

CONCURSO PÚBLICO



PREVIDÊNCIA SOCIAL

DATAPREV



DATA: 11/01/2009 - DOMINGO - TARDE

CARGO: Analista de Tecnologia da Informação

PERFIL: S01 - Análise de Informações

A T E N Ç Ã O

O **Caderno de Questões** contém 60 questões de múltipla-escolha, cada uma com 5 opções (A, B, C, D e E) e 01 questão discursiva.

1. Ao receber o material, verifique no **Cartão de Respostas** e na **Folha de Resposta da Prova Discursiva** seu nome, número de inscrição, data de nascimento e cargo. Qualquer irregularidade comunique imediatamente ao fiscal de sala. Não serão aceitas reclamações posteriores.
2. As provas objetiva e discursiva terão juntas duração de 4 horas e 30 minutos, incluídos neste tempo o preenchimento do **Cartão de Respostas** e da **Folha de Resposta da Prova Discursiva**.
3. Leia atentamente cada questão e assinale no **Cartão de Respostas** a opção que responde corretamente a cada uma delas. O **Cartão de Respostas** será o único documento válido para a correção eletrônica. O preenchimento do **Cartão de Respostas** e a respectiva assinatura serão de inteira responsabilidade do candidato. Não haverá substituição do **Cartão de Respostas** e da **Folha de Resposta da Prova Discursiva**, por erro do candidato.
4. Observe as seguintes recomendações relativas ao **Cartão de Respostas**:
 - A maneira correta de marcação das respostas é cobrir, fortemente, com esferográfica de tinta azul ou preta, o espaço correspondente à letra a ser assinalada.
 - Outras formas de marcação diferentes implicarão a rejeição do **Cartão de Respostas**.
 - Será atribuída nota zero às questões não assinaladas ou com falta de nitidez, ou com marcação de mais de uma opção, e as emendadas ou rasuradas.
5. O fiscal de sala não está autorizado a alterar qualquer destas instruções. Em caso de dúvida, solicite a presença do coordenador local.
6. Você só poderá retirar-se definitivamente do recinto de realização da prova após 60 minutos contados do seu efetivo início, **sem levar o Caderno de Questões**.
7. Você só poderá levar o próprio **Caderno de Questões** faltando uma hora para o término do horário da prova, conforme Edital do Concurso.
8. Por motivo de segurança, só é permitido fazer anotação durante a prova no **Caderno de Questões**.
9. Após identificado e instalado na sala, você não poderá consultar qualquer material, enquanto aguarda o horário de início da prova.
10. Os três últimos candidatos deverão permanecer na sala até que o último candidato entregue o **Cartão de Respostas** e a **Folha de Resposta da Prova Discursiva**.
11. Ao terminar a prova, é de sua responsabilidade entregar ao fiscal o **Cartão de Respostas** e a **Folha de Resposta da Prova Discursiva**. Não esqueça seus pertences.
12. O **Gabarito Oficial da Prova Objetiva** será disponibilizado no site www.concursos.uff.br, no dia 13/01/2009, conforme estabelecido no Cronograma Previsto.

CADERNO DE QUESTÕES

Realização:



BOA PROVA

CONHECIMENTO GERAL

1. "O executivo americano Ken Musgrave, coloca o pequeno retângulo de acrílico vermelho, um PC de mesa, ao lado de um PC normal, uma trivial caixa de alumínio cinza-chumbo. "Ambos têm as mesmas funcionalidades e capacidade de processamento", afirma diante das duas máquinas. "A diferença é que o modelo novo tem quase um quinto do tamanho do antigo e, portanto, precisa de muito menos material para ser fabricado. Fora isso, consome quase 70% menos energia".

O computador apresentado por Musgrave é considerado o que há de mais inovador do ponto de vista ambiental. "É o equipamento que melhor traduz o conceito de *green IT* (tecnologia da informação verde)", (...) o produto é diferenciado por reduzir o impacto ambiental de variadas formas - na menor quantidade de energia que será consumida durante sua vida útil, na redução de materiais que o compõem e até na embalagem, que é 95% reciclável".

(Adaptado de <http://info.abril.com.br/professional/ti-verde/quero-ser-verde.shtml>)

A sustentabilidade é um importante aspecto da relação do homem moderno com o seu ambiente, que, no caso acima, envolve diretamente:

- A) a utilização de matéria prima "verde" em sua fabricação;
- B) o investimento em *hardwares* esteticamente menos conspícuos;
- C) o uso de *softwares* de gerenciamento ambiental;
- D) a redução do impacto ambiental pelo uso da bioinformática;
- E) a redução da produção de lixo e do impacto ambiental decorrente.

2. "Depois de quase um ano de tentativas, os cientistas conseguiram transformar as células-tronco da gordura em células musculares humanas no corpo dos camundongos. E junto com elas, veio também um resultado impressionante.

Os camundongos tinham distrofia muscular, uma doença genética grave que afeta a força dos músculos aos poucos e que ainda não tem cura. Um camundongo que não recebeu células-tronco, por causa da distrofia muscular, não conseguia se segurar no arame e caía. Já o camundongo que recebeu as células-tronco de gordura se agarrava com firmeza ao arame. As células de gordura viraram músculos que funcionam."

(Disponível em <http://g1.globo.com/Noticias/Ciencia/0,,MUL913218-5603,00-CIENTISTAS+TRANSFORMAM+CELULASTRONCO+DE+GORDURA+EM+CELULAS+MUSCULARES.html>)

As células-tronco podem ser muito úteis na medicina, pois:

- A) são capazes de se modificarem em células especializadas;
- B) devem ser usadas para a preservação da biodiversidade;
- C) podem facilitar as pesquisas do genoma humano;
- D) são incapazes de sofrer qualquer tipo de transformação;
- E) não sofrem nenhum tipo de diferenciação celular.

3. Exemplos de materiais semicondutores inorgânicos responsáveis pelo grande desenvolvimento tecnológico atual na área da microeletrônica e na fabricação de *microchips* são:

- A) ouro e prata;
- B) silício e germânio;
- C) hidrogênio e nitrogênio;
- D) gálio e urânio;
- E) carvão e hélio.

4. Sobre o item Meu Computador, analise:

- I. Pode-se copiar arquivos e pastas, mas nunca atalhos.
- II. É possível apagar arquivos enviando a lixeira.
- III. Só é possível mover arquivos e atalhos.

Dos itens acima mencionados:

- A) I, apenas, está correto;
- B) II, apenas, está correto;
- C) III, apenas, está correto;
- D) I e II, apenas, estão corretos;
- E) I, II e III estão corretos.

5. Para selecionar várias células de uma mesma coluna numa Planilha utilizando o MS-Excel, deve-se manter pressionada a tecla (ou combinação de teclas):

- A) ALT
- B) CTRL
- C) SHIFT
- D) CTRL + SHIFT
- E) ESC

6. Conflitos de hardware ocorrem quando dois ou mais dispositivos utilizam o mesmo recurso. A opção que possui os recursos de hardware que podem entrar em conflito é:

- A) endereços de memória, endereços de E/S, linhas de interrupções (IRQ) e canais de DMA;
- B) canais de impressão, endereços de memória, monitor;
- C) mouse, impressora, monitor;
- D) endereços de memória, canais de impressão e interrupções (IRC);
- E) considerar que as opções acima são incorretas.

7. No Menu Exibir do Windows Explorer, a opção que faz com que o modo exibição dos arquivos e pastas forneça informações tais como nome, tipo, tamanho e data de modificação é:

- A) Miniaturas.
- B) Lado a lado.
- C) Ícones.
- D) Lista.
- E) Detalhes.

8. A única das proposições abaixo que pode ser considerada uma negação lógica da proposição: "**Se é feriado, então vou à praia**", é:

- A) Não é feriado e não vou à praia.
- B) Se não vou à praia, é feriado.
- C) Não vou à praia ou é feriado.
- D) Não vou à praia é feriado.
- E) É feriado e não vou à praia.

9. Sejam V (verdadeiro) e F (falso) os valores lógicos associados às proposições compostas a seguir.

- I. O cachorro é mamífero ou vaca voa. ()
- II. Se cachorro é mamífero, então vaca voa. ()
- III. O cachorro é mamífero e vaca voa. ()
- IV. O cachorro é mamífero, se e somente se vaca voa. ()

A seqüência ordenada dos valores lógicos obtidos é:

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

10. Dentre os argumentos lógicos apresentados abaixo, o único que é um **silogismo** é:

- A) $1 < 4$ ou $7 < 4$. $7 \geq 4$, logo, $1 < 4$.
- B) Se $1 < 4$, então $7 \geq 4$.
- C) $1 < 4$ ou $7 \geq 4$. Logo, $1 < 4$.
- D) $1 < 4$; $7 \geq 4$; logo $1 \neq 4$
- E) $1 < 4$, se e somente se $7 \geq 4$.

LÍNGUA INGLESA

Gadget Designers Push the Limits of Size, Safety

By Brian X. Chen, August 28, 2008

Just as small, fast-moving mammals replaced lumbering dinosaurs, pocketable gadgets are evolving to fill niches that larger, deskbound computers can't reach. But as they shrink, these gadgets are faced with problems mammals face, too, such as efficiently dissipating heat.

The recent example of Apple's first-generation iPod nanos causing fires in Japan raises the question of whether increasingly innovative product designs are impinging on safety. The nano incident illustrates how risk can increase as devices decrease in size, says Roger Kay, an analyst at Endpoint Technologies.

"As [gadgets] get smaller, the tradeoffs become more difficult, the balance becomes more critical and there's less room for error," Kay said. "I'm not surprised it's happening to the nano because that's the small one. You're asking it to do a lot in a very, very small package and that's pushing the envelope."

There's no question that industrial designers' jobs have become much more difficult as the industry demands ever more powerful and smaller gadgets. With paper-thin subnotebooks, ultrasmall MP3 players, and pinkie finger-sized Bluetooth headsets becoming increasingly popular, it's questionable where exactly designers draw the line between innovation and safety.

11. In the sentence: "But as they shrink, these gadgets with problems mammals face too, such as dissipating heat." First paragraph, the author of the text believes that...

- A) like the early mammals, gadgets will have to be able to maintain heat;
- B) like early mammals, nano technology will have to develop better condition of safety to survive;
- C) pocketable gadgets are evolving to overtake deskbound computers positions;
- D) people are in danger when they use their newest gadgets;
- E) gadgets may be compared to lumbering dinosaurs and have no survival chance in the future.

12. The title of the text let us know that there must be a close relation between ...

- A) design and safety;
- B) technology and safety;
- C) nano technology and design;
- D) nano technology and safety;
- E) design and technology.

13. The reason given in the third paragraph of the text for the problem with smaller gadgets is that...

- A) it gets hard to deal with anything in a tiny space;
- B) gadgets designer will never know how to deal with nano technology;
- C) gadgets designer are not able to deal with high technology;
- D) science hasn't reach the necessary safety;
- E) it is very difficult to work with high technology at any rate.

14. According to the text industrial designers' jobs have become more difficult because they...

- A) require more responsibility;
- B) require a production of smaller and safer gadgets;
- C) require more skills;
- D) demand a whole lot of patience;
- E) demand too much time of research and of tests.

15. The expression *such as* in the last line of the first paragraph could be changed by ____ and its meaning would not altered.

- A) by;
- B) like;
- C) although;
- D) but;
- E) because.

Choose the best option.

16. Dumping your entire music collection ____ your iPod is a simple, one click process. But what about getting your music ____ your iPod?

- A) in / out;
- B) in / off;
- C) on / off;
- D) onto / off;
- E) into / out.

17. Boeing ____ this week that it ____ successfully ____ a manned airplane powered ____ hydrogen fuel cells.

- A) announced / has / flown / on;
- B) has announced / had / flown / by;
- C) announced / has / flown / by;
- D) have announced / has / flown / on;
- E) announced / had / flown / on.

18. Robots ____ have the brains to "intelligently and autonomously search ____ objects" ____ their own.

- A) didn't/for/in;
- B) don't/of/in;
- C) haven't/for/on;
- D) don't/for/on;
- E) doesn't/for/on.

19. If this scientific breakthrough is reliable, after years testing it, it ____ lead to the cure of this type of cancer in the future.

- A) must;
- B) should;
- C) might;
- D) ought to;
- E) could.

20. He was very precise about the proper procedure and the material to be used, he said he didn't want anything especial, only the regular, ____ one.

- A) unique;
- B) ordinary;
- C) rare;
- D) exquisite;
- E) unknown.

23. A soma dos desvios ao quadrado de um grupo de medidas em relação à média dividido pelo total de valores, é igual :

- A) ao desvio-médio;
- B) à mediana;
- C) ao desvio-padrão;
- D) à variância;
- E) à média harmônica.

24. Determine a mediana do conjunto: 8; 16; 14; 6; 10; 12.

- A) 10;
- B) 11;
- C) 12;
- D) 13;
- E) 14.

25. Considere que os eventos A, B, C, D, E e F são mutuamente exclusivos, calcule $P(B \cap D)$.

Eventos	Probabilidade
A	0,387
B	0,345
C	0,116
D	0,033
E	0,079
F	0,040

- A) 0,010;
- B) 0,622;
- C) 0,543;
- D) 0,189;
- E) 0,378.

CONHECIMENTO ESPECÍFICO

Considere a seguinte distribuição de frequência relativa acumulada. Considere 100 anos como a idade máxima da distribuição. Com base na tabela responda às questões 21 e 22.

Idade (em anos)	Frequência relativa acumulada
16 — 26	0,05
26 — 34	0,15
34 — 43	0,35
43 — 52	0,55
52 — 61	0,85
61 — 70	0,95
> 70	1,00

21. A idade média é aproximadamente:

- A) 56 anos;
- B) 49 anos;
- C) 55 anos;
- D) 44 anos;
- E) 60 anos.

22. Sobre a distribuição, é correto afirmar que:

- A) possui assimetria à esquerda;
- B) é simétrica;
- C) a idade modal situa-se entre 61 e 70 anos;
- D) o segundo quartil situa-se entre 26 e 34 anos;
- E) a diferença interquartilica é superior a 50 anos.

26. Dois dados não viciados são lançados. Considere X_1 = face do dado 1, X_2 = face do dado 2 e os eventos:

$$A \quad X_1, X_2 \mid X_1 \quad X_2 \quad 10$$

$$B \quad X_1, X_2 \mid X_1 \quad X_2$$

Calcule $P(B/A)$.

- A) 12/15;
- B) 5/12;
- C) 1/3;
- D) 1/15;
- E) 1/12.

27. A urna I contém 9 bolas: 3 pretas, 2 brancas e 4 vermelhas. A urna II contém 8 bolas: 4 pretas, 1 branca e 3 vermelhas. A urna III contém 9 bolas: 1 preta, 3 brancas e 5 vermelhas. Escolhe-se uma urna ao acaso e dela extrai-se uma bola também ao acaso, sabendo que a bola sorteada foi branca, a probabilidade de ter vindo da urna II é de:

- A) 1/6;
- B) 6/26;
- C) 25/59;
- D) 9/49;
- E) 1/8.

28. Uma variável aleatória X tem função de densidade de probabilidade dada por:

$$f(x) = \begin{cases} x/2^2, & 0 \leq x \leq 3 \\ 0, & \text{nos demais casos} \end{cases}$$

A esperança matemática é aproximadamente:

- A) 5,06;
- B) 2,25;
- C) 3,00;
- D) 1,62;
- E) 4,25.

29. O atributo $Z=(X-2)/3$ tem média amostral 20 e variância amostral 9. Indique o coeficiente de variação amostral de X:

- A) 0,284;
- B) 0,077;
- C) 0,145;
- D) 0,060;
- E) 0,052.

30. Considere a variável aleatória X que assume valores inteiros no intervalo [0,5] com a seguinte função densidade de probabilidade:

$$f(x) = \begin{cases} 0,05x, & 0 \leq x \leq 5 \\ 0, & \text{nos demais casos} \end{cases}$$

Calcule a $P(2 < X \leq 5)$.

- A) 0,925;
- B) 0,825;
- C) 0,725;
- D) 0,625;
- E) 0,525.

31. Sobre a distribuição Binomial, é correto afirmar que é:

- A) freqüentemente utilizada para modelar eventos discretos ocorridos com baixa freqüência no tempo ou no espaço, por vezes é chamada de distribuição de eventos raros;
- B) uma seqüência de n ensaios de Bernoulli, cada um com probabilidade de sucesso p;
- C) a única distribuição de probabilidades que admite valores não inteiros;
- D) uma distribuição contínua limitada por valores discretos;
- E) totalmente descrita por seus dois parâmetros: média e desvio-padrão.

32. Se X_1 tem distribuição $N(17,9)$ e X_2 tem distribuição $N(34,16)$, sendo X_1 e X_2 independentes, determine $\sigma(X_1+X_2)$.

- A) 12;
- B) 25;
- C) 5;
- D) 18;
- E) 51.

33. Um dispositivo eletrônico recentemente desenvolvido apresenta vida média de 80 horas. Considerando o comportamento segundo a distribuição exponencial, a probabilidade desse dispositivo durar mais de 100 horas é:

- A) $e^{5/4}$;
- B) $1+e^{-5/4}$;
- C) $1-e^{-5/4}$;
- D) $e^{-5/4}$;
- E) 5/4.

34. A desigualdade de _____ nos permite dizer que para qualquer número k _____ ou igual a 1, pelo menos _____ das medidas no conjunto de dados se encontram a até _____ desvios-padrão de suas médias.

Os termos que completam adequadamente o trecho acima são, respectivamente:

- A) Markov / maior / $[1-(1/k)^2] / k$;
- B) Chebychev / maior / $[1-(1/k)^2] / k$;
- C) Chebychev / menor / $[1-(1/k)^{1/2}] / k^2$;
- D) Chebychev / maior / $[1-(1/k)^2] / 2k$;
- E) Markov / menor / $[1-(1/k)^{1/2}] / 2k^2$.

35. O Teorema central do limite aplica-se a qualquer população com _____ finito(a), _____ da forma da distribuição original. Entretanto, quanto mais a população original se _____ da distribuição normal, _____ é o tamanho da amostra necessário para assegurar a normalidade da distribuição amostral.

Os termos que completam adequadamente o trecho acima são, respectivamente:

- A) desvio-padrão / independente / afasta / maior;
- B) média / dependente / afasta / maior;
- C) desvio-padrão / dependente / aproxima / menor;
- D) variância / dependente / afasta / maior;
- E) média / independente / aproxima / menor.

36. Considere as afirmativas:

- I. Se considerarmos somente estimadores não tendenciosos de um parâmetro, aquele com a menor variância é dito eficiente.
- II. Um estimador é considerado não tendencioso se ele se aproxima do valor real do parâmetro estudado, conforme as amostras tornam-se maiores.
- III. Um estimador é dito linear se ele é uma função linear das observações amostrais.

Das mencionadas acima, apenas:

- A) I é verdadeira;
- B) II é verdadeira;
- C) III é verdadeira;
- D) I e III são verdadeiras;
- E) I e II são verdadeiras.

37. Para uma amostra aleatória de 225 trabalhadores de uma empresa, que emprega 12.000 pessoas, 45 preferem escolher seu próprio plano de aposentadoria. Encontre o erro máximo aproximado do intervalo de confiança de 95%, para a proporção de todos os trabalhadores da empresa que preferem escolher seu próprio plano de aposentadoria. Considere $Z_{/2} = 1,96$.

- A) 0,15;
- B) 0,10;
- C) 0,05;
- D) 0,01;
- E) 0,005.

38. Deseja-se estimar com uma margem de erro de 3% na estimativa, a percentagem de beneficiários que possuem plano de previdência privada, porém não existe nenhuma informação que possa sugerir um valor da proporção estimada. Considerando um nível de confiança de 95%, a quantidade de beneficiários que devem ser entrevistados é de: Considere $Z_{/2} = 1,96$.

- A) 305;
- B) 384;
- C) 635;
- D) 956;
- E) 1.068.

39. Para uma amostra aleatória de tamanho 144 selecionada da população de um município brasileiro, a média de um atributo é 400. Com base na população, sabe-se que o valor médio desse atributo é 415 e variância 16. O valor da estatística mais adequada ao teste é:

Considere $H_0: \mu = 415$
 $H_1: \mu \neq 415$

- A) 60;
- B) 3,75;
- C) -75;
- D) -845;
- E) -45.

40. Os motivos de contato a um canal de atendimento foram classificados como reclamação e não reclamação, conforme o status do cliente na base de dados.

Status	Reclamação	
	Não	Sim
Não aposentado	100	250
Aposentado	50	150

O valor calculado da estatística qui-quadrado para testar independência vale:

- A) 11/672;
- B) 275/336;
- C) 1/50;
- D) 23/55;
- E) 12/15.

41. O nível de significância de um teste de hipóteses é a probabilidade de:

- A) não rejeitar H_0 quando H_0 é verdadeira;
- B) não rejeitar H_0 quando H_0 não é verdadeira;
- C) rejeitar H_0 quando H_0 não é verdadeira;
- D) rejeitar H_0 quando H_0 é verdadeira;
- E) H_0 ser verdadeira.

42. Considere a seguinte distribuição conjunta de probabilidades, onde $(X;Y)$ é uma variável bidimensional discreta. Calcule a covariância entre X e Y.

X \ Y	-1	1
0	0,1	0,2
1	0,3	0,1
2	0,2	0,1

- A) 1,0;
- B) 0,25;
- C) -0,4;
- D) -0,2;
- E) 0,6.

43. Baseada numa amostra aleatória de 212 famílias, a seguinte regressão foi obtida pelo método de mínimos quadrados ordinários: $\hat{y}_i = 1,924 - 0,19\hat{x}_i$

Onde \hat{y}_i representa os impostos anuais pagos por família e \hat{x}_i a renda familiar mensal, ambas medidas em milhares de real.

Com base nessas informações, é possível afirmar que:

- A) mantendo as influências externas constantes, para cada incremento de R\$1.000 na renda familiar, a taxa paga pelos impostos sobe em média 19%;
- B) existe heterocedasticidade na distribuição dos resíduos;
- C) se a renda familiar é de R\$400, o valor dos impostos estimados é de aproximadamente R\$115;
- D) essa é uma regressão linear múltipla;
- E) o intercepto da regressão vale 0,19.

44. Considere X e Y variáveis aleatórias e $r_{X,Y}$, a correlação entre X e Y. Pode-se afirmar que:

- A) se $r_{X,Y} > 0$, a correlação é dita positiva;
- B) se $r_{X,Y} = 1$, então o coeficiente de determinação é igual a zero;
- C) $r_{X,Y} = r_{Y,X}$;
- D) $r_{X,Y}$ tem a unidade de medida igual a unidade de medida de X ao quadrado;
- E) se $r_{X,Y} = 1$, não há relação linear entre X e Y.

45. Considere \mathbf{a} , o ponto onde uma reta ajustada corta o eixo da variável Y e \mathbf{b} , a tangente do ângulo que a reta forma com uma paralela ao eixo da variável X. Se $n=500$, $\sum_{i=1}^n x_i = 250$, $\sum_{i=1}^n y_i = 750$ e $\mathbf{a}=0,15$, então \mathbf{b} vale:

- A) 0,3;
- B) 0,4;
- C) 1,4;
- D) -1,2;
- E) 2,7.

46. O poder explicativo de uma regressão linear múltipla é determinado pelo(a):

- A) coeficiente de determinação;
- B) correlação linear entre a variável de maior coeficiente e a variável explicada;
- C) teste de Durbin-Watson;
- D) soma dos resíduos;
- E) estatística F da análise da variância.

47. A seguinte tabela de análise da variância foi disponibilizada, porém algumas informações estão faltando. O valor calculado da estatística F é:

Fonte de variação	Soma de quadrados	Graus de liberdade	Quadrados médios	F calculado
Regressão				
Resíduo	52,52	13		
Total	58,72	15		

- A) 3,91;
- B) 0,77;
- C) 1,30;
- D) 7,14;
- E) 0,48;

48. Quando é dito que um modelo de regressão linear é heterocedástico, quer-se dizer que:

- A) a soma das diferenças entre os valores observados e seus resíduos não é nula;
- B) todos os estimadores do modelo são eficientes;
- C) a variância dos resíduos não é constante;
- D) existe forte correlação entre as variáveis explicativas;
- E) existe autocorrelação entre os resíduos.

49. A instabilidade dos coeficientes estimados em uma regressão e de seus erros-padrão, os quais, particularmente, tornam-se muito grandes, é um sintoma de um determinado problema ao estimar uma regressão. Trata-se:

- A) da variância dos resíduos não constantes;
- B) do baixo coeficiente de determinação;
- C) da heterocedasticidade;
- D) da autocorrelação dos resíduos;
- E) da colinearidade.

50. Se X tem distribuição normal multivariada, é possível observar as propriedades:

- I. Combinações lineares das componentes de X são normalmente distribuídas;
- II. Todos os subconjuntos das componentes de X têm distribuição normal;
- III. Covariância nula entre componentes implica que estas são independentemente distribuídas;
- IV. A distribuição condicional das componentes é normal multivariada.

Das propriedades acima:

- A) I, apenas, está correta;
- B) I e III, apenas, estão corretas;
- C) I, II e IV, apenas, estão corretas;
- D) III e IV, apenas, estão corretas;
- E) I, II, III e IV estão corretas.

51. Em uma regressão logística múltipla, se uma variável explicativa X_i é dicotômica, seu coeficiente $\hat{\beta}_i$ pode ter uma interpretação especial: a razão de chance (*odds ratio*) estimada da resposta para dois níveis possíveis de X_i . Pode-se definir a razão de chance como:

- A) antilogaritmo de $\hat{\beta}_i$;
- B) logaritmo de $\hat{\beta}_i$;
- C) razão entre $e^{\hat{\beta}_i}$ e $e^{-\hat{\beta}_i}$;
- D) soma dos quadrados de $\hat{\beta}_i$;
- E) razão entre $e^{\hat{\beta}_i}$ e $e^{-\hat{\beta}_i}$.

52. Em análise de conglomerados, por muitos chamada de análise de *clusters*, cita-se dois métodos conhecidos: o método hierárquico e o k-médias (*k-means*). Sobre o método hierárquico, pode-se afirmar:

- A) é um método mais subjetivo que o k-médias, o analista precisa ter excelente conhecimento do fenômeno estudado;
- B) exige recursos computacionais elevados para ser trabalhado com grandes amostras;
- C) analisa o relacionamento entre as variáveis dependentes e independentes;
- D) é fundamental que o pesquisador defina *a priori* a quantidade de agrupamentos a serem analisados, sem essa definição, não é possível fazer quaisquer análises;
- E) não é conveniente padronizar as variáveis.

53. O objetivo principal da Análise de Componentes Principais é:

- A) obtenção de um pequeno número de combinações lineares, de um conjunto de variáveis, que retenham o máximo possível da informação contida nas variáveis originais;
- B) calcular, para cada indivíduo da amostra, a probabilidade associada a um determinado fenômeno;
- C) testar, por meio de sub-amostras, a distribuição de probabilidades do conjunto de dados estudados;
- D) explicar a correlação ou covariância, entre um conjunto de variáveis, em termos de um número limitado de variáveis não-observáveis;
- E) identificar se há multicolinearidade nos resíduos, após a estimação de uma curva de regressão.

54. Deseja-se selecionar uma amostra aleatória contendo 800 indivíduos de uma população dividida em dois estratos. O estrato I é composto de 100.000 indivíduos e o II, de 20.000. Estudos passados revelaram que as variâncias da variável de interesse nos estratos I e II são, respectivamente, 225 e 625. Utilizando o método de alocação ótima de Neyman, é correto afirmar que os tamanhos das amostras a serem extraídas nos estratos I e II, respectivamente, são:

- A) 833 e 167;
- B) 750 e 250;
- C) 643 e 357;
- D) 545 e 455;
- E) 500 e 500.

55. Um profissional deseja estimar a renda média de uma população. O número de indivíduos que devem ser selecionados ($n > 30$), se o profissional deseja ter 95% de confiança para que a média amostral esteja a mais ou a menos \$500 da verdadeira média populacional, é de: (por um estudo prévio, para tais rendas, $Z_{\alpha/2} = 1,96$).

- A) 465;
- B) 43;
- C) 234;
- D) 129;
- E) 354.

56. Se uma amostra de tamanho n é desejada e a fração amostral da população, n/N , não é um inteiro, seu valor é arredondado para o número inteiro mais próximo. A seguir, conforme a lista é percorrida, cada k -ésima unidade consecutiva é selecionada. A lista NÃO pode estar ordenada por atributos.

As "dicas" referem-se ao tipo de amostragem :

- A) aleatória simples;
- B) estratificada;
- C) sistemática;
- D) conglomerados;
- E) não probabilística.

57. "Mining é parte de um processo maior de conhecimento, que o processo consiste, fundamentalmente, na estruturação do banco de dados; na seleção, preparação e pré-processamento dos dados; na transformação, adequação e redução da dimensionalidade dos dados; e nas análises, assimilações, interpretações e uso do conhecimento extraído do banco de dados."

O processo maior citado no início do texto é denominado:

- A) Data mining;
- B) Data mart;
- C) Data warehouse;
- D) KDD;
- E) Segmentação de dados.

58. Uma ferramenta OLAP nos permite:

- A) manipular e analisar um largo volume de dados sob múltiplas perspectivas;
- B) ter uma visão fixa dos dados;
- C) armazenar informações;
- D) disparar requisições diretamente ao servidor de Banco de Dados multidimensional, tendo um ganho de desempenho;
- E) registrar todas as transações contidas em uma determinada operação organizacional.

59. Sobre a relação DW e OLAP, pode-se afirmar:

- I. DW é utilizado para armazenar informações e o OLAP para recuperá-las, ambos são especializados para exercer suas funções de forma eficiente;
- II. DW e OLAP são poderosas tecnologias independentes e não complementares;
- III. Para a exploração completa do DW, a ferramenta OLAP irá extrair e alavancar as informações contidas nele.

Dos itens acima mencionados, apenas:

- A) I é verdadeiro;
- B) II é verdadeiro;
- C) III é verdadeiro;
- D) I e III são verdadeiros;
- E) II e III são verdadeiros.

60. Sobre modelagem multidimensional, o cubo:

- I. é uma representação intuitiva, pois todas as dimensões coexistem para todo ponto no cubo e são independentes umas das outras;
- II. é, de fato, apenas uma metáfora visual;
- III. serve para descrever requisitos funcionais.

Acerca dos itens acima mencionados, apenas:

- A) I e III estão corretos;
- B) I está correto;
- C) I e II estão corretos;
- D) II e III estão corretos;
- E) III está correto.

DISCURSIVA

Elabore um texto dissertativo-argumentativo, em torno de 20 a 30 linhas, a respeito do seguinte tema: "Data warehouse e data mining".

Pontos relevantes a serem abordados:

- 1. Conceito de DW, benefícios da tecnologia e os cuidados necessários em sua construção;
- 2. Relação DW e data mining;
- 3. Elenco dos principais algoritmos de data mining, ilustrando seus mecanismos de funcionamento e, sempre que possível, relacionando-os às técnicas estatísticas que possam vir a servir de embasamento aos algoritmos exemplificados.

RASCUNHO