

## -- CONHECIMENTOS ESPECÍFICOS --

### BLOCO I

Em um processo em que se utiliza a ciência de dados, o número de variáveis necessárias para a realização da investigação de um fenômeno é direta e simplesmente igual ao número de variáveis utilizadas para mensurar as respectivas características desejadas; entretanto, é diferente o procedimento para determinar o número de variáveis explicativas, cujos dados estejam em escalas qualitativas.

Considerando esse aspecto dos modelos de regressão, julgue o item a seguir.

**51** Para evitar um erro de ponderação arbitrária, deve-se recorrer ao artifício de uso de variáveis *dummy*, o que permitirá a estratificação da amostra da maneira que for definido um determinado critério, evento ou atributo, para então serem inseridas no modelo em análise; isso permitirá o estudo da relação entre o comportamento de determinada variável explicativa qualitativa e o fenômeno em questão, representado pela variável dependente.

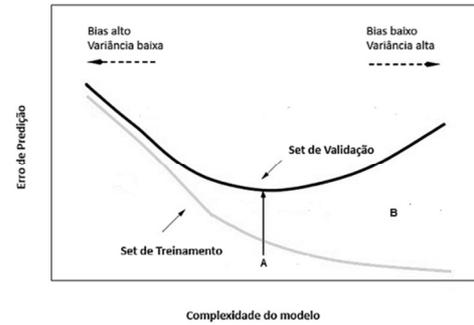
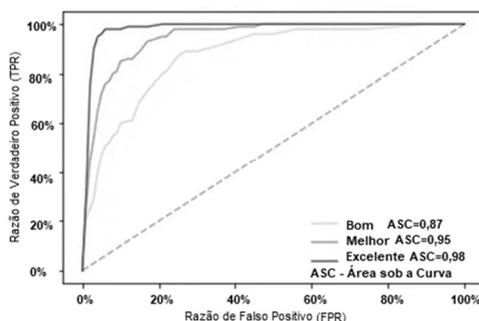
As métricas de avaliação de desempenho de um modelo de aprendizado de máquina, que é um componente integrante de qualquer projeto de ciência de dados, destinam-se a estimar a precisão da generalização de um modelo sobre os dados futuros (não vistos ou fora da amostra). Dentre as métricas mais conhecidas, estão a matriz de confusão, precisão, *recall*, pontuação, especificidade e a curva de características operacionais do receptor (ROC).

Acerca das características específicas dessas métricas, julgue os próximos itens.

**52** A matriz de confusão a seguir apresenta três rótulos de classe; os elementos diagonais representam o número de pontos para os quais o rótulo previsto é igual ao rótulo verdadeiro, enquanto qualquer coisa fora da diagonal teve um rótulo atribuído erroneamente pelo classificador. Quanto menores forem os valores diagonais da matriz de confusão, melhor o modelo adotado.

		Legume	Hortaliça	Fruta
Valor Apurado	Legume	13		
	Hortaliça		15	4
	Fruta			18
		Valor Preditivo		

**53** As curvas ROC a seguir mostram a taxa de especificidade (verdadeiros positivos) *versus* a taxa de sensibilidade (falsos positivos) do modelo adotado; a linha tracejada é a linha de base da métrica de avaliação e define uma adivinhação aleatória.



Julgue os próximos itens, considerando o gráfico precedente, que representa as regiões de *overfitting* e de *underfitting*, permitindo uma avaliação do relacionamento da complexidade do modelo de aprendizagem de máquina adotado, com o erro de predição.

**54** A região do gráfico entre as duas curvas, indicada pela letra B, mostra a região de erro de generalização para o modelo de aprendizado de máquina.

**55** O Set de Treinamento é usado para qualificar o desempenho do modelo, enquanto o Set de Validação é utilizado para criar o modelo de aprendizado de máquina.

**56** Considerando que a variância é um erro de sensibilidade para pequenas flutuações no conjunto de treinamento, infere-se que um baixo nível de variância pode fazer que o algoritmo associado a um modelo de aprendizado de máquina perca as relações relevantes entre os atributos de entrada e a variável de saída, caracterizando o erro de *overfitting*, percebido na região à direita do ponto A.

**57** Quando se verifica um alto erro no treinamento com valor próximo ao erro na validação, percebido na região à esquerda do ponto A, tem-se um clássico problema de *underfitting*, caracterizado por alto valor do *bias*.

Uma determinada repartição pública fez um levantamento do tempo  $y$ , em minutos, que os cinco funcionários de uma sessão gastam para chegar ao trabalho em função da distância  $x$ , em quilômetros, de suas residências. O resultado da pesquisa realizada com cada um deles é apresentado na tabela a seguir, em que  $\bar{x}$  e  $\bar{y}$  são, respectivamente, as médias amostrais das variáveis  $x$  e  $y$ .

$i$	tempo $y_i$	distância $x_i$	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) \times (y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	10	5	-4	-7	28	16
2	20	5	-4	3	-12	16
3	15	10	1	-2	-2	1
4	10	10	1	-7	-7	1
5	30	15	6	13	78	36
<b>média</b>	17	9				

Com base nos dados dessa tabela, julgue os próximos itens.

**58** Uma forma de melhorar o modelo de regressão linear para a situação em questão é utilizar o modelo de regressão logística, uma vez que a variável dependente se apresenta de forma quantitativa.

**59** Pelo modelo de regressão linear simples, a equação que expressa o relacionamento ajustado entre a variável  $y$  em função de  $x$  é  $\hat{y}_i = \frac{85}{70}x_i + \alpha$ , em que  $\alpha$  é uma constante.

Uma árvore de decisão representa um determinado número de caminhos possíveis de decisão e os resultados de cada um deles, apresentando muitos pontos positivos, ou seja, são fáceis de entender e interpretar. Elas têm processo de previsão completamente transparente e lidam facilmente com diversos atributos numéricos, assim como atributos categóricos, podendo até mesmo classificar dados sem atributos definidos.

De acordo com os aspectos construtivos de uma árvore de decisão, julgue os itens a seguir.

- 60** Se o processo adotado para a construção de árvores de decisão for determinístico, uma forma de obtenção de árvores aleatórias, que compõem as florestas aleatórias, pode ser realizada por meio do *bootstrap* dos dados, em que cada árvore é treinada com base no resultado de *bootstrap\_sample* (*inputs*).
- 61** A entropia de uma árvore de decisão aborda o aspecto da quantidade de informações que está associada às respostas que podem ser obtidas às perguntas formuladas, representando o grau de incerteza associado aos dados.

As máquinas de vetores de suporte (SVMs) são originalmente utilizadas para a classificação de dados em duas classes, ou seja, na geração de dicotomias. Nas SVMs com margens rígidas, conjuntos de treinamento linearmente separáveis podem ser classificados. Acerca das características das SVMs com margens rígidas, julgue o item a seguir.

- 62** Um conjunto linearmente separável é composto por exemplos que podem ser separados por pelo menos um hiperplano. As SVMs lineares buscam o hiperplano ótimo segundo a teoria do aprendizado estatístico, definido como aquele em que a margem de separação entre as classes presentes nos dados é minimizada.

Os modelos ditos fracos, também chamados modelos de base, muitas vezes são combinados com o objetivo de se construir um modelo mais forte, no qual a variância e o viés atinjam equilíbrio satisfatório. Esse procedimento, denominado *ensembles*, é muito utilizado em ciência de dados e aprendizado de máquinas. Quanto às formas de *ensembles*, julgue o próximo item.

- 63** O *ensemble* denominado *bagging* tem como foco principal a redução do viés e não da variância, treinando-se os modelos em sequência, tal que os erros dos primeiros modelos treinados são utilizados para o ajuste nos pesos matemáticos dos próximos modelos.

Os algoritmos de aprendizado supervisionado partem de um conjunto de dados rotulados para fazer previsões sobre novos dados não rotulados. O Python *scikit-learn* é uma biblioteca de código aberto utilizada para codificações de rotinas em aprendizado de máquina supervisionado; ela oferece ainda uma série de ferramentas utilizadas no ajuste de modelos e no pré-processamento de dados, para a seleção e avaliação de modelos.

Tendo como referência essas informações, julgue os itens a seguir.

- 64** No código a seguir, *DecisionTreeClassifier* é um classificador que recebe como entrada dois *arrays*: um *array* X, de valores inteiros, contendo os rótulos de classe para as amostras de treinamento; e um *array* Y, esparso ou denso, contendo as amostras de treinamento.

```
>>> from sklearn import tree
>>> X = [[0, 0], [1, 1]]
>>> Y = [0, 1]
>>> clf = tree.DecisionTreeClassifier()
>>> clf = clf.fit(X, Y)
```

- 65** SVC, NuSCV e LinearSVC são classes do *scikit-learn* capazes de realizar classificação binária e multiclasse em um conjunto de dados.

Os hiperparâmetros de um modelo são todos os parâmetros que podem ser definidos antes do início do treinamento, diferentemente dos parâmetros do modelo, que são aprendidos durante o treino do modelo. A busca por hiperparâmetros de determinado algoritmo de aprendizado de máquina que retorne o melhor desempenho medido em um conjunto de validação deu origem ao conceito de otimização de hiperparâmetros.

Acerca dos conceitos de otimização de hiperparâmetros de modelos de aprendizado de máquinas, julgue o item que se segue.

- 66** A otimização bayesiana se utiliza do conceito de probabilidade para encontrar o valor de entrada de uma função que possa retornar o menor valor de saída possível. Nesse método, o número de iterações de pesquisa pode ser reduzido a partir da escolha dos valores de entrada, levando em consideração os resultados anteriores, o que caracteriza um processo iterativo.

Com respeito a análise de componentes principais, mistura de gaussianas e agrupamento *k-means*, julgue os itens que se seguem.

- 67** Se a matriz de variância-covariância referente a três variáveis for

$$\Sigma = \begin{bmatrix} 4 & -0,5 & 0,2 \\ -0,5 & 4 & 0,5 \\ 0,2 & 0,5 & 2 \end{bmatrix},$$

e se o menor autovalor dessa matriz for igual a 1,84, então as duas primeiras componentes principais explicam 81,6% da variação total referente a essas variáveis.

- 68** Considere que, em uma análise de agrupamentos por meio de mistura de gaussianas, três distribuições normais com médias  $\mu_1, \mu_2$  e  $\mu_3$  se referem, respectivamente, aos *clusters* 1, 2 e 3. Nessa situação, pelo modelo de mistura de gaussianas, o *cluster* 1 é constituído pelas observações mais próximas da média  $\mu_1$ .

- 69** Considere a tabela abaixo que mostra as distâncias entre cada observação de um conjunto de dados hipotético e os vetores médios (centroides) do *cluster* correspondente ao final da aplicação do algoritmo de agrupamento *k-means*. Com base nessa tabela, infere-se que o *cluster* 1 é constituído pelas observações 2, 5 e 10.

observação	cluster			
	1	2	3	4
1	3	8	7	12
2	12	7	5	10
3	7	3	8	10
4	4	8	12	8
5	12	5	7	7
6	2	7	6	10
7	7	12	9	1
8	8	10	4	9
9	5	2	6	6
10	9	8	5	2

Julgue os próximos itens, relativos a redes neurais artificiais (RNA).

- 70 Uma RNA é formada por unidades que fazem operações a partir das entradas (sinais) recebidas pelas suas conexões; cada sinal é multiplicado por um peso e, após a soma ponderada dos sinais, caso o nível de atividade atinja o *threshold*, a unidade produz uma determinada resposta de saída.
- 71 Em RNA formada unicamente de *perceptron*, uma pequena alteração nos pesos de um único *perceptron* na rede pode ocasionar grandes mudanças na saída desse *perceptron*; mesmo com a inserção das funções de ativação, não é possível controlar o nível da mudança, por isso, essas redes são voltadas para a resolução de problemas específicos, tais como regressão e previsão de séries temporais.
- 72 As funções de ativação são elementos importantes nas redes neurais artificiais; essas funções introduzem componente não linear nas redes neurais, fazendo que elas possam aprender mais do que relações lineares entre as variáveis dependentes e independentes, tornando-as capazes de modelar também relações não lineares.
- 73 O algoritmo de *backpropagation* consiste das fases de propagação e de retro propagação: na primeira, as entradas são passadas através da rede e as previsões de saída são obtidas; na segunda, se calcula o termo de correção dos pesos e, por conseguinte, a atualização dos pesos.
- 74 Rede neural recorrente é uma arquitetura similar à *feedforward*; a diferença é que a cada nova camada oculta (*hidden layer*) é acrescentada outra camada recorrente à arquitetura conectada à camada anterior, duplicando assim a quantidade de camadas.
- 75 Em RNA, o uso de *early stopping*, ainda que não evite o *overfitting*, permite calcular com mais precisão a classificação nos dados de validação e, assim, melhorar a acurácia do treinamento.
- 76 As redes neurais convolucionais se utilizam de uma arquitetura especial que é adaptada para classificar imagens por meio de algoritmo de aprendizado profundo que pode captar uma imagem de entrada, atribuir importância por meio de pesos e ser capaz de diferenciar um do outro.
- 77 Uma rede neural convolucional é composta por camadas convolucionais, unidades de processamento não linear e camadas de subamostragem (*pooling*); ela possui como característica a habilidade em explorar correlações temporais e espaciais nos dados.
- 78 No código abaixo, escrito em Python, o método `fit()` retorna o valor de perda e os valores de métricas para o modelo, no modo de teste, tendo como referência o número de eras (*epochs*).

```
data_test      =      model.fit(x,      y,
batch_size=64,                               epochs=2,
validation_data=(x_val, y_val))
```

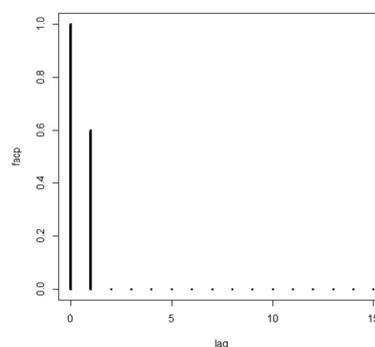
## BLOCO II

Com respeito a *machine learning* aplicado, julgue os próximos itens.

- 79 Classificação de imagens é um método de aprendizado não supervisionado no qual se aplica um modelo de treinamento para o reconhecimento de padrões gráficos presentes em amostras de imagens.
- 80 *Mask RCNN* (*region-based convolutional neural network*) é um método para segmentação de objetos e instâncias que se baseia em detecção, enquanto o método SSAP (*single-shot instance segmentation*) se baseia em *pixels*.
- 81 *Stop-words* constituem um conjunto de palavras que proporcionam pouca informação para o significado de uma frase.
- 82 O CBOW é um modelo de aprendizado de máquina desenhado para prever contexto com base em determinada palavra.
- 83 Suponha que a palavra `amor` ocorra 1.000 vezes no último livro escrito por certo autor, que escreveu, no total, 10 livros. Nesse caso, se a palavra `amor` for encontrada em todos os livros desse autor, então o valor do TF-IDF (*term frequency-inverse document frequency*) referente à palavra `amor` no último livro escrito será igual a 1/1.000.

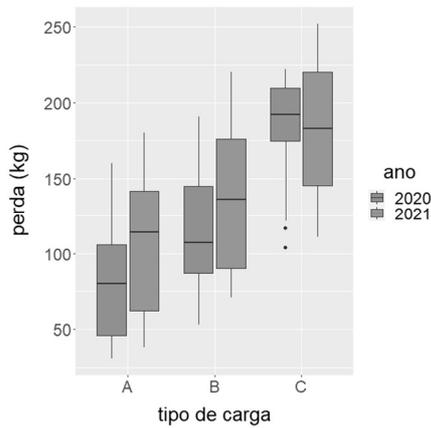
Considerando uma série temporal representada por  $\{X_t\}$ , julgue os itens a seguir.

- 84 Se a figura abaixo apresenta a forma da função de autocorrelação parcial (facp) da série temporal  $\{X_t\}$ , na qual as correlações parciais são nulas nos *lags* iguais ou superiores a 2, então a autocorrelação entre  $X_t$  e  $X_{t-4}$  é igual a zero.



- 85 Se a série temporal for gerada por um processo na forma
- $$X_t = \epsilon_t - X_{t-1},$$
- no qual  $\epsilon_t$  representa um ruído branco com média zero e desvio padrão igual a 1, então a variância de  $X_t$  será igual a 0,5.

Espaço livre



Considerando a figura precedente, que mostra desenhos esquemáticos das distribuições das quantidades de cargas perdidas nos anos de 2020 e 2021, segundo o tipo de carga transportada por uma mineradora, julgue os itens que se seguem.

- 86** No que se refere à distribuição da quantidade de carga do tipo B perdida em 2021, observa-se que o valor da perda mínima foi superior a  $Q_1 - 1,5D_q$ , no qual  $Q_1$  representa o primeiro quartil e  $D_q$  denota o intervalo interquartil da distribuição em tela.
- 87** Na distribuição da quantidade de carga do tipo A perdida em 2020, observa-se que o primeiro quartil foi superior a 100 kg, enquanto o terceiro quartil foi inferior a 50 kg.
- 88** No desenho esquemático referente à distribuição da quantidade de carga do tipo C perdida em 2020, os dois pontos exteriores representam as observações destoantes das demais, que podem ou não podem ser considerados *outliers*.
- 89** Suponha que os valores das quantidades de carga perdida sejam submetidos a uma normalização numérica com base no critério do *Z-score* da forma

$$Z_{a,t} = \frac{X_{a,t} - \mu_{a,t}}{\sigma_{a,t}},$$

em que  $X_{a,t}$  denota a quantidade de carga do tipo  $t$  perdida no ano  $a$ ,  $\mu_{a,t}$  representa a quantidade média de carga do tipo  $t$  perdida no ano  $a$ , e  $\sigma_{a,t}$  refere-se ao desvio padrão da distribuição da quantidade de carga do tipo  $t$  perdida no ano  $a$ . Como resultado dessa normalização, a média da soma

$$Z_{2020,A} + Z_{2020,B} + Z_{2020,C}$$

será igual à média da soma

$$Z_{2021,A} + Z_{2021,B} + Z_{2021,C}.$$

Com respeito a métodos para imputação de dados, julgue os seguintes itens.

- 90** O método de imputação K-NN (*k-nearest neighbours*) leva em consideração os padrões de similaridade presentes no conjunto de dados para prever os valores faltantes. No entanto, a escolha da função de distância para a aplicação desse método, como, por exemplo, HEOM (*heterogeneous euclidean-overlap metric*) ou HVDM (*heterogeneous value difference metric*), pode influenciar significativamente nos resultados da imputação.
- 91** Um dos passos para tratar com dados faltantes é avaliar o tipo de dado perdido; assim, por exemplo, o método MICE (*multivariate imputation by chained equations*) não seria aplicável para dados perdidos do tipo MAR (*missing at random*).

A tabela abaixo apresenta parte de um conjunto de dados referentes a uma variável categórica chamada opinião que possui quatro categorias de resposta: muito satisfeito, satisfeito, insatisfeito e muito insatisfeito.

observação	opinião
1	satisfeito
2	muito satisfeito
3	insatisfeito
4	insatisfeito
5	muito insatisfeito
6	satisfeito

Com base nessas informações, julgue o próximo item.

- 92** Para lidar numericamente com os dados categóricos, uma codificação binária proporciona uma conversão de cada categoria de resposta para uma sequência de dígitos binários, em que cada dígito binário representa uma variável numérica que assume valores 0 ou 1. Na situação em tela, uma possível codificação binária é exemplificada na tabela abaixo.

observação	D1	D2	D3
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	1
5	0	0	0
6	1	0	0

Um usuário deseja importar para o Python, por meio do módulo Pandas, o arquivo no formato csv de nome pesquisa.csv, cujo conteúdo é mostrado a seguir. Esse arquivo se encontra no diretório (ou pasta ou caminho) D: / .

05/06/2008,16:00,2.5788,2.5813,,,
05/06/2008,16:15,2.5768,2.5793,,,
05/06/2008,16:30,2.5744,2.5769,,,
05/06/2008,16:45,2.574,2.5765,,,
05/06/2008,17:00,2.571,2.5735,,,
05/06/2008,17:15,2.5733,2.5758,,,
05/06/2008,17:30,2.5728,2.5753,,,
05/06/2008,17:45,2.5727,2.5752,,,
05/06/2008,18:00,2.5788,2.5813,,,
05/06/2008,18:15,2.5731,2.5756,,,
05/06/2008,18:30,2.5743,2.5768,,,

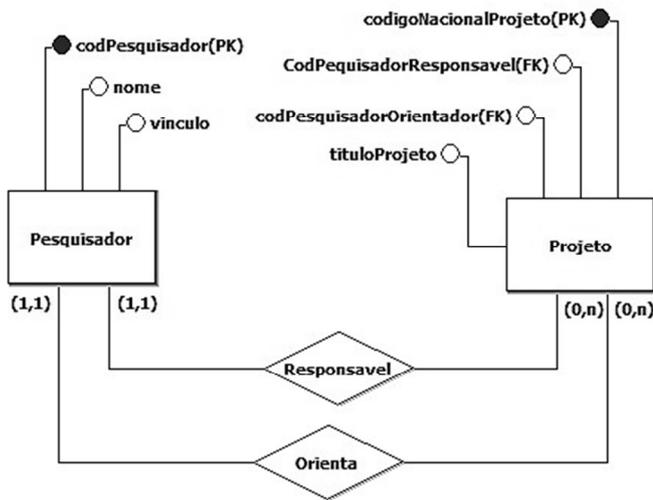
Considerando essas informações, julgue os próximos itens.

- 93** O código abaixo permite importar corretamente o arquivo pesquisa.csv para ser manuseado como um *data frame* no Python.

```
import os
os.chdir('D:/')
import pandas
dados = pandas.read_csv('pesquisa.csv')
```

- 94** O código abaixo permite importar apenas a última linha das duas primeiras colunas do arquivo pesquisa.csv.

```
import os
os.chdir('D:/')
import pandas
dados=pandas.read_csv('pesquisa.csv',
header=None,usecols=[1,2],
skiprows = range(1,9))
```



Com base no modelo entidade-relacionamento (MER) precedente, que apresenta a representação das regras de uma instituição de pesquisa, existe um Pesquisador cadastrado com o nome Pedro. Todos os atributos do MER são do tipo caractere e um dos comandos SQL usados para a construção do modelo é mostrado a seguir.

```
create table Projeto
(codNacionalProjeto char(2),
codPesquisadorResponsavel char(2),
codPesquisadorOrientador char(2),
tituloProjeto char(50),
primary key(codNacionalProjeto));
```

A partir das informações constantes no modelo e dos dados sobre o conteúdo dos atributos, julgue os itens subsequentes.

**95** Após a criação da tabela Projeto, a criação das chaves estrangeiras (FK) do relacionamento Orienta pode ser feita corretamente conforme a seguir.

```
ALTER TABLE Projeto ADD CONSTRAINT
id_fk_responsavel FOREIGN
KEY(codPesquisadorResponsavel)
REFERENCES Pesquisador(codPesquisador);
```

**96** Projeto é uma entidade fraca em relação à entidade Pesquisador, considerando o relacionamento identificado como Responsavel e os atributos do MER.

**97** Por meio do comando SQL a seguir, é possível recuperar o nome dos pesquisadores responsáveis por projetos, seguido pelo nome de seu orientador, mas apenas os projetos orientados por Pedro.

```
select responsavel.nome nomeresponsavel,
orientador.nome nomeorientador,
tituloProjeto
from Pesquisador responsavel,
Pesquisador orientador, Projeto
where orientador.nome = 'Pedro'
andcodPesquisadorResponsavel = codPesqui
sador
and codPesquisadorOrientador =
codPesquisador;
```

Com referência aos conceitos de banco de dados e *data warehouse*, julgue os itens seguintes.

**98** Em sistemas NoSQL baseados em armazenamento de chave-valor, a chave é multidimensional e composta pela combinação do nome de tabela com a chave linha-coluna e com o rótulo de data e hora.

**99** O *Hadoop Distributed File System* (HDFS) é construído usando a linguagem Java, o que permite que sua arquitetura mestre/escravo seja implementada em uma ampla variedade de máquinas.

### BLOCO III

O volume instantâneo de petróleo, em metros cúbicos por hora ( $m^3/h$ ), produzido por um poço no instante  $t$  horas, dentro das primeiras 24 horas de operação, é dado pela função  $f(t) = \alpha \cdot (300 - 12t^2 + t^3)$ , com  $0 \leq t \leq 24$ , em que  $0 < \alpha < 1$  é uma constante positiva e  $t = 0$  corresponde ao instante inicial em que o poço iniciou a sua produção.

Com base nessas informações hipotéticas, julgue os próximos itens.

**100** O volume instantâneo que o poço produz está aumentando no instante  $t = 10$  horas.

**101** Nas primeiras dez horas de produção, foram extraídos desse poço mais de 1.500 metros cúbicos de petróleo.

**102** O volume instantâneo mínimo ocorre após  $t = 6$  horas.

No que concerne à teoria de funções deriváveis de duas variáveis reais, julgue o item seguinte.

**103** Se  $P(x,y) = x^3/y^2$  representa o consumo de gasolina em uma cidade em função do preço  $y$  e da quantidade  $x$  de veículos da cidade, a taxa instantânea de variação do consumo quando  $y$  varia e  $x$  é mantido fixo é igual a  $x^3/2y$ .

Considere três campos de exploração de petróleo, A, B e C, e  $v = (n_A, n_B, n_C)$  como um vetor em que as 3 coordenadas representam, respectivamente, as quantidades de poços de extração nesses campos. Suponha, ainda, que os vetores na tabela a seguir representem as respectivas médias diárias de barris extraídos por poço em cada um dos campos nos 3 primeiros meses de 2021.

janeiro/2021	$u_1 = (4, 5, 4)$
fevereiro/2021	$u_2 = (5, 3, 2)$
março/2021	$u_3 = (3, 6, 4)$

Com relação a essa situação hipotética, julgue os itens a seguir.

**104** Se o produto escalar entre  $v$  e  $u_1$  é igual a 100, então os três campos, juntos, possuem menos de 25 poços para extração de petróleo.

**105** O produto vetorial entre os vetores  $u_1 - u_3$  e  $u_2$  é um vetor perpendicular ao vetor  $w = (2, 2, 1)$ .

**106** Se  $e_1 = (1, 0, 0)$ ,  $e_2 = (0, 1, 0)$ ,  $e_3 = (0, 0, 1)$  e

$T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  é transformação linear que satisfaz

$$T(e_1) = u_1, T(e_2) = u_2, T(e_3) = u_3,$$

então  $T(u_1) = u_2$ .

**107** Se  $\|\cdot\|_1$  e  $\|\cdot\|_2$  representarem, respectivamente, as normas  $L_1$  e  $L_2$  em  $\mathbb{R}^3$ , então  $2\|u_1 - u_2\|_2 < \|u_2 - u_1\|_1$ .

Considerando  $np$  a biblioteca *numpy* do Python, julgue os itens a seguir.

108 Se  $A = np.array([[6],[2],[-1],[0],[-5],[6],[7],[9],[2]])$ ,

$A_1 = A[3:7,0]$ , e  $A_2 = A_1.reshape(2,2)$ , então  $np.linalg.det(A_2) > 20$ .

109 Se  $a_1 = np.linspace(1,9,5)$ ,  $a_2 = np.linspace(3,7,5)$  e

$b = np.concatenate((a_2, a_1))$ , então  $b[-4] + b[1] > 10$ .

Considerando que uma amostra aleatória simples  $X_0, X_1, \dots, X_n$  seja retirada de uma distribuição com média  $\mu$  e variância  $\sigma^2$ , com respeito à soma ponderada

$$S_n = \sum_{k=0}^n \phi^k X_k,$$

na qual  $|\phi| < 1$ , julgue os itens que se seguem.

110 Se a soma for representada na forma  $S_n = S_{n-1} + \phi^n X_n$ , em que  $n \geq 1$ , então a correlação de Pearson entre  $S_{n-1}$  e  $X_n$  será igual a  $\phi^n$ .

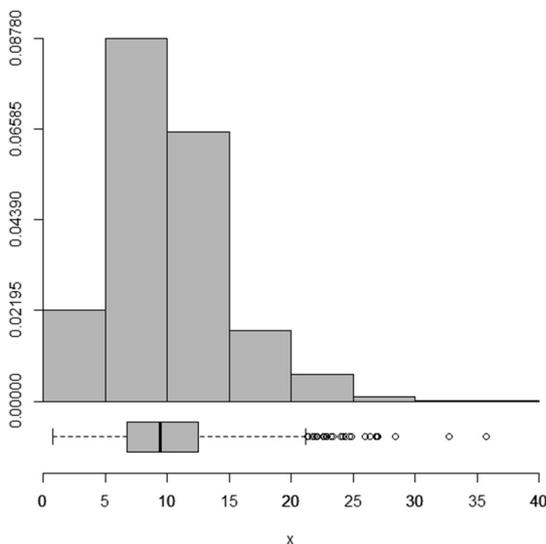
111 A variância de  $S_n$  é inferior a  $n\sigma^2$ .

112 Com base no teorema do limite central, é correto concluir que a variável padronizada

$$\frac{S_n - E[S_n]}{\sqrt{\text{Var}[S_n]}}$$

converge em distribuição para uma distribuição normal padrão quando  $n \rightarrow +\infty$ .

A figura seguinte mostra o histograma como uma estimativa da função de densidade de uma distribuição  $X$ , juntamente com o diagrama *boxplot* correspondente a esse conjunto de dados.



tamanho da amostra	1.000
média amostral	10
desvio padrão amostral	4,7

Considerando a figura e as informações apresentadas no quadro, julgue os itens que se seguem.

113 O primeiro decil da distribuição do conjunto de dados em tela é igual ou inferior a 5.

114 O percentual das observações tais que  $[0,6 \leq x \leq 19,4]$  é igual ou inferior a 70%.

115 A diferença entre a média amostral e a mediana amostral é superior a 0.

Considere uma variável aleatória  $Y_n$ , com média zero e variância 1, e uma função real  $g$ , tal que o valor esperado de  $g(Y_n)$  possa ser escrito como

$$E[g(Y_n)] = g(0) + \frac{g'(0)}{2} + \mathcal{O}(n^{-3/2}),$$

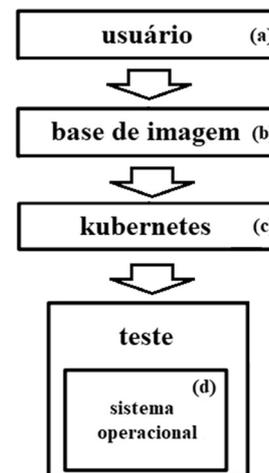
em que  $g'(0)$  representa o valor da primeira derivada da função  $g$  no ponto zero, e  $n \in \{1,2,3, \dots\}$ . Com relação à notação assintótica *big O*, julgue o próximo item.

116  $\mathcal{O}(n^{-3/2})$  significa que existe uma constante real  $c$  tal que  $n^{3/2} \mathcal{O}(n^{-3/2}) < c$  para todo  $n \in \{1,2,3, \dots\}$ .

Um aplicativo para edição de textos foi disponibilizado para seus usuários sob a forma de computação em nuvem. Esses usuários podem se conectar virtualmente, colaborando mutuamente para a elaboração de documentos. Tais documentos, bem como os respectivos históricos de versões anteriores, são armazenados na nuvem.

Com referência a essa situação hipotética, julgue o próximo item.

117 A situação descreve um modelo de nuvem denominado PaaS (*platform as a service*).



Considerando que a figura anterior mostra o *layout* de um *container*, julgue os próximos itens.

118 O método CI/CD refere-se a um processo de automação para os usuários de um sistema que enviam, de forma contínua, *feedbacks* para os desenvolvedores desse sistema.

119 Na figura em tela, o *container* se encontra no retângulo (d).

120 Na situação em tela, a cultura DevOps permite simplificar a implantação e o gerenciamento eficiente de *containers*.

Espaço livre