

**MINISTÉRIO PÚBLICO DO DISTRITO FEDERAL E TERRITÓRIOS
EDITAL N.º 01/2024**

NÍVEL SUPERIOR - ESTATÍSTICA

CADERNO DE PROVA OBJETIVA

ORIENTAÇÕES GERAIS SOBRE A PROVA OBJETIVA

1. A prova objetiva será composta de 20 (VINTE) questões de múltipla escolha, extraídas do conteúdo programático publicado no edital complementar de abertura 001/2024, sendo que, para cada uma das questões objetivas, são apresentadas 5 (CINCO) alternativas classificadas com as letras (A, B, C, D e E), mas somente uma responde adequadamente à questão proposta.
 2. Ao ser autorizado o início da prova, verifique, no Caderno de Questões, se a numeração das questões e a paginação estão corretas e se não há falhas, manchas ou borrões. Se algum desses problemas for detectado, solicite ao fiscal outro caderno completo. Não serão aceitas reclamações posteriores.
 3. Iniciada a Prova, nenhum candidato poderá retirar-se da sala antes de decorrido 2h (duas horas) de prova, devendo, ao sair, entregar ao fiscal de sala, obrigatoriamente, Cartão Respostas da Prova Objetiva.
 4. O candidato poderá levar o Caderno de Questões decorrido o prazo de duas horas do início da prova.
 5. podendo o candidato levar consigo o Caderno de Questões.
 6. A prova objetiva terá início às 14 da tarde, com duração de 3 (três) horas, e término impreterivelmente às 17 horas.
 7. O cartão de resposta será entregue após o início da prova e deverá ser conferido e assinado à vista dos fiscais.
 8. O preenchimento do cartão de resposta deverá ser feito com caneta esferográfica de cor azul ou preta, fabricada em material transparente.
 9. Será desclassificado o candidato que, durante a realização da prova escrita, for surpreendido portando, em local diverso do indicado pelos fiscais, equipamento eletrônico e/ou material de uso não autorizado, ainda que desligado.
 10. De igual forma, será desclassificado o candidato cujo equipamento eletrônico e/ou material de uso não autorizado que estiver em local indicado emitir qualquer tipo de ruído, alerta ou vibração.
 11. O cartão de resposta deverá ser entregue pelo candidato ao fiscal de sala no momento da saída do candidato.
 12. O candidato que entregar o cartão de resposta não poderá retornar ao recinto.
 13. Em hipótese alguma haverá substituição do cartão de resposta por erro do candidato.
 14. O candidato não poderá levar seu caderno de prova objetiva.
 15. O candidato não deverá amassar, molhar, dobrar, rasgar, manchar ou, de qualquer modo, danificar o seu caderno de respostas, sob pena de arcar com os prejuízos advindos da impossibilidade de realização da leitura.
- Os 3 (três) últimos candidatos deverão permanecer em sala, sendo liberados somente quando todos tiverem concluído a prova e terem seus nomes registrados em ata e com suas respectivas assinatura.

LÍNGUA PORTUGUESA

Leia o texto a seguir para responder às questões de 01 a 08.

ENTRE O MITO E O HOMEM

Jornalista e cronista Paulo Pestana

O escritor britânico George Orwell sabia muito sobre animais, mas muito mais sobre os homens. Usou de fábula para falar de igualdade social e liberdade em *A Revolução dos Bichos*, de ficção distópica para mostrar o autoritarismo oficial em *1984* e de vivência pessoal para criar *Na Pior em Paris e Londres*, ainda hoje uma das mais radicais experiências literárias.

São relatos feitos entre 150 e 400 páginas, mas Orwell, sob o pseudônimo de Eric Arthur Blair, reuniu todo o seu sarcasmo e sabedoria numa simples frase de um artigo: “a história é escrita pelos vencedores”. Faltou dizer que os vencedores de hoje podem ser os derrotados de amanhã e, daí, a história vai sendo reescrita.

Um exemplo é a história do Brasil, por séculos contada a partir dos relatos oficiais de Francisco Adolfo de Varnhagen, o visconde de Porto Seguro, que em 1854 publicou o primeiro dos dois volumes de *História Geral do Brasil*. São dois calhamaços de mais de 500 páginas cada, em que o esforço do autor para se manter imparcial sucumbe a cada página.

Calcada em documentos oficiais, é uma história contada a partir dos gabinetes e que reflete a vontade de quem mandava, além de preconceitos expostos em opiniões pessoais – notadamente de raça – que fazem da caudalosa obra uma ode sobre si mesma. O historiador brasileiro Capistrano de Abreu foi preciso ao determinar a personalidade do Visconde: “sensível ao *vitupério como ao louvor” – um orgulhoso. (*vitupério=ofensa)

Da obra ainda se aproveitam trechos de documentos e citações de personalidades, mas a história em si vem sendo toda reescrita, até por irreverentes – e pouco precisos – programas exibidos no YouTube.

A história de Brasília, acompanhada de perto por testemunhas que ainda estão entre nós, também vem sofrendo revisões periódicas. Mesmo com registros diários da epopeia da construção da capital no meio do ermo, é eivada de lendas, muitas criadas pelo jornalismo romântico que tinha a necessidade de criar heróis.

Um desses homens foi *Bernardo Sayão, engenheiro que teve a tarefa de transformar os desenhos criados nas pranchetas em realidade. Se contava que ele mesmo pegava no cabo do facão mateiro para abrir picadas no cerrado, chamava os peões pelo nome, era comparado a um Hércules pela força e entusiasmo; um homem a quem era impossível dizer não. (*Bernardo Sayão – eng. agrônomo, político, iniciou a construção de Brasília).

A morte de Sayão, vitimado por um galho caído de uma enorme árvore, foi cercada de misticismo desde a primeira notícia. Ele não era invulnerável? Quis o destino

que o corpo de Sayão fosse o primeiro a ser sepultado no Campo da Esperança.

A história desse homem, se não recontada, está sendo esmiuçada em um livro diferente, sensível e amoroso. *Caminhos, Afetos, Cidades* (edição do autor) abraça o primeiro mito criado por Brasília a partir do relato de Sérgio de Sá, neto do engenheiro, professor da UnB, jornalista.

Sá não conheceu Sayão, morto 11 anos antes dele nascer, mas cresceu ouvindo histórias sobre o superavô. O personagem, as histórias, corroboradas por intensa pesquisa e método fazem uma leitura cativante e envolvente, que elevam o homem sem matar o mito.

É uma história que precisava ser (re)contada. Não é o caso de outras tentativas de (re)escrita.

Publicado no Correio Braziliense em 10 de setembro de 2023.

01) A leitura do texto de Paulo Pestana revela que seu principal objetivo é:

- A) promover uma ironia na relação entre o homem e o mito.
- B) questionar a autenticidade de vídeos no YouTube.
- C) descrever a história do Brasil a partir de historiadores.
- D) mostrar a importância de Sayão na história de Brasília.
- E) valorizar as experiências literárias do escritor George Orwell.

02) São elementos presentes no texto, exceto:

- A) tempo.
- B) espaço.
- C) enredo.
- D) narrador personagem.
- E) narrador observador.

03) “Da obra ainda se aproveitam trechos de documentos e citações de personalidades, mas a história em si vem sendo toda reescrita [...]” 5º§

Ao se reescrever essa frase, a alternativa que altera o sentido textual é:

- A) Da obra ainda se aproveitam trechos de documentos e citações de personalidades, porém a história em si vem sendo toda reescrita [...].
- B) Da obra ainda se aproveitam trechos de documentos e citações de personalidades, contudo a história em si vem sendo toda reescrita [...].
- C) Da obra ainda se aproveitam trechos de documentos e citações de personalidades, porque a história em si vem sendo toda reescrita [...].
- D) Da obra ainda se aproveitam trechos de documentos e citações de personalidades, todavia a história em si vem sendo toda reescrita [...].
- E) Da obra ainda se aproveitam trechos de documentos e citações de personalidades, entretanto a história em si vem sendo toda reescrita [...].

04) “[...] que fazem da caudalosa obra uma ode sobre si mesma.” 4º§

É sinônimo da palavra destacada nessa frase:

- A) sucinta.
- B) exígua.
- C) desprovida.
- D) subjetiva.
- E) copiosa.

05) “A história de Brasília [...] também vem sofrendo revisões periódicas.” 6º§

O plural do verbo destacado, mantendo-se o modo e o tempo, é:

- A) vêm.
- B) veem.
- C) vêem.
- D) vieras.
- E) vieram.

06) “Faltou dizer que os vencedores de hoje podem ser os derrotados de amanhã [...]” 2º§

Nesse trecho do texto, encontra-se a seguinte figura de linguagem:

- A) personificação.
- B) antítese.
- C) eufemismo.
- D) metáfora.
- E) elipse.

07) “A história desse homem, se não recontada, está sendo esmiuçada em um livro diferente [...]” 9º§

As vírgulas nesse período do texto separam

- A) orações coordenadas assindéticas.
- B) oração principal antecipada.
- C) oração subordinada antecipada.
- D) oração coordenada intercalada.
- E) oração subordinada intercalada.

08) Assinale a alternativa em que a(s) palavra(s) destacada(s) não representa(m) um substantivo.

- A) “O escritor britânico George Orwell sabia muito sobre animais [...]” 1º§
- B) “A morte de Sayão, vitimado por um galho caído de uma enorme árvore [...]” 8º§
- C) “O personagem, as histórias, corroboradas por intensa pesquisa e método [...]” 10º§
- D) “[...] reuniu todo o seu sarcasmo e sabedoria numa simples frase de um artigo [...]” 2º§
- E) “[...] muitas criadas pelo jornalismo romântico que tinha a necessidade de criar heróis.” 6º§

09) Marque a alternativa que preenche corretamente as lacunas da frase a seguir:

“Já _____ anos, _____ nesta capital poucos imigrantes residentes, porém hoje _____ muitos estrangeiros.”

- A) fazem – havia – existem.
- B) faz – havia – existem.
- C) faz – havia – existe.
- D) faz – haviam – existem.
- E) fazem – haviam – existem.

10) De acordo com a norma-padrão da Língua Portuguesa, está correta a seguinte afirmativa:

- A) são palavras grafadas com “s”: pretensão e pretensioso.
- B) o plural de “cidadão” é “cidadões”.
- C) a palavra “onde” indica ideia de movimento.
- D) recebem hífen as palavras “anti-racista” e “auto-conhecimento”.
- E) a locução “à partir” é grafada com crase.

CONHECIMENTOS ESPECÍFICOS

11) Na análise de uma rede neural profunda, um estatístico se depara com a otimização de uma função de custo não-linear altamente complexa. Considerando que a rede neural é projetada para classificação de imagens médicas, onde precisão e especificidade são críticas, o estatístico opta por utilizar um algoritmo de gradiente conjugado pré-condicionado para acelerar a convergência. Este algoritmo é aplicado no espaço de uma matriz Hessiana aproximada, H , cuja inversa é estimada para melhorar a direção de busca em cada iteração. Se a matriz Hessiana, H , é aproximadamente calculada pela fórmula de Broyden-Fletcher-Goldfarb-Shanno (BFGS) e o pré-condicionador, M , é escolhido para ser a matriz de covariância das entradas da rede, qual das seguintes afirmações melhor descreve o impacto esperado dessa escolha no processo de otimização, dada a natureza dos dados de imagens médicas e a estrutura da rede neural?

- A) A convergência do algoritmo será prejudicada pela alta dimensionalidade dos dados, tornando a escolha do pré-condicionador ineficaz.
- B) A utilização da matriz de covariância como pré-condicionador não é adequada para redes neurais profundas, pois pode levar a direções de busca que ignoram a estrutura de camadas da rede.
- C) O pré-condicionador escolhido é eficaz para dados de alta dimensão, como imagens médicas, pois pode reduzir o condicionamento da matriz Hessiana e acelerar a convergência.
- D) A escolha de M como a matriz de covariância introduzirá um viés no algoritmo de otimização, favorecendo direções de busca que minimizam a variação nas entradas em detrimento da redução do erro de classificação.
- E) O uso da matriz de covariância das entradas como pré-condicionador não afetará significativamente a otimização, pois a inversa da matriz Hessiana estimada pela BFGS já fornece uma aproximação suficiente para a direção de busca ótima.

12) Um grupo de estatísticos está investigando a eficácia de diferentes políticas públicas no controle de uma epidemia através de um modelo matemático de transmissão de doenças. O modelo é representado pela equação diferencial $\frac{dI}{dt} = \beta SI - \gamma I$, onde I é o número de indivíduos infectados, S é o número de indivíduos suscetíveis à infecção, β é a taxa de transmissão da doença, e γ é a taxa de recuperação. Assumindo que a população total $N = S + I$ é constante e que todas as variáveis e parâmetros são positivos, qual estratégia seria mais apropriada para avaliar o impacto de uma política de vacinação que reduz o número de suscetíveis S sem afetar diretamente β ou γ ?

- A) Solucionar analiticamente a equação diferencial para $I(t)$ com condições iniciais dadas e comparar o número de infectados no pico da epidemia para diferentes valores de S .

B) Utilizar o conceito de número básico de reprodução $R_0 = \frac{\beta S}{\gamma}$ para analisar como a redução de S afeta a capacidade de transmissão da doença e determinar o limiar de imunidade coletiva.

C) Aplicar métodos de simulação Monte Carlo para modelar o curso da epidemia sob diferentes cenários de vacinação, variando o número inicial de suscetíveis S .

D) Implementar uma análise de sensibilidade para estimar como pequenas variações nos parâmetros β e γ influenciam o pico de infectados, assumindo diferentes valores de S .

E) Calcular a integral da equação diferencial ao longo do tempo para determinar o total acumulado de indivíduos infectados e comparar os resultados para diferentes níveis de vacinação.

13) Em um projeto de pesquisa focado em entender os efeitos de políticas de trabalho remoto na produtividade dos funcionários de uma grande corporação, dados foram coletados sobre horas trabalhadas, produtividade mensurada por entregas concluídas, satisfação do funcionário, e número de reuniões virtuais por semana. Com o objetivo de explorar as dinâmicas subjacentes e identificar possíveis correlações e tendências ocultas entre estas variáveis, qual das seguintes técnicas de Análise Exploratória de Dados (AED) os estatísticos deveriam priorizar para obter insights iniciais significativos?

A) Criação de gráficos de pareto para cada variável, enfatizando as maiores fontes de variação na produtividade e satisfação dos funcionários.

B) Utilização de gráficos de contorno para visualizar a densidade de pontos nos espaços bidimensionais formados por pares de variáveis, identificando áreas de maior concentração.

C) Execução de análises de componentes principais (PCA) para reduzir a dimensionalidade dos dados e descobrir as principais direções de variação nas métricas coletadas.

D) Construção de gráficos de violino para as variáveis de satisfação do funcionário e produtividade, oferecendo uma visão detalhada da distribuição dos dados e potenciais outliers.

E) Aplicação de análise bivariada, utilizando coeficientes de correlação de Pearson, para quantificar a força e a direção das relações lineares entre variáveis contínuas.

14) Em uma análise de redes sociais, um grupo de estatísticos investiga padrões de conexão entre usuários para identificar influenciadores chave dentro de uma plataforma. Eles modelam a rede utilizando grafos, onde os usuários são representados por vértices e as conexões por arestas. A equipe está particularmente interessada em medir a centralidade dos vértices para determinar a influência. Eles consideram o grau, a intermediação (betweenness) e a proximidade (closeness) como métricas de centralidade. Após a coleta e modelagem dos dados, qual das seguintes afirmações reflete mais adequadamente a importância de combinar essas métricas de centralidade para identificar influenciadores na rede?

- A) A centralidade de grau é suficiente por si só, pois usuários com muitas conexões são invariavelmente os mais influentes na rede.
 B) A centralidade de intermediação é a mais crítica, indicando usuários que agem como pontes entre grupos distintos, independentemente do número de conexões diretas.
 C) A centralidade de proximidade deve ser a métrica primária, pois usuários mais próximos a todos os outros têm maior potencial de influência, independentemente de sua posição como intermediários.
 D) Combinar as métricas de centralidade é crucial, pois diferentes influenciadores podem ter papéis distintos na rede, seja como hubs de conexão, pontes entre comunidades ou vértices centralmente localizados.
 E) Nenhuma métrica de centralidade é eficaz para identificar influenciadores, pois a influência não pode ser quantificada apenas por padrões de conexão na rede.

15) Avalie as seguintes afirmações sobre conceitos avançados em Cálculo, determinando se são verdadeiras (V) ou falsas (F).

- () Uma função $f(x)$ definida e contínua em todo o intervalo real pode ter um número infinito de máximos e mínimos locais.
 () A integral $\int e^{-x^2} dx$ pode ser expressa em termos de funções elementares.
 () O Teorema de Green é aplicável para calcular a circulação de um campo vetorial ao redor de um caminho fechado, somente se este caminho estiver no plano xy .
 () A série de Fourier de uma função periódica e par $f(x)$ contém apenas coeficientes de cossenos.
 () O limite $\lim_{x \rightarrow 0} \frac{\sin(x)}{x}$ pode ser usado para provar que a derivada de $\sin(x)$ em relação a x é $\cos(x)$.

A sequência correta é:

- A) V – F – V – V – V.
 B) V – F – V – F – V.
 C) F – V – F – V – F.
 D) V – F – F – V – V.
 E) F – V – V – F – F.

16) Durante um estudo longitudinal sobre o impacto de intervenções de saúde pública na prevenção de uma doença específica em uma comunidade, um estatístico decide modelar a ocorrência de novos casos da doença dentro de grupos etários específicos usando a distribuição binomial negativa. A distribuição é escolhida devido à variabilidade no número de casos e ao interesse na contagem do número de ensaios até o r -ésimo sucesso, onde r representa o número de casos antes de uma intervenção ser considerada eficaz. Supondo que a probabilidade de sucesso (prevenção eficaz da doença) em cada ensaio é $p = 0.2$, e que o interesse é calcular a probabilidade de que mais de 10 ensaios sejam necessários para observar o terceiro sucesso ($r=3$). Qual das seguintes opções representa corretamente essa probabilidade?

- A) $P(X > 10) = 1 - \sum_{k=3}^{10} \binom{k-1}{2} (0.2)^3 (0.8)^{k-3}$
 B) $P(X > 10) = \sum_{k=11}^{\infty} \binom{k-1}{2} (0.2)^3 (0.8)^{k-3}$
 C) $P(X > 10) = 1 - \sum_{k=3}^{10} \binom{k+2}{2} (0.2)^3 (0.8)^{k-3}$
 D) $P(X > 10) = \sum_{k=11}^{\infty} \binom{k+2}{2} (0.2)^3 (0.8)^{k-3}$
 E) $P(X > 10) = 1 - \sum_{k=11}^{\infty} \binom{k-1}{3} (0.2)^3 (0.8)^{k-3}$

17) Em um estudo sobre o impacto de um novo medicamento na redução dos níveis de colesterol, um grupo de estatísticos planeja realizar um teste de hipóteses para comparar a média dos níveis de colesterol antes e depois da administração do medicamento em uma amostra de pacientes. A hipótese nula (H_0) afirma que não há diferença na média dos níveis de colesterol antes e depois do tratamento, enquanto a hipótese alternativa (H_1) sugere que há uma redução nos níveis de colesterol após o tratamento. Os estatísticos decidem usar um nível de significância de 0.05 para o teste. Se o valor-p calculado após a análise for de 0.03, qual das seguintes ações é mais apropriada?

- A) Rejeitar (H_0) e concluir que o medicamento não tem efeito sobre os níveis de colesterol.
 B) Não rejeitar (H_0) porque o valor-p é maior que 0.01, indicando evidência insuficiente para concluir que o medicamento é eficaz.
 C) Rejeitar (H_0) e concluir que há evidências suficientes para afirmar que o medicamento reduz os níveis de colesterol.
 D) Não rejeitar (H_0) devido à possibilidade de erro tipo II, sugerindo que mais dados são necessários.
 E) Rejeitar tanto (H_0) quanto (H_1), pois o valor-p não fornece informações conclusivas sobre a eficácia do medicamento.

18) Analise as seguintes afirmações sobre métodos numéricos.

I – O método de Newton-Raphson pode convergir para uma raiz mesmo se a estimativa inicial estiver distante da raiz real, desde que a função seja contínua e diferenciável no intervalo considerado.

II – A interpolação polinomial de Lagrange garante que o erro de interpolação é zero nos pontos de interpolação fornecidos.

III – Métodos de integração numérica, como a regra de Simpson, são sempre mais precisos que a regra do trapézio, independentemente da função ser avaliada.

IV – O método de decomposição LU é aplicável apenas a matrizes quadradas simétricas positivo-definidas.

V – Em análise numérica, a estabilidade de um algoritmo refere-se à sua capacidade de produzir resultados precisos mesmo quando sujeito a pequenas variações nos dados de entrada.

Estão corretas as afirmativas:

- A) I, II e III, apenas.
- B) II, III e IV, apenas.
- C) III, IV e V, apenas.
- D) I, IV e V, apenas.
- E) I, II e V, apenas.

19) Em uma análise avançada de modelos de filas, um grupo de estatísticos está interessado no estudo do processo de chegadas de clientes a um banco que possui um único caixa. Eles modelam as chegadas usando um processo de Poisson com taxa λ clientes por hora. Supondo que o serviço do caixa possa ser descrito por um processo exponencial com taxa μ clientes por hora, onde $\mu > \lambda$, os estatísticos desejam determinar a distribuição do número de clientes no sistema em estado estacionário. Qual das seguintes afirmações melhor descreve essa distribuição?

- A) A distribuição do número de clientes no sistema em estado estacionário segue uma distribuição Poisson com parâmetro λ/μ .
- B) O número de clientes no sistema em estado estacionário é descrito por uma distribuição exponencial com taxa $\mu - \lambda$.
- C) A distribuição do número de clientes no sistema em estado estacionário segue uma distribuição geométrica com parâmetro $p = \lambda/\mu$.
- D) O número de clientes no sistema em estado estacionário é modelado por uma distribuição binomial, onde o número de tentativas é igual a λ e a probabilidade de sucesso é $p = \mu/\lambda$.
- E) A distribuição do número de clientes no sistema em estado estacionário não pode ser determinada sem informações adicionais sobre a distribuição de chegadas e serviço.

20) Um grupo de estatísticos está analisando dados de expressão gênica de pacientes para identificar padrões associados a diferentes tipos de câncer. Os dados consistem em medições de expressão para milhares de genes em centenas de amostras. Dada a alta dimensionalidade dos dados, os estatísticos decidem aplicar Análise de Componentes Principais (PCA) para reduzir a dimensionalidade antes de realizar a clusterização. Após a aplicação da PCA, eles planejam usar o método K-means para identificar grupos de amostras com perfis de expressão gênica similares. Qual das seguintes afirmações melhor descreve um desafio crítico e uma consideração importante neste estudo?

- A) A PCA pode resultar na perda de informações críticas sobre interações gênicas complexas, o que pode ser crucial para a interpretação correta dos grupos de câncer identificados.
- B) O método K-means pode ser aplicado diretamente aos dados de expressão gênica sem a necessidade de redução de dimensionalidade, pois é insensível à maldição da dimensionalidade.
- C) A escolha do número de componentes principais a serem retidos após a PCA não influencia a interpretação dos clusters resultantes do K-means.
- D) A normalização dos dados de expressão gênica antes da aplicação da PCA não é necessária, pois a PCA é robusta a diferenças de escala entre os genes.
- E) A escolha do número de clusters no método K-means deve ser baseada exclusivamente em critérios estatísticos, sem considerar conhecimento biológico prévio.