



Nome do candidato:

Número do documento de identidade:

Número de inscrição:

Sala:

Seqüencial:

CONCURSO PÚBLICO NÍVEL SUPERIOR

Cargo: ANALISTA DE TECNOLOGIA DA INFORMAÇÃO

PERFIL ANÁLISE DE INFORMAÇÕES

Aplicação: 21/5/2006



LEIA COM ATENÇÃO AS INSTRUÇÕES ABAIXO.

- 1 Ao receber este caderno, confira inicialmente os seus dados pessoais transcritos acima. Em seguida, verifique se ele contém cento e vinte itens, correspondentes às provas objetivas, corretamente ordenados de 1 a 120, seguidos da prova discursiva.
- 2 Caso os dados pessoais constantes neste caderno não correspondam aos seus, ou, ainda, caso o caderno esteja incompleto ou tenha qualquer defeito, solicite ao fiscal de sala mais próximo que tome as providências cabíveis.
- 3 O espaço para rascunho da prova discursiva é de uso opcional; não contará, portanto, para efeito de avaliação.
- 4 Não utilize lápis, lapiseira, borracha e(ou) qualquer material de consulta que não seja fornecido pelo CESPE/UnB.
- 5 Não serão distribuídas folhas suplementares para rascunho nem para texto definitivo.
- 6 Não se comunique com outros candidatos nem se levante sem autorização do chefe de sala.
- 7 Nas provas objetivas, recomenda-se não marcar ao acaso: em cada item, se a resposta divergir do gabarito oficial definitivo, o candidato receberá pontuação negativa, conforme consta em edital.
- 8 Na prova discursiva, não será avaliado texto escrito a lápis, texto escrito em local indevido ou texto que tenha identificação fora do local apropriado.
- 9 A duração das provas é de **quatro horas e trinta minutos**, já incluído o tempo destinado à identificação — que será feita no decorrer das provas —, ao preenchimento da folha de respostas e à transcrição do texto definitivo da prova discursiva para a folha de texto definitivo.
- 10 Você deverá permanecer obrigatoriamente em sala por, no mínimo, uma hora após o início das provas.
- 11 Ao terminar as provas, chame o fiscal de sala mais próximo, devolva-lhe a sua folha de respostas e a folha de texto definitivo da prova discursiva e deixe o local de provas.
- 12 A desobediência a qualquer uma das determinações constantes no presente caderno, na folha de respostas ou na folha de texto definitivo da prova discursiva poderá implicar a anulação das suas provas.

AGENDA

- I 23/5/2006, após as 19 h (horário de Brasília) – Gabaritos oficiais preliminares das provas objetivas: Internet — www.cespe.unb.br/concursos/dataprev2006.
- II 24 e 25/5/2006 – Recursos (provas objetivas): exclusivamente no Sistema Eletrônico de Interposição de Recurso, Internet — www.cespe.unb.br/concursos/dataprev2006, mediante instruções e formulários que estarão disponíveis nesse endereço.
- III 13/6/2006 – Resultados final das provas objetivas e provisório da prova discursiva: Diário Oficial da União e Internet — www.cespe.unb.br/concursos/dataprev2006.
- IV 14 e 15/6/2006 – Recursos (prova discursiva): em locais e horários que serão informados na divulgação do resultado provisório.
- V 27/6/2006 – Resultados finais da prova discursiva e do concurso: locais mencionados no item III.

OBSERVAÇÕES

- Não serão objeto de conhecimento recursos em desacordo com o item 11 do Edital n.º 1/2006 – DATAPREV, de 13/3/2006.
- Informações adicionais: telefone 0(XX) 61 3448-0100; Internet — www.cespe.unb.br/concursos/dataprev2006.
- É permitida a reprodução deste material apenas para fins didáticos, desde que citada a fonte.

ITEM	RESPOSTA														
1	C E	16	C E	31	C E	46	C E	61	C E	76	C E	91	C E	106	C E
2	C E	17	C E	32	C E	47	C E	62	C E	77	C E	92	C E	107	C E
3	C E	18	C E	33	C E	48	C E	63	C E	78	C E	93	C E	108	C E
4	C E	19	C E	34	C E	49	C E	64	C E	79	C E	94	C E	109	C E
5	C E	20	C E	35	C E	50	C E	65	C E	80	C E	95	C E	110	C E
6	C E	21	C E	36	C E	51	C E	66	C E	81	C E	96	C E	111	C E
7	C E	22	C E	37	C E	52	C E	67	C E	82	C E	97	C E	112	C E
8	C E	23	C E	38	C E	53	C E	68	C E	83	C E	98	C E	113	C E
9	C E	24	C E	39	C E	54	C E	69	C E	84	C E	99	C E	114	C E
10	C E	25	C E	40	C E	55	C E	70	C E	85	C E	100	C E	115	C E
11	C E	26	C E	41	C E	56	C E	71	C E	86	C E	101	C E	116	C E
12	C E	27	C E	42	C E	57	C E	72	C E	87	C E	102	C E	117	C E
13	C E	28	C E	43	C E	58	C E	73	C E	88	C E	103	C E	118	C E
14	C E	29	C E	44	C E	59	C E	74	C E	89	C E	104	C E	119	C E
15	C E	30	C E	45	C E	60	C E	75	C E	90	C E	105	C E	120	C E

De acordo com o comando a que cada um dos itens de 1 a 120 se refira, 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, caso deseje, o rascunho acima e, posteriormente, a **folha de respostas**, único documento válido para a correção das suas provas.

CONHECIMENTOS BÁSICOS

Texto para os itens de 1 a 7

Em numerosas reportagens de jornais e televisões, temos lido que as emissoras de TV defendem a escolha do padrão japonês de modulação da TV digital (ISDB), porque este seria o único padrão que lhes permitiria fazer transmissão para recepção móvel, usando a banda do espectro eletromagnético reservada para o UHF. No caso do padrão europeu (DVB), por exemplo, a transmissão para recepção móvel teria de usar a banda reservada para a telefonia celular, o que incluiria as empresas de telefonia no núcleo central da operação de TV. Recosas dessa concorrência, as emissoras, então, preferem a modulação japonesa. Antes de prosseguir, algumas ressalvas:

- ▶ o padrão de modulação brasileiro, desenvolvido pela PUC-RS, conhecido como SORCER, também permite a transmissão para recepção móvel. Portanto, mesmo aceitando o argumento das emissoras, poderíamos adotar uma modulação com tecnologia brasileira;
- ▶ todos os padrões de modulação (japonês, europeu, norte-americano e brasileiro, além do chinês, que está em desenvolvimento) permitem transmitir em SDTV, EDTV e HDTV, ou seja, para essa questão específica, a escolha da modulação é indiferente;
- ▶ igualmente, todos os padrões permitem que se desenvolva uma série de serviços interativos, como governo eletrônico, *e-learning*, *e-bank*, telemedicina etc. Novamente, nessa questão específica, a escolha da modulação é indiferente.

Gustavo Gindre, coordenador-geral do Instituto de Estudos e Projetos em Comunicação e Cultura (INDECS). Internet: <www.oppi.org.br> (com adaptações).

Julgue os itens a seguir, a respeito das idéias e das estruturas lingüísticas do texto.

- 1 De acordo com o texto, é indiferente a escolha de um padrão de modulação para TV digital, porque todos permitem a transmissão para recepção móvel.
- 2 Preservam-se o sentido e a correção gramatical, se for empregada a preposição **em** antes de “que” (l.4).
- 3 A forma verbal “usando” (l.5) se refere à expressão “o único padrão” (l.4).
- 4 Na expressão “o que” (l.9), o termo sublinhado retoma coesivamente o trecho “No caso do padrão europeu (...) telefonia celular” (l.6-9).
- 5 Com o emprego do termo “também” (l.14) soma-se mais um argumento à idéia de que a transmissão para recepção móvel não é exclusividade do padrão japonês de modulação.
- 6 Nas linhas de 13 a 27, a primeira palavra de cada ressalva pode ser reescrita com letra inicial maiúscula sem prejuízo da correção gramatical.
- 7 Seria sintaticamente correto substituir-se o trecho “que se desenvolva uma série de serviços interativos” (l.23-24) por: que uma série de serviços interativos seja desenvolvida.

1 Inclusão digital é, entre outras coisas, alfabetização
digital, ou seja, é a aprendizagem necessária ao indivíduo
para circular e interagir no mundo das mídias digitais como
4 consumidor e produtor de seus conteúdos e processos. Para
isso, computadores conectados em rede e *software* são
instrumentos técnicos imprescindíveis. Mas são apenas isso,
7 suportes técnicos às atividades a serem realizadas a partir
deles no universo da educação, no mundo do trabalho, nos
novos cenários de circulação das informações e nos
10 processos comunicativos.

Dizer que inclusão digital é somente oferecer
computadores seria análogo a afirmar que as salas de aula,
13 cadeiras e quadro-negro garantiriam a escolarização e o
aprendizado dos alunos. Sem a inteligência profissional dos
professores e sem a sabedoria de uma instituição escolar que
16 estabelecesse diretrizes de conhecimento e trabalho nesses
espaços, as salas seriam inúteis. Portanto, a oferta de
computadores conectados em rede é o primeiro passo, mas
19 não é o suficiente para se realizar a pretensa inclusão digital.

Elizabeth Rondelli. *Revista I-Coletiva*, 24/6/2003 (com adaptações).

Acerca da organização das idéias e das estruturas lingüísticas no
texto acima, julgue os itens subseqüentes.

- 8 Do ponto de vista da construção textual, a expressão
“alfabetização digital” (l.1-2) e o segmento “a aprendizagem
necessária ao indivíduo (...) processos” (l.2-4) estabelecem
uma relação semântica de identificação com “Inclusão
digital” (l.1).
- 9 A expressão “ou seja” (l.2) introduz uma idéia retificadora
do que foi dito na oração anterior.
- 10 As palavras “conteúdos” e “inúteis” são acentuadas com
base na mesma regra de acentuação gráfica.
- 11 A conjunção “Mas” (l.6) inicia um período cujo sentido
aponta para a insuficiência dos suportes técnicos como
recursos capazes de promover o aprendizado na educação,
no trabalho e nos meios de circulação das informações.
- 12 Mantêm-se a correção gramatical e o sentido original do
texto, se o vocábulo “às” (l.7) for substituído por a.
- 13 Na linha 15, o vocábulo “que” tem como referente semântico
o termo “sabedoria”.

1 Sendo a informação um bem que agrega valor a uma
empresa ou a um indivíduo, é preciso fazer uso de recursos
de tecnologia da informação de maneira apropriada, ou seja,
4 é preciso utilizar ferramentas, sistemas ou outros meios que
façam das informações um diferencial competitivo.

Além disso, é necessário buscar soluções que
7 tragam bons resultados, mas que tenham o menor custo
possível. A questão é que não existe fórmula mágica para se
determinar como utilizar da melhor maneira as informações.
10 Tudo depende da cultura, do mercado, do segmento e de
outros aspectos de uma empresa. As escolhas precisam ser
bem feitas. Do contrário, gastos desnecessários ou, ainda,
13 perda de desempenho podem ocorrer. Por exemplo, se uma
empresa renova sua base de computadores comprando
máquinas com processadores velozes, muita memória e placa
de vídeo 3D para serem utilizadas por empregados que
16 apenas precisam acessar a Internet ou trabalhar com pacotes
de escritório, a companhia faz gastos desnecessários.

19 Comprar máquinas de boa qualidade não significa
comprar as mais caras, mas aquelas que possuam os recursos
necessários. Por outro lado, imagine que uma empresa tenha
22 compre computadores com vídeo integrado à placa-mãe e
monitor de 15 polegadas para profissionais que trabalham
com Autocad. Para estes, o ideal é comprar computadores
25 que suportem aplicações pesadas e um monitor de, pelo
menos, 17 polegadas. Máquinas mais baratas certamente
conseguiriam rodar o Autocad, porém com lentidão, e o
28 monitor com área de visão menor daria mais trabalho aos
profissionais. Nesse caso, a aquisição das máquinas tem
reflexo direto no desempenho dos empregados. Por isso, é
31 preciso saber quais as necessidades de cada usuário.

Emerson Alecrim. Internet: <www.infowester.com> (com adaptações).

De acordo com as idéias e os aspectos sintático-semânticos do
texto acima, julgue os próximos itens.

- 14 Nas linhas 1 e 2, a conjunção “ou” liga dois termos que se
alternam e não se excluem: “a uma empresa” e “a um
indivíduo”.
- 15 Segundo o texto, as empresas, na busca de soluções para
seus problemas, precisam em primeiro lugar usar recursos
de tecnologia da informação que tenham o menor custo.
- 16 Se determinada empresa adquirir equipamentos e
suprimentos de informática sem levar em consideração o
melhor aproveitamento desses recursos por seus
empregados, ela, necessariamente, terá gastos e,
conseqüentemente, haverá perda de desempenho.
- 17 As expressões “de boa qualidade” (l.19) e “que possuam os
recursos necessários” (l.20-21) se referem a “máquinas”
(l.19).
- 18 Com o emprego das formas verbais “conseguiriam” (l.27) e
“daria” (l.28), sugere-se, no texto, que certamente há
problemas no uso de equipamentos de informática mais
baratos que não atendam às necessidades de trabalho dos
usuários.

Julgue os itens seguintes, referentes a redação de
correspondências oficiais.

- 19 O memorando é a comunicação feita apenas entre unidades
administrativas de mesmo órgão que estejam
hierarquicamente no mesmo nível.
- 20 O vocativo a ser empregado em comunicações dirigidas aos
chefes dos poderes da República é Ilustríssimo Senhor.

This text refers to items 21 through 30

1 When we think of the people who make our lives
miserable by hacking into computers, or spreading malicious
viruses, most of us imagine an unpopular teenage boy,
4 brilliant but geeky, venting his frustrations* from the safety
of a suburban bedroom.

Actually, these stereotypes are just that —
7 stereotypes — according to Sarah Gordon, an expert in
computer viruses and security technology, and a Senior
Research Fellow with Symantec Security Response. Since
10 1992, Gordon has studied the psychology of virus writers.
“A hacker or a virus writer is just as likely to be the guy next
door to you,” she says, “or the kid at the checkout line
13 bagging your groceries. Your average hacker is not
necessarily some Goth type dressed entirely in black and
sporting a nose ring: she may very well be a 50-year-old
16 female”.

The virus writers Gordon has come to know have
varied backgrounds; while predominately male, some are
19 female. Some are solidly academic, while others are athletic.
Many have friendship with members of the opposite sex,
good relationships with their parents and families; most are
22 popular with their peers. They don’t spend all their time in
the basement. One virus writer volunteers in his local library,
working with elderly people. One of them is a poet and a
25 musician, another is an electrical engineer, and others work
for a university quantum physics department.

Hackers and virus writers are actually very different,
28 distinct populations. “Hackers tend to have a more thorough
knowledge of systems and a more highly developed skill
set,” Gordon says, “whereas virus writers generally take a
shallower approach to what they’re doing.” Hackers tend to
31 have a much deeper knowledge of individual applications
and are still regarded as being somewhat “sexy” in today’s
34 counterculture, while virus writing is looked down upon,
mostly for its random damage and lack of required skill.

* **venting his frustrations** – getting rid of feelings of anger or resentment.

Neil Anderson. *Active skills for reading: Book 4*.
Thomson/Heinle, 2002, p. 17 (with adaptations).

Based on the text above, judge the following items.

- 21 Research confirms common belief: most hackers are teenage boys.
- 22 Sarah Gordon’s research main focus was on stereotypes.
- 23 Sarah Gordon has been studying virus writers’ behavior for more than a decade.
- 24 Your neighbor could very well be a hacker.
- 25 An average hacker would never be a female in her fifties.
- 26 There seems to be nothing in particular that could easily identify a virus writer.
- 27 The virus writers may spend some time in the basement.
- 28 Virus writers know more about computers than hackers.

In the text,

- 29 “parents” (l.21) refers to **mother and father**.
- 30 “volunteers” (l.23) is a **noun**.

Risk identification is a systematic attempt to specify threats to a project plan (estimates, schedule, resource loading, etc.). By identifying known and predictable risks, the project manager takes a first step toward avoiding them when possible and controlling them when necessary.

There are two distinct types of risks for each of the categories: generic risks and product-specific risks. Generic risks are a potential threat to every software project. Product-specific risks can only be identified by those with a clear understanding of the technology, the people, and the environment that is specific to the project at hand. To identify product-specific risks, the project plan and the software statement of scope are examined and an answer to the following question is developed: what special characteristics of this product may threaten our project plan?

R. S. Pressman. *Software engineering – A practitioner’s approach*.
4th Edition, The McGraw-Hill Companies, Inc., 1997 (with adaptations).

From the text above, it can be concluded that

- 31 the identification of risks is a hazardous attempt to prevent menaces to the project plan.
- 32 systematic identification helps to avoid foreseen risks.
- 33 the project manager is in charge of controlling and forestalling risks.
- 34 generic risks and product-specific risks both refer to software projects.
- 35 the project plan and the software statement of scope are product-specific risks.

In a small software development project a single person can analyze requirements, perform design, generate code, and conduct tests. As the size of a project increases, more people must become involved — we can rarely afford the luxury of approaching a ten person-year effort with one person working for ten years!

There is a common myth that is still believed by many managers who are responsible for software development effort: “if we fall behind schedule, we can always add more programmers and catch up later in the project”.

Idem, ibidem (with adaptations).

Based on the text above, judge the following items.

- 36 Small software projects usually require just one person to perform different tasks.
- 37 The bigger the project, the fewer people are demanded.
- 38 We can often afford to have a ten person-year effort or one person working for ten years when developing a software project.
- 39 A lot of project managers tend to believe in the same myth.
- 40 To update a software project is just a matter of hiring more people.

CONHECIMENTOS ESPECÍFICOS

RASCUNHO

Um fabricante de impressoras possui três fornecedores — I, II e III — de um certo circuito eletrônico. Para a produção de um lote de 100 impressoras, a fábrica dispõe de 50, 30 e 20 circuitos fornecidos, respectivamente, por I, II e III. As probabilidades de que um circuito fornecido por I, II ou III apresente defeito são, respectivamente, iguais a 0,01, 0,03 e 0,05. Depois da produção do lote, m impressoras serão selecionadas aleatoriamente para testes de qualidade. Um indicador de qualidade da empresa é a razão $f = n/m$, em que n é o número observado de impressoras com defeitos no circuito.

Considerando as informações acima, julgue os itens a seguir.

- 41 Dos 20 circuitos fornecidos por III, a probabilidade de que apenas um deles apresente defeito é igual a $\frac{1}{20}$.
- 42 Dos circuitos disponíveis no estoque, o número esperado de circuitos defeituosos é menor ou igual a 2.
- 43 A probabilidade de uma das impressoras selecionadas ao acaso do lote produzido apresentar defeito no referido circuito é igual ou superior a 0,03.
- 44 Como a amostragem não é estratificada segundo o fornecedor, a razão f é um estimador tendencioso para a proporção de impressoras com defeitos no mencionado circuito.
- 45 A razão f é um estimador cujo erro padrão é maior ou igual a $\frac{0,1}{\sqrt{m}}$.
- 46 Entre as m impressoras testadas, o número esperado de impressoras que apresentarão defeitos no circuito é igual a $f \times m$.
- 47 Dado que uma impressora testada tenha apresentado defeito no circuito, a probabilidade de que esse circuito tenha sido fornecido por I é maior ou igual a 0,17.
- 48 Suponha que 3 impressoras serão testadas. Nessa situação, a moda de f é igual a zero.
- 49 A probabilidade de n ser igual a m é nula.

Considere que o tempo T , em segundos, que cada cliente aguarda em uma fila seja uma variável aleatória cuja função de distribuição acumulada é $F(t) = 0$, se $t < 0$, e $F(t) > 0$, se $t \geq 0$. Considere ainda que a média, a mediana e o desvio-padrão de T sejam, respectivamente, iguais a 30, 40 e 20 segundos. Nessa situação, julgue os itens a seguir.

- 50 Pela desigualdade de Markov, é correto afirmar que $F(90) \geq \frac{2}{3}$.
- 51 Pela desigualdade de Chebychev, espera-se que até 25% dos clientes aguardarão pelo menos 70 segundos na fila.
- 52 Em um grupo de 100 clientes, 50 aguardarão pelo menos 40 segundos na fila.
- 53 O coeficiente de variação da distribuição do tempo T é igual a 1,5.
- 54 A distribuição exponencial é um candidato para descrever a distribuição do tempo T . Nesse caso, a função de distribuição acumulada seria $F(t) = 0$, se $t < 0$, e $F(t) = 1 - e^{-\frac{t}{30}}$, se $t \geq 0$.

Em uma pesquisa de satisfação pelos serviços realizados em 2000, entrevistaram 400 usuários do setor público e 100 usuários do setor privado. Cada usuário entrevistado foi selecionado, ao acaso, em uma grande população de usuários. Nessa população, 60% dos usuários são do setor público e os restantes, do setor privado. Os resultados da pesquisa mostram que 80% dos usuários do setor público estão plenamente satisfeitos com os serviços. Dos usuários do setor privado, 70% estão plenamente satisfeitos com os serviços. Os testes de hipóteses apresentados na tabela a seguir foram considerados. Na tabela, P é o percentual populacional dos usuários do setor público que estão plenamente satisfeitos com os serviços e R é o percentual populacional dos usuários do setor privado que estão plenamente satisfeitos com os serviços.

teste	hipótese nula (H_0)	hipótese alternativa (H_A)
A	$P \geq 90\%$	$P < 90\%$
B	$R = 80\%$	$R \neq 80\%$
C	$P = R$	$P \neq R$
D	$P = 90\%$ e $R = 80\%$	$P \neq 90\%$ e(ou) $R \neq 80\%$
E	a satisfação não depende do usuário (público/privado)	a satisfação depende do usuário (público/privado)
F	as distribuições da satisfação são as mesmas nos dois grupos de clientes	as distribuições da satisfação não são as mesmas nos dois grupos de clientes

Considerando essas informações e os dados da tabela, disponível ao final das provas objetivas, de distribuição normal padrão, julgue os itens a seguir.

- 55 O intervalo $80\% \pm 5,16\%$ é um intervalo com pelo menos 80% de confiança para o percentual populacional dos usuários do setor público que estão plenamente satisfeitos com os serviços.
- 56 Na amostra, 75% dos usuários, de um modo geral, estão plenamente satisfeitos com os serviços.
- 57 Se a amostragem foi estratificada, em que cada grupo de usuários (público/privado) constitui um estrato, então a estimativa do percentual populacional de usuários que estão plenamente satisfeitos com os serviços é igual a 76%.
- 58 Se a amostragem foi aleatória simples, então a estimativa do percentual populacional de usuários que estão plenamente satisfeitos com os serviços é igual a 74%.
- 59 Com 95% de confiança, a margem de erro da pesquisa é igual a 5%.
- 60 A hipótese nula do teste A é rejeitada ao nível de significância de 1%.
- 61 O nível descritivo para o teste B (ou P -valor) é igual a 0,0124.
- 62 A potência (ou poder) do teste B ao nível de significância de 4,56%, quando o percentual populacional dos usuários do setor privado que estão plenamente satisfeitos com os serviços é igual a 88%, é maior ou igual a 0,5.
- 63 O erro do tipo II torna-se menos freqüente à medida que o tamanho da amostra aumenta.
- 64 A probabilidade de ocorrência de um erro do tipo I depende da especificação da hipótese nula e é conhecida como função característica de operação do teste.
- 65 Em um teste de hipóteses, a probabilidade de se cometer o erro do tipo I ou erro do tipo II é igual a 1.

- 66 A estatística qui-quadrado do teste E é igual a $\frac{2.000}{429}$.
- 67 A hipótese nula do teste C é rejeitada ao nível de significância de 5%.
- 68 A estatística qui-quadrado do teste D é igual a $\frac{1.825}{36}$.
- 69 O teste qui-quadrado de homogeneidade (teste F) é equivalente aos testes C e E.
- 70 Considere que uma nova pesquisa de satisfação dos serviços deva ser realizada entrevistando-se, desta vez, o total de 440 usuários. Nessa situação, caso $P = 80\%$ e $R = 50\%$, então, pela alocação ótima de Neyman, deverão ser entrevistados 240 usuários do setor público.

RASCUNHO

projeto i	preço (Y_i) (em R\$ mil)	número total de horas (X_i)
1	3	7
2	4	9
3	9	10
4	8	12
5	10	15
6	9	15
7	12	16
8	12	16
9	15	30
10	16	30
11	20	40
12	38	100

RASCUNHO

O quadro acima mostra os preços cobrados em função do número de horas trabalhadas em 12 projetos realizados por uma empresa de consultoria. Considere o modelo de regressão linear simples na forma $Y_i = \alpha + \beta X_i + \varepsilon_i$, em que ε_i é o erro aleatório. Algumas estatísticas são dadas nas expressões a seguir.

$$\sum_{i=1}^{12} \frac{Y_i}{12} = 13; \quad \sum_{i=1}^{12} \frac{X_i}{12} = 25; \quad \sum_{i=1}^{12} Y_i X_i = 6.442;$$

$$\sum_{i=1}^{12} Y_i^2 = 2.964; \quad \sum_{i=1}^{12} X_i^2 = 14.736.$$

Considerando as informações acima, julgue os itens a seguir.

- 71 A mediana amostral do preço é um valor entre R\$ 10 mil e R\$ 12 mil.
- 72 O primeiro quartil da distribuição dos preços é um valor entre R\$ 8 mil e R\$ 9 mil.
- 73 A moda da distribuição dos preços é igual a R\$ 10,5 mil.
- 74 A variância amostral dos preços é um valor menor ou igual a 90.
- 75 A estimativa de mínimos quadrados para β é $\frac{6.442}{14.736}$.
- 76 Pelo menos 90% da variação total dos preços é explicada pelo número total de horas.
- 77 Dado um projeto com $X = 20$ horas, a estimativa do preço médio é maior ou igual a R\$ 13 mil.
- 78 A estimativa da variância do erro aleatório é maior ou igual a 4.
- 79 Considere que, para avaliar se o 12.º par de observações é um ponto de alavanca, um analista retira esse par do conjunto de dados e o modelo é ajustado novamente. Nesse caso, as medidas de influência C de Cook e C_p de Mallow baseiam-se na comparação entre os resultados do modelo inicial com o modelo ajustado sem o 12.º par de observações.
- 80 No modelo de regressão linear, é correto afirmar que $\hat{\alpha} + 25\hat{\beta} = 13$, em que $\hat{\alpha}$ e $\hat{\beta}$ são, respectivamente, as estimativas de mínimos quadrados para α e β .

Considere que o desempenho de um tipo de sistema seja uma variável aleatória X cuja função de distribuição é dada por $F(x) = \frac{1}{1 + \exp(-(x-\theta)/\lambda)}$, em que $-\infty < \theta < +\infty$ e $\lambda > 0$ são os parâmetros do modelo. Considere ainda que $D = \{x_1, \dots, x_n\}$ seja um conjunto de dados cujos elementos são n realizações independentes de X , e que Q_1 , Q_2 e Q_3 são, respectivamente, o primeiro, a mediana e o terceiro quartil do conjunto de dados D . Com relação a essas informações, julgue os itens a seguir.

- 81 Q_2 é uma estimativa para θ .
- 82 A média $\frac{Q_1 + Q_3}{2}$ é uma estimativa para θ .
- 83 A média aritmética $\sum_{i=1}^n \frac{x_i}{n}$ é a estimativa não tendenciosa para θ .
- 84 A estimativa de mínima variância para θ é a média aritmética $\sum_{i=1}^n \frac{x_i}{n}$.
- 85 O parâmetro λ pode ser estimado em função do intervalo entre quartis $Q_3 - Q_1$.
- 86 A média amostral $\sum_{i=1}^n \frac{x_i}{n}$ segue uma distribuição aproximadamente normal.
- 87 Uma estimativa de momentos para λ pode ser obtida em função da variância amostral.

Os clientes que utilizam um sistema podem ser classificados em três estados: -1 , 0 e 1 . Essa classificação é uma variável aleatória X cuja função de distribuição de probabilidade é definida como:

$$P(x=k) = \left(\frac{\theta}{2}\right)^{|k|} (1-\theta)^{1-|k|}, \text{ em que } k = -1, 0 \text{ ou } 1 \text{ e } 0 < \theta < 1.$$

Considerando uma seqüência de variáveis aleatórias independentes X_1, \dots, X_n , identicamente distribuídas como X , julgue os itens a seguir.

- 88 Não há um estimador de momentos para θ .
- 89 A média aritmética $\sum_{i=1}^n \frac{x_i}{n}$ é o estimador de máxima verossimilhança para θ .
- 90 Se $\hat{\theta}$ é o estimador de máxima verossimilhança para θ , então $\ell_n\left(\frac{1-\hat{\theta}}{\hat{\theta}}\right)$ é o estimador de máxima verossimilhança para $\ell_n\left(\frac{1-\theta}{\theta}\right)$.
- 91 A média $\sum_{i=1}^n \frac{X_i^2}{n}$ é um estimador não tendencioso para θ .
- 92 A média aritmética $\sum_{i=1}^n \frac{x_i}{n}$ é o estimador de mínimos quadrados para θ .
- 93 Pela lei dos grandes números, $\sum_{i=1}^n \frac{x_i}{n}$ tende para θ à medida que n aumenta.
- 94 Considere a variável aleatória Y_i definida da seguinte maneira: $Y_i = 2$, se $X_i = 1$ e $Y_i = 0$, se $X_i \neq 1$. Nessa situação, $\sum_{i=1}^n \frac{Y_i}{n}$ é um estimador não tendencioso para θ .

Um estudo acerca da satisfação de clientes considerou quatro variáveis X , Y , Z e W . A matriz de covariância entre essas variáveis é $\Sigma = 5 \times \begin{bmatrix} 1 & 0,6 & 0,6 & 0,6 \\ 0,6 & 1 & 0,6 & 0,6 \\ 0,6 & 0,6 & 1 & 0,6 \\ 0,6 & 0,6 & 0,6 & 1 \end{bmatrix}$ e o vetor de médias é

$$m = \begin{bmatrix} 6 \\ 5 \\ 5 \\ 4 \end{bmatrix}.$$

Com base nessas informações e com o auxílio da tabela de distribuição normal padrão, julgue os itens a seguir.

- 95 A primeira componente principal explica 36% da variação total.
- 96 A primeira componente principal P_1 é dada por $P_1 = \frac{1}{2}(X+Y+Z+W)$.
- 97 A covariância entre X e Y é maior que 1,8.
- 98 Mais de 40% da variação total de Z é explicada por W .
- 99 O intervalo de 95% de confiança para o segundo autovalor da matriz de covariância é $3 \pm 1,96\sqrt{\frac{2}{n}}$, em que n é o número de observações do conjunto de dados.
- 100 A distância de Mahalanobis é definida como $m^T \Sigma^{-1} m$, em que m^T é o vetor transposto de m .

RASCUNHO

Mineração de dados é o processo de empregar uma ou mais técnicas de aprendizagem em computador para, automaticamente, analisar e extrair conhecimentos de dados contidos em uma base de dados. Julgue os itens seguintes, que versam sobre mineração de dados e KDD.

- 101** Existem diversas técnicas de mineração de dados, todas elas, usam exclusivamente a aprendizagem indutiva.
- 102** KDD é o processo não-trivial de identificação de padrões em um conjunto de dados. Tais padrões devem possuir as seguintes características: validade (aplicarem-se a novos dados com algum grau de certeza ou probabilidade), novidade (não terem sido detectados por nenhuma outra abordagem), utilidade potencial (poderem ser utilizados para a tomada de decisões úteis, medidas por alguma função) e serem assimiláveis (ao conhecimento humano).
- 103** O processo de KDD é iterativo e cíclico, podendo a saída de uma etapa requerer revisão em etapa anterior. Nesse contexto, a mineração de dados pode ser entendida como uma etapa desse processo.
- 104** Os principais objetivos de alto nível da mineração de dados são previsão e descrição. A previsão envolve o uso de variáveis da base de dados para serem previstos valores desconhecidos ou futuros de variáveis de interesse. Com a descrição, objetiva-se encontrar padrões de descrição dos dados que sejam interpretáveis pelos seres humanos.
- 105** A metodologia para a realização de tarefas de mineração de dados, prescrita pelo modelo de referência proposto pelo Consórcio CRISP/DM, consiste nas seguintes fases: entendimento do negócio (*business understanding*), entendimento dos dados (*data understanding*), pré-processamento dos dados (*data preparation*), modelagem (*modeling*), avaliação dos modelos (*evaluation*) e colocação do modelo selecionado em uso (*deployment*).

Julgue os seguintes itens, relativos à tarefa de classificação, a qual consiste em aprender uma função (construir um modelo) que mapeie um item de dado para uma entre várias classes pré-definidas.

- 106** O aprendizado usado na tarefa de classificação é não-supervisionado.
- 107** Os atributos (variável de classe e variáveis independentes) podem ser categóricos ou numéricos.
- 108** Uma árvore de decisão corresponde a uma estrutura na forma de árvore, em que nós não-terminais representam testes em um ou mais atributos, e nós terminais representam as decisões de saída (classes da variável dependente).
- 109** Uma árvore de decisão pode ser representada na forma de um conjunto de regras inteligíveis, para se melhorar o entendimento de quem analisa os dados.
- 110** Algoritmos de árvore de decisão populares como o C4.5 e o CART examinam um atributo por vez, o que conduz a um particionamento dos dados em regiões de classificação retangulares, que sempre correspondem às distribuições reais das instâncias no espaço de decisão.
- 111** A acurácia de um classificador, em um conjunto de teste, pode ser sumarizada por uma matriz de confusão, em que os valores nas diagonais representam as classificações corretas.

Julgue os seguintes itens quanto a tarefa de associação, a qual consiste na aprendizagem de regras de produção a partir de uma base de dados, considerando a análise de afinidade entre atributos.

- 112** Uma regra de produção é uma regra da forma: se <condição antecedente ocorre> então <condição conseqüente>.
- 113** Uma regra de associação é uma regra de produção cujo conseqüente pode conter múltiplas condições ou relações entre atributos. Um atributo de saída em uma regra de associação pode ser um atributo de entrada em outra regra.
- 114** Algoritmos, como o *apriori*, para descoberta de regras de associação tendem a descobrir muitas regras. A significância de uma regra pode ser inferida por meio do suporte e da confiança. A confiança da regra corresponde ao percentual de exemplos da base de treinamento que levaram a determinada regra.
- 115** O suporte da regra é o percentual de casos que levam a determinada regra, em relação ao número de casos da base de dados que possuem as mesmas evidências de entrada na regra.
- 116** A descoberta de regras de associação corresponde a uma tarefa de treinamento supervisionado.

Julgue os seguintes itens, acerca de técnicas de mineração de dados.

- 117** Agrupamento (*clustering*) não-supervisionado pode ajudar a se detectarem instâncias de dados atípicas na base de dados.
- 118** O *K-means* (K-média) é um algoritmo de agrupamento estatístico que permite particionar-se um conjunto de dados em *K clusters* (grupos) disjuntos. Embora os centros iniciais dos *K clusters* sejam escolhidos aleatoriamente, eles apresentam bom desempenho.
- 119** OLAP (*On-Line Analytical Processing*) é ferramenta adequada para se visualizarem tendências no tempo de dados que representam informação em diversas dimensões.
- 120** *Outliers* são instâncias de dados (observações) atípicas porque se mantêm à distância anormal das outras instâncias em uma amostra aleatória representativa da população de onde as instâncias foram extraídas.

Distribuição Normal Padrão: Valores de p tais que

$$P(0 \leq Z \leq z_c) = p$$

Segunda casa decimal de z_c									
0	1	2	3	4	5	6	7	8	9
0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Tabela gerada pela função DIST.NORMP() do Excel

PROVA DISCURSIVA

- Nesta prova — que vale **30** pontos —, faça o que se pede, usando o espaço indicado no presente caderno para rascunho. Em seguida, transcreva o texto para a **FOLHA DE TEXTO DEFINITIVO DA PROVA DISCURSIVA**, no local apropriado, pois **não serão avaliados fragmentos de texto escritos em locais indevidos**.
- Qualquer fragmento de texto além da extensão máxima de **trinta** linhas será desconsiderado. Será também desconsiderado o texto que não for escrito na **folha de texto definitivo**.
- Na folha **texto definitivo**, identifique-se apenas no cabeçalho da primeira página, pois **não será avaliado** texto que tenha qualquer assinatura ou marca identificadora fora do local apropriado.

Redija um texto dissertativo, abordando a questão da determinação do tamanho da amostra para situações em que haja restrição de custo, limitação de tempo para se efetuar o levantamento e unidades amostrais que se recusam a participar da pesquisa. Apresente alguns exemplos como ilustração.

RASCUNHO

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	

