dessas dimensões, a correta é:

- (A) intervalo: as estimativas tendem a ter uma reduzida gama no início do projeto, quando não há muitas informações sobre o projeto, o escopo do produto, partes interessadas, requisitos, riscos e outras informações;
- (B) exatidão: refere-se à assertividade de uma estimativa e está ligada ao intervalo, pois quanto maior a exatidão, menor o intervalo de valores em potencial;
- (C) exatidão: uma estimativa no início do projeto terá mais exatidão do que se for elaborada quando da realização de 25% do trabalho do projeto;
- (D) precisão: refere-se ao grau de erro associado à estimativa; a precisão das estimativas deve ser compatível com a exatidão obtida;
- (E) confiança: aumenta com a experiência, e a experiência de trabalho em um projeto anterior semelhante pode ajudar com o nível de confiança necessário.

70

O uso de softwares de gerenciamento de projetos tem se disseminado desde a década de 1980, inicialmente em computadores pessoais isolados (PCs) e depois com solução em rede e armazenamento na nuvem. O Microsoft Project (MS Project) é um software surgido em 1985 que se tornou uma espécie de padrão para a área de projetos.

A respeito do planejamento de recursos humanos no MS Project, é correto afirmar que:

- (A) o “nivelamento de recursos” permite que todos os recursos tenham a mesma carga de trabalho ao longo do projeto;
- (B) o “nivelamento de recursos”, aliado ao cálculo das folgas das atividades do projeto, distribui as alocações individuais de trabalho, minimizando picos e períodos de ociosidade;
- (C) a “planilha de recursos” permite a visualização de todas as atividades alocadas a um recurso específico;
- (D) o “planejador de equipe” permite a visualização do custo previsto para todos os recursos do projeto;
- (E) a “planilha de recursos” permite a visualização do desempenho de cada recurso ao longo da execução do projeto.

RASCUNHO

RASCUNHO

Realização

