

- Cada um dos itens das provas objetivas está vinculado ao comando que imediatamente o antecede. De acordo com o comando a que cada um deles esteja vinculado, marque, na **Folha de Respostas**, para cada item: o campo designado com o código **C**, caso julgue o item **CERTO**; ou o campo designado com o código **E**, caso julgue o item **ERRADO**. A ausência de marcação ou a marcação de ambos os campos não serão apenadas, ou seja, não receberão pontuação negativa. Para as devidas marcações, use a **Folha de Respostas**, único documento válido para a correção das suas respostas.
- Nos itens que avaliem **conhecimentos de informática e(ou) tecnologia da informação**, a menos que seja explicitamente informado o contrário, considere que todos os programas mencionados estão em configuração padrão e que não há restrições de proteção, de funcionamento e de uso em relação aos programas, arquivos, diretórios, recursos e equipamentos mencionados.
- Eventuais espaços livres — identificados ou não pela expressão “**Espaço livre**” — que constarem deste caderno de provas poderão ser utilizados para rascunho.

## PROVAS OBJETIVAS

### -- CONHECIMENTOS BÁSICOS --

O texto mais célebre de **A República** é sem dúvida a **Alegoria da Caverna**, em que Platão, utilizando-se de linguagem alegórica, discute o processo pelo qual o ser humano pode passar da visão habitual que tem das coisas, “a visão das sombras”, unidirecional, condicionada pelos hábitos e preconceitos que adquire ao longo de sua vida, até a visão do Sol, que representa a possibilidade de alcançar o conhecimento da realidade em seu sentido mais elevado e compreendê-la em sua totalidade. A visão do Sol representa não só o alcance da Verdade e, portanto, do conhecimento em sua acepção mais completa, já que o Sol é “a causa de tudo”, mas também, como diz Sócrates na conclusão dessa passagem: “Nos últimos limites do mundo inteligível, aparece-me a ideia do Bem, que se percebe com dificuldade, mas que não se pode ver sem se concluir que ela é a causa de tudo o que há de reto e de belo. Acrescento que é preciso vê-la se se quer comportar-se com sabedoria, seja na vida privada, seja na vida pública.”.

De acordo com este texto, a possibilidade de um indivíduo tornar-se justo e virtuoso depende de um processo de transformação pelo qual deve passar. Assim, afasta-se das aparências, rompe com as cadeias de preconceitos e condicionamentos e adquire o verdadeiro conhecimento. Tal processo culmina com a visão da forma do Bem, representada pela matéria do Sol. O sábio é aquele que atinge essa percepção. Para Platão, conhecer o Bem significa tornar-se virtuoso. Aquele que conhece a justiça não pode deixar de agir de modo justo.

Daniilo Marcondes. **Textos básicos de ética**: de Platão a Foucault. 1ª ed. Rio de Janeiro: Jahar, 2007, p. 31 (com adaptações).

Em relação às ideias, aos sentidos e aspectos linguísticos do texto precedente, julgue os itens subsecutivos.

- 1 Dado o fato de o vocábulo “mas”, em sua primeira ocorrência no segundo período do primeiro parágrafo, ter sido empregado com sentidos adversativo, de oposição, os sentidos originais do texto seriam mantidos caso ele fosse substituído por **porém**.
- 2 Depreende-se do texto que as visões de Sócrates e Platão convergem ao relacionar a justiça ao conhecimento do Bem.
- 3 O pronome “este”, na expressão “De acordo com este texto”, que inicia o segundo parágrafo, remete a toda a ideia contida no parágrafo anterior.
- 4 O termo “Alegoria”, empregado no texto precedente, alude à ideia de **representação**.
- 5 Considerando-se as ideias de Platão e Sócrates apresentadas no texto, é possível concluir, se se acredita que pessoas que operam sua vivência unicamente a partir de redes sociais vivem de aparências, que tais pessoas não conhecem a Verdade.
- 6 Seria mantida a correção gramatical do texto caso o trecho “pelo qual” (primeiro período do primeiro parágrafo) fosse substituído por **porque**.

- 7 A supressão da vírgula empregada logo após o vocábulo “Assim”, que inicia o segundo período do segundo parágrafo, manteria a correção gramatical, embora alterasse o sentido original do texto.
- 8 No último período do texto, o termo “Aquele” exerce a função sintática de sujeito da oração “que conhece a justiça”.

Muito tem sido escrito e debatido sobre a afirmativa de que a “Internet é terra de ninguém”. Tal afirmativa não é de hoje, mas ainda alimenta uma sensação de impunidade ou de falsa responsabilidade do que é postado ou compartilhado na Internet e pelas redes sociais. A expressão *fakes news*, em particular, representa um estrangeirismo que mascara diversos crimes cometidos contra a honra, como injúria, calúnia e difamação. Sob um olhar semântico, dizer “compartilhei *fake news* de alguém” não carrega qualquer sentimento de culpa, ou se carrega, ela é mínima. Agora, dizer “cometi um crime contra honra” já traz outras implicações, não só de ordem jurídica, mas também de grande responsabilidade pessoal.

Marcelo Hugo da Rocha e Fernando Elias José. **Cancelado**: a cultura do cancelamento e o prejulgamento nas redes sociais. Belo Horizonte, MG: Letramento, 2021, p. 36 (com adaptações).

No que se refere às ideias, aos sentidos e às construções linguísticas do texto precedente, assim como a sua tipologia, julgue os itens a seguir.

- 9 No texto, predomina o tipo textual dissertativo-argumentativo.
- 10 No texto, as frases ‘compartilhei *fake news* de alguém’ e ‘cometi um crime contra honra’ são consideradas equivalentes, já que o compartilhamento de *fake news* pode ser crime.
- 11 A palavra “Agora”, no início do último período, introduz no texto um marco temporal, correspondendo, em sentido, a **Atualmente**.
- 12 Infere-se do emprego do trecho “um estrangeirismo que mascara diversos crimes” (terceiro período) que há outros estrangeirismos que não mascaram crimes ou a eles se associam.
- 13 Sem prejuízo dos sentidos originais e da correção gramatical do texto, o trecho “falsa responsabilidade do que é postado ou compartilhado” (segundo período) poderia ser substituído por **falsa responsabilidade do que se posta ou se compartilha**.
- 14 No início do texto, a forma verbal “escrito” poderia ser corretamente substituída por **escrito**.
- 15 Feitas as devidas alterações de maiúsculas e minúsculas, o ponto final empregado logo após “ela é mínima” (penúltimo período do texto) poderia ser corretamente substituído por ponto e vírgula.

A PETROBRAS responde por cerca de 80% dos combustíveis ofertados no Brasil. Para isso, muito foi investido em infraestrutura, com operações que consomem quase 100 bilhões de reais ao ano, conforme dados de 2021.

O caminho do petróleo do poço até virar combustível no carro das pessoas é longo e complexo. Começa na procura: acertar onde furar e encontrar petróleo exige conhecimento técnico de geólogos e geofísicos e bastante investimento. E, mesmo com um time de *experts* do mais alto nível, achar petróleo não é certo.

Transportar o petróleo do mar até as refinarias é também uma tarefa complexa, para a qual são utilizados dutos e navios. Em terra, ele é tratado em refinarias, que separam desse óleo as frações de gasolina, *diesel* e gás de cozinha, entre outros derivados. Os produtos são então disponibilizados às diversas distribuidoras que hoje atendem o mercado brasileiro, responsáveis por fazer chegar cada um deles aos consumidores finais.

Internet: <duvidasgasolina.petrobras.com.br> (com adaptações).

Considerando as ideias, os sentidos e aspectos linguísticos do texto precedente, julgue os itens subsequentes.

- 16 Infere-se do texto que achar petróleo não é uma certeza porque os erros na identificação do local de perfuração impedem o reconhecimento da localização exata do produto.
- 17 A informação apresentada no primeiro período do segundo parágrafo é desenvolvida no restante do segundo parágrafo e no terceiro parágrafo.
- 18 Se o trecho “A PETROBRAS responde por cerca de 80% dos combustíveis ofertados no Brasil” fosse reescrito como **A PETROBRAS está à frente de aproximadamente 80% dos combustíveis que produz-se no Brasil**, seriam mantidos os sentidos e a correção gramatical do texto.
- 19 No terceiro parágrafo, o trecho “que separam desse óleo as frações de gasolina, *diesel* e gás de cozinha, entre outros derivados” consiste em uma oração adjetiva restritiva, na medida em que delimita o tipo específico de refinarias a que se refere o texto.
- 20 No segundo parágrafo, as orações “onde furar” e “achar petróleo” desempenham a função sintática de sujeito nos períodos em que ocorrem.

Alguns linguistas acreditam que o *Homo erectus*, há mais ou menos 1 milhão e meio de anos, já tinha uma linguagem. Os argumentos que eles dão são que o *Homo erectus* tinha um cérebro relativamente grande e usava ferramentas de pedra primitivas, porém bastante padronizadas. Essa hipótese pode ser verdadeira, mas pode também estar bem longe do correto.

O uso de ferramentas certamente não requer linguagem. Chimpanzés usam galhos como ferramentas para caçar cupins, ou pedras para quebrar nozes. Obviamente, mesmo as ferramentas mais primitivas do *Homo erectus* (pedras lascadas) são muito mais sofisticadas que qualquer coisa usada por chimpanzés, mas ainda assim não há uma razão convincente para crer que essas pedras não pudessem ter sido produzidas sem linguagem.

O tamanho do cérebro é igualmente problemático como indicador da presença de linguagem, porque ninguém tem uma boa ideia de quanto cérebro exatamente é necessário para a linguagem. Além disso, a capacidade para a linguagem pode ter permanecido latente no cérebro por milhões de anos, sem ter sido de fato colocada em uso.

Guy Deutscher. **O desenrolar da linguagem**. Renato Basso e Guilherme Henrique May (Trad.). Campinas: Mercado de Letras, 2014, p. 28-29 (com adaptações).

A respeito das ideias, dos sentidos e aspectos linguísticos do texto precedente, julgue os itens que se seguem.

- 21 O autor do texto contesta os argumentos usados por alguns linguistas que defendem a ideia de que o *Homo erectus* apresentava linguagem.

- 22 O fato de que alguns animais, como chimpanzés, também utilizam ferramentas enfraquece o argumento de que se requer linguagem para usar ferramentas.
- 23 A correção gramatical do texto seria mantida caso o adjetivo “primitivas”, no trecho “ferramentas de pedra primitivas”, fosse flexionado no singular, embora o sentido original do trecho e as relações sintáticas nele estabelecidas fossem alterados: no original, o adjetivo qualifica o termo “ferramentas”; com o emprego do singular, o adjetivo qualificaria o termo “pedra”.
- 24 O uso do advérbio “Obviamente” (segundo parágrafo) desempenha importante papel na argumentação apresentada no texto, realçando uma informação que já é tomada como conhecimento geral.
- 25 Caso fosse suprimida a vírgula empregada logo antes da preposição “sem” (terceiro parágrafo), haveria prejuízo para a correção gramatical do texto, embora seu sentido original fosse mantido.
- 26 A expressão “Além disso” (terceiro parágrafo) introduz o argumento mais forte apresentado pelo autor do texto para comprovar sua tese acerca do surgimento da linguagem humana.

As tecnologias de contar e escrever histórias não seguiram um caminho linear. A própria escrita foi inventada pelo menos duas vezes, primeiro na Mesopotâmia e depois nas Américas. Os sacerdotes indianos se recusavam a escrever as histórias sagradas por medo de perder o controle sobre elas. Professores carismáticos (como Sócrates) se recusaram a escrever. Algumas invenções posteriores foram adotadas somente de forma seletiva, como quando os eruditos árabes usaram o papel chinês, mas não demonstraram nenhum interesse por outra invenção chinesa, a impressão. As invenções relacionadas à escrita tinham muitas vezes efeitos colaterais inesperados. Preservar textos antigos significava manter vivas artificialmente as línguas. Desde então, passou-se a estudar línguas mortas e alguns textos acabaram sendo declarados sagrados.

Martin Puchner. **O mundo da escrita: como a literatura transformou a civilização**. Pedro Maia Soares (Trad.). São Paulo: Companhia das Letras, 2019, p. 18 (com adaptações).

Julgue os itens seguintes, relativos à tipologia, aos sentidos e aspectos linguísticos do texto precedente.

- 27 O texto é predominantemente informativo.
- 28 Estaria mantida a correção gramatical do trecho “Os sacerdotes indianos se recusavam a escrever as histórias sagradas por medo de perder o controle sobre elas. Professores carismáticos (como Sócrates) se recusaram a escrever”, caso a posição do pronome “se”, em suas duas ocorrências, fosse alterada de proclítica — como está no texto — para enclítica.
- 29 O emprego predominante do pretérito perfeito no texto tem o propósito de apresentar fatos já ocorridos em determinado momento no passado e cujos efeitos, além de ainda serem sentidos no momento atual, afetam o tempo presente.
- 30 No trecho “As invenções relacionadas à escrita tinham muitas vezes efeitos colaterais inesperados”, o emprego do sinal indicativo de crase justifica-se pela fusão de preposição e artigo feminino em uma locução adverbial de modo.

The worn wooden floorboards squeak under Rafael Molina's heavy steps as he paces the saloon. Outside, the sound of galloping horses breaks the silence of the surrounding desert. All around him, the Old West town's empty shops and abandoned houses look as if they have just been ransacked by cowboy bandits.

"When I was a kid, I could only dream about all this," says the 68-year-old former actor and stuntman. "My aspiration was to see a film set firsthand. Today I own one of the most famous ones in the history of Western movies."

But this busy movie site is located in Spain — not Montana or Texas. It's one of three faux Old Western towns in the small village of Tabernas and the surrounding desert of the Almería province. Since the late 1950s, these rugged mountains, arid plains, and dry canyons have provided the backdrops for more than 170 movie Westerns, including **The Good, The Bad and The Ugly** (1966) and **Once Upon a Time in The West** (1968).

Molina belongs to a small community of local cowboy actors and stuntmen in Tabernas who have played a role in movies and TV shows since the first productions in the 1950s. They can perform anything from fistfights to horse drags. Knowledge and skills often pass from father to son, keeping tricks of the trade in the family. Steeped in the golden era of Westerns, these actors embody the values of their movie heroes: pride, bravado, freedom, and a trusting relationship with horses.

"I've always liked horses and the [U.S.] West," says 29-year-old Ricardo Cruz Fernández, a stuntman and cowboy who appeared in recent productions including **Game of Thrones**. Fernández started his career as a cowboy after completing a stuntman course a decade ago.

Between productions, he performs daily shows at Fort Bravo for thousands of tourists who visit the set each year. In one show, Fernández portrays a bank robber who absconds with some gold. Visitors encounter him in the saloon, fist and (fake) gun fighting with actors playing his double-crossing accomplices. The town also offers cancan dance shows and set tours by horse-drawn wagon.

"I prefer to play the bad guy, because it gives me a wider range of possibilities," says Fernández. "The good guy only has to keep things in order."

Almería has hosted more than 500 productions, including blockbuster films (**Patton**, **Terminator: Dark Fate**) and TV shows (**Doctor Who**). "Our landscapes are very convenient. We have sea, desert, and snowy mountains all within a short distance," says local producer Plácido Martínez. "We can serve as Texas, New Mexico, Arizona, California, and endless other natural settings."

Matteo Fagotto. **The Wild West lives on in southern Spain**.

In: **National Geographic**. Internet: <www.nationalgeographic.com> (adapted).

Concerning the previous text and its linguistic aspects, judge the following items.

- 31 There are three fake Old West towns in the province of Almería, in Spain.
- 32 In the fifth paragraph, the word "stuntman" means a man who performs a dangerous action which needs to be done by someone skilled, especially instead of an actor in a film or television programme.
- 33 According to the text, the actors of the small community in Tabernas avoid being like their movie heroes.
- 34 In the fragment 'The good guy only has to keep things in order', the word 'has' could be correctly replaced with **must**, without changing the meaning of the text.
- 35 According to the text, the Almería's actors have already played in Texas, New Mexico, Arizona, California, and endless other natural settings.

- 36 In the sentence "In one show, Fernández portrays a bank robber who absconds with some gold", the fragment "absconds with" could be correctly replaced with **keeps**, without changing the meaning of the text.
- 37 The fragment "Fernández started his career as a cowboy after completing a stuntman course a decade ago" can be correctly rewritten as **A decade ago, Fernández started his career as a cowboy and then he completed a stuntman course**, without changing the meaning of the text.
- 38 It can be inferred from the text that many different productions were made in Almería because its landscapes are very convenient and diversified.
- 39 Rafael Molina owns one of the most famous film sets in the history of Western movies.
- 40 According to the text, in Spain there are villages that were very important cities during the time of the American Wild West.

### Espaço livre

In a world where many of us are glued to our smartphones, Dulcie Cowling is something of an anomaly — she has ditched hers. The 36-year-old decided at the end of last year that getting rid of her handset would improve her mental health. So, over Christmas she told her family and friends that she was switching to an old Nokia phone that could only make and receive calls and text messages.

She recalls that one of the pivotal moments that led to her decision was a day at the park with her two boys, aged six and three: “I was on my mobile at a playground with the kids and I looked up and every single parent — there was up to 20 — were looking at their phones, just scrolling away,” she says.

“I thought ‘when did this happen?’. Everyone is missing out on real life. I don’t think you get to your death bed and think you should have spent more time on Twitter, or reading articles online.”

Ms Cowling, who is a creative director at London-based advertising agency Hell Yeah!, adds that the idea to abandon her smartphone had built up during the covid-19 lockdowns.

“I thought about how much of my life is spent looking at the phone and what else could I do. Being constantly connected to lots of services creates a lot of distractions, and is a lot for the brain to process.”

She plans to use the time gained from quitting her smartphone to read and sleep more.

About nine out of 10 people in the UK now own a smartphone, a figure broadly replicated across the developed world. And we are glued to them — one recent study found that the average person spends 4.8 hours a day on their handset.

Yet for a small, but growing number of people, enough is enough.

Alex Dunedin binned his smartphone two years ago. “Culturally we have become addicted to these tools,” says the educational researcher and technology expert. “They are blunting cognition and impeding productivity.”

He has become happier and more productive since he stopped using a smartphone, he says.

Mr Dunedin doesn’t even have an old-fashioned mobile phone or even a landline anymore. He is instead only electronically contactable via emails to his home computer.

“It has improved my life,” he says. “My thoughts are freed up from constantly being cognitively connected to a machine that I need to feed with energy and money. I think that the danger of technologies is that they are emptying our lives.”

Yet, while some worry about how much time they spend on their handset, for millions of others they are a godsend.

“More than ever, access to healthcare, education, social services and often to our friends and family is digital, and the smartphone is an essential lifeline for people,” says a spokesperson for UK mobile network Vodafone.

“We also create resources to help people get the most from their tech, as well as to stay safe when they’re online — that’s hugely important.”

Suzanne Beame. **The people deciding to ditch their smartphones.**  
Internet: <www.bbc.com> (adapted).

Considering the previous text, judge the following items.

- 41 According to the text, handsets are essential for people because sometimes healthcare, education and social services are offered only through smartphones.
- 42 In the sentence (thirteenth paragraph) “Yet, while some worry about how much time they spend on their handset, for millions of others they are a godsend.”, the word “Yet” is synonymous with **However**.
- 43 The only way Alex Dunedin connects electronically is via emails on his home computer.
- 44 People who are giving up on their mobile devices believe they were spending too much time with being connected and they were missing their real lives because of that.

45 It was when Dulcie Cowling was in the park with her two kids that she took her decision to ditch her smartphone and then she told her family and friends about it after Christmas.

46 In the sentence ‘They are blunting cognition and impeding productivity’ (ninth paragraph), the pronoun ‘They’ refers to the “nine out of 10 people in the UK who own a smartphone” (seventh paragraph).

47 Although there is a movement of people ditching their smartphones in order to have what they think is a better life quality, millions believe digital technology is essential to everyone’s lives.

Mars is the fourth planet from the Sun — a dusty, cold, desert world with a very thin atmosphere. Mars is also a dynamic planet with seasons, polar ice caps, canyons, extinct volcanoes, and evidence that it was even more active in the past.

No other planet has captured our collective imagination quite like Mars.

In the late 1800s when people first observed the canal-like features on Mars’ surface, many speculated that an intelligent alien species resided there. This led to numerous stories about Martians, some of whom invade Earth, like in the 1938 radio drama, **The War of the Worlds**. According to an enduring urban legend, many listeners believed the story to be real news coverage of an invasion, causing widespread panic.

Countless stories since have taken place on Mars or explored the possibilities of its Martian inhabitants. Movies like **Total Recall** (1990 and 2012) take us to a terraformed Mars and a struggling colony running out of air. A Martian colony and Earth have a prickly relationship in **The Expanse** television series and novels.

Internet: <www.solarsystem.nasa.gov> (adapted).

Judge the following items, based on the previous text.

- 48 In the last sentence of the last paragraph, the word “prickly” means **peaceful**.
- 49 **The War of the Worlds** was a radio drama that told the real story of an invasion from Mars, panicking countless people.
- 50 According to the text, the speculations about extraterrestrial life started in the late 1800s due to canal-like features observed on Mars.

## -- 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.

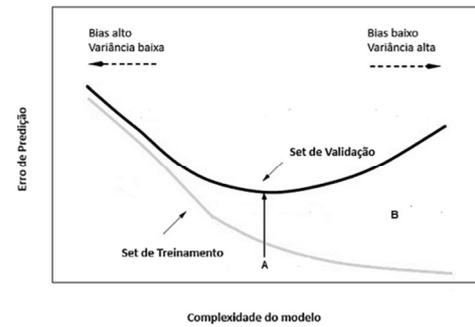
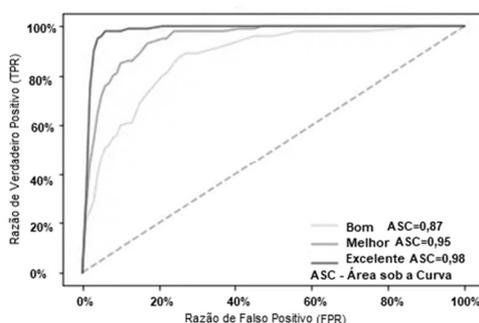
As métricas de avaliação de desempenho de um modelo de aprendizado de máquina, que é um componente integrante de qualquer projeto de ciência de dados, destinam-se a estimar a precisão da generalização de um modelo sobre os dados futuros (não vistos ou fora da amostra). Dentre as métricas mais conhecidas, 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))
```

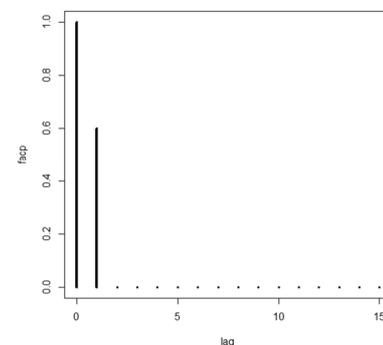
## BLOCO II

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.

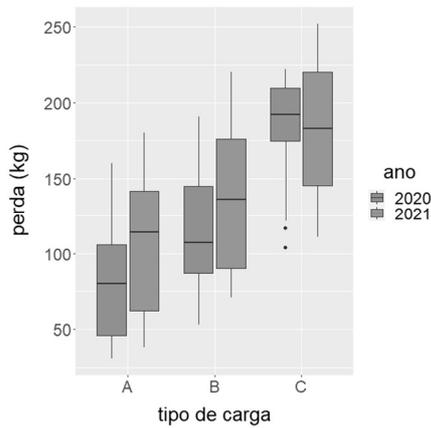
Considerando uma série temporal representada por  $\{X_t\}$ , julgue os itens a seguir.

- 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,,,

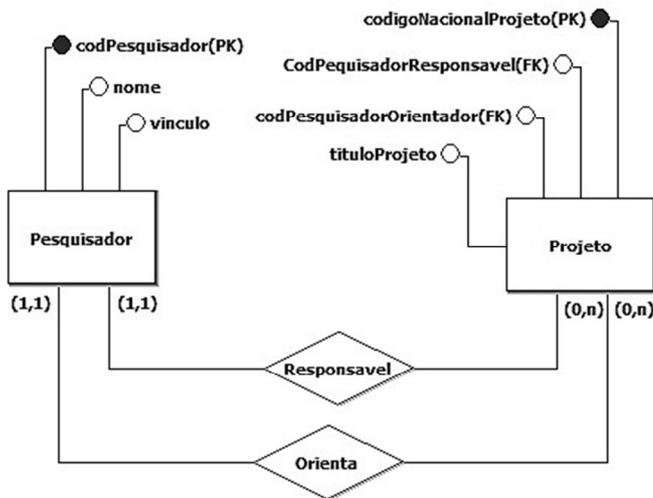
Considerando essas informações, julgue os próximos itens.

- 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 subsecutivos.

**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  $\phi^n$ .

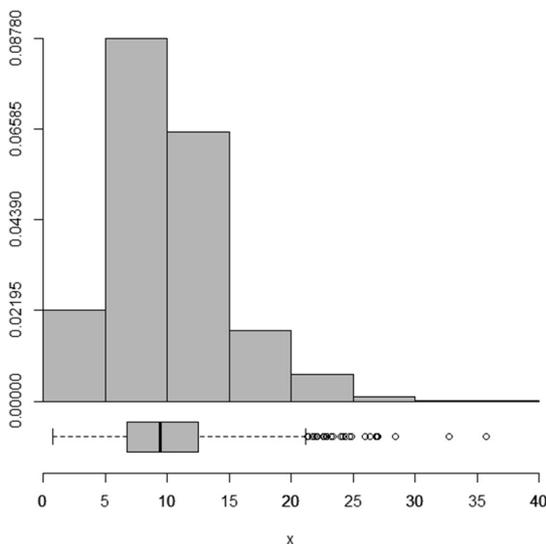
111 A variância de  $S_n$  é inferior a  $n\sigma^2$ .

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

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

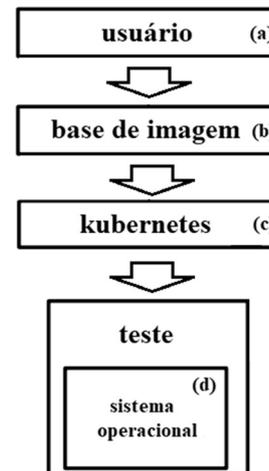
em que  $g'(0)$  representa o valor da primeira derivada da função  $g$  no ponto zero, e  $n \in \{1, 2, 3, \dots\}$ . Com relação à notação assintótica *big O*, julgue o próximo item.

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