

Prova Objetiva – Nível Superior

## TECNOLOGISTA - ESTATÍSTICA

Tipo 1 – BRANCA



### SUA PROVA

Além deste caderno de prova, contendo setenta questões objetivas, você receberá do fiscal de sala:

- uma folha destinada às respostas das questões objetivas



### TEMPO

- **4 horas** é o período disponível para a realização da prova, já incluído o tempo para a marcação da folha de respostas da prova objetiva
- **2 horas** após o início da prova é possível retirar-se da sala, sem levar o caderno de prova
- **1 hora** antes do término do período de prova é possível retirar-se da sala levando o caderno de prova



### NÃO SERÁ PERMITIDO

- Qualquer tipo de comunicação entre os candidatos durante a aplicação da prova
- Levantar da cadeira sem autorização do fiscal de sala
- Usar o sanitário ao término da prova, após deixar a sala



### INFORMAÇÕES GERAIS

- As questões objetivas têm cinco alternativas de resposta (A, B, C, D, E) e somente uma delas está correta
- Verifique se seu caderno está completo, sem repetição de questões ou falhas. Caso contrário, notifique imediatamente o fiscal da sala, para que sejam tomadas as devidas providências
- Confira seus dados pessoais, especialmente nome, número de inscrição e documento de identidade e leia atentamente as instruções para preencher a folha de respostas
- Use somente caneta esferográfica, fabricada em material transparente, com tinta preta ou azul
- Assine seu nome apenas nos espaços reservados
- Marque na folha de respostas o campo relativo à confirmação do tipo/cor de prova, conforme o caderno recebido
- O preenchimento das respostas da prova objetiva é de sua responsabilidade e não será permitida a troca da folha de respostas em caso de erro
- Reserve tempo suficiente para o preenchimento de suas respostas. Para fins de avaliação, serão levadas em consideração apenas as marcações realizadas na folha de respostas da prova objetiva, não sendo permitido anotar informações relativas às respostas em qualquer outro meio que não seja o caderno de prova
- A FGV coletará as impressões digitais dos candidatos
- Os candidatos serão submetidos ao sistema de detecção de metais quando do ingresso e da saída de sanitários durante a realização das provas
- Boa Sorte!



## Conhecimentos Básicos

Texto – A eficácia das palavras certas

Havia um cego sentado numa calçada em Paris. A seus pés, um boné e um cartaz em madeira escrito com giz branco gritava: “Por favor, ajude-me. Sou cego”. Um publicitário da área de criação, que passava em frente a ele, parou e viu umas poucas moedas no boné. Sem pedir licença, pegou o cartaz e com o giz escreveu outro conceito. Colocou o pedaço de madeira aos pés do cego e foi embora.

Ao cair da tarde, o publicitário voltou a passar em frente ao cego que pedia esmola. Seu boné, agora, estava cheio de notas e moedas. O cego reconheceu as pegadas do publicitário e perguntou se havia sido ele quem reescrevera o cartaz, sobretudo querendo saber o que ele havia escrito.

O publicitário respondeu: “Nada que não esteja de acordo com o conceito original, mas com outras palavras”. E, sorrindo, continuou o seu caminho. O cego nunca soube o que estava escrito, mas seu novo cartaz dizia: “Hoje é primavera em Paris e eu não posso vê-la”. (*Produção de Texto*, Maria Luíza M. Abaurre e Maria Bernadete M. Abaurre)

**1**

O título dado ao texto:

- (A) resume a história narrada no corpo do texto;
- (B) afirma algo que é contrariado pela narrativa;
- (C) indica um princípio que é demonstrado no texto;
- (D) mostra um pensamento independente do texto;
- (E) denuncia um princípio negativo de convencimento.

**2**

A frase abaixo que exemplifica uma incoerência é:

- (A) “O que vem fácil, vai fácil”. (Geoffrey Chaucer);
- (B) “Se você deseja atingir o ponto mais alto, comece pelo mais baixo”. (Ciro, o Jovem);
- (C) “Perseverança não é uma corrida longa, são muitas corridas curtas, uma após a outra”. (Walter Elliot);
- (D) “Nossa maior glória não é nunca cair, mas sim levantar toda vez que caímos”. (Oliver Goldsmith);
- (E) “Seja breve, não importa quanto tempo isto leve”. (Saul Gorn).

**3**

“Havia um cego sentado numa calçada em Paris. A seus pés, um boné e um cartaz em madeira escrito com giz branco gritava: “Por favor, ajude-me. Sou cego”. Um publicitário da área de criação, que passava em frente a ele, parou e viu umas poucas moedas no boné. Sem pedir licença, pegou o cartaz e com o giz escreveu outro conceito. Colocou o pedaço de madeira aos pés do cego e foi embora”.

O texto pertence ao modo narrativo de organização discursiva, caracterizado pela evolução cronológica das ações. O segmento que comprova essa evolução é:

- (A) “Havia um cego sentado numa calçada em Paris. A seus pés, um boné e um cartaz em madeira escrito com giz branco gritava”;
- (B) “Por favor, ajude-me. Sou cego”;
- (C) “Um publicitário da área de criação, que passava em frente a ele”;
- (D) “parou e viu umas poucas moedas no boné”;
- (E) “Sem pedir licença, pegou o cartaz”.

**4**

A frase abaixo em que o emprego do demonstrativo sublinhado está inadequado é:

- (A) “As capas deste livro que você leva são muito separadas”. (Ambrose Bierce);
- (B) “Quando alguém pergunta a um autor o que este quis dizer, é porque um dos dois é burro”. (Mário Quintana);
- (C) “Claro que a vida é bizarra. O único modo de encarar isso é fazer pipoca e desfrutar o show”. (David Gerrold);
- (D) “Não há nenhum lugar nessa Terra tão distante quanto ontem”. (Robert Nathan);
- (E) “Escritor original não é aquele que não imita ninguém, é aquele que ninguém pode imitar”. (Chateaubriand).

**5**

“Havia um cego sentado numa calçada em Paris. A seus pés, um boné e um cartaz em madeira escrito com giz branco gritava: “Por favor, ajude-me. Sou cego”.

A respeito dos componentes e do sentido desse segmento do texto, é correto afirmar que:

- (A) o cego gritava para ser ouvido pelos transeuntes;
- (B) as palavras gritadas pelo cego tentavam convencer o público que passava;
- (C) as palavras do cartaz apelavam para a caridade religiosa das pessoas;
- (D) a segunda frase do cartaz do cego funciona como consequência da primeira;
- (E) o cartaz “gritava” porque o giz branco se destacava no fundo preto.

**6**

A frase abaixo em que a substituição de uma oração reduzida por uma desenvolvida equivalente é inadequada é:

- (A) “Sou como uma planta do deserto. Uma única gota de orvalho é suficiente para me alimentar”. (Lyonel Brizola) / para que eu me alimente;
- (B) “Você nunca realmente perde até parar de tentar”. (Mike Ditka) / até que pare de tentar;
- (C) “Uma rua sem saída é apenas um bom lugar para se dar a volta”. (Naomi Judd) / para que se dê a volta;
- (D) “Amor é um truque sujo que nos impuseram para obter a continuidade de nossa espécie”. (Somerset Maugham) / para que se obtivesse a continuidade de nossa espécie;
- (E) “O amor é a asa que Deus deu ao homem para voar até Ele”. (Roger Luján) / para que voe até Ele.

**7**

“Por favor, ajude-me. Sou cego”; reescrevendo as duas frases em uma só, de forma correta e respeitando-se o sentido original, a estrutura adequada é:

- (A) Embora seja cego, por favor, ajude-me;
- (B) Me ajude, por favor, pois sou cego;
- (C) Ajude-me já que sou cego, por favor;
- (D) Por favor, ainda que seja cego, ajude-me;
- (E) Ajude-me, por favor, contanto que sou cego.

**8**

“Sem pedir licença, pegou o cartaz e com o giz escreveu outro conceito”; a oração “Sem pedir licença” pode ser adequadamente substituída pela seguinte oração desenvolvida:

- (A) Sem que pedisse licença;
- (B) Sem o pedido de licença;
- (C) Sem que peça licença;
- (D) Sem a petição de licença;
- (E) Sem que havia pedido licença.

**9**

A nova forma do cartaz apela para:

- (A) a intimidação das pessoas pelo constrangimento;
- (B) o racionalismo típico dos franceses;
- (C) a inteligência culta dos transeuntes;
- (D) o sentimentalismo diante da privação do cego;
- (E) a sedução das pessoas pelo orgulho da ajuda prestada.

**10**

A frase abaixo, de Millôr Fernandes, que exemplifica o emprego da vírgula por inserção de um segmento entre sujeito e verbo é:

- (A) “O difícil, quando forem comuns as viagens interplanetárias, será a gente descobrir o planeta em que foram parar as bagagens”;
- (B) “Quando um quer, dois brigam”;
- (C) “Para compreender a situação do Brasil, já ninguém discorda, é necessário um certo distanciamento. Que começa abrindo uma conta numerada na Suíça”;
- (D) “Pouco a pouco o carnaval se transfere para Brasília. Brasília já tem, pelo menos, o maior bloco de sujos”;
- (E) “Mal comparando, Platão era o Pelé da Filosofia”.

**11**

O termo em função adjetiva sublinhado que está substituído por um adjetivo inadequado é:

- (A) “A arte da previsão consiste em antecipar o que irá acontecer e depois explicar por que não aconteceu”. (anônimo) / divinatória;
- (B) “Por mais numerosos que sejam os meandros do rio, ele termina por desembocar no mar”. (Provérbio hindu) / pluviais;
- (C) “A morte nos ensina a transitoriedade de todas as coisas”. (Leo Buscaglia) / universal;
- (D) “Eu não tenho problemas com igrejas, desde que elas não interfiram no trabalho de Deus”. (Brooks Atkinson) / divino;
- (E) “Uma escola de domingo é uma prisão onde as crianças pagam penitência pela consciência pecadora de seus pais”. (H. L. Mencken) / dominical.

**12**

A polissemia – possibilidade de uma palavra ter mais de um sentido – está presente em todas as frases abaixo, EXCETO em:

- (A) Não deixe para amanhã o que pode fazer hoje;
- (B) CBN: a rádio que toca a notícia;
- (C) Na vida tudo é passageiro, menos o motorista;
- (D) Os dentes do pente mordem o couro cabeludo;
- (E) Os surdos da bateria não escutam o próprio barulho.

**13**

A frase em que a redundância está ausente é:

- (A) “Ninguém jamais se afogou em seu próprio suor”. (Ann Landers);
- (B) “Embora ninguém possa voltar atrás e fazer um novo começo, qualquer um pode começar agora e fazer um novo fim”. (Chico Xavier);
- (C) “Espero que sua vida seja tão inteira como duas metades”. (anônimo);
- (D) “Todos os funcionários receberam um prêmio adicional extra por seu desempenho”. (Cartaz em lanchonete);
- (E) “Os cemitérios estão cheios de gente insubstituível”. (Charles De Gaulle).

**14**

A frase em que o vocábulo *mas* tem valor aditivo é:

- (A) “Perseverança não é só bater em porta certa, mas bater até abrir”. (Guy Fawks);
- (B) “Nossa maior glória não é nunca cair, mas sim levantar toda vez que caímos”. (Oliver Goldsmith);
- (C) “Eu caminho devagar, mas nunca caminho para trás”. (Abraham Lincoln);
- (D) “Não podemos fazer tudo imediatamente, mas podemos fazer alguma coisa já”. (Calvin Coolidge);
- (E) “Ele estudava todos os dias do ano, mas isso contribuía para seu progresso”. (Nouailles).

**15**

Em todas as frases abaixo o verbo *ter* foi empregado no lugar de outros com significado mais específico. A frase em que a substituição por esses verbos mais específicos foi feita de forma adequada é:

- (A) “Nunca é tarde para ter uma infância feliz”. (Tom Robbins) / desfrutar de;
- (B) “Você pode aprender muito com crianças. Quanta paciência você tem, por exemplo”. (Franklin P. Jones) / você oferece;
- (C) “O maior recurso natural que qualquer país pode ter são suas crianças”. (Danny Kaye) / usar;
- (D) “Acreditar que basta ter filhos para ser pai é tão absurdo quanto acreditar que basta ter instrumentos para ser um músico”. (Mansour Challita) / originar;
- (E) “A família é como a varíola: a gente tem quando criança e fica marcado para o resto da vida”. (Sartre) / sofre.

## READ TEXT I AND ANSWER QUESTIONS 16 TO 20

## TEXT I

**Will computers ever truly understand what we're saying?**

Date: January 11, 2016

Source University of California - Berkeley

**Summary:**

*If you think computers are quickly approaching true human communication, think again. Computers like Siri often get confused because they judge meaning by looking at a word's statistical regularity. This is unlike humans, for whom context is more important than the word or signal, according to a researcher who invented a communication game allowing only nonverbal cues, and used it to pinpoint regions of the brain where mutual understanding takes place.*

From Apple's Siri to Honda's robot Asimo, machines seem to be getting better and better at communicating with humans. But some neuroscientists caution that today's computers will never truly understand what we're saying because they do not take into account the context of a conversation the way people do.

Specifically, says University of California, Berkeley, postdoctoral fellow Arjen Stolk and his Dutch colleagues, machines don't develop a shared understanding of the people, place and situation - often including a long social history - that is key to human communication. Without such common ground, a computer cannot help but be confused.

"People tend to think of communication as an exchange of linguistic signs or gestures, forgetting that much of communication is about the social context, about who you are communicating with," Stolk said.

The word "bank," for example, would be interpreted one way if you're holding a credit card but a different way if you're holding a fishing pole. Without context, making a "V" with two fingers could mean victory, the number two, or "these are the two fingers I broke."

"All these subtleties are quite crucial to understanding one another," Stolk said, perhaps more so than the words and signals that computers and many neuroscientists focus on as the key to communication. "In fact, we can understand one another without language, without words and signs that already have a shared meaning."

(Adapted from <http://www.sciencedaily.com/releases/2016/01/160111135231.htm>)

**16**

The title of Text I reveals that the author of this text is:

- (A) unsure;
- (B) trustful;
- (C) careless;
- (D) annoyed;
- (E) confident.

**17**

Based on the summary provided for Text I, mark the statements below as TRUE (T) or FALSE (F).

- ( ) Contextual clues are still not accounted for by computers.
- ( ) Computers are unreliable because they focus on language patterns.
- ( ) A game has been invented based on the words people use.

The statements are, respectively:

- (A) F – T – T;
- (B) T – F – T;
- (C) F – F – T;
- (D) F – T – F;
- (E) T – T – F.

**18**

According to the researchers from the University of California, Berkeley:

- (A) words tend to have a single meaning;
- (B) computers can understand people's social history;
- (C) it is easy to understand words even out of context;
- (D) people can communicate without using actual words;
- (E) social context tends to create problems in communication.

**19**

If you are holding a fishing pole, the word "bank" means a:

- (A) safe;
- (B) seat;
- (C) boat;
- (D) building;
- (E) coastline.

**20**

The word "so" in "perhaps more so than the words and signals" is used to refer to something already stated in Text I. In this context, it refers to:

- (A) key;
- (B) crucial;
- (C) subtleties;
- (D) understanding;
- (E) communication.

READ TEXT II AND ANSWER QUESTIONS 21 TO 25:

TEXT II

### The backlash against big data

[...]

Big data refers to the idea that society can do things with a large body of data that weren't possible when working with smaller amounts. The term was originally applied a decade ago to massive datasets from astrophysics, genomics and internet search engines, and to machine-learning systems (for voice-recognition and translation, for example) that work well only when given lots of data to chew on. Now it refers to the application of data-analysis and statistics in new areas, from retailing to human resources. The backlash began in mid-March, prompted by an article in *Science* by David Lazer and others at Harvard and Northeastern University. It showed that a big-data poster-child—Google Flu Trends, a 2009 project which identified flu outbreaks from search queries alone—had overestimated the number of cases for four years running, compared with reported data from the Centres for Disease Control (CDC). This led to a wider attack on the idea of big data.

The criticisms fall into three areas that are not intrinsic to big data per se, but endemic to data analysis, and have some merit. First, there are biases inherent to data that must not be ignored. That is undeniably the case. Second, some proponents of big data have claimed that theory (ie, generalisable models about how the world works) is obsolete. In fact, subject-area knowledge remains necessary even when dealing with large data sets. Third, the risk of spurious correlations—associations that are statistically robust but happen only by chance—increases with more data. Although there are new statistical techniques to identify and banish spurious correlations, such as running many tests against subsets of the data, this will always be a problem.

There is some merit to the naysayers' case, in other words. But these criticisms do not mean that big-data analysis has no merit whatsoever. Even the Harvard researchers who decried big data "hubris" admitted in *Science* that melding Google Flu Trends analysis with CDC's data improved the overall forecast—showing that big data can in fact be a useful tool. And research published in PLOS Computational Biology on April 17th shows it is possible to estimate the prevalence of the flu based on visits to Wikipedia articles related to the illness. Behind the big data backlash is the classic hype cycle, in which a technology's early proponents make overly grandiose claims, people sling arrows when those promises fall flat, but the technology eventually transforms the world, though not necessarily in ways the pundits expected. It happened with the web, and television, radio, motion pictures and the telegraph before it. Now it is simply big data's turn to face the grumblers.

(From <http://www.economist.com/blogs/economist-explains/2014/04/economist-explains-10>)

**21**

The use of the phrase "the backlash" in the title of Text II means the:

- (A) backing of;
- (B) support for;
- (C) decision for;
- (D) resistance to;
- (E) overpowering of.

**22**

The three main arguments against big data raised by Text II in the second paragraph are:

- (A) large numbers; old theories; consistent relations;
- (B) intrinsic partiality; outdated concepts; casual links;
- (C) clear views; updated assumptions; weak associations;
- (D) objective approaches; dated models; genuine connections;
- (E) scientific impartiality; unfounded theories; strong relations.

**23**

The base form, past tense and past participle of the verb "fall" in "The criticisms fall into three areas" are, respectively:

- (A) fall-fell-fell;
- (B) fall-fall-fallen;
- (C) fall-fell-fallen;
- (D) fall-falled-fell;
- (E) fall-felled-falling.

**24**

When Text II mentions "grumblers" in "to face the grumblers", it refers to:

- (A) scientists who use many tests;
- (B) people who murmur complaints;
- (C) those who support large data sets;
- (D) statisticians who promise solid results;
- (E) researchers who work with the internet.

**25**

The phrase "lots of data to chew on" in Text II makes use of figurative language and shares some common characteristics with:

- (A) eating;
- (B) drawing;
- (C) chatting;
- (D) thinking;
- (E) counting.

## Conhecimentos Específicos

**26**

Adotando-se para as estatísticas de posição de uma dada distribuição de frequências as convenções,  $Q_k$  = Quartil de ordem  $k$ ,  $D_k$  = Decil de ordem  $k$ ,  $Qt_k$  = Quintil de ordem  $k$  e  $P_k$  = Percentil de ordem  $k$ , é correto afirmar que:

- (A)  $Q_3 \geq D_6 \geq Qt_4 = P_{80}$ ;
- (B)  $Qt_2 \leq P_{55} \leq D_6 \leq Q_3$ ;
- (C)  $D_9 \geq P_{85} \geq Q_3 = Qt_3$ ;
- (D)  $Q_1 \geq Qt_2 = P_{20} \leq D_3$ ;
- (E)  $D_6 \leq Q_3 = P_{75} \leq Qt_3$ .

27

Com a finalidade de estimar a proporção  $p$  de indivíduos de certa população, com determinado atributo, através da proporção amostral  $\hat{p}$ , é extraída uma amostra de tamanho  $n$ , grande, compatível com um erro amostral de  $\epsilon$  e com um grau de confiança de  $(1-\alpha)$ . Assim, é correto afirmar que:

- (A) uma redução de  $\alpha$  pode ser compensado por uma redução de  $\epsilon$ , compatíveis com o mesmo tamanho de amostra  $n$ ;
- (B) quanto maior a variância verdadeira de  $\hat{p}$ , menor poderá ser a amostra capaz de assegurar a manutenção  $\epsilon$  e  $(1-\alpha)$ ;
- (C) se a variância de  $\hat{p}$  for máxima, o erro  $\epsilon$  for 5% e a amostra tiver tamanho  $n = 1000$ , o nível de significância será de 5%;
- (D) fixos  $\alpha$  e  $p$ , quanto maior a amostra menores serão os ganhos de precisão (redução de  $\epsilon$ ) gerados por seus incrementos;
- (E) fixo  $\epsilon$ , quanto menor a proporção populacional ( $p$ ) e maior o nível de significância ( $\alpha$ ), maior deverá ser a amostra ( $n$ ).

28

As médias harmônica (H), geométrica (G) e aritmética (A) de dois números  $x_1$  e  $x_2$  positivos quaisquer mantem entre si uma relação. Nesse sentido, pode-se garantir que:

- (A)  $A \geq H \geq G$ ;
- (B)  $G = H \cdot A^2$ ;
- (C) as raízes do polinômio  $x^2 - 2Ax + G^2$  são  $x_1$  e  $x_2$ ;
- (D)  $H = G/A$ ;
- (E)  $G \leq A \leq H$ .

29

Sejam  $x_1, x_2, x_3, x_4, \dots, x_n$  determinações distintas de variáveis representativas de uma amostra de tamanho  $n$  com média igual a  $\bar{x}$ . Além disso, sabe-se que sua moda,  $Mo(x_i)$ , é o dobro da média. Sendo  $Var(x_i)$  a variância, é correto afirmar que:

- (A) a distribuição dos  $x_i$  é assimétrica à esquerda;
- (B)  $\sum_{i=1}^n (x_i - (\bar{x} + a))^2 \leq nVar(x_i)$ , sendo  $a$  uma constante;
- (C)  $\sum_{i=1}^n (x_i - (\bar{x} - a))^2 \geq nVar(x_i)$ , sendo  $a$  uma constante;
- (D) a distribuição dos  $x_i$  é assimétrica à direita;
- (E) a mediana dos  $x_i$ ,  $Me(x_i)$  é tal que  $\sum_{i=1}^n (x_i - Me(x_i)) = 0$ .

30

As principais medidas de dispersão utilizadas na estatística são a amplitude (A), a variância (Var), o desvio padrão (DP), o coeficiente de variação (CV) e o desvio-interquartilico (DI).

Sobre o tema, é correto afirmar que:

- (A) as medidas acima listadas têm seus valores dependentes, na íntegra, dos valores da distribuição amostral;
- (B) a variância apresenta a vantagem de ser diretamente comparável com os valores da distribuição;
- (C) é possível afirmar que  $Var(X) \geq DP(X)$ ;
- (D) o desvio-interquartilico é sempre superior ou no mínimo igual à amplitude;
- (E) o coeficiente de variação é uma medida invariante às mudanças de escala.

31

Considere a distribuição de frequências abaixo, apresentada de forma incompleta, sabendo-se não haver valores iguais aos extremos dos intervalos de classe.

Classes	0 - 10	10 - 20	20 - 30	30 - 40
Frequências	3	5	6	Y

Entretanto, antes de se perder o registro de Y, e trabalhando sempre com os dados grupados, a média da distribuição foi calculada, sendo igual a 25. Apesar disso, é correto afirmar que:

- (A) a mediana pertence a 3ª classe de frequências;
- (B) a moda da distribuição de frequências é igual a 25;
- (C) a distribuição de frequências é assimétrica à direita;
- (D) o primeiro quartil ocupa a 1ª classe de frequências;
- (E) a estatística do máximo é igual a 40.

32

A partir de uma amostra de tamanho  $2n+1$ , sendo  $n$  um número inteiro, elaborou-se a distribuição de frequência de tal forma que apenas os dados grupados ficaram disponíveis. Apesar disso, é possível determinar com certeza a classe à qual pertence o valor exato:

- (A) da moda;
- (B) da mediana;
- (C) da média;
- (D) dos quartis;
- (E) do desvio padrão.

33

A elaboração do Plano Amostral de uma pesquisa de campo demanda três especificações: a unidade amostral, a forma de seleção da amostra e o tamanho da amostra. Para seleções de natureza aleatórias, existem algumas alternativas, sobre as quais é correto afirmar que:

- (A) a amostragem estratificada divide a população em grupos que devem ser os mais heterogêneos possíveis para algum atributo da população;
- (B) a amostragem sistemática é realizada em dois estágios, sendo que apenas no segundo se chega a unidade amostral;
- (C) na amostragem estratificada proporcional, o número de elementos da amostra, em cada estrato, é proporcional ao tamanho do estrato;
- (D) na amostra aleatória simples, as probabilidades de seleção dos indivíduos da população podem ser diferentes, porém devem ser conhecidas;
- (E) na amostragem por cotas, a probabilidade de seleção deve ser proporcional ao tamanho da cota, essa determinada de forma aleatória.

**34**

Existem diversas situações que, se observadas na prática, são indicativas da oportunidade de emprego da amostragem por conglomerados. Entre os requisitos e/ou características para sua correta aplicação, está:

- (A) a realização de dois estágios de seleção, com vantagem de que a segunda etapa dispensa a existência de um cadastro;
- (B) a divisão da amostra em grupos de indivíduos para fins de seleção, que devem ser os mais homogêneos possíveis internamente;
- (C) a vantagem de produzir resultados, via de regra, mais precisos do que os obtidos nos métodos estratificado e sistemático;
- (D) que os conglomerados sejam definidos geograficamente, mas sem dispensar a existência de um cadastro das unidades secundárias;
- (E) ser mais econômica do que por outros métodos.

**35**

Existem dois métodos relativamente usuais para identificar, num conjunto de dados, valores não aderentes, denominados outliers. Um deles utiliza uma distribuição teórica, enquanto o outro emprega duas medidas descritivas, uma de posição e outra de dispersão. A propósito:

- (A) o método Box Plot é baseado em estatísticas de ordem;
- (B) as distribuições das estatísticas do máximo e do mínimo são empregadas no método do z score;
- (C) o método Box Plot emprega intervalos de confiança ao redor da média amostral para identificar outliers;
- (D) o método do z score utiliza como régua para a identificação dos outliers o desvio interquartilico;
- (E) o método da distribuição Qui-quadrado é empregado para testar a aderência de valores espúrios.

**36**

Existem duas medidas de probabilidade, frequentemente empregadas, que apropriam dois conceitos bem distintos, o conceito clássico e o conceito frequencial. Entre as principais diferenças está o fato de que:

- (A) o clássico se aplica no caso de experimentos com espaço amostral não enumerável e o conceito frequencial não;
- (B) o segundo pode ser empregado observando-se apenas as condições iniciais do experimento aleatório;
- (C) para espaços amostrais finitos, a medida pelo conceito frequencial é determinada de forma única, com valor fixo;
- (D) mesmo que o experimento seja não aleatório, o conceito frequencial é aplicável, sendo mais preciso quanto maior for a amostra;
- (E) o conceito clássico utiliza, em muitos casos, técnicas de contagem para o cálculo das probabilidades.

**37**

A teoria das probabilidades está apoiada em um conjunto de três axiomas, atribuídos a Kolmogorov. Sendo  $S$  o espaço amostral,  $A$  e  $B$  dois eventos,  $\emptyset$  do vazio e  $P(\cdot)$  a medida de probabilidade, os axiomas estabelecem que:

- (A)  $P(S) = 1$ ,  $P(A) \geq 0$  e  $P(A \cap B) = P(A) \cdot P(B)$ ;
- (B)  $P(\emptyset) = 0$ ,  $P(A) \leq 1$  e  $P(A \cup B) = P(A) + P(B)$ ;
- (C)  $P(A) \geq 0$ ;  $P(A) = 1 - P(A^c)$  e  $P(S) = 1$ ,  $A^c$  = Complementar de  $A$ ;
- (D)  $P(A) \geq 0$ ;  $P(S) = 1$  e  $P(A \cup B) = P(A) + P(B)$  com  $A \cap B = \emptyset$ ;
- (E)  $P(A) \leq 1$ ;  $P(S) = 1$  e  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

**38**

Sejam  $X$ ,  $Z$  e  $W$  variáveis aleatórias tais que,  $\text{Var}(W) = 16$ ,  $\rho(X, Z) = 1$ ,  $\text{Var}(3 \cdot Z + 2 \cdot X) = 144$ ,  $\text{Cov}(W, Z) = 4$  e  $\rho(W, Z) = 0,5$ .

Então a variância de  $X$  é:

- (A) 4;
- (B) 9;
- (C) 16;
- (D) 25;
- (E) 36.

**39**

Considere a variável aleatória bidimensional  $(X, Y)$  cuja função de densidade conjunta é dada por:

$$f_{X,Y}(x, y) = \frac{3}{4} \cdot y \cdot x^2, 0 < x < 2 \text{ e } 0 < y < 1 \text{ e zero caso contrário.}$$

Então:

- (A)  $X$  e  $Y$  não são variáveis aleatórias independentes;
- (B)  $E(X/Y=0,5) = 1,75$  e  $E(Y/X=1) = 2/3$ ;
- (C)  $\rho(X, Y)$  = coeficiente de correlação entre  $X$  e  $Y$  é igual a  $0,8$ ;
- (D)  $P(X > Y) = 19/20$ ;
- (E)  $\text{Var}(X/Y < 0,5) > \text{Var}(X/Y > 0,5)$ .

**40**

Considere os eventos  $A$  e  $B$  quaisquer de um mesmo espaço amostral  $S$  de um experimento aleatório  $\varepsilon$ . Caso  $P(A) = 0,40$  então é possível supor que:

- (A)  $P(B) = 0,7$  e  $P(A \cup B) = 0,8$ , se  $A$  e  $B$  forem independentes;
- (B)  $P(B) = 0,8$ , se  $A$  e  $B$  forem mutuamente exclusivos;
- (C)  $P(B) = 0,6$  e  $P(A \cup B) = 0,76$ , se  $A$  e  $B$  forem independentes;
- (D)  $P(A \cup B) = 0,3$ , se  $A$  e  $B$  forem mutuamente exclusivos;
- (E)  $P(B) = 0,7$  e  $P(A \cap B) = 0,30$ , se  $A$  e  $B$  forem independentes.

**41**

Sobre os eventos mutuamente exclusivos e/ou independentes de um mesmo espaço amostral  $S$ , a partir de um experimento aleatório  $\epsilon$  é correto afirmar que:

- (A) se dois eventos são mutuamente exclusivos, então também serão independentes;
- (B) se dois eventos são independentes, então também serão mutuamente exclusivos;
- (C) um evento qualquer não pode ser independente de si mesmo;
- (D) se dois eventos são mutuamente exclusivos, então certamente não poderão ser independentes;
- (E) se dois eventos são independentes, então é possível que sejam mutuamente exclusivos.

**42**

A presente prova de estatística está sendo aplicada a uma população de candidatos composta por 70% de indivíduos bem preparados, 20% de medianos e 10% de insuficientes. Os mais aptos têm probabilidade de 80% de acertar qualquer questão, sendo essa probabilidade 25% menor no caso dos medianos e outros 50% menor no caso dos insuficientes, com relação aos medianos. Um candidato é escolhido ao acaso. A probabilidade de que ele acerte determinada questão é de:

- (A) 0,34
- (B) 0,54;
- (C) 0,66;
- (D) 0,71;
- (E) 0,83.

**43**

Suponha que, por coincidência, as 12 pessoas que estão numa sala de espera, aguardando por uma chamada, nasceram todas no mês de agosto. Então a probabilidade de que não haja sequer uma coincidência entre os dias do mês de nascimento é de:

- (A)  $\frac{31!}{19!(31)^{12}}$ ;
- (B)  $\frac{31!}{(31)^{12}}$ ;
- (C)  $\frac{31!}{12!(31)^{12}}$ ;
- (D)  $\frac{31!}{19!12!}$ ;
- (E)  $\frac{30!}{12!(30)^{12}}$ .

**44**

Um experimento é realizado a partir de três urnas, contendo bolas brancas e pretas com a seguinte composição:

Urna I = 3 Brancas e 4 Pretas

Urna II = 5 Brancas e 3 Pretas

Urna III = 2 Brancas e 3 Pretas

A realização consiste em, a partir da Urna III, sortear uma bola e colocar na Urna I, caso seja branca, ou na Urna II caso seja preta. Em seguida é escolhida, aleatoriamente, uma bola da urna que foi abastecida. Se ao final do experimento a bola sorteada foi branca, a probabilidade de que a primeira bola sorteada tenha sido preta é igual a:

- (A)  $\frac{3}{5}$ ;
- (B)  $\frac{3}{8}$ ;
- (C)  $\frac{5}{8}$ ;
- (D)  $\frac{7}{15}$ ;
- (E)  $\frac{8}{15}$ .

**45**

A microcefalia tem, em síntese, duas causas, a contaminação pelo zika vírus, transmitido pelo *aedes aegypti*, além de um conjunto de outras origens. Entre a população feminina de grávidas, sabe-se que 5% foi picada pelo mosquito, enquanto 10% está sujeita as outras origens, não havendo interseção entre esses dois grupos. O desenvolvimento da doença não é certo, acontecendo em 80% das picadas do mosquito e em 30% na eventualidade das outras origens.

Se uma mulher, sorteada aleatoriamente entre as grávidas, carrega um feto que apresenta o problema, a probabilidade de que ela NÃO tenha sido picada pelo mosquito é de:

- (A) 3%;
- (B) 7%;
- (C)  $\frac{3}{4}$ ;
- (D)  $\frac{3}{7}$ ;
- (E)  $\frac{4}{7}$ .

**46**

Em um jogo de azar disputado por dois indivíduos, através de uma sequência de rodadas, vencerá aquele que ganhar, antes do que o outro, uma das rodadas. A chance de que cada um vença qualquer rodada é de  $\frac{2}{9}$  e  $\frac{1}{3}$ . Assim a probabilidade de que cada jogador vença o jogo, são respectivamente:

- (A)  $\left(\frac{2}{9}\right)$  e  $\left(\frac{1}{3}\right)$ ;
- (B)  $\left(\frac{4}{9}\right)$  e  $\left(\frac{5}{9}\right)$ ;
- (C)  $\left(\frac{2}{5}\right)$  e  $\left(\frac{3}{5}\right)$ ;
- (D)  $\left(\frac{5}{9}\right)$  e  $\left(\frac{4}{9}\right)$ ;
- (E)  $\left(\frac{3}{5}\right)$  e  $\left(\frac{2}{5}\right)$ .

**47**

Suponha que o número de chegada de pessoas a uma fila é uma variável aleatória, cuja média, proporcional ao tempo, é de seis pessoas por hora. As chegadas são independentes e a probabilidade de que duas pessoas cheguem à fila num mesmo segundo é nula. Logo, uma aproximação para a probabilidade de que duas pessoas entrem na fila no período de meia hora é:

- (A)  $(\frac{1}{3})^2$  ;
- (B)  $6 \cdot e^{-4,5}$ ;
- (C)  $(\frac{1}{3})^2 (\frac{2}{3})^4$  ;
- (D)  $3 \cdot e^{-6}$ ;
- (E)  $4,5 \cdot e^{-3}$ .

**48**

Seja X uma variável aleatória mista com função densidade de probabilidade dada por:

$f_X(x) = \frac{1}{x^2}$  para  $1 < x \leq 4$ ,  $P(X = 1) = 0,25$ , sendo igual a zero caso contrário.

Então os valores de  $P(X \leq 2)$  e  $E(X^2)$ , esperança matemática de X ao quadrado, são respectivamente iguais a:

- (A) 0,25 e 2,50;
- (B) 0,50 e 2,50;
- (C) 0,50 e 3,25;
- (D) 0,75 e 2,50;
- (E) 0,75 e 3,25.

**49**

A capacidade de um time de futebol de marcar gols em uma única partida é uma variável aleatória. A tabela a seguir apresenta a probabilidade de certo time marcar um número mínimo (Y) de gols em uma partida:

Y	1	2	3	4	5
Probabilidade(Y)	0,85	0,43	0,16	0,01	0,00

Isso significa que o número médio de gols marcados por esse time em uma única partida de futebol é igual a:

- (A) 1,28;
- (B) 1,45;
- (C) 1,64;
- (D) 1,72;
- (E) 2,23.

**50**

Seja X uma variável aleatória contínua e  $Y = G(X)$  uma função de X tal que, no domínio da  $f_X(x)$ , densidade da X, as derivadas de 1ª e de 2ª ordem da G(X) são estritamente negativas. Considerando,

$f_Y(y)$  = função densidade de probabilidade de Y;

$f_X^{-1}(x)$  = função inversa da densidade de X;

$\frac{\partial f(x)}{\partial x}$  = derivada de f(x) com respeito à x;

$E(X)$  = esperança matemática de X;

$h[f(X)]$  = função composta de f com h.

Então é correto afirmar que:

- (A)  $f_Y(y) = f_X[G^{-1}(y)] \cdot \frac{\partial G^{-1}(y)}{\partial y}$  e  $E(G(X)) \leq G(E(X))$ ;
- (B)  $f_Y(y) = -f_X[G^{-1}(y)] \cdot \frac{\partial G^{-1}(y)}{\partial y}$  e  $E(G(X)) \geq G(E(X))$ ;
- (C)  $f_Y(y) = f_X[G^{-1}(y)] \cdot \frac{\partial G^{-1}(y)}{\partial y}$  e  $E(G(X)) \geq G(E(X))$ ;
- (D)  $f_Y(y) = -f_X[G^{-1}(y)] \cdot \frac{\partial G^{-1}(y)}{\partial y}$  e  $E(G(X)) \leq G(E(X))$ ;
- (E)  $f_Y(y) = -f_X[G^{-1}(y)] \cdot \frac{\partial G^{-1}(y)}{\partial y}$  e  $E(G(X)) = G(E(X))$ .

**51**

De uma população arbitrariamente grande é extraída uma pequena amostra de tamanho  $n = 5$ , com o objetivo de avaliar o apoio a um dirigente político. Se forem verdadeiros os rumores de que tal indivíduo tem o apoio de apenas 10% da população, então a probabilidade de que dois se declarem favoráveis a ele é de:

- (A) 0,729%;
- (B) 1,458%;
- (C) 1,620%.
- (D) 7,290%;
- (E) 14,580%;

**52**

Suponha que uma amostra aleatória de tamanho  $n = 3$  será extraída de uma população cuja variável a ser observada é X, tendo função de densidade teórica  $f_X(x) = 2x$ , para  $0 < x < 1$  e zero caso contrário. A extração é feita com a ajuda de uma tabela de números aleatórios, com valores convertidos aos valores amostrais de X através da transformação integral de  $Y = F_X(x)$ , que é a função distribuição acumulada de X. Se os valores lidos na tabela de aleatórios forem 0,25, 0,49 e 0,81, a média amostral será igual a:

- (A) 0,30;
- (B) 0,49;
- (C) 0,51;
- (D) 0,63;
- (E) 0,70.

**53**

Um fabricante de equipamentos de informática, que conhece a distribuição do tempo de vida útil dos HDs externos, precisa avaliar os gastos com serviços de garantia. Essa distribuição é exponencial com média  $\beta = 15$  anos, sendo que os HDs já vendidos têm, por hipótese, 3 anos de uso, sem apresentar defeitos. Supondo que a garantia é de 12 anos, a probabilidade de que ele tenha que prestar assistência a um determinado HD entre os vendidos é:

- (A)  $1 - e^{-0,6}$ ;
- (B)  $e^{-0,75}$ ;
- (C)  $e^{-0,6}$ ;
- (D)  $1 - e^{-0,75}$ ;
- (E)  $e^{-0,25}$ .

**54**

Suponha que uma amostra de tamanho  $n = 5$  é extraída de uma população Normal, com média desconhecida, obtendo as seguintes observações:

$$X_1 = 3, X_2 = 5, X_3 = 6, X_4 = 9 \text{ e } X_5 = 12$$

São dados ainda os seguintes valores, retirados da tabela da distribuição Qui-Quadrado:

$$P(\chi_4^2 < 5) \cong 0,713 \text{ e } P(\chi_4^2 < 12,5) \cong 0,986$$

$$P(\chi_5^2 > 5) \cong 0,854 \text{ e } P(\chi_5^2 > 12,5) \cong 0,971$$

Se a população tem variância verdadeira  $\sigma^2 = 4$ , em nova amostra ( $n=5$ ), a probabilidade de se observar uma variância amostral maior do que a anterior é de:

- (A) 0,014;
- (B) 0,029;
- (C) 0,146;
- (D) 0,287;
- (E) 0,713.

**55**

A distribuição das alturas dos indivíduos de uma população é aproximadamente Normal, com média 1,70 m e variância 0,01. Adicionalmente, não havendo, na população, pessoas com alturas inferiores a 1,50 m nem superiores a 1,90 m, essa distribuição é truncada nos extremos.

São fornecidas também as seguintes informações:

$$\phi(1) \cong 0,84 \text{ e } \phi(2) \cong 0,98$$

$\phi(z)$  = função distribuição acumulada da Normal Padrão

Então a probabilidade de que um indivíduo da população, sorteado ao acaso, tenha altura entre 1,60 m e 1,80 m é:

- (A)  $\left(\frac{23}{24}\right)$ ;
- (B)  $\left(\frac{21}{24}\right)$ ;
- (C)  $\left(\frac{21}{23}\right)$ ;
- (D)  $\left(\frac{17}{24}\right)$ ;
- (E)  $\left(\frac{17}{23}\right)$ .

**56**

Para estimar, por máxima verossimilhança (MV) ou pelo método dos momentos (MM), o único parâmetro de dada distribuição de probabilidades, seleciona-se uma amostra de tamanho  $n$ .

A função densidade da distribuição é:

$$f_X(x) = \theta x^{\theta-1}, \text{ para } 0 < x < 1 \text{ e zero caso contrário.}$$

Além disso, considere:  $\bar{X}_A = \frac{\sum_{i=1}^n X_i}{n}$  e  $\bar{X}_G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots x_n}$

Então, os estimadores de MV e de MM (com base na média da distribuição) para  $\theta$  são, respectivamente:

- (A)  $\hat{\theta}_{MV} = \ln \bar{X}_G$  e  $\hat{\theta}_{MM} = \frac{1}{\bar{X}_A}$
- (B)  $\hat{\theta}_{MV} = \frac{\ln \bar{X}_G}{n}$  e  $\hat{\theta}_{MM} = \frac{1 - \bar{X}_A}{\bar{X}_A}$
- (C)  $\hat{\theta}_{MV} = \frac{\ln \bar{X}_G}{n}$  e  $\hat{\theta}_{MM} = \frac{1}{\bar{X}_A}$
- (D)  $\hat{\theta}_{MV} = \frac{-1}{\ln \bar{X}_G}$  e  $\hat{\theta}_{MM} = \frac{\bar{X}_A}{1 - \bar{X}_A}$
- (E)  $\hat{\theta}_{MV} = \frac{\ln \bar{X}_G}{\bar{X}_A}$  e  $\hat{\theta}_{MM} = \frac{1 - \bar{X}_G}{\bar{X}_G}$

**57**

Considere os estimadores a seguir, tendo em vista a média populacional  $\mu$ , a partir de uma amostra de tamanho  $n$ .

$$\hat{\mu} = \frac{\sum_{i=1}^n X_i}{(n-1)} \text{ e } \tilde{\mu} = \frac{\sum_{i=2}^n X_i}{(n-1)} - \frac{X_1}{n}$$

Se a variância populacional é finita, sobre as propriedades de  $\hat{\mu}$  e  $\tilde{\mu}$  é correto afirmar que:

- (A) apenas um dos estimadores propostos é assintoticamente não tendencioso;
- (B) o estimador  $\hat{\mu}$  subestima enquanto  $\tilde{\mu}$  superestima o valor verdadeiro da média populacional  $\mu$ ;
- (C) a variância do estimador  $\hat{\mu}$  é maior do que a variância do estimador  $\tilde{\mu}$ ;
- (D) o estimador  $\tilde{\mu}$  é mais eficiente do que o estimador  $\hat{\mu}$ ;
- (E) em termos absolutos, a tendenciosidade de  $\hat{\mu}$  é menor do que a tendenciosidade de  $\tilde{\mu}$ .

**58**

Com o objetivo de estimar, por intervalo, a verdadeira média populacional de uma distribuição, é extraída uma amostra aleatória de tamanho  $n = 26$ . Sendo a variância desconhecida, calcula-se o valor de  $\hat{s}^2 = 100$ , além da média amostral de  $\bar{X} = 8$ . O grau de confiança pretendido é de 95%. Somam-se a todas essas informações os valores tabulados:

$$\phi(1,65) \cong 0,95 \quad \phi(1,96) \cong 0,975 \quad T_{25}(1,71) \cong 0,95$$

$$T_{26}(1,70) \cong 0,95 \quad T_{25}(2,06) \cong 0,975 \quad T_{26}(2,05) \cong 0,975$$

Onde,  $\hat{s}^2$  = estimador não-viesado da variância populacional;

$\phi(z)$  = fç distribuição acumulada da Normal-padrão;

$T_n(t)$  = fç distribuição acumulada da T-Student com  $n$  graus de liberdade.

Então os limites do intervalo de confiança desejado são:

- (A) 3,86 e 12,14;
- (B) 3,88 e 12,12;
- (C) 4,30 e 11,70;
- (D) 4,58 e 11,42;
- (E) 4,60 e 11,40.

**59**

Um teste de hipótese será feito com base numa distribuição normal, com média desconhecida e variância  $\sigma^2 = 64$ . Uma amostra de tamanho  $n = 16$  é extraída e sua média calculada, sendo  $\bar{X} = 7$ . O conjunto de hipóteses a ser testado é:

$$H_0: \mu \leq 5$$

$$H_a: \mu > 5$$

Considere ainda que a região crítica do teste é  $RC = (9, +\infty)$  e que, caso  $H_0$  seja falsa, o  $\mu$  verdadeiro seria igual a 8.

Além disso, são fornecidos os seguintes dados sobre a função distribuição acumulada da normal-padrão.

$$\Phi(0,5) \cong 0,69 \quad \Phi(1) \cong 0,84 \quad \Phi(1,5) \cong 0,93 \quad \Phi(2) \cong 0,98$$

Logo, as probabilidades dos erros do Tipo I, do Tipo II e do p-valor (bilateral) do teste são, respectivamente, iguais a:

- (A) 0,02, 0,69 e 0,32;
- (B) 0,07, 0,84 e 0,31;
- (C) 0,04, 0,16 e 0,62;
- (D) 0,02, 0,69 e 0,14;
- (E) 0,04, 0,84 e 0,32.

**60**

Os testes clássicos de inferência estão baseados na obtenção ou não de evidência estatística contrária à hipótese suposta, previamente, verdadeira. A construção está associada a uma série de conceitos e definições. Entre esses elementos estão:

- (A) a não ocorrência de um evento, supostamente, de alta probabilidade produz evidência à rejeição da hipótese nula;
- (B) a não rejeição da hipótese nula produz evidência estatística associada a um evento de baixa probabilidade;
- (C) a potência do teste depende da escolha do valor da amostra e do nível de significância adotado;
- (D) os erros do Tipo I e Tipo II, que têm probabilidades complementares;
- (E) a estatística do teste deve ser conhecida, além de depender, na sua expressão, do valor do parâmetro a ser testado.

**61**

Os principais métodos para a estimação de parâmetros em modelos de regressão linear são os de Mínimos Quadrados Ordinários (MQO), o do Melhor Estimador Linear Não Tendencioso (BLUE) e o de Máxima Verossimilhança (MV).

Sobre esses métodos, é correto afirmar que:

- (A) o MV tem a vantagem de prover inclusive um estimador não tendencioso para a variância dos termos de erro do modelo;
- (B) os estimadores BLUE mantêm suas propriedades mesmo que a matriz de variância covariância não seja diagonal;
- (C) o MQO consiste, em síntese, em fazer com que os valores da amostra atendam a alguns dos pressupostos no modelo;
- (D) uma das principais vantagens do estimador BLUE é que, por construção, seus estimadores têm propriedades assintóticas;
- (E) o MV é o mais exigente dentre os mencionados, pois requer que se conheça a distribuição dos erros e o valor de seus parâmetros.

**62**

Um economista resolve propor e estimar um modelo de regressão linear simples como forma de estimar o efeito da temperatura sobre o volume de venda de sorvetes. Emprega, para esse fim, a formulação:

$$QS_i = \alpha + \beta \cdot T_i + \varepsilon_i \quad i = 1, 2, 3, 4, \dots, 20.$$

Onde QS é a quantidade de sorvetes (em milhares), T é a temperatura (celsius) e  $\varepsilon$  é o termo de erro do modelo.

Apenas estatísticas descritivas básicas sobre QS e T são dadas, como  $\sigma_T^2 = 64$ ,  $\sigma_{QS}^2 = 100$ ,  $\mu_T = 24$ ,  $\mu_{QS} = 58$  e  $\sigma_{T,QS} = 48$

Onde, variâncias ( $\sigma^2$ ), médias ( $\mu$ ) e covariância ( $\sigma_{T,QS}$ ).

Supondo-se válidos todos os pressupostos clássicos, a partir das informações disponíveis, verifica-se que:

- (A) estimativas para a quantidade de sorvetes podem ser obtidas através da relação estimada  $\widehat{QS}_i = 41,2 + 0,7 \cdot T_i$ ;
- (B) o coeficiente de correlação entre T e QS é igual a 0,75;
- (C) se na amostra à temperatura de 40 graus o consumo foi de 72 mil sorvetes, o resíduo produzido pelo modelo é  $\hat{\varepsilon}(40) = 2$ ;
- (D) a estatística do teste t-Student de significância para o estimador de  $\beta$  é igual a 1;
- (E) o desvio-padrão estimado para os erros amostrais é igual a 6.

**63**

Após estimar um modelo de regressão linear múltipla, por MQO, um economista repara que, por algum motivo, a tabela contendo os resultados da análise da variância ficou incompleta, conforme abaixo:

Fonte	S. Quadrados	G. Liberdade	Q. Médio	F-Snedecor	p-valor
Equação	1800	5	360	4.5	1.05%
Resíduos	Y	X	W		
Modelo	3000	20	Z		

Apesar dos valores acima omitidos, é correto afirmar que:

- (A) a equação de regressão tem cinco variáveis explicativas;
- (B) o coeficiente de determinação  $R^2$  é igual a 0,8;
- (C) ao nível de significância de 2% não se rejeita a hipótese nula de que o modelo explica a variável dependente;
- (D) o tamanho da amostra é  $n = 20$ ;
- (E) a estimativa não tendenciosa da variância dos erros aleatórios do modelo  $\hat{\sigma}_\varepsilon^2$  é igual a 80.

**64**

Um modelo de regressão linear múltipla é estimado por MQO (Mínimos Quadrados Ordinários), conforme a equação:

$$Y_i = \alpha + \beta \cdot X_i + \gamma \cdot W_i + \varepsilon_i$$

As estimativas estão colocadas na tabela abaixo, com algumas omissões:

Parâmetro	Estimativa	Erro-Padrão	t-Student	p-valor
$\alpha$	A	0.75	2	6.39%
$\beta$	1.25	0.5	C	2.45%
$\gamma$	-0.75	B	-1.25	23.05%

Com base nas estatísticas disponíveis e no cálculo dos valores omitidos, é correto afirmar que:

- (A) o efeito fixo estimado pelo modelo é negativo;
- (B) a variância estimada do estimador do parâmetro da variável independente W é 0,60;
- (C) se X = 3 e W = 2 o valor estimado da variável dependente é igual a 2,75;
- (D) ao nível de significância de 10% rejeita-se a hipótese de que o efeito fixo é nulo;
- (E) ao nível de significância de 2% rejeita-se a hipótese nula do teste  $H_0: \beta \leq 0$  contra  $H_a: \beta > 0$ .

**65**

Para estimar a demanda por diárias em hotéis (DH) de um dado destino turístico é elaborado um modelo econométrico, tendo como variáveis explicativas o preço da diária (PD), a renda dos turistas nacionais (RN), a renda dos turistas internacionais (RI), o preço das diárias em destinos alternativos (PA) e uma variável qualitativa que reflete o momento de alta ou baixa estação. A forma funcional do modelo é a seguinte:

$$\ln DH_t = \alpha + \beta \cdot \ln PD_t + \beta^* \cdot (D_t \cdot \ln PD_t) + \gamma \cdot \ln RN_t + \delta \cdot \ln RI_t + \varphi \cdot \ln PA_t + \omega \cdot D_t + \varepsilon_t$$

Onde  $D_t = 1$ , se for alta temporada e  $D_t = 0$ , na baixa.

$\varepsilon_t$  é termo de erro e  $\ln$  é o logaritmo neperiano.

Após a aplicação de MQO foram obtidas as estimativas a seguir.

Parâmetro	$\alpha$	$\beta$	$\beta^*$	$\gamma$	$\delta$	$\varphi$	$\omega$
Estimativa	8,0	-0,7	0,5	0,74	0,81	0,21	2,0

Considerando que os preços e rendas estão em reais (R\$) e que as estimativas acima são significativas, é correto afirmar que:

- (A) tudo o mais constante, o número médio de diárias contratadas na baixa temporada é 20% menor do que na alta temporada;
- (B) variações idênticas na renda dos turistas internacionais e nacionais interna, uma para cima e a outra para baixo, vão provocar uma variação de 7% na quantidade de diárias;
- (C) na alta temporada, para cada 1% de variação no preço das diárias, a quantidade demandada estimada deverá cair 0,2%;
- (D) o impacto de variações na renda dos turistas internacionais sobre a demanda por diárias é estatisticamente superior ao da renda dos turistas nacionais;
- (E) caso o preço das diárias sofra um aumento de 1%, a demanda por diárias estimada deverá recuar 0,7%.

**66**

Em modelos de regressão linear a qualidade das estimativas depende, sobretudo, da verificação de pressupostos sobre o comportamento da componente de erro do modelo. Entre os principais pressupostos, estão o da homocedasticidade e o da ortogonalidade dos erros com relação às variáveis explicativas do modelo. No caso de violação dessas condições, os estimadores de MQO poderão perder, respectivamente, as propriedades da:

- (A) não-tendenciosidade e consistência;
- (B) consistência e suficiência;
- (C) robustez e não-tendenciosidade;
- (D) suficiência e eficiência;
- (E) eficiência e consistência.

**67**

Para modelar o comportamento de determinada proporção é proposto um modelo de regressão com variável dependente do tipo qualitativa. A forma funcional apresentada é:

$$p_i = \frac{1}{1 + e^{-(\alpha + \beta x_i)}} + \varepsilon_i$$

Sobre esse tipo de modelo e formulação, é correto afirmar que:

- (A) por hipótese, a probabilidade segue o padrão de uma curva logit, tomando valores em todo o domínio dos Reais;
- (B) a função adotada irá exigir procedimentos de correção para a autocorrelação serial dos erros;
- (C) a distribuição de probabilidade do termo de erro  $\varepsilon_i$  deve ser suposta Normal, para aplicação de MQO;
- (D) com hipóteses sobre os erros, estimativas podem ser obtidas por meio de um modelo linear para  $\ln\left(\frac{p_i^*}{1-p_i^*}\right)$ , onde  $p_i^*$  é a probabilidade observada;
- (E) por hipótese, a probabilidade segue o padrão de uma curva probit, tomando valores no intervalo (0,1).

**68**

Para explicar o comportamento de um processo estocástico, após uma análise das funções de autocorrelação, o seguinte modelo foi proposto e estimado:

$$y_t = 6 + 0,85y_{t-1} - 0,15y_{t-2} + \hat{\varepsilon}_t$$

Considerando válidas as estimativas, é correto afirmar que:

- (A) o processo estocástico é estacionário de segunda ordem, apresentando um padrão de amortecimento cíclico;
- (B) os autovalores do polinômio característico correspondente são iguais a:  $x_1 = 2,5$  e  $x_2 = 4$ ;
- (C) o processo contém pelo menos uma raiz unitária;
- (D) a esperança matemática incondicional do processo é  $\bar{Y} = 10$ ;
- (E) o processo estocástico  $y_t$  não é invertível.

## 69

A aplicação das técnicas de Box-Jenkins, para reconhecer a forma funcional de um modelo de tempo, com base numa série amostrada, indicou a conveniência de se adotar:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$$

Onde,  $\varepsilon_t$  é um ruído branco.

Se após a estimação dos parâmetros a inferência estatística apontar para a adequação do modelo, será possível afirmar que:

- (A) se  $q = 0$  e as raízes do polinômio  $1 - \alpha_1 L + \dots + \alpha_p L^p$  estiverem fora do círculo unitário o processo é não estacionário;
- (B) se  $p = 0$  e as raízes do polinômio  $1 + \beta_1 L + \dots + \beta_q L^q$  estiverem no interior do círculo unitário o processo será não estacionário;
- (C) se  $z_t = y_t - y_{t-1} = \Delta z_t$ , o processo estocástico  $z_t$  é gerado por um ARIMA (q,1,p);
- (D) se  $p = 0$ ,  $q = +\infty$  e  $\sum_{i=1}^{\infty} \beta_i^2$  tem soma finita o processo é estacionário;
- (E) se  $p = 0$  e as raízes do polinômio  $1 + \beta_1 L + \dots + \beta_q L^q$  estiverem dentro do círculo unitário o processo será invertível.

## 70

Um modelo de regressão linear simples, supondo válidos todos os pressupostos clássicos, é estimado por Mínimos Quadrados Ordinários, obtendo os seguintes resultados:

$$y_t = 24 + 0,94 y_{t-1} + 0,075t + \varepsilon_t$$

(8,41)    (0,08)    (0,02)

$$DW = 0,87 \quad R^2 = 0,93 \quad \tau_{5\%}^{DF} = -2,78 \quad T_{10\%}^{tab} = 2,17 \quad T = 40$$

Onde, DW é o valor observado da Estatística Durbin-Watson

$R^2$  é o Coeficiente de Determinação

$\tau_{5\%}^{DF}$  é o valor tabelado da estatística Dickey-Fuller

$T_{10\%}^{tab}$  é o valor da distribuição acumulada da t-Student

$T$  = tamanho da amostra

Os números entre parênteses, abaixo das estimativas dos parâmetros, são os valores estimados dos erros padrão correspondentes. O tamanho da amostra é  $n = 100$ . Com tais informações, é correto afirmar que:

- (A) a comparação entre os valores de DW e de  $R^2$  é indicativa de que a regressão estimada não deve ser espúria;
- (B) ao testar o parâmetro da variável que captura a tendência não estocástica, observa-se que esse é não significativo, ao nível de significância de 10%;
- (C) ao nível de significância de 5%, através do teste Dickey-Fuller, não é possível rejeitar a existência de uma raiz unitária;
- (D) a distribuição t-Student deve ser empregada para detectar a existência ou não de raiz unitária na série temporal  $y_t$ ;
- (E) a média condicional de  $y_t$  para  $t = 8$  é  $E(y_t | t = 8) = 3$ .



Realização

