Estratégia Concursos

Aula 00

Desenvolve SP (Economista) Estatística - 2024 (Pós-Edital)

Autor:
Equipe Exatas Estratégia Concursos

27 de Fevereiro de 2024

Índice

AVISO IMPORTANTE!

Olá, Alunos (as)!

Passando para informá-los a respeito da **disposição das questões** dentro do nosso material didático. Informamos que a escolha das bancas, dentro dos nossos Livros Digitais, é feita de maneira estratégica e pedagógica pelos nossos professores a fim de proporcionar a melhor didática e o melhor direcionamento daquilo que mais se aproxima do formato de cobrança da banca do seu concurso.

Assim, o formato de questões divididas por tópico facilitará o seu processo de estudo, deixando mais alinhado às disposições constantes no edital.

No mais, continuaremos à disposição de todos no Fórum de dúvidas!

Atenciosamente,

Equipe Exatas

Estratégia Concursos

APRESENTAÇÃO DO CURSO

Olá, pessoal! Tudo bem?

É com grande satisfação damos início ao nosso curso!

Os professores **Eduardo Mocellin, Francisco Rebouças, Luana Brandão, Djefferson Maranhão** e **Vinicius Veleda** ficarão responsáveis pelo **Livro Digital**.

Antes de continuarmos, vamos apresentar os professores do material escrito:

Eduardo Mocellin: Fala, pessoal! Meu nome é Eduardo Mocellin, sou professor de Matemática e de Raciocínio Lógico do Estratégia Concursos e engenheiro Mecânico-Aeronáutico pelo Instituto Tecnológico de Aeronáutica (ITA). Sinto-me feliz em poder contribuir com a sua aprovação! Não deixe de me seguir no Instagram: **@edu.mocellin**

Francisco Rebouças: Fala, alunos! Aqui é o Francisco Rebouças, professor de Matemática do Estratégia Concursos. Sou Engenheiro Aeroespacial formado pelo Instituto Tecnológico de Aeronáutica (ITA). Saiba que será uma honra fazer parte da sua jornada rumo à aprovação e que estaremos sempre aqui para auxiliá-los com o que precisarem. Um grande abraço e nos vemos nas aulas!

Luana Brandão: Oi, pessoal! O meu nome é Luana Brandão e sou professora de Estatística do Estratégia Concursos. Sou Graduada, Mestre e Doutora em Engenharia de Produção, pela Universidade Federal Fluminense. Passei nos concursos de Auditor Fiscal (2009/2010) e Analista Tributário (2009) da Receita Federal e de Auditor Fiscal do Estado do Rio de Janeiro (2010). Sou Auditora Fiscal do Estado do RJ desde 2010. Vamos juntos nesse caminho até a aprovação? **@professoraluanabrandao**

Djefferson Maranhão: Olá, amigos do Estratégia Concursos, tudo bem? Meu nome é Djefferson Maranhão, professor de Estatística do Estratégia Concursos. Sou Graduado em Ciência da Computação pela Universidade Federal do Maranhão (UFMA). Desde 2015, sou Auditor da Controladoria Geral do Estado do Maranhão (2015 - 5º lugar). Antes, porém, exerci os cargos de Analista de Sistemas na UFMA (2010 - 1º lugar) e no TJ-MA (2011 - 1º lugar). Já estive na posição de vocês e sei o quanto a vida de um concurseiro é um tanto atribulada! São vários assuntos para se dominar em um curto espaço de tempo. Por isso, contem comigo para auxiliá-los nessa jornada rumo à aprovação. Um grande abraço.

Vinicius Veleda: Olá, caros alunos! Sou Auditor Fiscal do Estado do Rio Grande do Sul. Professor de Matemática e Matemática Financeira do Estratégia Concursos. Aprovado nos Concursos de Auditor Fiscal da Secretaria da Fazenda dos Estados do Rio Grande do Sul (SEFAZ RS - 2019), Santa Catarina (SEFAZ SC - 2018) e Goiás (SEFAZ GO - 2018). Formado em Engenharia de Petróleo pela Universidade Federal do Rio de Janeiro (UFRJ) com graduação sanduíche em Engenharia Geológica pela Universidade Politécnica de Madrid (UPM). Pela UFRJ, fui campeão sulamericano do Petrobowl (Buenos Aires) e, posteriormente, Campeão Mundial (Dubai). Cursei meu ensino médio na Escola Preparatória de Cadetes do Exército (EsPCEx). Contem comigo nessa trajetória! **@viniciusveleda**

O material escrito em **PDF** está sendo construído para ser sua fonte **autossuficiente** de estudos. Isso significa que o livro digital será **completo** e **voltado para o seu edital**, justamente para que você não perca o seu precioso tempo "caçando por aí" o conteúdo que será cobrado na sua prova. Ademais, sempre que necessário, você poderá fazer perguntas sobre as aulas no **fórum de dúvidas**. **Bons estudos!**

Olá, amigo(a)!

Nessa aula, vamos estudar uma parte muito importante da Estatística Inferencial, que é a **estimação** por ponto e por intervalo. Primeiro, vamos começar vendo a distribuição dos estimadores que será o fundamento para a estimação intervalar.

Vamos começar?!

Luana Brandão

Doutora em Engenharia de Produção (UFF)

Auditora Fiscal da SEFAZ-RJ

Se tiver alguma dúvida, entre em **contato** comigo!

 *professoraluanabrandao@gmail.com*

 *@professoraluanabrandao*

"Não importa o quão devagar você vá, desde que não pare."

Confúcio

DISTRIBUIÇÃO AMOSTRAL

Nesta seção, estudaremos a distribuição de probabilidade dos principais estimadores, chamada de **Distribuição Amostral**. Vale ressaltar que os **estimadores** são **variáveis aleatórias** e por isso apresentam distribuições de probabilidade.

Inicialmente, é importante pontuar que a distribuição dos elementos de uma **amostra** aleatória qualquer segue a **mesma** distribuição **populacional**.

Por exemplo, vamos considerar uma moeda com 2 faces, que vamos chamar de face 0 e face 1. Tratando-se de uma moeda equilibrada, a probabilidade de cada face é de $\frac{1}{2} = 0{,}5$ e a esperança e variância são, respectivamente:

$$E(X) = \sum x.P(X = x)$$

$$E(X) = 0 \times 0{,}5 + 1 \times 0{,}5 = 0{,}5$$

$$V(X) = \sum (x - \mu)^2.P(X = x)$$

$$V(X) = (0 - 0{,}5)^2 \times \frac{1}{2} + (1 - 0{,}5)^2 \times \frac{1}{2} = \frac{0{,}25 + 0{,}25}{2} = 0{,}25$$

Agora, suponha que vamos extrair uma **amostra** aleatórias de tamanho 3, ou seja, vamos lançar a moeda 3 vezes. Podemos representar essa amostra por X_1, X_2, X_3.

Considerando que os possíveis resultados das amostras são os mesmos da população (0 ou 1) e com as mesmas probabilidades de $\frac{1}{2} = 0{,}5$ para cada face, então temos:

$$E(X_1) = E(X_2) = E(X_3) = 0 \times 0{,}5 + 1 \times 0{,}5 = 0{,}5$$

$$V(X_1) = V(X_2) = V(X_3) = (0 - 0{,}5)^2 \times \frac{1}{2} + (1 - 0{,}5)^2 \times \frac{1}{2} = \frac{0{,}25 + 0{,}25}{2} = 0{,}25$$

Ou seja, a esperança e a variância de cada amostra são **iguais** às da população:

$$E(X_i) = E(X) = \mu$$

$$V(X_i) = V(X) = \sigma^2$$

Na verdade, toda a distribuição de probabilidade da **amostra** é **igual** à distribuição de probabilidade da **população** (as esperanças e as variâncias são iguais como consequência desse fato).

Quando as variáveis $X_1, X_2, X_3, \dots X_n$, que representam os elementos da **amostra**, são **independentes**, dizemos que elas **independentes e identicamente distribuídas** (**i.i.d.**), isto é, são independentes e apresentam a mesma distribuição (igual à da população).

Isso ocorre quando a **população** é **infinita** (ou muito grande em comparação com o tamanho da amostra) **ou** quando a **amostra** é extraída **com reposição**.

(FGV/2021 – FunSaúde/CE - *Adaptada*) Se X_1, X_2, ... X_n é uma amostra aleatória simples extraída de uma população infinita com determinada distribuição de probabilidades f(x), avalie se as afirmativas a seguir estão corretas.

I. X_1, X_2, ... X_n são independentes.

II. X_1, X_2, ... X_n são identicamente distribuídos.

III. Nem sempre cada X_i, i = 1,..., n, tem distribuição f(x).

Está correto o que se afirma em

a) I, apenas.

b) I e II, apenas.

c) I e III, apenas.

d) II e III, apenas.

e) I, II e III.

Comentários:

Essa questão trabalha com os conceitos fundamentais envolvendo distribuição amostral.

Sendo a população infinita, então os elementos da amostra (que o enunciado chamou de X_1, X_2, ... X_n) são variáveis aleatórias **independentes** que apresentam a **mesma distribuição** da população.

Assim, as afirmativas I e II estão corretas, enquanto a afirmativa III está incorreta, pois cada variável X_i apresenta a mesma distribuição f(x) da população (sempre).

Gabarito: B

Agora, estudaremos a distribuição dos estimadores mais utilizados, quais sejam, a **média**, a **proporção** e a **variância** amostrais.

Distribuição Amostral da Média

Para estimarmos a **média da população μ**, utilizamos como estimador a **média amostral $\bar{X}$**.

Sendo $X_1, X_2, \dots, X_n$ os valores observados da amostra, a média amostral é a **razão** entre a soma dos valores observados, $X_1 + X_2 + \cdots + X_n$, e o número de elementos observados, n:

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Assim como os demais estimadores, a média amostral é uma **variável aleatória**, uma vez que $\bar{X}$ **varia** de acordo com os valores observados da amostra $X_1, X_2, \dots, X_n$.

Vamos ao exemplo da moeda lançada 3 vezes. Se o resultado for {0, 0, 1}, a média amostral será:

$$\bar{X} = \frac{0 + 0 + 1}{3} = \frac{1}{3} \cong 0{,}33$$

Se o resultado for {0, 1, 1}, por exemplo, a média amostral será:

$$\bar{X} = \frac{0 + 1 + 1}{3} = \frac{2}{3} \cong 0{,}67$$

*E qual seria a esperança desse estimador? Bem, sabendo que as faces possíveis são 0 e 1, cada uma com 50% de chance, esperamos que as **médias** desse experimento estejam em torno de 0,5.*

Ou seja, a **esperança da média amostral** é igual à **média populacional**?

$$E(\bar{X}) = \mu$$

E quanto à variância? A **variância da média amostral** é dada por:

$$V(\bar{X}) = \frac{V(X)}{n} = \frac{\sigma^2}{n}$$

Para o exemplo dos 3 lançamentos da moeda, em que V(X) = 0,25 e n = 3, a variância da média amostral é:

$$V(\bar{X}) = \frac{0{,}25}{3} \cong 0{,}08$$

A **variância** da **média amostral** $\boldsymbol{V(\bar{X})}$ é **menor** do que a **variância populacional** $\boldsymbol{V(X)}$.

Isso ocorre porque a **média** das observações $\bar{X}$ tende a ser um valor bem mais **próximo** da média populacional μ do que as observações **individuais** da amostra X_i. Afinal, no cálculo da média amostral, os valores acima da média **compensam** os valores abaixo da média.

Em outras palavras, as médias amostrais $\bar{X}$ **variam menos** do que as observações individuais X_i. Lembrando que as observações individuais da amostra seguem a mesma distribuição da população, então concluímos que a **variância da média amostral** $V(\bar{X})$ é **menor** do que a **variância da população** $\boldsymbol{V(X)}$.

Além disso, quanto **maior o tamanho da amostra** $\boldsymbol{n}$, **menor será a variância da média amostral**.

O **desvio padrão** (raiz quadrada da variância) de um estimador pode ser chamado de **erro padrão**.

O **erro padrão** (ou **desvio padrão**) da **média amostral** é dado por:

$$EP(\bar{X}) = \sqrt{V(\bar{X})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Ou seja, o erro padrão (ou desvio padrão) da média amostral pode ser calculado como a **raiz quadrada** da **variância da média amostral** $\boldsymbol{\sqrt{V(\bar{X})}}$, ou como a **razão** entre o **desvio padrão populacional** e a **raiz do número de elementos da amostra** $\boldsymbol{\frac{\sigma}{\sqrt{n}}}$.

Também podemos denotar o erro padrão (ou desvio padrão) da média amostral por $\boldsymbol{\sigma_{\bar{X}}}$.

Para o nosso exemplo, o erro padrão da média amostral pode ser calculado como:

$$EP(\bar{X}) = \sqrt{V(\bar{X})} = \sqrt{\frac{0{,}25}{3}} \cong 0{,}29$$

As fórmulas da esperança e variância podem ser obtidas pelas respectivas **propriedades**.

Sendo $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$, a **esperança** $E(\bar{X})$ é dada por:

$$E(\bar{X}) = E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = E\left(\frac{X_1}{n}\right) + E\left(\frac{X_2}{n}\right) + \cdots + E\left(\frac{X_n}{n}\right)$$

$$E(\bar{X}) = \frac{1}{n}E(X_1) + \frac{1}{n}E(X_2) + \cdots + \frac{1}{n}E(X_n)$$

Vimos que a esperança amostral é igual à esperança populacional, $E(X_i) = E(X) = \mu$:

$$E(\bar{X}) = \frac{1}{n}\mu + \frac{1}{n}\mu + \cdots + \frac{1}{n}\mu = n \times \frac{1}{n} \times \mu = \mu$$

Considerando que as variáveis $X_1, X_2, \ldots, X_n$ são **independentes**, a **variância** $V(\bar{X})$ é:

$$V(\bar{X}) = V\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = V\left(\frac{X_1}{n}\right) + V\left(\frac{X_2}{n}\right) + \cdots + V\left(\frac{X_n}{n}\right)$$

$$V(\bar{X}) = \frac{1}{n^2}V(X_1) + \frac{1}{n^2}V(X_2) + \cdots + \frac{1}{n^2}V(X_n)$$

Sabendo que a variância amostral é igual à variância populacional, $V(X_i) = V(X) = \sigma^2$:

$$\boldsymbol{V(\bar{X}) = \frac{1}{n^2}\sigma^2 + \frac{1}{n^2}\sigma^2 + \cdots + \frac{1}{n^2}\sigma^2 = n \times \frac{1}{n^2} \times \sigma^2 = \frac{\sigma^2}{n}}$$

A **variância** dos **elementos da amostra** X_i é **igual** à **variância populacional** $V(X_i) = \sigma^2$. O que é **diferente** é a **variância** da **média amostral**, $V(\bar{X}) = \frac{\sigma^2}{n}$.

Se a variância da população **não** for conhecida, ela precisa ser estimada a partir da amostra (variância amostral):

$$s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

Nessa situação, a estimativa para a **variância da média amostral** será:

$$V(\bar{X}) = \frac{s^2}{n}$$

(CESPE/2016 – TCE/PA) Uma amostra aleatória, com n = 16 observações independentes e identicamente distribuídas (IID), foi obtida a partir de uma população infinita, com média e desvio padrão desconhecidos e distribuição normal. Tendo essa informação como referência inicial, julgue o seguinte item.

Para essa amostra aleatória simples, o valor esperado da média amostral é igual à média populacional.

Comentários:

De fato, o **valor esperado da média amostral** (para uma amostra aleatória) é igual à **média populacional**.

Gabarito: Certo.

(2019 – UEPA) Considere uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população normal de média μ e variância $\sigma^2 = 9$. Então, a média e a variância de $\overline{X_n} = \frac{X_1, X_2, \ldots, X_n}{n}$, são, respectivamente,

a) μ e $\frac{3}{n}$.

b) $\frac{\mu}{n}$ e $\frac{9}{n}$.

c) μ e $\frac{9}{n}$.

d) μ e $\frac{n}{9}$.

Comentários:

A **média** (ou esperança) e a **variância** da média amostral são, respectivamente:

$$E(\bar{X}) = \mu$$

$$V(\bar{X}) = \frac{V(X)}{n} = \frac{9}{n}$$

Gabarito: C.

(VUNESP/2015 – TJ-SP) Resultados de uma pesquisa declaram que o desvio padrão da média amostral é 32. Sabendo que o desvio padrão populacional é 192, então o tamanho da amostra que foi utilizada no estudo foi

a) 6.

b) 25.

c) 36.

d) 49.

e) 70.

Comentários:

O desvio padrão (ou erro padrão) da média amostral pode ser calculado como a razão entre o desvio padrão da população e a raiz do número de elementos da amostra:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

O enunciado informa que o desvio padrão da média amostral é $\sigma_{\bar{X}} = 32$ e o desvio padrão populacional é $\sigma = 192$. Logo:

$$32 = \frac{192}{\sqrt{n}}$$

$$\sqrt{n} = \frac{192}{32} = 6$$

$$n = (6)^2 = 36$$

Gabarito: C.

(FCC/2012 – TRE-SP) Uma variável aleatória U tem distribuição uniforme contínua no intervalo $[\boldsymbol{\alpha}, \boldsymbol{3\alpha}]$. Sabe-se que U tem média 12. Uma amostra aleatória simples de tamanho n, com reposição, é selecionada da distribuição de U e sabe-se que a variância da média dessa amostra é 0,1. Nessas condições, o valor de n é

a) 80.

b) 100.

c) 120.

d) 140.

e) 150.

Comentários:

O que essa questão exige, em relação à matéria que acabamos de estudar, é a fórmula do desvio padrão da média amostral:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Assim, podemos escrever o tamanho amostral como:

$$\sqrt{n} = \frac{\sigma}{\sigma_{\bar{X}}}$$

$$n = \left(\frac{\sigma}{\sigma_{\bar{X}}}\right)^2 = \frac{\sigma^2}{\sigma_{\bar{X}}^2}$$

Ou seja, o tamanho amostral é a razão entre a variância populacional σ^2 (quadrado do desvio padrão populacional σ) e a variância da média amostral $\sigma_{\bar{X}}^2$ (quadrado do desvio padrão da média amostral $\sigma_{\bar{X}}$).

O enunciado informa que a variância da média amostral é $\sigma_{\bar{X}}^2 = 0{,}1$.

Pronto! A matéria desta aula acabou. Agora, para calcular a variância populacional, precisamos saber calcular a média e a variância da distribuição contínua uniforme.

A média (esperança) dessa distribuição é igual à média aritmética dos limites do intervalo, a e b. Sabendo que a = $\boldsymbol{\alpha}$, b = 3.$\boldsymbol{\alpha}$ e E(X) = 12, conforme dados do enunciado, temos:

$$E(X) = \frac{a+b}{2}$$

$$12 = \frac{\alpha + 3\alpha}{2} = \frac{4\alpha}{2} = 2\alpha$$

$$\alpha = 6$$

Ou seja, a = 6 e b = 3x6 = 18. Então, a variância é dada por:

$$V(X) = \frac{(b-a)^2}{12} = \frac{(18-6)^2}{12} = \frac{(12)^2}{12} = 12$$

Voltando à nossa fórmula para encontrar o tamanho amostral, temos:

$$n = \frac{\sigma^2}{\sigma_{\bar{X}}^2} = \frac{12}{0{,}1} = 120$$

Gabarito: C

Fator de Correção para População Finita

Os resultados da variância e do desvio padrão da média amostral são válidos para variáveis $X_1, X_2, \ldots, X_n$ **independentes**, ou seja, quando a população é **infinita ou** quando as amostras são extraídas **com reposição**.

Quando isso não ocorre, ou seja, quando a **população é finita e** as amostras são extraídas **sem reposição**, precisamos fazer um **ajuste**.

Sendo n o tamanho da amostra e N o tamanho da população, precisamos multiplicar a **variância** da média amostral, pelo **fator de correção de população finita** $\frac{N-n}{N-1}$.

$$V(\overline{X_*}) = \frac{\sigma^2}{n} \times \frac{N-n}{N-1}$$

Observe que o fator de correção é **menor** do que **1**, pois $N - n < N - 1$. Assim, o ajuste **diminui** a variância da média amostral.

Sabendo que o erro (ou desvio) padrão é a raiz quadrada da variância, podemos ajustá-lo para populações finitas e amostras extraídas sem reposição, multiplicando-o pela raiz quadrada do fator de correção:

$$EP(\overline{X_*}) = \sqrt{V(\overline{X_*})} = \sqrt{\frac{\sigma^2}{n} \times \frac{N-n}{N-1}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

Distribuição da Média Amostral e a Curva Normal

Quando a **população** segue distribuição **normal (ou gaussiana)** com variância conhecida, a **média amostral** também seguirá **distribuição normal**.

Como vimos anteriormente, a esperança, a variância e o desvio (ou erro) padrão da média amostral são:

$$E(\bar{X}) = \mu$$

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

$$EP(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Assim, para calcular as **probabilidades** envolvendo a média amostral, nessa situação, utilizamos a seguinte **transformação** para a **normal padrão**:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Por exemplo, vamos supor uma população com média $\mu = 20$ e desvio padrão $\sigma = 2$. Para calcular a probabilidade de a **média de uma amostra** de tamanho $n = 16$ ser maior que $\bar{x} = 21$, fazemos:

$$z = \frac{21 - 20}{\frac{2}{\sqrt{16}}} = \frac{1}{\frac{2}{4}} = 2$$

Assim, a probabilidade $P(\bar{X} > 21)$ é igual à probabilidade $P(Z > 2)$, que pode ser obtida a partir da tabela normal padrão.

De modo equivalente, podemos dizer que a seguinte variável segue distribuição **normal padrão (ou reduzida)**, isto é, com média igual a 0 e desvio padrão igual a 1, que representamos como $N(0,1)$:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Ainda que a população **não** siga distribuição normal com variância conhecida, pelo **Teorema Central do Limite**, é possível **aproximar** a distribuição da média amostral a uma **normal**, também com média $E(\bar{X}) = \mu$, variância $V(\bar{X}) = \frac{\sigma^2}{n}$ e desvio padrão (ou erro padrão) $EP(\bar{X}) = \frac{\sigma}{\sqrt{n}}$, quando o tamanho da amostra n for **grande** o bastante.

A possibilidade de tal aproximação **depende** do **tamanho da amostra** e da **distribuição da população**.

Quanto **maior** o tamanho da amostra e quanto mais **simétrica** for a distribuição da população, **melhor** será a aproximação. Usualmente, considera-se que para uma amostra grande, com $n \geq 30$, a aproximação será satisfatória, para **qualquer distribuição** populacional.

A distribuição dos **quantis amostrais** (mediana, quartis, decis, percentis etc.) também pode ser aproximada a uma **normal**, à medida que o tamanho da amostra aumenta, em decorrência do Teorema Central do Limite.

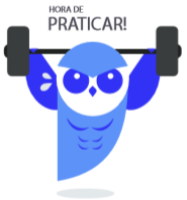
A média da distribuição é igual ao quantil populacional e a variância é:

$$Var(\widetilde{x_q}) = \frac{q(1-q)}{n.[f(x_q)]^2}$$

Sendo q a proporção da distribuição delimitada pelo quantil x_q, $P(X < x_q) = q$, e $f(x_q)$o valor da função densidade da variável no ponto x_q.

- Por exemplo, para uma distribuição uniforme no intervalo [0,1], temos $f(x) = 1$. Assim, a variância da mediana amostral ($q = 0{,}5$) para n suficientemente grande é:

$$Var(\widetilde{Md}) = \frac{0{,}5 \times 0{,}5}{n.1^2} = \frac{1}{4n}$$



(CESPE/2014 – ANATEL) Com base no teorema limite central, julgue o item abaixo.

Sendo uma amostra aleatória simples retirada de uma distribuição X com média μ e variância 1, a distribuição da média amostral dessa amostra, $\bar{X}$, converge para uma distribuição normal de média nμ e variância 1, à medida que n aumenta.

Comentários:

A distribuição da média amostral $\bar{X}$ converge para uma distribuição normal à medida que n aumenta. Porém, a média dessa distribuição é $E(\bar{X}) = \mu$ e variância $V(\bar{X}) = \frac{V(X)}{n}$. Sabendo que $V(X) = 1$, a variância da média amostral é $V(\bar{X}) = \frac{1}{n}$ (e não 1).

Gabarito: Errado.

(CESPE 2018/PF) O tempo gasto (em dias) na preparação para determinada operação policial é uma variável aleatória X que segue distribuição normal com média M, desconhecida, e desvio padrão igual a 3 dias. A observação de uma amostra aleatória de 100 outras operações policiais semelhantes a essa produziu uma média amostral igual a 10 dias. Com referência a essas informações, julgue o item que se segue, sabendo que $P(Z > 2) = 0{,}025$, em que Z denota uma variável aleatória normal padrão.

O erro padrão da média amostral foi inferior a 0,5 dia.

Comentários:

O erro padrão da média amostral é:

$$EP(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Em que n é o tamanho da amostra ($n = 100$); e σ é o desvio padrão amostral ($\sigma = 3$), logo:

$$EP(\bar{X}) = \frac{3}{\sqrt{100}} = \frac{3}{10} = 0{,}3$$

Gabarito: Certo.

(CESPE/2019 – Analista Judiciário TJ) Um pesquisador deseja comparar a diferença entre as médias de duas amostras independentes oriundas de uma ou duas populações gaussianas. Considerando essa situação hipotética, julgue o próximo item.

Para que a referida comparação seja efetuada, é necessário que ambas as amostras tenham N ≥ 30.

Comentários:

Quando a população segue uma distribuição normal (ou gaussiana), a média amostral também seguirá uma distribuição normal, **independentemente do tamanho da amostra**. Logo, o item está errado. Para fins de complementação, se a amostra for grande o suficiente (normalmente, consideramos isso para $n \geq 30$), a média amostral seguirá aproximadamente uma distribuição normal, mesmo que a população não siga distribuição normal.

Gabarito: Errado.

(FCC 2015/SEFAZ-PI) Instrução: Para responder à questão utilize, dentre as informações dadas a seguir, as que julgar apropriadas. Se Z tem distribuição normal padrão, então: P(Z < 0,4) = 0,655; P(Z < 1,2) = 0,885; P(Z< 1,6) = 0,945; P(Z < 1,8) = 0,964; P(Z < 2) = 0,977.

Uma auditoria feita em uma grande empresa considerou uma amostra aleatória de 64 contas a receber. Se a população de onde essa amostra provém é infinita e tem distribuição normal com desvio padrão igual a R$ 200,00 e média igual a R$ 950,00, a probabilidade da variável aleatória média amostral, usualmente denotada por $\boldsymbol{X}$, estar situada entre R$ 980,00 e R$ 1.000,00 é dada por

a) 18,4%

b) 9,2%

c) 28,5%

d) 47,7%

e) 86,2%

Comentários:

O enunciado informa que a população segue distribuição normal, logo, a média amostral também terá distribuição normal. Assim, utilizamos a seguinte transformação para a normal padrão:

$$Z_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Sabemos que o tamanho da amostra é n = 64, logo, $\sqrt{n} = 8$; que a média populacional é $\mu = 950$ e que o desvio padrão populacional é $\sigma = 200$. Substituindo esses valores, a transformação para $\overline{X_{inf}} = 980$ é:

$$Z_{\bar{X}} = \frac{980 - 950}{\frac{200}{8}} = 30 \times \frac{8}{200} = \frac{6}{5} = 1{,}2$$

Para $\overline{X_{sup}} = 1000$, temos:

$$Z_{\bar{X}} = \frac{1000 - 950}{\frac{200}{8}} = 50 \times \frac{8}{200} = 2$$

Ou seja, a probabilidade desejada corresponde a $P(1{,}2 < Z < 2)$, que pode ser calculada pelos dados fornecidos no enunciado:

$$P(1{,}2 < Z < 2) = P(Z < 2) - P(Z < 1{,}2) = 0{,}977 - 0{,}885 = 0{,}092 = 9{,}2\%$$

Gabarito: B

Distribuição Amostral da Proporção

Agora, vamos trabalhar com uma população em que **determinada característica** está presente em uma **proporção p** dessa população, por exemplo, 15% da população apresenta olhos azuis; 20% da população está doente, 1% da produção apresenta defeito, etc.

Um elemento qualquer da população X pode **apresentar** a característica estudada, o que chamamos de **sucesso** ($X = 1$), ou **não**, o que chamamos de **fracasso** ($X = 0$). A probabilidade de sucesso é p e a probabilidade de fracasso é $q = 1 - p$.

Essa população apresenta uma distribuição de **Bernoulli**, com parâmetro p.

Sendo essa proporção populacional desconhecida, precisamos **estimá-la** a partir da proporção de sucessos encontrados na **amostra**, que indicamos por $\hat{p}$.

Considerando que cada observação X_i da amostra será $X_i = 0$ ou $X_i = 1$ (assim como para a população), então a proporção de sucessos na amostra pode ser calculada como:

$$\hat{p} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Vamos supor que queremos estimar a proporção de **defeitos** em uma produção de medicamentos. Para isso, extraímos uma amostra de 10 medicamentos, que apresentou o seguinte resultado, em que 0 representa um item não defeituoso e 1 representa um item defeituoso:

$$\{0, 0, 1, 0, 0, 0, 1, 0, 0, 0\}$$

Logo, a proporção encontrada nessa amostra é:

$$\hat{p} = \frac{0+0+1+0+0+0+1+0+0+0}{10} = \frac{2}{10} = 0{,}2$$

A partir da proporção amostral, podemos estimar o total de elementos que apresentam a característica desejada (sucesso) na população, multiplicando-a pelo tamanho total da população N:

$$\hat{\tau} = N \times \hat{p}$$

Note que o estimador $\hat{p}$ é calculado da mesma forma que a média amostra $\bar{X}$ que vimos anteriormente. Logo, a **esperança** de $\hat{p}$ é calculada da mesma forma que para $\bar{X}$, utilizando p no lugar de μ:

$$E(\hat{p}) = p$$

Ou seja, a **esperança** do estimador é igual à **proporção populacional**.

Em outras palavras, a proporção amostral **tende** à **proporção populacional**.

A **variância** de $\hat{p}$ também é calculada de forma análoga à de $\bar{X}$:

$$V(\hat{p}) = \frac{V(p)}{n}$$

Sabendo que a população segue distribuição de Bernoulli, temos $V(p) = p.q$, então:

$$V(\hat{p}) = \frac{p.q}{n}$$

E o **erro padrão** (ou desvio padrão) para $\hat{p}$, raiz quadrada da sua variância é dado por:

$$EP(\hat{p}) = \sqrt{\frac{p.q}{n}}$$

Se a proporção populacional p for **desconhecida**, utilizamos a proporção amostral $\hat{p}$ para **estimar** a variância populacional.

A estimativa da **variância populacional** é dada por:

$$V(p) = \hat{p}.\hat{q}$$

Em que $\hat{q} = 1 - \hat{p}$.

E a estimativa da **variância do estimador $\hat{p}$** é:

$$V(\hat{p}) = \frac{\hat{p}.\hat{q}}{n}$$

Para o nosso exemplo, em que encontramos $\hat{p} = 0{,}2$ (logo, $\hat{q} = 1 - \hat{p} = 0{,}8$). A estimativa da **variância** da proporção **populacional** é:

$$V(p) = \hat{p}.\hat{q} = 0{,}2 \times 0{,}8 = 0{,}16$$

E a estimativa da **variância** da proporção **amostral** é:

$$V(\hat{p}) = \frac{\hat{p}.\hat{q}}{n} = \frac{0{,}2 \times 0{,}8}{10} = 0{,}016$$

Logo, a estimativa para o **erro padrão** (ou desvio padrão) da **proporção amostral** é:

$$EP(\hat{p}) = \sqrt{V(\hat{p})} = \sqrt{0{,}016} \cong 0{,}126$$

Considerando que cada elemento da população segue distribuição de Bernoulli, então o número de elementos com o atributo sucesso encontrados em uma **amostra** de tamanho n segue uma distribuição **binomial**, com parâmetros n e p. Para o nosso exemplo, temos uma distribuição binomial com $n = 10$ e proporção estimada $\hat{p} = 0{,}2$.

Porém, também é possível aproximar, pelo **Teorema Central do Limite**, a distribuição da proporção amostral a uma **normal**, com média $E(\hat{p}) = p$, variância $V(\hat{p}) = \frac{p.q}{n}$ e desvio (ou erro) padrão $EP(\hat{p}) = \sqrt{\frac{p.q}{n}}$, quando o tamanho da amostra n for suficientemente **grande**.

Assim, para calcular as probabilidades envolvendo a proporção amostral, utilizamos a **transformação** para a **normal padrão**:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p.q}{n}}}$$

Por exemplo, supondo que a proporção de sucesso de uma população seja $p = 0{,}5$ (e proporção de fracasso $q = 1 - p = 0{,}5$), vamos calcular a probabilidade de observar uma proporção maior que $\hat{p} = 0{,}6$ em uma amostra de tamanho $n = 25$:

$$z = \frac{0{,}6 - 0{,}5}{\sqrt{\frac{0{,}5 \times 0{,}5}{25}}} = \frac{0{,}1}{\frac{0{,}5}{5}} = \frac{0{,}1}{0{,}1} = 1$$

Assim, a probabilidade $P(\hat{p} > 0{,}6)$ é aproximadamente igual à probabilidade $P(Z > 1)$, que pode ser calculada pela tabela da normal padrão.

Se a população for **finita** e a amostra for extraída **sem reposição**, será necessário aplicar o fator de **correção** para população finita, multiplicando a **variância** da proporção amostral por $\frac{N-n}{N-1}$:

$$V(\hat{p}) = \frac{p.q}{n} \times \frac{N-n}{N-1}$$

E o erro (ou desvio) padrão do estimador, com a correção para população finita, é igual à raiz quadrada:

$$EP(\hat{p}) = \sqrt{\frac{p.q}{n} \times \frac{N-n}{N-1}}$$

(CESPE 2016/TCE-PA) Em estudo acerca da situação do CNPJ das empresas de determinado município, as empresas que estavam com o CNPJ regular foram representadas por 1, ao passo que as com CNPJ irregular foram representadas por 0.

Considerando que a amostra {0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1} foi extraída para realizar um teste de hipóteses, julgue o item subsequente.

A estimativa pontual da proporção de empresas da amostra com CNPJ regular é superior a 50%.

Comentários:

O estimador da proporção $\hat{p}$ pode ser calculado pela soma dos valores dos elementos, dividida pelo número de elementos na amostra:

$$\hat{p} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$\hat{p} = \frac{0+1+1+0+0+1+0+1+0+1+1+0+0+1+1+0+1+1+1+1}{20} = \frac{12}{20} = 0{,}6$$

Logo, a proporção é de 60%.

Gabarito: Certo

(CESPE/2019 – TJ-AM) Para estimar a proporção de menores infratores reincidentes em determinado município, foi realizado um levantamento estatístico. Da população-alvo desse estudo, constituída por 10.050 menores infratores, foi retirada uma amostra aleatória simples sem reposição, composta por 201 indivíduos. Nessa amostra foram encontrados 67 reincidentes. Com relação a essa situação hipotética, julgue o seguinte item.

A estimativa do erro padrão da proporção amostral foi inferior a 0,04.

Comentários:

O erro padrão da proporção é dada pela relação:

$$EP(\hat{p}) = \sqrt{\frac{p \times q}{n}}$$

A questão nos diz que a amostra é composta por 201 indivíduos, sendo 67 deles reincidentes. Assim, temos $n = 201$ e $p = \frac{67}{201} = \frac{1}{3}$. Logo, $q = 1 - p = \frac{2}{3}$. Logo, o erro padrão é:

$$E = \sqrt{\frac{\frac{1}{3} \times \frac{2}{3}}{201}} = \frac{1}{3} \times \sqrt{\frac{2}{201}} \cong \frac{1}{3} \times \frac{1}{10} \cong 0{,}033$$

Gabarito: Certo.

Estimador para a média: $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$

Esperança: $E(\bar{X}) = \mu$; Variância: $V(\bar{X}) = \frac{V(X)}{n}$

Estimador para a proporção: $\hat{p} = \frac{X_1+X_2+\cdots+X_n}{n}$

Esperança: $E(\hat{p}) = p$; Variância: $V(\hat{p}) = \frac{p.q}{n}$

Distribuição Amostral da Variância

Quando a variância da população é desconhecida, precisamos estimá-la a partir da amostra, assim como fizemos com a média e a proporção.

O **estimador da variância** que utilizamos para uma amostra de tamanho n é:

$$s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

Utilizamos esse estimador, com a divisão por $\boldsymbol{n-1}$, porque a sua **esperança** é igual à variância populacional, como veremos posteriormente, o que **não** ocorre com o estimador com a divisão por n.

Vamos supor que a variância da altura de determinado grupo de adultos seja desconhecida, assim como a sua média. Para estimar esses parâmetros, obtemos a seguinte amostra de 5 pessoas:

{1,65; 1,75; 1,8; 1,85; 1,95}

Primeiro, precisamos calcular a média da amostra:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{1{,}65+1{,}75+1{,}8+1{,}85+1{,}95}{5} = \frac{9}{5} = 1{,}8$$

Agora, calculamos o estimador da variância, somando os desvios em relação à média, elevados ao quadrado, e dividindo o somatório por $\boldsymbol{n-1}$:

$$s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

$$s^2 = \frac{(1{,}65-1{,}8)^2+(1{,}75-1{,}8)^2+(1{,}8-1{,}8)^2+(1{,}85-1{,}8)^2+(1{,}95-1{,}8)^2}{4}$$

$$s^2 = \frac{(-0{,}15)^2+(-0{,}05)^2+(0)^2+(0{,}05)^2+(0{,}15)^2}{4}$$

$$s^2 = \frac{0{,}0225+0{,}0025+0{,}0025+0{,}0225}{4} = \frac{0{,}05}{4} = 0{,}0125$$

Uma maneira alternativa de calcular a variância **populacional** é:

$$\sigma^2 = E(X^2) - [E(X)]^2$$

Para a variância **amostral**, também podemos utilizar uma fórmula similar a essa, mas com as devidas adaptações. No lugar de $E(X)$, utilizamos a média amostral $\bar{X}$; e, no lugar de $E(X^2)$, utilizamos $\overline{X^2}$, que é a média dos valores elevados ao quadrado:

$$\overline{X^2} = \frac{\sum X_i^2}{n}$$

Para o exemplo anterior, teríamos:

$$\overline{X^2} = \frac{(1{,}65)^2+(1{,}75)^2+(1{,}8)^2+(1{,}85)^2+(1{,}95)^2}{5} = \frac{16{,}25}{5} = 3{,}25$$

Por fim, fazemos um ajuste. Para calcular a variância populacional, dividimos por n e para a variância amostral, dividimos por $\boldsymbol{n-1}$.

Logo, precisamos multiplicar o resultado por $\frac{n}{\boldsymbol{n-1}}$ para obter a variância amostral:

$$s^2 = \left[\overline{X^2} - (\bar{X})^2\right] \times \frac{n}{\boldsymbol{n-1}}$$

Para o nosso exemplo, em que $\overline{X^2} = 3{,}25$, $\bar{X} = 1{,}8$ e $n = 5$, a variância amostral pode ser calculada como:

$$s^2 = [3{,}25 - (1{,}8)^2] \times \frac{5}{4} = [3{,}25 - 3{,}24] \times \frac{5}{4} = \frac{0{,}01\times5}{4} = 0{,}0125$$

Para a variância amostral, a sua **esperança** é igual à **variância populacional**, analogamente ao que ocorreu com os demais estimadores. A sua variância e erro padrão são dados por:

$$E(s^2) = \sigma^2$$

$$V(s^2) = \frac{2.\sigma^4}{n-1}$$

$$EP(s^2) = \sqrt{\frac{2}{n-1}}\sigma^2$$

Se a variância populacional σ^2 for **desconhecida**, utilizamos, no lugar de σ^2, a própria estimativa s^2 para estimar a média $E(s^2)$, a variância $V(s^2)$ e o erro padrão $EP(s^2)$.

Para o nosso exemplo, em que calculamos $s^2 = 0{,}0125$, para a amostra com $n = 5$ observações, o erro padrão de s^2 pode ser estimado como:

$$EP(s^2) = \sqrt{\frac{2}{n-1}}s^2 = \sqrt{\frac{2}{4}} \times 0{,}0125 \cong 0{,}009$$

Se a população seguir uma distribuição **normal**, então o estimador s^2, multiplicado pelo fator $\frac{n-1}{\sigma^2}$, segue uma distribuição **qui-quadrado** com $\boldsymbol{n-1}$ graus de liberdade:

$$\mathcal{X}^2_{n-1} = \left(\frac{n-1}{\sigma^2}\right).s^2$$

Em outras palavras, o estimador s^2 é uma variável com distribuição qui-quadrado, com $n-1$ graus de liberdade, multiplicada por $\frac{\sigma^2}{n-1}$:

$$s^2 = \left(\frac{\sigma^2}{n-1}\right).\mathcal{X}^2_{n-1}$$

Vamos supor que, para uma população normal, a variância populacional seja $\sigma^2 = 1$, e que vamos extrair amostras de tamanho $n = 5$. Nesse caso, a variância amostral s^2 terá a seguinte distribuição:

$$s^2 = \left(\frac{1}{5-1}\right).\mathcal{X}^2_{5-1} = \frac{\mathcal{X}^2_4}{4}$$

Ou seja, a variância amostral seguirá uma distribuição qui-quadrado com $n-1 = 4$ graus de liberdade, dividida por 4.

A seguir, consta a tabela da distribuição qui-quadrado com 4 graus de liberdade, que apresenta os valores de probabilidade $P(\mathcal{X}^2_4 < x)$ e os respectivos valores de x. Como a variância amostral segue essa distribuição, dividida por 4, criamos uma terceira coluna, dividindo os valores de x por 4.

$P(\mathcal{X}^2_4 < x)$	**0,005**	**0,01**	**0,025**	**0,05**	**0,1**	**0,25**	**0,5**	**0,75**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
x	0,21	0,30	0,48	0,71	1,06	1,92	3,36	5,39	7,78	9,49	11,14	13,28	14,86
$x/4$	0,05	0,07	0,12	0,18	0,27	0,48	0,84	1,35	1,94	2,37	2,79	3,32	3,72

Por exemplo, a probabilidade de a variância amostral observada ser inferior a 0,48 é:

$$P(s^2 < 0{,}48) = \mathbf{0{,}25}$$

A partir desse resultado, podemos calcular a esperança e a variância do estimador. A **esperança** é dada por:

$$s^2 = \left(\frac{\sigma^2}{n-1}\right).\mathcal{X}^2_{n-1}$$

$$E[s^2] = E\left[\left(\frac{\sigma^2}{n-1}\right).\mathcal{X}^2_{n-1}\right] = \left(\frac{\sigma^2}{n-1}\right).E[\mathcal{X}^2_{n-1}]$$

Considerando que a média de uma distribuição qui-quadrado com k graus de liberdade é igual a k, então fazendo $k = n - 1$, temos:

$$E[\mathcal{X}^2_{n-1}] = k = n - 1$$

Substituindo no resultado anterior, temos:

$$E[s^2] = \left(\frac{\sigma^2}{n-1}\right).(n-1) = \sigma^2$$

Esse é o resultado que vimos no início da seção. A **variância** do estimador é:

$$V[s^2] = V\left[\left(\frac{\sigma^2}{n-1}\right)\mathcal{X}^2_{n-1}\right] = \left(\frac{\sigma^2}{n-1}\right)^2.V[\mathcal{X}^2_{n-1}] = \left(\frac{\sigma^4}{(n-1)^2}\right).V[\mathcal{X}^2_{n-1}]$$

Considerando que a variância de uma distribuição qui-quadrado com k graus de liberdade é igual a $2k$, então fazendo $k = n - 1$, temos:

$$V[\mathcal{X}^2_{n-1}] = 2.k = 2.(n-1)$$

Substituindo no resultado anterior, temos:

$$V[s^2] = \left(\frac{\sigma^4}{(n-1)^2}\right).2.(n-1) = \frac{2.\sigma^4}{n-1}$$

Vale acrescentar que uma população normal depende dos dois parâmetros, variância e média, os quais são independentes. Consequentemente, os estimadores correspondentes também serão **independentes**.

(FGV/2016 – IBGE) Suponha que uma amostra de tamanho $\boldsymbol{n = 5}$ é extraída de uma população normal, com média desconhecida, obtendo as seguintes observações:

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

São dados ainda os seguintes valores, retirados da tabela da distribuição qui-quadrado:

- $P(\chi_4^2 < \mathbf{5}) \cong \mathbf{0{,}713}$
- $P(\chi_4^2 < \mathbf{12{,}5}) \cong \mathbf{0{,}986}$
- $P(\chi_5^2 > \mathbf{5}) \cong \mathbf{0{,}854}$
- $P(\chi_5^2 > \mathbf{12{,}5}) \cong \mathbf{0{,}971}$

Se a população tem variância verdadeira $\boldsymbol{\sigma^2 = 4}$, em nova amostra ($\boldsymbol{n = 5}$), 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

Comentários:

Para resolver essa questão, vamos primeiro calcular a variância amostral obtida nessa primeira amostra:

$$s^2 = \frac{\sum(X_i - \bar{X})^2}{\boldsymbol{n - 1}}$$

Para isso, precisamos da média amostral:

$$\bar{X} = \frac{3 + 5 + 6 + 9 + 12}{5} = \frac{35}{5} = 7$$

Agora, podemos calcular a variância dessa primeira amostra:

$$s_1^2 = \frac{(3-7)^2 + (5-7)^2 + (6-7)^2 + (9-7)^2 + (12-7)^2}{4}$$

$$s_1^2 = \frac{(-4)^2 + (-2)^2 + (-1)^2 + (2)^2 + (5)^2}{4} = \frac{16 + 4 + 1 + 4 + 25}{4} = \frac{50}{4} = 12{,}5$$

Para calcular a probabilidade de a variância amostral ser $s^2 > 12{,}5$, consideramos que esse estimador segue uma distribuição qui-quadrado com n – 1 graus de liberdade, multiplicada por $\left(\frac{\sigma^2}{n-1}\right)$:

$$s^2 = \left(\frac{\sigma^2}{n-1}\right).\chi_{n-1}^2$$

Sabendo que a variância populacional é $\sigma^2 = 4$ e que o tamanho da amostra é $n = 5$, então:

$$s^2 = \left(\frac{4}{5-1}\right).\mathcal{X}^2_{5-1} = \mathcal{X}^2_4$$

Portanto, a variância amostral segue a mesma distribuição de $\mathcal{X}^2_4$. A probabilidade de $s^2 > 12{,}5$ é, portanto:

$$P(s^2 > 12{,}5) = P(\mathcal{X}^2_4 > 12{,}5)$$

O enunciado informa que $P(\mathcal{X}^2_4 < 12{,}5) = 0{,}986$. A probabilidade $P(\mathcal{X}^2_4 > 12{,}5)$ é complementar:

$$P(\mathcal{X}^2_4 > 12{,}5) = 1 - P(\mathcal{X}^2_4 < 12{,}5) = 1 - 0{,}986 = 0{,}014$$

Gabarito: A

Distribuições para Amostragem Estratificada

Para uma **amostragem estratificada**, com k estratos, a **média amostral** será calculada como:

$$\bar{X} = \sum_{i=1}^{k} \frac{N_i}{N} \bar{x_i}$$

Nessa expressão, N_i é o tamanho de cada **estrato**; N é o tamanho total da **população**; e $\bar{x_i}$, a **média amostral** observada para cada estrato.

Ou seja, calculamos a média $\bar{x_i}$ para cada estrato i, multiplicamos pelo tamanho do estrato N_i e dividimos pelo tamanho total N.

Para ilustrar, vamos supor uma população dividida em 3 estratos, com os seguintes tamanhos N_i e os seguintes valores de média amostral $\bar{x_i}$ para cada estrato i:

Estrato	N_i	n_i	$\bar{x_i}$
1	50	5	2
2	30	3	3
3	20	2	4

A média amostral para toda a população corresponde às médias amostrais dos estratos, ponderadas pelos respectivos tamanhos dos estratos:

$$\bar{X} = \frac{50 \times 2 + 30 \times 3 + 20 \times 4}{50 + 30 + 20} = \frac{100 + 90 + 80}{100} = 2{,}7$$

A **razão** entre o tamanho do estrato N_i e o tamanho total da população N pode ser chamada de **peso** do estrato (W_i).

Nessa situação, para calcular a média amostral, basta multiplicar a média de cada estrato pelo respectivo peso e somar todos os resultados:

$$\bar{X} = \sum_{i=1}^{k} W_i \times \bar{x_i}$$

Podemos calcular, ainda, a **variância da média amostral**:

$$V(\bar{X}) = V\left(\sum_{i=1}^{k} \frac{N_i}{N} \bar{x_i}\right)$$

Pelas propriedades da variância, temos:

$$V(\bar{X}) = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 V(\bar{x_i})$$

Em que a variância da média amostral de cada estrato $V(\bar{x_i})$ é calculada pela **razão** entre a variância **populacional** do estrato e o **tamanho da amostra** do estrato:

$$V(\bar{x_i}) = \frac{V(X_i)}{n_i}$$

Em uma amostragem **proporcional**, o tamanho amostral do estrato n_i pode ser calculado pela **razão** entre o tamanho do estrato populacional e o tamanho total da população, **multiplicada** pelo tamanho total da amostra:

$$n_i = \frac{N_i}{N}.n$$

Vamos supor as seguintes variâncias para os estratos da população, que vimos anteriormente:

Estrato	N_i	n_i	$V(X_i)$
1	50	5	4
2	30	3	2
3	20	2	1

As variâncias das **médias amostrais** dos estratos são dadas por:

$$V(\overline{x_1}) = \frac{V(X_1)}{n_1} = \frac{4}{5} = 0{,}8$$

$$V(\overline{x_2}) = \frac{V(X_2)}{n_2} = \frac{2}{3} \cong 0{,}67$$

$$V(\overline{x_3}) = \frac{V(X_3)}{n_3} = \frac{1}{2} = 0{,}5$$

E a variância da média amostral global é dada por:

$$V(\bar{X}) = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 V(\overline{x_i}) = \left(\frac{N_1}{N}\right)^2 . V(\overline{x_1}) + \left(\frac{N_2}{N}\right)^2 . V(\overline{x_2}) + \left(\frac{N_3}{N}\right)^2 . V(\overline{x_3})$$

$$V(\bar{X}) = \left(\frac{50}{100}\right)^2 \times 0{,}8 + \left(\frac{30}{100}\right)^2 \times 0{,}67 + \left(\frac{20}{100}\right)^2 \times 0{,}5 = 0{,}25 \times 0{,}8 + 0{,}09 \times 0{,}67 + 0{,}04 \times 0{,}5 = 0{,}28$$

A variância da média de uma amostra estratificada proporcional é **menor ou igual** do que a variância da média de uma amostra aleatória simples (AAS). Afinal, a amostragem estratificada proporcional é **mais precisa** do que a AAS[1].

[1] A variância da média de uma AAS é a razão entre a variância da população como um todo e o tamanho da amostra; e a variância de toda a população é **maior** (ou igual) às variâncias dos estratos ponderadas pelos respectivos tamanhos:

$$V(\bar{X}_{AAS}) = \frac{1}{n} \times V(X) \geq \frac{1}{n} \times \sum_{i=1}^{k} \frac{N_i}{N} . V(X_i)$$

A expressão à direita corresponde justamente à variância da média amostral global em uma amostragem estratificada proporcional. Sabendo que $V(\bar{X}_{Est}) = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 V(\overline{x_i})$, que $V(\overline{x_i}) = \frac{V(X_i)}{n_i}$ e que, na alocação proporcional, $n_i = \frac{N_i}{N} . n$, então:

$$V(\bar{X}_{EP}) = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 \frac{V(X_i)}{n_i} = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 \frac{V(X_i).N}{N_i.n} = \sum_{i=1}^{k} \frac{N_i}{N} \frac{V(X_i)}{n}$$

Sendo a variância populacional de cada estrato **desconhecida**, precisamos estimá-la a partir da variância amostral de cada estrato $s^2_{\bar{x}_i}$. Substituindo, na fórmula acima, $V(\bar{X})$ por $s^2_{\bar{x}}$ e $V(\overline{x_i})$ por $s^2_{\bar{x}_i}$, temos:

$$s^2_{\bar{x}} = \sum_{h=1}^{k} \left(\frac{N_i}{N}\right)^2 \times s^2_{\bar{x}_i}$$

Em que a estimativa da variância da média amostral para cada estrato $s^2_{\bar{x}_i}$, considerando uma amostra de tamanho n_i para esse estrato, com a **correção para população finita**, é dada por:

$$s^2_{\bar{x}_i} = \frac{s^2_{x_i}}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right)$$

Vamos supor que as estimativas da variância para cada estrato sejam:

Estrato	N_i	n_i	$s^2_{x_i}$
1	50	5	2
2	30	3	1,5
3	20	2	1

Assim, as variâncias das médias amostrais para cada estrato, com o fator de correção, são:

$$s^2_{\bar{x}_1} = \frac{2}{5}\left(\frac{50-5}{50-1}\right) \cong 0{,}367$$

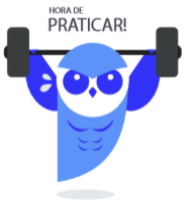
$$s^2_{\bar{x}_2} = \frac{1{,}5}{3}\left(\frac{30-3}{30-1}\right) \cong 0{,}466$$

$$s^2_{\bar{x}_3} = \frac{1}{2}\left(\frac{20-2}{20-1}\right) \cong 0{,}474$$

E a variância da média amostral global é dada por:

$$s^2_{\bar{x}} = \left(\frac{50}{100}\right)^2 \times 0{,}367 + \left(\frac{30}{100}\right)^2 \times 0{,}466 + \left(\frac{20}{100}\right)^2 \times 0{,}474$$

$$s^2_{\bar{x}} = 0{,}25 \times 0{,}367 + 0{,}09 \times 0{,}466 + 0{,}04 \times 0{,}474 = 0{,}092 + 0{,}042 + 0{,}019 = 0{,}153$$



(CESPE/2018 – STM) Um estudo acerca do tempo (x, em anos) de guarda de autos findos em determinada seção judiciária considerou uma amostragem aleatória estratificada. A população consiste de uma listagem de autos findos, que foi segmentada em quatro estratos, segundo a classe de cada processo (as classes foram estabelecidas por resolução de autoridade judiciária.)

A tabela a seguir mostra os tamanhos populacionais (N) e amostrais (n), a média amostral ($\bar{x}$) e a variância amostral dos tempos (s^2) correspondentes a cada estrato.

Estratos	Tamanhos Populacionais (N)	Tamanhos Amostra (n)	$\bar{x}$	s^2
A	30.000	300	20	3
B	40.000	400	15	16
C	50.000	500	10	5
D	80.000	800	5	8
Total	200.000	2.000	-	-

Considerando que o objetivo do estudo seja estimar o tempo médio populacional (em anos) de guarda dos autos findos, julgue os itens a seguir.

(CESPE/2018 – STM) A estimativa do tempo médio populacional da guarda dos autos findos é maior ou igual a 12 anos.

Comentários:

Em uma amostra aleatória por estratificação, a média amostral é dada por:

$$\bar{X} = \sum_{i=1}^{k} \frac{N_i}{N} \bar{x}_i$$

Em que k é a quantidade de estratos; N_i o tamanho de cada estrato; e $\bar{x}_i$ a média de cada estrato.

Pelos dados de N e $\bar{x}$ fornecidos na tabela, temos:

$$\bar{X} = \frac{N_A \times \bar{x_A} + N_B \times \bar{x_B} + N_C \times \bar{x_C} + N_D \times \bar{x_D}}{N}$$

$$\bar{X} = \frac{30000 \times 20 + 40000 \times 15 + 50000 \times 10 + 80000 \times 5}{200000}$$

$$\bar{X} = \frac{60 + 60 + 50 + 40}{20} = 10{,}5$$

Gabarito: Errado.

(CESPE/2018 – STM) Combinando-se todos os estratos envolvidos, a estimativa da variância do tempo médio amostral da guarda dos autos findos é inferior a $0{,}005\ ano^2$.

Comentários:

Em uma amostragem estratificada, a estimativa da variância da média amostral é dada por:

$$s_{\bar{x}}^2 = \sum_{i=1}^{k} \left(\frac{N_i}{N}\right)^2 \times s_{\bar{x}_i}^2$$

Em que $s_{\bar{x}_i}^2 = \frac{s_{\bar{x}_i}^2}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right)$. Pelos valores apresentados na tabela, temos:

$$s_{\bar{x}_A}^2 = \frac{3}{300}\left(\frac{30000 - 300}{30000 - 1}\right) = 0{,}0099$$

$$s_{\bar{x}_B}^2 = \frac{16}{400}\left(\frac{40000 - 400}{40000 - 1}\right) = 0{,}0396$$

$$s_{\bar{x}_C}^2 = \frac{5}{500}\left(\frac{50000 - 500}{50000 - 1}\right) = 0{,}0099$$

$$s_{\bar{x}_D}^2 = \frac{8}{800}\left(\frac{80000 - 800}{80000 - 1}\right) = 0{,}0099$$

Além disso, temos que:

$$\left(\frac{N_A}{N}\right)^2 = \left(\frac{30000}{200000}\right)^2 = 0{,}0225$$

$$\left(\frac{N_B}{N}\right)^2 = \left(\frac{40000}{200000}\right)^2 = 0{,}04$$

$$\left(\frac{N_C}{N}\right)^2 = \left(\frac{50000}{200000}\right)^2 = 0{,}0625$$

$$\left(\frac{N_D}{N}\right)^2 = \left(\frac{80000}{200000}\right)^2 = 0{,}16$$

Portanto:

$$s_{\bar{x}}^2 = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \times s_{\bar{x}_h}^2$$

$$s_{\bar{x}}^2 = (0{,}0225 \cdot 0{,}0099 + 0{,}04 \cdot 0{,}0396 + 0{,}0625 \cdot 0{,}0099 + 0{,}16 \cdot 0{,}0099)$$

$$s_{\bar{x}}^2 = (0{,}0225 + 0{,}0625 + 0{,}16) \cdot 0{,}0099 + 0{,}04 \cdot 0{,}0396 \cong 0{,}0041.$$

Gabarito: Certo.

ESTIMAÇÃO PONTUAL

A **estimação pontual** é o **valor** (número) calculado para o estimador. Os principais estimadores são a média amostral $\bar{X}$, a proporção amostral $\hat{p}$ e a variância amostral s^2.

Agora, vamos estudar as **propriedades** dos estimadores e os **métodos** utilizados para obtê-los.

Propriedades dos Estimadores

Um estimador é uma **função** dos dados observados (na amostra), o que chamamos de **estatística**.

Para que uma estatística possa ser considerada um bom estimador, ela deve apresentar determinadas características desejáveis, o que chamamos de **propriedades**.

Suficiência

Uma **estatística** é considerada **suficiente** se ela captura, a partir da amostra obtida, **toda a informação possível** sobre o parâmetro populacional desconhecido, de modo que qualquer outra informação associada à amostra não contribuirá com a estimação do parâmetro populacional.

Em outras palavras, uma estatística suficiente é uma forma de **resumir** as informações presentes na amostra, **sem perder as informações** necessárias para estimar o parâmetro populacional.

Por exemplo, a **média amostral** captura toda a informação, disponível na amostra, necessária para estimar a média populacional. Qualquer **outra** informação, como a média geométrica ou a variância da amostra, **não** influencia na estimativa da média populacional.

Ademais, se uma estatística T apresenta o **mesmo valor** para duas amostras X e Y, ou seja, $T(X) = T(Y)$, essa estatística será considerada suficiente se a **inferência** sobre o parâmetro populacional θ for a **mesma**, independente da amostra X ou Y.

Por exemplo, a **soma** dos valores da amostra também é uma estatística suficiente para a média populacional, pois dispomos de **toda** a informação necessária para estimar o parâmetro desconhecido - basta dividir essa estatística pelo tamanho da amostra n (que está previamente definido).

Não importa quais sejam os valores exatos de cada variável presente na amostra: se a soma (valor da estatística) for a mesma, a estimativa para a média populacional (parâmetro) será sempre a mesma.

Todos os estimadores que mencionamos anteriormente (média amostral $\bar{X}$, proporção amostral $\hat{p}$ e variância amostral s^2) são considerados suficientes.

Para resolver algumas questões sobre esse assunto, precisamos nos aprofundar um pouco mais. Se esse é um dos seus primeiros contatos com a matéria, não se preocupe muito com essa parte.

A definição exata de estatística suficiente é a seguinte:

> *Sendo $X_1, X_2, \dots X_n$ uma amostra aleatória de uma população com função de probabilidade $f(x)$ dependente de θ; uma estatística $T(X)$ será **suficiente** para θ se, e somente se, a distribuição conjunta de $X_1, X_2, \dots X_n$, condicionada a um valor da estatística $T(X) = t$, **não depender de θ**, para qualquer valor possível de t.*

Vamos supor uma população de Bernoulli com parâmetro θ (probabilidade de sucesso), em que cada elemento assume $X_i = 1$, com probabilidade θ, ou $X_i = 0$, com probabilidade $1 - \theta$.

A **função de probabilidade** de cada elemento é dada por:

$$f(x_i, \theta) = \theta^{x_i}.(1-\theta)^{1-x_i}$$

E a função de probabilidade **conjunta** para uma amostra independente de tamanho n é:

$$f(x_1, x_2, \dots, x_n, \theta) = f(x_1) \times f(x_2) \dots \times f(x_n)$$

$$f(x_1, x_2, \dots, x_n) = \theta^{x_1}.(1-\theta)^{1-x_1} \times \theta^{x_2}.(1-\theta)^{1-x_2} \times \dots \times \theta^{x_n}.(1-\theta)^{1-x_n}$$

Quando multiplicamos potências de mesma base, somamos os expoentes:

$$f(x_1, x_2, \dots, x_n) = \theta^{x_1+x_2+\cdots+x_n}.(1-\theta)^{n-(x_1+x_2+\cdots+x_n)} = \theta^{\sum_{i=1}^{n} x_i}.(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Vamos considerar a **estatística** $T(X) = \sum_{i=1}^{n} x_i$, que representa a soma dos sucessos em uma amostra de tamanho n. Essa variável segue distribuição binomial, com parâmetros n e θ, cuja função de probabilidade é dada por:

$$f(t) = C_{n,t}.\theta^t.(1-\theta)^{n-t}$$

Por fim, calculamos a função conjunta da amostra **condicionada** ao valor da estatística $t = \sum_{i=1}^{n} x_i$:

$$f(x_1, x_2, \dots, x_n | T = t) = \frac{f(x_1, x_2, \dots, x_n, t = \sum_{i=1}^{n} x_i)}{f(t)}$$

$$f(x_1, x_2, \dots, x_n | T = t) = \frac{\theta^t.(1-\theta)^{n-t}}{C_{n,t}.\theta^t.(1-\theta)^{n-t}} = \frac{1}{C_{n,t}}$$

Como essa função **não depende de θ**, concluímos que $T(X) = \sum_{i=1}^{n} x_i$ é estatística **suficiente** para estimar a probabilidade de sucesso de uma distribuição de Bernoulli.

Agora, vamos supor uma estatística que **não** considere todos os resultados da amostra. Por exemplo, uma amostra de tamanho $n = 3$ e uma estatística $T(X) = X_1 + X_3$.

Nesse caso, a função de probabilidade conjunta da amostra é:

$$f(x_1, x_2, x_3) = \theta^{x_1+x_2+x_3}.(1-\theta)^{3-(x_1+x_2+x_3)}$$

Sabendo que a estatística $T(X) = X_1 + X_3$ segue distribuição binomial com parâmetro 2 e θ, a sua função de probabilidade é:

$$f(t) = C_{2,t}.\theta^t.(1-\theta)^{2-t}$$

E a função conjunta da amostra condicionada ao valor da estatística $t = x_1 + x_3$ é:

$$f(x_1, x_2, \dots, x_n | T = t) = \frac{f(x_1, x_2, \dots, x_n, t = x_1 + x_3)}{f(t)}$$

$$f(x_1, x_2, \dots, x_n | T = t) = \frac{\theta^{t+x_2}.(1-\theta)^{3-(t+x_2)}}{C_{2,t}.\theta^t.(1-\theta)^{2-t}} = \frac{\theta^{x_2}.(1-\theta)^{1-x_2}}{C_{2,t}}$$

Que **depende de θ**. Assim, concluímos que a estatística $t = X_1 + X_3$ **não é suficiente** para estimar a proporção de sucessos.

Em alguns casos, o cálculo da função $f(x_1, x_2, \dots, x_n | T = t)$ pode ser muito trabalhoso. Nesses casos, podemos aplicar o **Teorema de Neyman-Fisher** ou **Teorema da Fatorização**:

> *Sendo $X_1, X_2, \dots X_n$ uma amostra aleatória de uma população com função de probabilidade $f(x, \theta)$ dependente de θ; uma estatística $T(X) = t$ será **suficiente** para θ se, e somente se, existir uma função $g(t, \theta)$ e uma função $h(x)$, tal que:*
>
> $$f(x, \theta) = g(t, \theta) \times h(x)$$
>
> *para todo valor possível de t.*

Ou seja, se você conseguir transformar a **função conjunta da amostra** f, que depende dos valores da amostra e do parâmetro θ, no **produto** de uma função g que dependa apenas da **estatística** t e do **parâmetro** θ, sem depender de qualquer outro valor da amostra, com uma outra função h que **não** dependa do **parâmetro** θ, podemos concluir que a estatística será **suficiente**.

Em relação ao exemplo da população de Bernoulli, vimos que a **função conjunta da amostra** é:

$$f(x, \theta) = \theta^{\sum_{i=1}^{n} x_i}.(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Se a estatística for $T(X) = \sum_{i=1}^{n} x_i$, temos:

$$f(x, \theta) = \theta^t . (1 - \theta)^{n-t}$$

Que **não** depende de qualquer valor da **amostra** além da estatística. Assim, temos:

$$f(x, \theta) = g(\theta, t) \times h(x)$$

tal que:

$$g(\theta, t) = \theta^t . (1 - \theta)^{n-t}$$

$$h(x) = 1$$

Como $h(x) = 1$ não depende do parâmetro θ, concluímos que $T(X) = \sum_{i=1}^{n} x_i$ é **suficiente** para estimar a proporção de sucessos de uma distribuição de Bernoulli.

Há, ainda, as seguintes definições relacionadas à suficiência de uma estatística:

- **Suficiente minimal**: se a estatística for suficiente e uma **função** de alguma outra estatística suficiente.
- **Anciliar**: se a estatística **não trouxer informação** alguma a respeito do parâmetro θ, isto é, se ela for **independente** de θ.
- **Completa**: se o fato de a **esperança** de uma função g da estatística ser igual a **zero** para todo θ, $E[g(T)] = 0$, implicar necessariamente em $\boldsymbol{g(T) = 0}$ para todo θ.

Por exemplo, para uma amostra $X_1, X_2, \dots, X_n$ com distribuição de Bernoulli, a estatística $T = X_1 - X_2$ **não é completa**.

Isso porque a esperança da estatística é igual a zero:

$$E[T] = E[X_1 - X_2] = E[X_1] - E[X_2] = p - p = 0$$

Porém, o valor da estatística $T = X_1 - X_2$ pode ser diferente de **zero**. Por exemplo, sendo $X_1 = 1$ e $X_2 = 0$, temos $T = 1$; sendo $X_1 = 0$ e $X_2 = 1$, temos $T = -1$.

(CESPE/2018 – EBSERH) X_1, X_2, ..., X_{10} representa uma amostra aleatória simples retirada de uma distribuição normal com média μ e variância σ^2, ambas desconhecidas. Considerando que $\hat{\mu}$ e $\widehat{\sigma^2}$ representam os respectivos estimadores de máxima verossimilhança desses parâmetros populacionais, julgue o item subsecutivo.

A soma $X_1 + X_2 + ... + X_{10}$ é uma estatística suficiente para a estimação do parâmetro μ.

Comentários:

A soma dos elementos captura todas as informações disponíveis na amostra para a estimação da média, de modo que **qualquer outra informação** (média, variância,...) a respeito dessa amostra **não irá influenciar na estimação do parâmetro populacional**. Afinal, a estimativa para o parâmetro μ será a **mesma**, sempre que a **soma** das 10 variáveis for a **mesma**, independentemente dos seus resultados exatos. Logo, a soma dos elementos é uma estatística **suficiente**.

Gabarito: Certo.

(CESPE/2013 – Telebras) A respeito de inferência estatística, julgue o item que se segue.

Considerando uma amostra aleatória simples X_1, X_2, X_3 retirada de determinada distribuição de Bernoulli, com parâmetro p desconhecido, é correto afirmar que $X_1+X_2.X_3$ é estatística suficiente.

Comentários:

Ao multiplicarmos os resultados $X_2.X_3$, há uma possibilidade de **perda de informação**, pois sendo um dos elementos iguais a zero, ficamos sem a informação do outro elemento. Por exemplo, se tivermos X_1=1, X_2=1 e X_3=0 e, portanto, $X_1 + X_2.X_3 = 1$, a estimativa para a proporção p será p = 2/3; e se tivermos X_1=1, X_2=0 e X_3=0, e, portanto, o mesmo valor para a estatística, $X_1 + X_2.X_3 = 1$, a estimativa para p será p = 1/3.

Ou seja, a **estimação** do parâmetro p desconhecido pode ser **diferente**, mesmo se o resultado da **estatística** for **igual**. Logo, essa estatística **não é suficiente**.

Alternativamente, podemos verificar o Teorema de Neyman-Fisher. A função conjunta da amostra é:

$$f(x,\theta) = f(x_1).f(x_2).f(x_3) = \theta^{x_1+x_2+x_3}.(1-\theta)^{3-(x_1+x_2+x_3)}$$

Sendo $t = x_1 + x_2.x_3 \rightarrow x_1 = t - x_2.x_3$*, então a função conjunta pode ser representada como:*

$$f(x,\theta) = \theta^{t-x_2.x_3+x_2+x_3}.(1-\theta)^{3-(t-x_2.x_3+x_2+x_3)}$$

Essa função corresponde ao seguinte produto:

$$f(x,\theta) = \theta^{t}.(1-\theta)^{3-t} \times \theta^{-x_2.x_3+x_2+x_3}.(1-\theta)^{x_2.x_3-x_2-x_3}$$

A primeira função $g = \theta^{t}.(1-\theta)^{3-t}$ *depende apenas de* θ *e* t*; no entanto, a segunda função* $h = \theta^{-x_2.x_3+x_2+x_3}.(1-\theta)^{x_2.x_3-x_2-x_3}$ *também depende de* θ*. Ou seja, não conseguimos transformar a função conjunta* $f(x,\theta)$ *em um produto de* $g(\theta,t)$ *por* $h(x)$ *e assim concluímos que a estatística não é suficiente.*

Gabarito: Errado.

(FGV/2022 – TRT/PB) Suponha que $X_1, X_2, \dots X_n$ seja uma amostra aleatória de densidade $f(x;\theta) = \theta.x^{\theta-1}$, se $0 < x < 1$, $f(x) = 0$ nos demais casos, $\theta > 0$. Uma estatística suficiente para θ é dada por:

a) $S = \sum_{i=1}^{n} X_i$

b) $S = \frac{1}{\sum_{i=1}^{n} X_i}$

c) $S = \sum_{i=1}^{n} X_i^2$

d) $S = \prod_{i=1}^{n} X_i$

e) $S = máx\{X_i\}$

Comentários:

Para resolver essa questão, vamos utilizar o teorema de Neyman-Fisher. A função conjunta da amostra é:

$$f(x,\theta) = \theta.x_1^{\theta-1} \times \dots \times \theta.x_n^{\theta-1} = \theta^n(x_1 \times \dots \times x_n)^{\theta-1}$$

Para a estatística $S = \prod_{i=1}^{n} X_i = x_1 \times \dots \times x_n$, temos:

$$f(x,\theta) = \theta^n.S^{\theta-1}$$

Que **não** depende de qualquer valor da **amostra** além da estatística. Assim, temos $f(x,\theta) = g(\theta,S) \times h(x)$, tal que:

$$g(\theta,S) = \theta^n.S^{\theta-1}$$

$$h(x) = 1$$

Assim concluímos que $S = \prod_{i=1}^{n} X_i$ é uma estatística suficiente para θ. Com qualquer outra estatística indicada dentre as alternativas, não será possível aplicar o teorema.

Gabarito: D

Não viés

Dizemos que um estimador $\hat{\theta}$ é **não viesado** (também chamado de **não viciado** ou **não tendencioso**) quando a sua **esperança** é igual ao **parâmetro populacional** θ sendo estimado:

$$E(\widehat{\theta_{NT}}) = \theta$$

Em relação aos principais estimadores que mencionamos anteriormente, temos:

$$E(\bar{X}) = \mu$$

$$E(\hat{p}) = p$$

$$E(s^2) = \sigma^2$$

Logo, podemos afirmar que a média amostral $\bar{X}$ é um estimador não viesado (ENV) para a média populacional; o parâmetro amostral $\hat{p}$ é um ENV para o parâmetro populacional; e s^2 é um ENV para a variância populacional.

TOME NOTA!

O estimador **não tendencioso** para a variância populacional é definido como:

$$s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

Já o estimador $s_*^2 = \frac{\sum(X_i - \bar{X})^2}{n}$, com a divisão por n em vez de $n-1$, é **tendencioso**, ou seja:

$$E(s_*^2) \neq \sigma^2$$

Apesar de s^2 ser um estimador não tendencioso para a variância populacional, a sua raiz quadrada $s = \sqrt{s^2}$ é um estimador **tendencioso** para o desvio padrão populacional, ou seja, a esperança de s é **diferente** do desvio padrão populacional[1], embora seja o estimador mais comum.

Outra forma de descrever essa propriedade é com base no **erro**, que é a **diferença** entre o **estimador** e o **parâmetro** populacional:

$$e = \hat{\theta} - \theta$$

A **esperança** do erro é chamada de **viés do estimador** (ou **tendenciosidade**), denotado por $b(\hat{\theta})$:

$$b(\hat{\theta}) = E(e)$$

O viés pode ser calculado pelas propriedades da esperança:

$$b(\hat{\theta}) = E(e) = E(\hat{\theta} - \theta) = E(\hat{\theta}) - E(\theta)$$

$$b(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Assim, o viés representa a diferença entre a média (ou esperança) da estimativa e o parâmetro populacional estimado. Para um estimador $\widehat{\theta_{NT}}$ **não tendencioso**, a sua esperança é igual ao parâmetro populacional $E(\widehat{\theta_{NT}}) = \theta$. Logo, o **viés do estimador** (**esperança do erro**) é **nulo**:

$$b(\widehat{\theta_{NT}}) = E(\widehat{\theta_{NT}}) - \theta = \theta - \theta = 0$$

[1] Essa aparente discrepância decorre do fato de que a esperança da raiz quadrada é **diferente** da raiz quadrada da esperança:

$$E\left(\sqrt{s^2}\right) \neq \sqrt{E(s^2)} = \sqrt{\sigma^2} = \sigma$$

Cálculo do Viés para a Média

Para calcularmos o viés de um estimador para a **média populacional μ**, calculamos a **esperança** do estimador e subtraímos a média populacional μ.

No cálculo da esperança do estimador, utilizamos as **propriedades** da esperança, em especial:

- $E(X+Y) = E(X) + E(Y)$
- $E(k.X) = k.E(X)$
- Se X e Y forem independentes, então $E(X.Y) = E(X).E(Y)$

Também devemos considerar que a distribuição dos elementos da **amostra** é **igual** à distribuição da **população**. Em particular, temos $E(X_i) = \mu$.

Para uma amostra de n = 5 elementos X_1, X_2, X_3, X_4 e X_5, vamos calcular o **viés** do estimador $\hat{\theta} = X_1 + X_2 + X_3 - X_4 - X_5$ para a **média μ**. O primeiro passo é calcular a esperança:

$$E(\hat{\theta}) = E(X_1 + X_2 + X_3 - X_4 - X_5)$$

Pelas propriedades da esperança, temos:

$$E(\hat{\theta}) = E(X_1) + E(X_2) + E(X_3) - E(X_4) - E(X_5)$$

Como a amostra segue a mesma distribuição da população, temos $E(X_i) = E(X) = \mu$:

$$E(\hat{\theta}) = \mu + \mu + \mu - \mu - \mu = \mu$$

Como a esperança desse estimador é **igual** ao parâmetro estimado, o seu **viés é nulo** e o estimador é **não tendencioso**:

$$b(\hat{\theta}) = \mu - \mu = 0$$

Agora, vejamos **outro** estimador para a média μ: $\widehat{\theta'} = X_1 - X_2$. A sua esperança é:

$$E(\widehat{\theta'}) = E(X_1 - X_2) = E(X_1) - E(X_2) = \mu - \mu = 0$$

Como a esperança desse estimador é nula e, portanto, **diferente** do parâmetro estimado, o **viés é diferente de zero** e o estimador é **tendencioso**:

$$b(\hat{\theta}) = 0 - \mu = -\mu$$

Podemos também **combinar estimadores**, formando um **novo** estimador, por exemplo:

$$\check{\theta} = \hat{\theta} + \hat{\theta'}$$

Como saberemos se o novo estimador é tendencioso ou não?

Como fizemos anteriormente, aplicando a fórmula do viés e as propriedades da esperança:

$$b(\check{\theta}) = E(e) = E(\check{\theta}) - \theta$$

Em relação ao estimador $\check{\theta} = \hat{\theta} + \hat{\theta'}$ para a média populacional μ, o seu viés é dado por:

$$b(\check{\theta}) = E(e) = E(\check{\theta}) - \mu = E(\hat{\theta} + \hat{\theta'}) - \mu$$

Sabendo que a esperança da soma é a soma das esperanças (propriedade), temos:

$$b(\check{\theta}) = E(\hat{\theta}) + E(\hat{\theta'}) - \mu$$

Para esse exemplo, vimos que a esperança do primeiro estimador é $E(\hat{\theta}) = \mu$ e que a esperança do segundo estimador é $E(\hat{\theta'}) = 0$, logo:

$$b(\check{\theta}) = \mu + 0 - \mu = 0$$

Como o viés é nulo, esse estimador é **não tendencioso**!

Cálculo do Viés para a Variância

Para calcularmos o viés de um estimador para a variância populacional σ^2, também calculamos a sua **esperança** e subtraímos o parâmetro sendo estimado, qual seja a variância populacional σ^2.

Os estimadores para a variância são normalmente baseados no **quadrado** dos elementos da amostra X_i^2 e no **quadrado** da **média** amostral $\bar{X}^2$.

Nesse sentido, é importante pontuar que a **esperança** de $\boldsymbol{X_i^2}$ pode ser calculada pela fórmula da variância:

$$V(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = V(X) + [E(X)]^2$$

Sabendo que os elementos da amostra seguem a **mesma** distribuição da população, com média $E(X) = \mu$ e variância $V(X) = \sigma^2$, então:

$$\boldsymbol{E(X_i^2) = \sigma^2 + \mu^2}$$

Por exemplo, vamos supor que vamos extrair uma amostra de uma população com variância $\sigma^2 = 4$ e média $\mu = 5$. A **esperança** do **quadrado** de cada elemento da amostra é:

$$E(X_i^2) = 4 + 5^2 = 4 + 25 = 29$$

Similarmente, podemos calcular a **esperança** de $\boldsymbol{\bar{X}^2}$:

$$E(\bar{X}^2) = V(\bar{X}) + [E(\bar{X})]^2$$

Sabendo que a esperança da média amostral é $E(\bar{X}) = \mu$ e que a variância da média amostral é $V(\bar{X}) = \frac{\sigma^2}{n}$:

$$\boldsymbol{E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2}$$

Vamos supor que a amostra extraída da mesma população com variância $\sigma^2 = 4$ e média $\mu = 5$ tenha tamanho $n = 16$. A **esperança** do **quadrado** da **média amostral** é:

- $$E(X_i^2) = \frac{4}{16} + 5^2 = 0{,}25 + 25 = 25{,}25$$

Vamos calcular o **viés** do seguinte estimador para a **variância** populacional σ^2, a partir de uma amostra de n = 2 elementos X_1 e X_2, sendo $\bar{X}$ a média amostral:

$$\tilde{\theta} = X_1^2 + X_2^2 - 2\bar{X}^2$$

A **esperança** desse estimador é:

$$E(\tilde{\theta}) = E(X_1^2 + X_2^2 - 2\bar{X}^2) = E(X_1^2) + E(X_2^2) - 2.E(\bar{X}^2)$$

Sabemos que $E(X_i^2) = \sigma^2 + \mu^2$ e que $E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$, sendo $n = 2$, logo:

$$E(\tilde{\theta}) = \sigma^2 + \mu^2 + \sigma^2 + \mu^2 - 2\left(\frac{\sigma^2}{2} + \mu^2\right) = 2.\sigma^2 + 2.\mu^2 - \sigma^2 - 2.\mu^2 = \sigma^2$$

Como a esperança do estimador é **igual** à variância populacional, que é o parâmetro sendo estimado, o viés é nulo $b(\tilde{\theta}) = E(\tilde{\theta}) - \sigma^2 = 0$ e o estimador é não tendencioso.

Erro Quadrático Médio

A **variância do erro** é chamada de **Erro Quadrático Médio** (EQM) e corresponde à soma da **variância** do estimador com o **quadrado do viés** do estimador.

$$EQM(\hat{\theta}) = V(e) = V(\hat{\theta}) + [b(\hat{\theta})]^2$$

Ou, simplesmente:

$$EQM(\hat{\theta}) = \sigma^2 + b^2$$

Para estimadores **não viesados**, o viés do estimador é $b(\widehat{\theta_{NT}}) = 0$, logo, o EQM é igual à **variância** do estimador:

$$EQM(\widehat{\theta_{NT}}) = V(\widehat{\theta_{NT}})$$

Vamos supor a seguinte expressão para o **EQM** de um estimador, em que a primeira parcela corresponde à **variância** do estimador $V(\hat{\theta})$ e a segunda parcela corresponde ao **quadrado do viés** do estimador $[b(\hat{\theta})]^2$ (exemplo extraído da prova FGV/2017 – IBGE):

$$EQM(\hat{\theta}_1) = \frac{3.\sigma^2}{n} + \left(\frac{\theta - 1}{n}\right)^2$$

Podemos observar que o **viés** $\left(\frac{\theta-1}{n}\right)$ **não é nulo**, logo, concluímos que o estimador é **tendencioso**.

Agora, vamos calcular o EQM dos estimadores para a média μ, que vimos anteriormente. Para isso, precisamos da variância e do viés do estimador. Como já calculamos o viés, falta calcular a sua **variância**.

$$\hat{\theta} = X_1 + X_2 + X_3 - X_4 - X_5$$

$$V(\hat{\theta}) = V(X_1 + X_2 + X_3 - X_4 - X_5)$$

Se as amostras forem **independentes** (população infinita ou amostra extraída com reposição), então, pelas propriedades da variância, temos:

$$V(\hat{\theta}) = V(X_1) + V(X_2) + V(X_3) + V(X_4) + V(X_5)$$

Considerando que a amostra segue a mesma distribuição da população, temos $V(X_i) = V(X) = \sigma^2$, logo:

$$V(\hat{\theta}) = \sigma^2 + \sigma^2 + \sigma^2 + \sigma^2 + \sigma^2 = 5\sigma^2$$

Como o viés é nulo, o EQM desse estimador é igual variância:

$$EQM(\hat{\theta}) = V(\hat{\theta}) = 5\sigma^2$$

Para o outro exemplo de estimador, a variância é (sendo as amostras **independentes**):

$$\widehat{\theta'} = X_1 - X_2$$

$$V(\widehat{\theta'}) = V(X_1 - X_2) = V(X_1) + V(X_2) = \sigma^2 + \sigma^2 = 2\sigma^2$$

Já calculamos o viés do estimador $b(\widehat{\theta'}) = -\mu$. Portanto, o EQM é:

$$EQM(\widehat{\theta'}) = V(\hat{\theta}) + [b(\hat{\theta})]^2 = 2\sigma^2 + [-\mu]^2 = 2\sigma^2 + \mu^2$$

Se as amostras **não** forem **independentes**, a variância da soma e da subtração de duas variáveis é dada por:

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2.Cov(X_1, X_2)$$

$$V(X_1 - X_2) = V(X_1) + V(X_2) - 2.Cov(X_1, X_2)$$

(2019/Analista Censitário – *Adaptada*) É conhecido que o erro quadrático médio (EQM($\hat{\theta}$)) mede, em média, quão perto um estimador $\hat{\theta}$ chega ao valor real do parâmetro θ. Diante do exposto, julgue os itens seguintes:

I – EQM($\hat{\theta}$) é uma medida que combina viés e variância de um parâmetro;

II – para estimadores viesados, EQM($\hat{\theta}$) é igual à variância;

III – EQM($\hat{\theta}$) = $\hat{\theta}$ – Viés($\hat{\theta}$).

Comentários:

Em relação ao primeiro item, o EQM($\hat{\theta}$) combina viés e variância de um **estimador** (não do parâmetro populacional), logo o item I está **incorreto**.

Em relação ao segundo item, o EQM($\hat{\theta}$) é igual à variância para um estimador **não** viesado, logo o item II está **incorreto**.

Em relação ao terceiro item, a fórmula do erro quadrático médio é:

$$EQM(\hat{\theta}) = V(\hat{\theta}) + [Viés(\hat{\theta})]^2$$

Logo, o item III está **incorreto**.

Resposta: todos os itens errados.

(FGV/2022 - EPE – *Adaptada*) Considere uma amostra aleatória de n variáveis X_1, X_2, ..., X_n, normalmente distribuídas com média μ e variância σ^2. Considere o seguinte estimador da média populacional:

$$\bar{T} = \frac{1}{n^2}\sum_{i=1}^{n} X_i$$

Sobre as propriedades desse estimador, julgue o seguinte item.

O estimador é não-tendencioso.

Comentários:

Para um estimador não tendencioso, a sua esperança é igual ao parâmetro populacional estimado, no caso, a **média** μ. Vamos, então, calcular a esperança do estimador $\bar{T}$:

$$E(\bar{T}) = E\left(\frac{1}{n^2}\sum_{i=1}^{n} X_i\right)$$

Quando multiplicamos uma variável por uma constante, a sua esperança será multiplicada por essa constante (propriedade da esperança), logo:

$$E(\bar{T}) = \frac{1}{n^2}.E\left(\sum_{i=1}^{n} X_i\right)$$

Ademais, a esperança da soma de variáveis é igual à soma das esperanças (outra propriedade da esperança):

$$E(\bar{T}) = \frac{1}{n^2}.\sum_{i=1}^{n} E(X_i)$$

Considerando que cada variável da amostra segue a mesma distribuição da população, temos $E(X_i) = \mu$. Portanto, a soma $\sum_{i=1}^{n} E(X_i)$ para as n variáveis corresponde ao produto $n.\mu$:

$$E(\bar{T}) = \frac{1}{n^2}.n.\mu = \frac{\mu}{n}$$

Que é **diferente** de μ. Logo, o estimador $\bar{T}$ é **tendencioso**.

Resposta: Errado.

(FGV/2022 - TRT/MA) Suponha uma amostra X_1, X_2, X_3, X_4 de uma variável populacional com média μ e variância σ^2. Se $\bar{X}$ é a média amostral, assinale a opção que apresenta uma estatística não tendenciosa para σ^2:

a) $(X_1^2 + X_2^2 + X_3^2 + X_4^2 - 4.\bar{X}^2)/3$

b) $(X_1^2 + X_2^2 + X_3^2 + X_4^2)/3$

c) $(X_1^2 + X_2^2 + X_3^2 + X_4^2)/4$

d) $(X_1^2 + X_2^2 + X_3^2 + X_4^2 - \bar{X}^2)/3$

e) $(X_1^2 + X_2^2 + X_3^2 + X_4^2 - 4.\bar{X}^2)/4$

Comentários:

Para identificar o estimador não tendencioso para a **variância** σ^2, precisamos calcular a sua esperança.

A esperança do estimador da alternativa A é:

$$E(\theta_A) = E\left(\frac{X_1^2 + X_2^2 + X_3^2 + X_4^2 - 4.\bar{X}^2}{3}\right) = \frac{E(X_1^2) + E(X_2^2) + E(X_3^2) + E(X_4^2) - 4.E(\bar{X}^2)}{3}$$

Sendo $E(X_i^2) = \sigma^2 + \mu^2$ e $E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$, com $n = 4$, então:

$$E(\theta_A) = \frac{\sigma^2 + \mu^2 + \sigma^2 + \mu^2 + \sigma^2 + \mu^2 + \sigma^2 + \mu^2 - 4\left(\frac{\sigma^2}{4} + \mu^2\right)}{3}$$

$$E(\theta_A) = \frac{4.\sigma^2 + 4.\mu^2 - \sigma^2 - 4.\mu^2}{3} = \frac{3.\sigma^2}{3} = \sigma^2$$

Que é **igual** ao parâmetro σ^2 estimado. Logo, esse estimador é **não tendencioso**.

Mas, vejamos as demais alternativas. A esperança do estimador da alternativa B é:

$$E(\theta_B) = E\left(\frac{X_1^2 + X_2^2 + X_3^2 + X_4^2}{3}\right) = \frac{E(X_1^2) + E(X_2^2) + E(X_3^2) + E(X_4^2)}{3}$$

Sendo $E(X_i^2) = \sigma^2 + \mu^2$:

$$E(\theta_B) = \frac{\sigma^2 + \mu^2 + \sigma^2 + \mu^2 + \sigma^2 + \mu^2 + \sigma^2 + \mu^2}{3} = \frac{4.\sigma^2 + 4.\mu^2}{3} = \frac{4}{3}(\sigma^2 + \mu^2)$$

Que é **diferente** de σ^2. Logo, esse estimador é **tendencioso**.

Em relação ao estimador da alternativa C, temos:

$$E(\theta_C) = E\left(\frac{X_1^2 + X_2^2 + X_3^2 + X_4^2}{4}\right) = \frac{E(X_1^2) + E(X_2^2) + E(X_3^2) + E(X_4^2)}{4}$$

Sendo $E(X_i^2) = \sigma^2 + \mu^2$:

$$E(\theta_C) = \frac{\sigma^2 + \mu^2 + \sigma^2 + \mu^2 + \sigma^2 + \mu^2 + \sigma^2 + \mu^2}{4} = \frac{4.\sigma^2 + 4.\mu^2}{4} = \sigma^2 + \mu^2$$

Que também é **diferente** de σ^2. Logo, esse estimador é **tendencioso**. Vejamos o estimador da alternativa D:

$$E(\theta_D) = E\left(\frac{X_1^2 + X_2^2 + X_3^2 + X_4^2 - \bar{X}^2}{3}\right) = \frac{E(X_1^2) + E(X_2^2) + E(X_3^2) + E(X_4^2) - E(\bar{X}^2)}{3}$$

Sendo $E(X_i^2) = \sigma^2 + \mu^2$ e $E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$, com $n = 4$, então:

$$E(\theta_D) = \frac{\sigma^2 + \mu^2 + \sigma^2 + \mu^2 + \sigma^2 + \mu^2 + \sigma^2 + \mu^2 - \left(\frac{\sigma^2}{4} + \mu^2\right)}{3}$$

$$E(\theta_D) = \frac{4.\sigma^2 + 4.\mu^2 - \frac{\sigma^2}{4} - 4.\mu^2}{3} = \frac{4.\sigma^2}{3} - \frac{\sigma^2}{12} = \frac{16.\sigma^2 - \sigma^2}{12} = \frac{15.\sigma^2}{12} = \frac{5}{4}\sigma^2$$

Que é **diferente** de σ^2. Logo, esse estimador é **tendencioso**.

Por fim, a esperança do estimador da alternativa E é:

$$E(\theta_E) = E\left(\frac{X_1^2 + X_2^2 + X_3^2 + X_4^2 - 4.\bar{X}^2}{4}\right) = \frac{E(X_1^2) + E(X_2^2) + E(X_3^2) + E(X_4^2) - 4.E(\bar{X}^2)}{4}$$

Sendo $E(X_i^2) = \sigma^2 + \mu^2$ e $E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$, com $n = 4$, então:

$$E(\theta_D) = \frac{\sigma^2 + \mu^2 + \sigma^2 + \mu^2 + \sigma^2 + \mu^2 + \sigma^2 + \mu^2 - 4\left(\frac{\sigma^2}{4} + \mu^2\right)}{4}$$

$$E(\theta_A) = \frac{4.\sigma^2 + 4.\mu^2 - \sigma^2 - 4.\mu^2}{4} = \frac{3.\sigma^2}{4}$$

Que também é **diferente** de σ^2. Logo, esse estimador também é **tendencioso**.

Gabarito: A

(CESPE 2020/TJ-PA) Um estimador que fornece a resposta correta em média é chamado não enviesado. Formalmente, um estimador é não enviesado caso seu valor esperado seja igual ao parâmetro que está sendo estimado. Os possíveis estimadores para a média populacional (μ) incluem β, média de uma amostra, α, a menor observação da amostra, e π, a primeira observação coletada de uma amostra. Considerando essas informações, julgue os itens subsequentes.

I A média de uma amostra (β) é exemplo de um estimador enviesado para a média populacional (μ), pois seu valor esperado é igual à média populacional, ou seja, E(β) = μ.

II A menor observação da amostra (α) é um exemplo de estimador não enviesado, pois o valor da menor observação da amostra deve ser inferior à média da amostra; portanto, E(α) < μ.

III A primeira observação coletada de uma amostra equivale a tomar ao acaso uma amostra aleatória da população de tamanho igual a um e, portanto, é considerado um estimador não enviesado.

Assinale a opção correta.

a) Nenhum item está certo.

b) Apenas o item I está certo.

c) Apenas o item II está certo.

d) Apenas o item III está certo.

e) Todos os itens estão certos.

Comentários:

A questão trabalha com alguns estimadores: a média amostral β, que normalmente chamamos de $\bar{X}$; a menor observação da amostra, α; e a primeira observação, π.

Em relação ao item I, o valor esperado da média amostral é **igual** à média populacional, E(β) = μ, o que define um estimador **não** enviesado. Logo, o item I está errado.

Em relação ao item II, o valor esperado da **menor** observação da amostra é inferior à média, E(α) < μ, o que define um estimador **enviesado**. Logo, o item II está errado.

Em relação ao item III, a primeira observação pode ser considerada uma amostra com tamanho **n = 1**, de fato, sendo E(π) = μ. Logo, esse estimador é **não enviesado**. Logo, o item III está **certo**.

Gabarito: D.

Eficiência

Vimos que, para um estimador não tendencioso, o erro quadrático médio é igual à sua **variância**:

$$EQM(\widehat{\theta_{NT}}) = V(\widehat{\theta_{NT}})$$

Assim, a **variância** de um estimador não tendencioso está inversamente relacionada à **precisão** da estimativa: quanto **menor** a variância, **maior** a precisão.

Por sua vez, a **precisão** da estimativa está diretamente relacionada à **eficiência** do estimador. Ou seja, quanto **menor** a **variância** do estimador, **maior** será sua **eficiência**.

Essa comparação pode ser feita entre estimadores não tendenciosos, ou entre estimadores com o **mesmo viés**. Não podemos comparar a eficiência entre estimadores com viés diferente.

Dizemos que um estimador é **eficiente** se for **não viesado** e apresentar a **menor variância possível**.

Tanto a **média amostral** quanto a **proporção amostral** são estimadores **eficientes**.

Um estimador é dito **assintoticamente eficiente** quando a sua matriz de variância-covariância assintótica não é maior que a de qualquer outro estimador, ou seja, quando a sua variância **converge mais rapidamente**.

Para resolver algumas questões, precisamos nos aprofundar um pouco mais. Se esse é um dos seus primeiros contatos com a matéria, não se preocupe com a parte a seguir.

Existe um **limite inferior** para a variância de um estimador **não tendencioso**, chamado limite de **Cramér-Rao**. Nenhum estimador não tendencioso terá uma variância inferior a esse limite. Quando um estimador não tendencioso apresenta variância **igual** ao limite de Cramér-Rao, podemos concluir que tal estimador é **eficiente**, pois apresenta a menor variância possível.

Por exemplo, para uma população com distribuição normal, o limite de Cramér-Rao para estimadores não tendenciosos da média populacional é $\frac{\sigma^2}{n}$, que é justamente a variância da **média amostral** $\bar{X}$. Por se tratar de um estimador não tendencioso com a menor variância possível, concluímos que esse estimador é **eficiente**.

No entanto, nem sempre o limite de Cramér-Rao **pode** ser atingido por um estimador não tendencioso. Por exemplo, para uma população com distribuição normal, o limite de Cramér-Rao para estimadores da variância populacional é $\frac{2.\sigma^4}{n}$, porém **não** há estimador não tendencioso com essa variância - todos apresentam variância maior que esse limite.

Em outras palavras, a variância de um estimador eficiente, isto é, de um estimador não tendencioso com a menor variância possível, **não** será necessariamente igual ao limite de Cramér-Rao.

O limite inferior de Cramér-Rao é o inverso da chamada **Informação de Fisher** I:

$$Var(\hat{\theta}) \geq \frac{1}{I}$$

em que $I = E\left[\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2\right]$, sendo $f(x,\theta)$ a função de probabilidade conjunta da amostra e $\ln f(x,\theta)$ o seu logaritmo natural, chamado função **log-verossimilhança**. A esperança E é calculada para todos os possíveis valores de x, ponderados pela função de probabilidade.

Quando algumas condições de regularidade estiverem presentes, a Informação de Fisher pode ser calculada como:

$$I = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f(x,\theta)\right]$$

Por exemplo, para a distribuição **exponencial**, a função densidade de probabilidade de cada elemento é:

$$f(x,\theta) = \theta . e^{-\theta x}$$

E a **função de probabilidade conjunta** para uma amostra de tamanho n (função de verossimilhança) é:

$$f(x,\theta) = \theta . e^{-\theta x_1} \times \theta . e^{-\theta x_2} \times \dots \times \theta . e^{-\theta x_n} = \theta^n . e^{-\theta \sum_{i=1}^{n} x_i}$$

A função log-verossimilhança é o **logaritmo natural** dessa função:

$$\ln f(x,\theta) = \ln\left(\theta^n . e^{-\theta \sum_{i=1}^{n} x_i}\right)$$

Sabendo que o logaritmo do produto é igual à soma dos logaritmos ($\ln a.b = \ln a + \ln b$); que o logaritmo da potência é o produto do expoente pelo logaritmo ($\ln a^b = b.\ln a$); e que $\ln e = 1$, temos:

$$\ln f(x,\theta) = n.\ln\theta - \theta \sum_{i=1}^{n} x_i . \ln e = n.\ln\theta - \theta \sum_{i=1}^{n} x_i$$

Agora, **derivamos** essa função em relação a θ, sabendo que $\frac{d}{dx}\ln x = x^{-1}$:

$$\frac{\partial}{\partial\theta}\ln f(x,\theta) = n.\theta^{-1} - \sum_{i=1}^{n} x_i$$

E **derivamos novamente** para utilizar o cálculo **alternativo** da Informação de Fisher, uma vez que as condições de regularidade estão presentes:

$$\frac{\partial^2}{\partial\theta^2}\ln f(x,\theta) = -n.\theta^{-2}$$

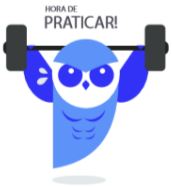
Como esse resultado é **independente de x**, a sua esperança é igual à própria expressão, logo:

$$I = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f(x,\theta)\right] = \frac{\partial^2}{\partial\theta^2}\ln f(x,\theta) = -(-n.\theta^{-2}) = \frac{n}{\theta^2}$$

E o **limite mínimo** de Cramér-Rao para a variância de um estimador não tendencioso para a distribuição exponencial é o inverso desse resultado:

$$Var(\hat{\theta}) \geq \frac{1}{I} = \frac{\theta^2}{n}$$

No entanto, não há estimador não tendencioso que atinja esse limite.



(CESPE/2019 – TJ-AM) Com relação aos parâmetros estatísticos e suas estimativas, julgue o item que se segue.

Entre dois estimadores, A e B, com consistência, viés e demais características iguais, o estimador mais útil é aquele que possui menor variância.

Comentários:

Para estimadores com todas as demais características iguais, o melhor estimador será aquele com a **menor variância**, que está associada ao erro quadrático médio (EQM).

Gabarito: Certo

(2018 – Petrobras) Seja (Y_1 , Y_2 , Y_3) uma amostra aleatória simples extraída de modo independente de uma população com média μ e variância σ^2, ambas desconhecidas. Considere os dois estimadores da média da população definidos abaixo

$\widehat{\mu_1} = \frac{Y_1+2.Y_2+3.Y_3}{6}$ e $\widehat{\mu_2} = \frac{Y_1+Y_2+Y_3}{3}$

Relativamente a esses dois estimadores, conclui-se que:

a) apenas o primeiro estimador é não tendencioso.

b) apenas o segundo estimador é não tendencioso.

c) os dois estimadores são não tendenciosos, mas a eficiência não pode ser determinada sem a estimação da variância σ^2.

d) os dois estimadores são não tendenciosos, mas o primeiro é mais eficiente por apresentar a variância inferior à do segundo estimador.

e) os dois estimadores são não tendenciosos, mas o segundo é mais eficiente por apresentar a variância inferior à do primeiro estimador.

Comentários:

O **viés** do estimador é a **esperança** do seu **erro**:

$$b(\hat{\theta}) = E(e) = E(\hat{\theta} - \theta) = E(\hat{\theta}) - \theta$$

O viés do primeiro estimador é:

$$b(\widehat{\mu_1}) = E\left(\frac{Y_1 + 2.Y_2 + 3.Y_3}{6}\right) - \mu$$

Pelas propriedades da esperança, temos:

$$b(\widehat{\mu_1}) = \frac{1}{6}[E(Y_1) + 2.E(Y_2) + 3.E(Y_3)] - \mu$$

Sabendo que a distribuição da amostra é igual à distribuição da população, temos $E(Y_1) = E(Y) = \mu$:

$$b(\widehat{\mu_1}) = \frac{1}{6}[\mu + 2.\mu + 3.\mu] - \mu = \frac{1}{6}[6.\mu] - \mu = \mu - \mu = 0$$

Logo, o primeiro estimador é não viesado (alternativa B errada).

Similarmente, o viés do segundo estimador é dado por:

$$b(\widehat{\mu_2}) = E\left(\frac{Y_1 + Y_2 + Y_3}{3}\right) - \mu$$

$$b(\widehat{\mu_2}) = \frac{1}{3}[E(Y_1) + E(Y_2) + E(Y_3)] - \mu$$

$$b(\widehat{\mu_2}) = \frac{1}{3}[\mu + \mu + \mu] - \mu = \frac{1}{3}[3.\mu] - \mu = \mu - \mu = 0$$

Ou seja, o segundo estimador também não é viesado (alternativa A errada).

A eficiência do estimador não viesado é medida por sua variância. A variância do primeiro estimador é:

$$V(\widehat{\mu_1}) = V\left(\frac{Y_1 + 2.Y_2 + 3.Y_3}{6}\right)$$

Pelas propriedades da variância, para variáveis independentes, temos:

$$V(\widehat{\mu_1}) = \frac{1}{6^2}[V(Y_1) + 2^2.V(Y_2) + 3^2.V(Y_3)] = \frac{1}{36}[V(Y_1) + 4.V(Y_2) + 9.V(Y_3)]$$

Sabendo que a distribuição da amostra é igual à distribuição da população, temos $V(Y_1) = V(Y) = \sigma^2$:

$$V(\widehat{\mu_1}) = \frac{1}{36}[\sigma^2 + 4.\sigma^2 + 9.\sigma^2] = \frac{1}{36}[14.\sigma^2] = \frac{7}{18}\sigma^2$$

Similarmente, a variância do segundo estimador é dada por:

$$V(\widehat{\mu_2}) = V\left(\frac{Y_1 + Y_2 + Y_3}{3}\right) = \frac{1}{3^2}[V(Y_1) + V(Y_2) + V(Y_3)] = \frac{1}{9}[V(Y_1) + V(Y_2) + V(Y_3)]$$

$$V(\widehat{\mu_2}) = \frac{1}{9}[\sigma^2 + \sigma^2 + \sigma^2] = \frac{1}{9}[3.\sigma^2] = \frac{3}{9}\sigma^2 = \frac{6}{18}\sigma^2$$

A variância do segundo estimador é **menor** que a do primeiro, então o segundo estimador é **mais eficiente**.

Gabarito: E

Consistência

Para um estimador consistente $\widehat{\theta_C}$, as suas estimativas **convergem** para o parâmetro populacional θ com o **aumento do tamanho amostral**.

A definição de **consistência** é:

$$\lim_{n\to\infty} P\left(\left|\widehat{\theta_C} - \theta\right| > \varepsilon\right) = 0$$

para algum valor $\varepsilon > 0$ pequeno.

Ou seja, quando o tamanho da amostra n tende a infinito, a probabilidade de um estimador consistente $\hat{\theta}_C$ **diferir** do parâmetro verdadeiro θ, em mais do que ε, é **nula**.

Todos os estimadores que mencionamos anteriormente (média amostral $\bar{X}$, proporção amostral $\hat{p}$ e variância amostral s^2) são consistentes.

Ademais, a estimativa tendenciosa para a variância amostral $s_*^2 = \frac{\sum(X_i - \bar{X})^2}{n}$ também é consistente.

O estimador $\hat{\theta}$ será consistente **se** a sua esperança $E(\hat{\theta})$ tende ao parâmetro populacional θ e sua variância $V(\hat{\theta})$ tende a zero, quando o **tamanho da amostra** n tende ao **infinito**:

$$\lim_{n\to\infty} E(\hat{\theta}) = \theta$$

$$\lim_{n\to\infty} V(\hat{\theta}) = 0$$

Podemos deduzir se essas duas características estão presentes ou não, a partir da fórmula do **Erro Quadrático Médio** do estimador:

$$EQM(\hat{\theta}) = V(\hat{\theta}) + \left[b(\hat{\theta})\right]^2$$

Vamos considerar o exemplo da prova da FGV/2017 (IBGE), em que a primeira expressão corresponde à variância do estimador $V(\widehat{\theta_1})$ e a segunda expressão ao quadrado do viés do estimador $\left[b(\widehat{\theta_1})\right]^2$:

$$EQM(\hat{\theta}_1) = \frac{3.\sigma^2}{n} + \left(\frac{\theta - 1}{n}\right)^2$$

Podemos deduzir que o viés é:

$$b(\hat{\theta}_1) = \frac{\theta - 1}{n}$$

Apesar de o viés não ser nulo, em sua expressão, consta uma divisão por n. Por isso, conforme n aumenta, o viés diminui. No limite, quando n tende a **infinito**, o viés é igual a **zero**, portanto, pode-se dizer que o estimador é **assintoticamente não tendencioso**.

E a variância é:

$$V(\hat{\theta}_1) = \frac{3.\sigma^2}{n}$$

Nessa expressão, também consta uma divisão por n. Logo, quando n tende a infinito, a sua variância é igual a zero.

Como o estimador apresenta as duas características, podemos concluir que é **consistente**.

Se o estimador **apresentar** essas duas propriedades, podemos concluir que é **consistente**.

Porém, se o estimador **não** apresentar essas propriedades assintóticas, **não** podemos concluir que o estimador é **inconsistente**.

Em outras palavras, um estimador pode **não** apresentar as duas propriedades e, ainda assim, ser **consistente**.

Ou seja, essas propriedades são condições **suficientes**, que nos permitem concluir que o estimador é consistente; porém, **não** são **necessárias** para que um estimador seja consistente.

(CESPE/2019 – TJ-AM) Com relação aos parâmetros estatísticos e suas estimativas, julgue o item que se segue.

Situação hipotética: A e B são dois estimadores não viciados e diferentes utilizados para estimar um mesmo parâmetro. A variância de A é menor que a variância de B. Assertiva: Em relação à consistência desses estimadores, é correto afirmar que o estimador A é mais consistente que o B.

Comentários:

Um estimador é dito consistente quando a esperança tende ao estimador e a variância tende a zero, quando a **amostra cresce**.

Logo, o fato de a variância de A ser menor do que a variância de B não nos permite afirmar que A é mais consistente que B.

Gabarito: Errado

(FGV/2019 – DPE-RJ – *Adaptada*) Sobre as principais propriedades dos estimadores pontuais, para pequenas e grandes amostras, julgue os itens a seguir.

I – Se $\lim_{n\to\infty} EQM(\hat{\theta}) = +\infty$, então $\lim_{n\to\infty} P\left(\left|\hat{\theta} - \theta\right| < \varepsilon\right) \neq 0$.

II – Um estimador que seja assintoticamente tendencioso não poderá ser consistente.

Comentários:

Esses 2 itens trabalham com o fato de que as propriedades assintóticas são condições **suficientes**, porém **não necessárias** para a consistência do estimador.

O item I descreve um estimador cujo erro quadrático médio (EQM) cresce com o aumento do tamanho amostral n.

Como o EQM é a soma da variância com o quadrado do viés, $EQM(\hat{\theta}) = V(\hat{\theta}) + \left[b(\hat{\theta})\right]^2$, então **ou** o viés cresce com o aumento de n **ou** a variância cresce com o aumento de n, ou ambos.

Portanto, **não** podemos concluir que o estimador é **consistente**, pois ele **não** segue as propriedades assintóticas.

Por outro lado, também **não** podemos concluir que o estimador é **inconsistente**, pois é possível que um estimador não siga as propriedades assintóticas, mas ainda assim seja consistente.

Ou seja, não podemos afirmar que $\lim_{n\to\infty} P\left(\left|\hat{\theta} - \theta\right| < \varepsilon\right) \neq 0$. Por isso, o item I está errado.

Similarmente, em relação ao item II, um estimador pode ser consistente, ainda que seja assintoticamente tendencioso.

Resposta: Ambos os itens errados.

ESQUEMATIZANDO

Estatística Suficiente: Resume **todas as informações** disponíveis na amostra para estimar o parâmetro populacional

Estimador Não Viesado (ENV), ou **Não Viciado**, ou **Não Tendencioso**: Esperança do estimador igual ao parâmetro; **Viés** (esperança do erro) igual a **zero**:

$$E(\hat{\theta}) = \theta \leftrightarrow b(\hat{\theta}) = E(e) = 0$$

$s_*^2 = \frac{\sum(X_i - \bar{X})^2}{n}$ é **tendencioso**; $\bar{X}$, $\hat{p}$ e s^2 são **não tendenciosos**

Estimador Eficiente: apresenta a **menor** **variância** $V(\hat{\theta})$ possível. $\bar{X}$ e $\hat{p}$ são **eficientes**

Estimador Consistente: estimativas convergem para parâmetro populacional se:

$$\lim_{n\to\infty} E(\hat{\theta}) = \theta, \lim_{n\to\infty} V(\hat{\theta}) = 0$$

Mas essas condições **não** são necessárias para o estimador ser consistente.

Os estimadores $\bar{X}$, $\hat{p}$, s^2 e s_*^2 (estimador tendencioso para a variância) são consistentes.

Métodos de Estimação

Os estimadores são obtidos pelos chamados **métodos de estimação**. Veremos alguns desses métodos agora.

Método dos Momentos

O Método dos Momentos se baseia nos **momentos teóricos e amostrais** das variáveis aleatórias.

O k-ésimo **momento teórico** da variável X, denotado por μ_k, é:

$$\mu_k = E(X^k)$$

Se a variável for discreta, temos:

$$\mu_k = E(X^k) = \sum x^k . P(X = x)$$

Em particular, o **primeiro momento teórico** é igual à **esperança** da variável:

$$\mu_1 = E(X) = \sum x.P(X = x)$$

Se a variável for contínua, substituímos o somatório pela integral e a função de probabilidade $P(X = x)$ pela função densidade de probabilidade $f(x)$.

Pontue-se que é possível que distribuições **distintas** apresentem os **mesmos** momentos teóricos.

Os **momentos amostrais** são calculados de maneira análoga. O k-ésimo **momento amostral**, denotado por m_k, é a soma dos valores observados elevados a k, dividida pelo tamanho da amostra, n:

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Em particular, o **primeiro momento amostral** ($k = 1$) é a **média amostral**:

$$m_1 = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

E o **segundo momento amostral** ($k = 2$) corresponde a $\overline{X^2}$:

$$m_2 = \frac{\sum_{i=1}^{n} X_i^2}{n} = \overline{X^2}$$

No exemplo das alturas, em que os $n = 5$ valores da amostra observada foram {1,65; 1,75; 1,8; 1,85; 1,95}, o primeiro momento amostral é igual a média amostral:

$$m_1 = \bar{X} = \frac{\sum X_i}{n} = \frac{1{,}65 + 1{,}75 + 1{,}8 + 1{,}85 + 1{,}95}{5} = \frac{9}{5} = 1{,}8$$

E o segundo momento amostral é a soma dos valores observados, elevados ao quadrado, divididos pelo tamanho da amostra:

$$m_2 = \overline{X^2} = \frac{\sum {X_i}^2}{n} = \frac{(1{,}65)^2 + (1{,}75)^2 + (1{,}8)^2 + (1{,}85)^2 + (1{,}95)^2}{5} = \frac{16{,}25}{5} = 3{,}25$$

Para calcular os estimadores pelo **método dos momentos (EMM)**, que podemos denotar por $\hat{\theta}_{MM}$, **igualamos** os **momentos teóricos (ou populacionais)** aos **momentos amostrais**:

$$\mu_k = m_k$$

Essa equação é feita para **todos os parâmetros** do modelo. Por exemplo, se houver 2 parâmetros, fazemos:

$$\mu_1 = m_1$$

$$\mu_2 = m_2$$

Em seguida, **resolvemos o sistema de equações**.

O EMM para a **média populacional** $\hat{\mu}_{MM}$ é dado por:

$$\hat{\mu}_{MM} = m_1 = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

Ou seja, a **média amostral** é o **estimador** para a **média populacional**, obtido pelo **método dos momentos**.

Em relação à variância, sabemos que $\sigma^2 = E(X^2) - [E(X)]^2$, ou seja, a variância populacional é a **diferença** entre o **segundo momento** e o **quadrado do primeiro momento**. Logo, o **EMM** para a **variância** é a **diferença** entre o **segundo momento amostral** e o **quadrado do primeiro momento amostral** (o qual corresponde à média amostral):

$$\hat{\sigma}^2_{MM} = m_2 - (m_1)^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}^2$$

Essa expressão pode ser trabalhada para obtermos o seguinte resultado:

$$\hat{\sigma}^2_{MM} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\boldsymbol{n}}$$

Entretanto, sabemos que esse estimador é **tendencioso**, ou seja, a esperança desse estimador é **diferente** do parâmetro populacional estimado (variância):

$$E(\hat{\sigma}^2_{MM}) \neq \sigma^2$$

Portanto, **nem sempre** o estimador obtido pelo método dos momentos apresenta **boas propriedades**. Veremos outros problemas relativos a esse método a seguir:

Vamos calcular o estimador de θ pelo método dos momentos para uma variável **uniforme** no intervalo $(0, \theta)$. Nesse caso, temos um único parâmetro, então basta calcularmos o primeiro momento:

$$\mu_1 = m_1 = \bar{X}$$

Para uma variável uniforme em um intervalo (a, b), a média populacional (ou esperança da variável) corresponde à média aritmética entre os extremos do intervalo:

$$\mu_1 = \frac{a+b}{2} = \frac{0+\theta}{2} = \frac{\theta}{2} = \bar{X}$$

$$\hat{\theta}_{MM} = 2.\bar{X}$$

Assim, o estimador para o parâmetro θ obtido pelo método dos momentos é o **dobro da média amostral**.

Esse estimador **não** é uma estatística **suficiente**, porque a informação do maior valor observado na amostra, necessária para estimar o parâmetro, é perdida.

Também é possível ter **mais de um estimador** para um **mesmo parâmetro** populacional. Para a distribuição de Poisson, por exemplo, temos $E(X) = V(X) = \lambda$. Assim, o parâmetro λ pode ser estimado tanto por $\hat{\mu}_{MM}$, quanto por $\hat{\sigma}^2_{MM}$, o que pode resultar em estimativas bem **diferentes**.

Ademais, é possível obter valores **negativos** para os parâmetros de uma distribuição binomial, segundo o método dos momentos.

Vamos entender como é possível obter estimadores negativos?

Como há 2 parâmetros para essa distribuição, precisamos de 2 momentos. O primeiro momento populacional é:

$$\mu_1 = E(X) = n.p$$

Já, a variância pode ser calculada como $V(X) = E(X^2) - [E(X)]^2$.

Sabendo que $V(X) = n.p.(1-p)$, podemos calcular $E(X^2)$ para essa distribuição:

$$\mu_2 = E(X^2) = V(X) + [E(X)]^2 = n.p.(1-p) + [n.p]^2$$

$$\mu_2 = E = n.p.(1-p) + n^2.p^2$$

Para calcular os EMMs, fazemos $m_1 = \mu_1$ e $m_2 = \mu_2$. Assim, precisamos resolver o seguinte sistema de equações, para uma amostra de tamanho k:

$$\left\{ \begin{array}{c} \bar{X} = n.p \\ \left(\frac{1}{k}\right)\sum_{i=1}^{k} {X_i}^2 = n.p.(1-p) + n^2.p^2 \end{array} \right\}$$

Resolvendo esse sistema (que não vamos demonstrar aqui), obtemos as seguintes expressões para os estimadores dos parâmetros n e p:

$$\hat{n} = \frac{\bar{X}^2}{\bar{X} - \left(\frac{1}{k}\right)\sum_{i=1}^{k}(X_i - \bar{X})^2} \qquad \hat{p} = \frac{\bar{X}}{\hat{n}}$$

Como o valor de $\left(\frac{1}{k}\right)\sum_{i=1}^{k}(X_i - \bar{X})^2$ pode superar o valor de $\bar{X}$, é possível que o denominador de $\hat{n}$ seja negativo e assim termos $\hat{n}$ e $\hat{p}$ negativos.

(2017 – TRF-2ª Região) Considerando o método de estimação conhecido como Método dos Momentos, assinale a afirmativa INCORRETA.

a) Estimadores obtidos utilizando o k-ésimo momento não têm propriedades assintóticas para k ≥ 2.

b) O k-ésimo momento amostral é dado por $m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ e o k-ésimo momento populacional é dado por $E(X^k)$.

c) Duas variáveis aleatórias com funções densidade de probabilidade distintas podem ter os mesmos momentos.

d) Este método iguala os momentos da população aos momentos amostrais. Os estimadores são obtidos resolvendo a equação ou o sistema de equações resultante.

Comentários:

Essa questão pede a alternativa **INCORRETA**.

Em relação à alternativa A, sabemos que o momento para k = 2 é $m_2 = \frac{\sum_{i=1}^{n} X_i^2}{n}$.

Conforme n aumenta, esse valor tende a zero. Logo, **não** podemos dizer que tais estimadores não possuem propriedades assintóticas. Assim, a alternativa A está INCORRETA.

Em relação à alternativa B, vimos que o k-ésimo momento amostral é, de fato, $m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ e que o k-ésimo momento populacional é $\mu_k = E(X^k)$. Logo, a alternativa B está CORRETA.

Em relação à alternativa C, de fato, distribuições **distintas** podem apresentar **momentos iguais**, logo a alternativa C está CORRETA.

Em relação à alternativa D, vimos que o método dos momentos iguala os momentos amostrais aos momentos teóricos (ou populacionais) aos momentos amostrais.

Em seguida, os estimadores são encontrados resolvendo a equação (quando houver apenas 1 parâmetro) ou o sistema de equações (quando houver mais parâmetros). Logo, a alternativa D está CORRETA.

Gabarito: A

(2007 – TCE/RO – *Adaptada*) Seja (y_1, y_2, y_3, y_4) uma amostra aleatória independente e identicamente distribuída, de tamanho n = 4, extraída de uma população cuja característica estudada possui distribuição de probabilidade

$$f_Y(y,\theta) = \begin{cases} 1/\theta, & 0 \le y \le \theta \\ 0, & caso\ contrário \end{cases}$$

Se a amostra selecionada foi (6, 2, 14, 8), a estimativa do parâmetro θ pelo método dos momentos é:

a) 7

b) 7,5

c) 14

d) 15

e) 14,5

Comentários:

Para calcular o estimador pelo método dos momentos, igualamos os momentos amostrais aos momentos teóricos:

$$\mu_k = m_k$$

Ou seja, o primeiro momento teórico, que corresponde à esperança (ou média) populacional, é igualado ao primeiro momento amostral, que corresponde à média amostral:

$$\bar{X} = \frac{\sum X}{n} = \frac{6 + 2 + 14 + 8}{4} = \frac{30}{4} = 7{,}5$$

Pela função densidade fornecida, podemos identificar que se trata de uma distribuição uniforme, no intervalo $(0, \theta)$.

A média dessa distribuição é a média aritmética dos extremos:

$$\mu = \frac{0 + \theta}{2} = \frac{\theta}{2}$$

Igualando os dois momentos, temos:

$$\frac{\widehat{\theta_{MM}}}{2} = 7{,}5$$

$$\widehat{\theta_{MM}} = 15$$

Resposta: D

Método da Máxima Verossimilhança

Para estimar o parâmetro populacional desejado, o **método da máxima verossimilhança** busca a estimativa para a qual a **probabilidade** de se obter os valores observados é a **maior** possível.

Em outras palavras, suponha que as observações tenham sido $X_1, X_2, \dots, X_n$.

Nesse caso, considerando o estimador de máxima verossimilhança, a probabilidade associada às observações $X_1, X_2, \dots, X_n$ é **maior** do que se considerássemos **qualquer outro estimador**.

Esse método terá uma função diferente, de acordo com a **distribuição** da população.

O estimador de máxima verossimilhança para um parâmetro θ é calculado a partir de uma função de **probabilidade** conhecida dependente desse parâmetro $f(\theta, x)$ e das observações de uma amostra aleatória $x_1, x_2, \dots, x_n$.

Para calcular o estimador, precisamos da **função de máxima verossimilhança** $L(\theta, x_i)$, dada pelo produto das funções de probabilidade, aplicadas para cada resultado da amostra (função densidade conjunta):

$$L(\theta, x_i) = \prod_{i=1}^{n} f(\theta, x_i) = f(\theta, x_1) \times f(\theta, x_2) \times \dots \times f(\theta, x_n)$$

O estimador de máxima verossimilhança será aquele que **maximizar** essa função. Para isso, **derivamos** a função em relação a θ e a igualamos a **zero**.

Normalmente, é mais simples trabalhar com o **logaritmo natural** da função $\ln L(\theta, x_i)$.

Para uma variável com **distribuição normal**, o estimador de máxima verossimilhança (EMV) para a média é:

$$\widehat{\mu_{MV}} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

Ou seja, para uma população com **distribuição normal**, a **média amostral** é o **estimador de máxima verossimilhança (EMV)** para a média populacional.

E o EMV para a **variância** de uma **distribuição normal**, é:

$$\widehat{\sigma^2_{MV}} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$$

Esse é o mesmo estimador obtido pelo método dos momentos, que é **tendencioso**:

$$E(\hat{\sigma}^2_{MV}) \neq \sigma^2$$

Para uma variável aleatória com distribuição de **Poisson**, o estimador de máxima verossimilhança (EMV) também é a **média amostral** (que é um estimador eficiente):

$$\widehat{\lambda_{MV}} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

Por exemplo, vamos supor que queiramos estimar o número médio (λ) de pessoas que chegam em um ponto de ônibus por hora, sabendo que essa variável segue distribuição de Poisson.

Para isso, vamos considerar uma amostra de $n = 6$ horas (ou seja, vamos ficar 6 horas observando o ponto de ônibus e anotando o número de pessoas que chegam a cada hora).

Vamos supor que os valores observados sejam os seguintes:

$$\{10, 12, 15, 10, 8, 5\}$$

Para estimar o parâmetro λ pelo método de máxima verossimilhança, calculamos a **média amostral**:

$$\widehat{\lambda_{MV}} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{10 + 12 + 15 + 10 + 8 + 5}{6} = \frac{60}{6} = 10$$

Ou seja, estimamos que chegam, em média, 10 pessoas por hora nesse ponto de ônibus.

Vamos calcular o estimador de máxima verossimilhança para a distribuição de **Poisson**. A função de distribuição de probabilidade dessa variável é $f(\theta, x) = \frac{e^{-\theta}.\theta^{x}}{x!}$.

A **função de máxima verossimilhança** $L(\theta, x_i)$ corresponde ao produto das funções de probabilidade, aplicadas para cada resultado da amostra:

$$L(\theta, x_i) = \prod_{i=1}^{n} f(\theta, x_i) = \frac{e^{-\theta}.\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}.\theta^{x_2}}{x_2!} \times \ldots \times \frac{e^{-\theta}.\theta^{x_n}}{x_n!}$$

Quando multiplicamos potências de mesma base, somamos os expoentes:

$$L(\theta, x_i) = \frac{e^{-n.\theta}.\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n}(x_i!)}$$

Para simplificar, vamos trabalhar com o **logaritmo natural** dessa função. O logaritmo do produto e do quociente correspondem à soma e à diferença dos logaritmos:

$$\ln L(\theta, x_i) = \ln\left(\frac{e^{-n.\theta}.\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n}(x_i!)}\right) = \ln(e^{-n.\theta}) + \ln\left(\theta^{\sum_{i=1}^{n} x_i}\right) - \sum_{i=1}^{n} \ln(x_i!)$$

E o logaritmo da potência é igual ao produto do logaritmo:

$$\ln L(\theta, x_i) = -n.\theta + \sum_{i=1}^{n} x_i \,.\ln\theta - \sum_{i=1}^{n} \ln(x_i!)$$

Agora, **derivamos** essa função em relação a θ, sabendo que a derivada de $\ln\theta$ é $\frac{1}{\theta}$:

$$\frac{\partial \ln L(\theta,x_i)}{\partial\theta} = -n + \frac{\sum_{i=1}^{n} x_i}{\theta}$$

E igualamos a derivada a **zero** (**equação de log-verossimilhança**):

$$\frac{\partial \ln L(\theta,x_i)}{\partial\theta} = -n + \frac{\sum_{i=1}^{n} x_i}{\theta} = 0$$

$$\frac{\sum_{i=1}^{n} x_i}{\theta} = n$$

$$\theta = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}$$

E assim confirmarmos que a estimativa MV da distribuição de Poisson é a **média amostral**.

Para uma variável com distribuição **exponencial**, a média é o inverso do parâmetro: $E(X) = \frac{1}{\lambda}$, então a estimativa de máxima verossimilhança para o parâmetro λ é o **inverso da média amostral** (que também é um estimador **eficiente**):

$$\widehat{\lambda_{MV}} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}$$

Por exemplo, vamos supor que queiramos estimar o tempo de duração de uma lâmpada incandescente. Foram selecionadas $n = 5$ lâmpadas e observadas as seguintes durações, em horas, {50, 52, 46, 48, 54}.

A média dessa amostra é:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{50 + 52 + 46 + 48 + 54}{5} = \frac{250}{5} = 50$$

E a estimativa do parâmetro dessa distribuição, por MMV, é o **inverso da média amostral**:

$$\widehat{\lambda_{MV}} = \frac{1}{\bar{X}} = \frac{1}{50} = 0{,}02\ por\ hora$$

Vamos calcular o estimador de máxima verossimilhança para a distribuição **exponencial**. A função densidade de probabilidade dessa variável é $f(\theta, x) = \theta . e^{-\theta . x}$.

Primeiro, calculamos a **função de máxima verossimilhança** $L(\theta, x_i)$, que corresponde ao produto das funções de probabilidade, aplicadas para os resultados da amostra:

$$L(\theta, x_i) = \theta . e^{-\theta . x_1} \times \theta . e^{-\theta . x_2} \times \ldots \times \theta . e^{-\theta . x_n} = \theta^n . e^{-\theta . \sum_{i=1}^{n} x_i}$$

Agora, aplicamos o **logaritmo natural** dessa função:

$$\ln L(\theta, x_i) = \ln\left(\theta^n . e^{-\theta . \sum_{i=1}^{n} x_i}\right) = n . \ln \theta - \theta . \sum_{i=1}^{n} x_i$$

Em seguida, igualamos a **derivada** dessa função a **zero** (**equação de log-verossimilhança**):

$$\frac{\partial \ln L(\theta, x_i)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i = 0$$

$$\frac{n}{\theta} = \sum_{i=1}^{n} x_i$$

$$\theta = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{X}}$$

Para uma variável aleatória com distribuição de **Bernoulli**, o estimador de máxima verossimilhança para p é a **proporção amostral** (que segue a mesma definição de média amostral e também é **eficiente**):

$$\widehat{p_{MV}} = \frac{\sum_{i=1}^{n} X_i}{n} = \hat{p}$$

Por exemplo, vamos supor que queiramos estimar a proporção p de peças defeituosas produzidas por uma fábrica. Para isso, foi selecionada uma amostra de $n = 10$ elementos que apresentou o seguinte resultado (em que 1 representa o defeito e 0 representa a peça boa):

$$\{0, 0, 1, 0, 0, 0, 0, 1, 0, 0\}$$

O estimador de máxima verossimilhança para essa distribuição é a proporção amostral:

$$\widehat{p_{MV}} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{0+0+1+0+0+0+0+1+0+0}{10} = \frac{2}{10} = 0{,}2$$

Para uma distribuição **geométrica**, que representa o número de tentativas até o primeiro sucesso, a média é o inverso da probabilidade de sucesso: $E(X) = \frac{1}{p}$, então o estimador de máxima verossimilhança para a proporção p é o **inverso da média amostral** (que é um estimador eficiente):

$$\widehat{p_{MV}} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}}$$

Vamos supor que 5 pessoas estejam jogando para ver quem consegue primeiro a face CARA lançando uma mesma moeda viciada (não equilibrada). Vamos considerar que os resultados foram {3, 4, 2, 5, 1}.

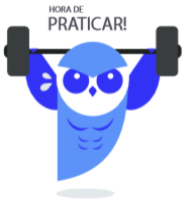
A média amostral desses resultados é:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{3+4+2+5+1}{5} = \frac{15}{5} = 3$$

Logo, a probabilidade de obter a face CARA nessa moeda, pelo método da máxima verossimilhança, é:

$$\widehat{p_{MV}} = \frac{1}{\overline{X}} = \frac{1}{3}$$

Para uma variável **uniforme** no intervalo $(0, \theta)$, o estimador de máxima verossimilhança para θ é o **maior valor observado na amostra**. (Observe como os estimadores são diferentes pelos diferentes métodos: lembra que o estimador de θ nesse caso pelo método dos momentos é o dobro da média amostral?)



(CESPE/2015 – Telebras) Considerando que os principais métodos para a estimação pontual são o método dos momentos e o da máxima verossimilhança, julgue o item a seguir.

Para a distribuição normal, o método dos momentos e o da máxima verossimilhança fornecem os mesmos estimadores aos parâmetros μ e σ.

Comentários:

Para uma distribuição **normal**, o estimador da **média**, segundo o método da máxima verossimilhança, é a **média amostral** $\bar{X}$, o mesmo estimador obtido pelo método dos momentos. E o estimador da **variância**, segundo o método da máxima verossimilhança, para uma população normal, é $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, que também é o mesmo estimador obtido pelo método dos momentos.

Gabarito: Certo.

(CESPE/2015 – Telebras) Considerando que os principais métodos para a estimação pontual são o método dos momentos e o da máxima verossimilhança, julgue o item a seguir.

O estimador da máxima verossimilhança para a variância da distribuição normal é expresso por $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ e este estimador é não viciado.

Comentários:

De fato, o estimador $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ é o estimador de máxima verossimilhança para a variância de uma população com distribuição normal, porém esse estimador é **viciado**.

Gabarito: Errado.

(CESPE/2016 – TCE/PA) Uma amostra aleatória com n = 16 observações independentes e identicamente distribuídas (IID) foi obtida a partir de uma população infinita, com média e desvio padrão desconhecidos e distribuição normal. Tendo essa informação como referência inicial, julgue o seguinte item.

Caso, em uma amostra aleatória de tamanho n = 4, os valores amostrados sejam A = {2, 3, 0, 1}, a estimativa de máxima verossimilhança para a variância populacional será igual a $\frac{5}{3}$.

Comentários:

A estimativa de máxima verossimilhança para a variância de uma população normal, é:

$$\widehat{\sigma^2_{MV}} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$$

Para calculá-la, precisamos primeiramente da média amostral, $\bar{X}$:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{2+3+0+1}{4} = \frac{6}{4} = 1{,}5$$

A estimativa é, portanto:

$$\widehat{\sigma^2_{MV}} = \frac{(2-1{,}5)^2 + (3-1{,}5)^2 + (0-1{,}5)^2 + (1-1{,}5)^2}{4} = \frac{(0{,}5)^2 + (1{,}5)^2 + (-1{,}5)^2 + (-0{,}5)^2}{4}$$

$$\widehat{\sigma^2_{MV}} = \frac{0{,}25 + 2{,}25 + 2{,}25 + 0{,}25}{4} = \frac{5}{4}$$

Que é diferente do valor indicado no item $\frac{5}{3}$, que corresponde ao estimador não tendencioso para a variância (variância amostral), e não aquele obtido pelo método da máxima verossimilhança.

Gabarito: Errado.

(FGV/2022 – SEFAZ/ES) Uma amostra aleatória simples de tamanho 4 de uma população normalmente distribuída forneceu os seguintes dados: 2,1 3,8 3,1 3,0. As estimativas de máxima verossimilhança da média e da variância populacionais são respectivamente

a) 3,0 e 0,486

b) 2,8 e 0,386

c) 2,8 e 0,535

d) 3,0 e 0,544

e) 3,0 e 0,365

Comentários:

Essa questão trabalha com o método de **máxima verossimilhança** para estimar a média e a variância de uma população com distribuição **normal**. Para estimar a média populacional, utilizamos a **média amostral**:

$$\bar{X} = \frac{\sum x}{n} = \frac{2{,}1 + 3{,}8 + 3{,}1 + 3{,}0}{4} = \frac{12}{4} = 3$$

Já a estimativa de máxima verossimilhança para a variância é o estimador **tendencioso**, em que dividimos a soma dos quadrados dos desvios por n, e não por $n-1$:

$$\widehat{\sigma^2_{MV}} = \frac{\sum(x_i - \bar{X})^2}{\boldsymbol{n}}$$

Para calcular o numerador, vamos construir uma tabela com os desvios:

x_i	2,1	3,8	3,1	3,0
$x_i - \bar{X}$	-0,9	0,8	0,1	0
$(x_i - \bar{X})^2$	0,81	0,64	0,01	0

E o numerador é a soma dos valores da última linha:

$$\sum(x_i - \bar{X})^2 = 0{,}81 + 0{,}64 + 0{,}01 + 0 = 1{,}46$$

Para calcular o estimador da variância, dividimos esse resultado por $n = 4$:

$$\widehat{\sigma^2_{MV}} = \frac{1{,}46}{4} = 0{,}365$$

Gabarito: E

(FGV/2017 – IBGE) Seja X uma variável aleatória com função de probabilidade dada por $P(X = x) = p.(1-p)^{x-1}$ para x = 1, 2, 3,..., onde p é um parâmetro desconhecido. Dispondo de uma amostra de tamanho n, x_1, x_2, x_3,...., x_n, o estimador de Máxima Verossimilhança de p é:

a) $\hat{p} = \sum x_i$

b) $\hat{p} = \frac{1}{\sum x_i}$

c) $\hat{p} = \frac{\sum x_i}{n}$

d) $\hat{p} = \frac{n}{\sum x_i}$

e) $\hat{p} = \sqrt[n]{\sum x_i}$

Comentários:

Podemos observar que a função de probabilidade fornecida na questão é de uma variável com distribuição **geométrica**. Para essa variável, o estimador de máxima verossimilhança é o **inverso da média amostral**:

$$\widehat{p_{MV}} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}$$

Gabarito: D

(CESPE/2013 – MPU) Suponha que $x_1, ..., x_n$ seja uma sequência de cópias independentes retiradas de uma distribuição com função densidade de probabilidade $f(x) = \alpha.x.e^{\frac{-\alpha x^2}{2}}$, em que $x \geq 0$ e $\alpha > 0$ é seu parâmetro.

Com base nessas informações, julgue o item a seguir.

Supondo que $(x_1, ..., x_5) = (3, 4, 4, 6, 6)$, a estimativa de máxima verossimilhança do parâmetro α é inferior a 1/10.

Comentários:

O primeiro passo é calcular a **função de máxima verossimilhança** $L(\alpha, x_i)$, dada pelo **produto** das funções de probabilidade, aplicadas para cada resultado da amostra:

$$L(\theta, x_i) = \prod_{i=1}^{n} f(\alpha, x_i) = f(\alpha, x_1) \times f(\alpha, x_2) \times \ldots \times f(\alpha, x_n)$$

Para o nosso caso, temos:

$$L(\alpha, x_i) = \prod_{i=1}^{n} \alpha.x_i.e^{\frac{-\alpha x_i^2}{2}} = \alpha^n.\left(\prod_{i=1}^{n} x_i\right).e^{\frac{-\alpha}{2}\sum_{i=1}^{n} x_i^2}$$

Agora, calculamos o **logaritmo natural** (na base e) dessa função, sabendo que o **logaritmo do produto** corresponde à **soma dos logaritmos**; e o logaritmo da potência é igual ao **produto do logaritmo**:

$$\ln L(\alpha, x_i) = \ln\left[\alpha^n.\left(\prod_{i=1}^{n} x_i\right).e^{\frac{-\alpha}{2}\sum_{i=1}^{n} x_i^2}\right] = n.\ln\alpha + \sum_{i=1}^{n} \ln x_i - \frac{\alpha}{2}.\sum_{i=1}^{n} x_i^2$$

Agora, calculamos a **derivada** dessa função em relação a α; e a igualamos a **zero** (**equação de log-verossimilhança**):

$$\frac{\partial \ln L(\alpha, x_i)}{\partial \alpha} = \frac{n}{\alpha} - \frac{\sum_{i=1}^{n} x_i^2}{2} = 0$$

$$\frac{n}{\alpha} = \frac{\sum_{i=1}^{n} x_i^2}{2}$$

$$\alpha = \frac{2n}{\sum_{i=1}^{n} x_i^2}$$

Encontramos o estimador de máxima verossimilhança para α. Para a amostra descrita no enunciado, temos:

$$\alpha = \frac{2 \times 5}{3^2 + 4^2 + 4^2 + 6^2 + 6^2} = \frac{10}{9 + 16 + 16 + 36 + 36} = \frac{10}{113}$$

Que é inferior a 1/10.

Gabarito: Certo.

Método de Mínimos Quadrados

O Método de Mínimos Quadrados busca a estimativa $\widehat{\theta_{MQ}}$ que resulta no **menor valor** para o **quadrado das diferenças** entre os valores observados X_i e o estimador $\widehat{\theta_{MQ}}$:

$$\text{Min} \sum_{i=1}^{n} \left(X_i - \widehat{\theta_{MQ}}\right)^2$$

Ou seja, o método busca **minimizar o erro quadrático total da amostra**. Assim, $\widehat{\theta_{MQ}}$ é chamado de Estimador de Mínimos Quadrados (EMQ) para o parâmetro populacional θ. Esse método é utilizado quando **não se conhece o tipo de distribuição da variável**.

O **estimador de mínimos quadrados** para a **média populacional**, $\widehat{\mu_{MQ}}$, é igual à **média amostral**:

$$\widehat{\mu_{MQ}} = \bar{X}$$

Similarmente, o **estimador de mínimos quadrados** para a **proporção populacional**, $\widehat{p_{MQ}}$, é a **proporção amostral**:

$$\widehat{p_{MQ}} = \hat{p}$$

A **média amostral** é o estimador para a média populacional obtido pelo **método dos momentos** (EMM) e pelo **método dos mínimos quadrados** (EMQ).

Se a população tiver distribuição **normal**, a **média amostral** também é o estimador obtido pelo **método da máxima verossimilhança** (EMV).

A **proporção amostral** é o estimador para a proporção populacional obtido pelo **método dos mínimos quadrados** (EMQ) e pelo **método da máxima verossimilhança** (EMV).

O estimador para a **variância** obtido pelo **método dos momentos** (EMM) e pelo **método da máxima verossimilhança** (EMV) é **tendencioso**.

ESTIMAÇÃO INTERVALAR

Após obtermos uma estimativa para o parâmetro populacional desejado (estimação pontual), calculamos um **intervalo** associado a esse parâmetro. Ou seja, na **estimação intervalar** (ou **estimação por intervalos**), a estimativa deixa de ser um ponto (isto é, um valor único) e passa a ser um **intervalo**. Esse intervalo, chamado **intervalo de confiança**, fornece uma noção da **precisão** da estimativa.

O **intervalo de confiança** é construído em torno da estimativa pontual $\hat{\theta}$, da forma $\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{E}; \widehat{\boldsymbol{\theta}} + \boldsymbol{E}\right)$, que também pode ser indicado como $\widehat{\boldsymbol{\theta}} \pm \boldsymbol{E}$. Esse intervalo indica que o parâmetro populacional θ deve estar entre o limite inferior, $\hat{\theta} - E$, e o limite superior $\hat{\theta} + E$. O valor E corresponde à **metade da amplitude** do intervalo, podendo ser chamado de **margem de erro**, **erro de precisão**, **erro máximo**.

Como assim o parâmetro "deve" estar no intervalo? É possível que o parâmetro θ esteja fora desse intervalo? Por se tratar de um intervalo construído em torno de uma **variável aleatória**, sim! Em Inferência Estatística, sempre convivemos com um nível de dúvida. Mas, felizmente, é possível dimensionar esse nível de dúvida.

Assim, atribuímos ao intervalo um **nível** (ou **grau**) **de confiança** $1 - \alpha$, indicado em forma percentual, por exemplo, 95%. A interpretação desse nível é a seguinte: repetindo o procedimento para a construção do intervalo muitas vezes, em $(1 - \alpha)\%$ (por exemplo, 95%) dessas vezes, o intervalo construído **incluirá** o parâmetro populacional.

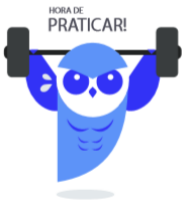
Ou seja, o **nível de confiança** é uma **probabilidade**. Porém, **não** se trata da probabilidade de o **parâmetro populacional pertencer ao intervalo**, uma vez que o parâmetro populacional é **fixo** (embora seja desconhecido). O nível de confiança representa a **probabilidade de o intervalo**, o qual é construído a partir de variáveis aleatórias, **incluir o parâmetro populacional**.

Quanto **maior o nível de confiança** $(1 - \alpha)$, ou seja, quanto maior for a probabilidade de o intervalo incluir o parâmetro populacional, maior será o **tamanho do intervalo**, se mantivermos as demais características iguais. Logo, maior será a **margem de erro** (isto é, a semi-amplitude do intervalo).

Se, para isso, a margem de erro não puder aumentar, então teremos que aumentar o **tamanho da amostra**. Ou seja, para termos uma estimativa mais precisa (com menor margem de erro e/ou com maior nível de confiança), teremos que investigar um número maior de elementos. Veremos as fórmulas dessas relações adiante, mas é importante entender essa lógica por trás delas.

E o valor de α? α, chamado de **nível de significância**, é o **complementar** do nível de confiança e corresponde à probabilidade de o intervalo de confiança **não** englobar o parâmetro populacional.

Nas próximas seções, veremos como construir o intervalo de confiança para os parâmetros populacionais (média, proporção e variância), a partir dos respectivos estimadores.



(CESPE/2019 – TJ/AM) Acerca de métodos usuais de estimação intervalar, julgue o item subsecutivo. Um intervalo de confiança de 95% descreve a probabilidade de um parâmetro estar entre dois valores numéricos na próxima amostra não aleatória a ser coletada.

Comentários:

Existem alguns erros neste item. A estimação, de modo geral, trabalha com amostras **aleatórias já coletadas**. Além disso, não se trata da probabilidade de o **parâmetro** estar entre dois valores, mas sim de os dois valores **englobarem o parâmetro**.

Gabarito: Errado.

(CESPE/2019 – TJ/AM) Acerca de métodos usuais de estimação intervalar, julgue o item subsecutivo. É possível calcular intervalos de confiança para a estimativa da média de uma distribuição normal, representativa de uma amostra aleatória.

Comentários:

De fato, é possível calcular intervalos de confiança para a média, a partir de uma amostra aleatória.

Gabarito: Certo.

(FGV/2019 – DPE-RJ – *Adaptada*) Para a aplicação de técnica de estimação por intervalos, há uma série de requisitos e recomendações. Sobre essas condições, julgue os seguintes.

I – A amplitude do intervalo varia positivamente com o grau de confiança e o tamanho da amostra.

II – A ideia da técnica é a da construção de um intervalo ao qual seja possível associar uma probabilidade, justamente aquela de que o parâmetro de interesse esteja nele contido.

III – É inquestionável que, antes da seleção da amostra, o grau de confiança é a probabilidade de o intervalo teórico conter de fato o verdadeiro valor do parâmetro de interesse.

Comentários:

Em relação ao item I, quanto **maior o nível de confiança**, **maior** será a **amplitude do intervalo**, logo essas grandezas, de fato, variam no mesmo sentido. Porém, quanto **maior o tamanho da amostra, menor** será a **amplitude** do intervalo. Logo, essas grandezas variam em sentidos **opostos** e o item I está errado.

Em relação ao item II, não se trata de uma probabilidade de o parâmetro de interesse (que é fixo) estar contido no intervalo construído, mas sim de o **intervalo** construído **conter** o parâmetro de interesse, logo o item II está errado.

Em relação ao item III, o grau de confiança $(1 - \alpha)$, de fato, representa a probabilidade de o intervalo conter o parâmetro de interesse.

Resposta: Itens I e II Errados; Item III Certo.

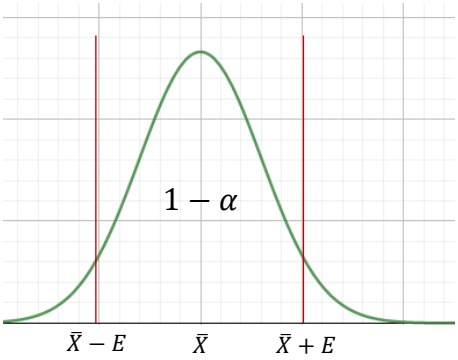
Intervalo de Confiança para a Média

Para estimarmos a média populacional, utilizamos a **média amostral** como estimador. Porém, para construirmos um intervalo de confiança em torno desse estimador, é necessário saber se a **variância** populacional é **conhecida ou não**.

População com Variância Conhecida

Se a população tiver **variância conhecida** σ^2 e **distribuição normal** (ou se apresentar outra distribuição, mas o tamanho da amostra for suficientemente **grande**), a média amostral $\bar{X}$ terá **distribuição normal**.

Assim, consideramos a curva normal para construir um intervalo que delimite uma probabilidade de $\boldsymbol{1-\alpha}$ em **torno de** $\boldsymbol{\bar{X}}$. Ou seja, devemos encontrar os valores de X que delimitam uma área sob a curva normal, correspondente a $1-\alpha$:



Para encontrar os limites, devemos utilizar a tabela normal padrão e a transformação para a distribuição normal padrão Z (com média 0 e desvio padrão igual a 1):

$$z = \frac{valor\ procurado - média\ da\ distribuição}{desvio\ padrão\ da\ distribuição}$$

No caso, os **valores procurados** são $X = \bar{X} + E$ e $X = \bar{X} - E$; a **média** da distribuição é $\bar{X}$; e o **desvio padrão** da distribuição é o erro padrão de $\bar{X}$:

$$EP(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Substituindo esses valores na fórmula da transformação, encontramos a expressão do erro:

$$z = \frac{\bar{X} + E - \bar{X}}{\frac{\sigma}{\sqrt{n}}} = \frac{E}{\frac{\sigma}{\sqrt{n}}}$$

$$E = z.\frac{\sigma}{\sqrt{n}}$$

O intervalo de confiança $\bar{X} \pm E$ é, portanto:

$$\left(\bar{X} - z.\frac{\sigma}{\sqrt{n}}; \bar{X} + z.\frac{\sigma}{\sqrt{n}}\right)$$

Esses limites podem ser chamados de **valores críticos**.

E qual é o valor de z? Depende do **nível de confiança**.

Por exemplo, suponha um nível de confiança de 95%. Para que toda a região delimitada pelo intervalo $\bar{X} \pm E$ delimite uma área de $1 - \alpha = 95\%$, então a área entre a média $\bar{X}$ e o limite superior $\bar{X} \pm E$ deve ser a **metade** (47,5%). Considerando a transformação para a normal padrão, temos $P(0 < Z < z) = 0{,}475$.

Z	...	0,05	0,06	0,07	...
...	...	...	...	...	...
1,8	...	0,4678	0,4686	0,4693	...
1,9	...	0,4744	0,475	0,4756	...
2	...	0,4798	0,4803	0,4808	...
...	...	...	...	...	...

Pela tabela da normal padrão, temos $z = 1{,}96$. Assim, se o nível de confiança for de 95%, o intervalo será:

$$\bar{X} \pm 1{,}96.\frac{\sigma}{\sqrt{n}}$$

Agora podemos entender melhor o raciocínio que vimos no início desta seção: quanto **maior o nível de confiança** (maior valor de $\boldsymbol{z}$), **maior** será a **margem de erro** $E = z.\frac{\sigma}{\sqrt{n}}$ e, consequentemente, **maior o tamanho** do intervalo $\bar{X} \pm E$, mantendo as demais características constantes.

E quanto **maior o tamanho da amostra** n, **menor** será a **margem de erro** $E = z.\frac{\sigma}{\sqrt{n}}$ e, consequentemente, **menor o tamanho** do intervalo $\bar{X} \pm E$, mantendo as demais características constantes.

Além disso, quanto **maior a variabilidade** da população (maior desvio padrão σ), **maior** será a **margem de erro** $E = z.\frac{\sigma}{\sqrt{n}}$ e, consequentemente, **maior o tamanho** do intervalo $\bar{X} \pm E$, mantendo as demais características constantes.

Para exemplificar, vamos supor uma população com média desconhecida e variância igual $\sigma^2 = 9$ (portanto, desvio padrão $\sigma = \sqrt{\sigma^2} = \sqrt{9} = 3$), sendo extraída uma amostra de 36 indivíduos. Considerando um nível de $1 - \alpha = 95\%$ de confiança (em que $z = 1{,}96$), a **margem de erro** é dada por:

$$E = 1{,}96.\frac{\sigma}{\sqrt{n}} = 1{,}96.\frac{3}{\sqrt{36}} = 1{,}96.\frac{3}{6} = \frac{1{,}96}{2} = 0{,}98$$

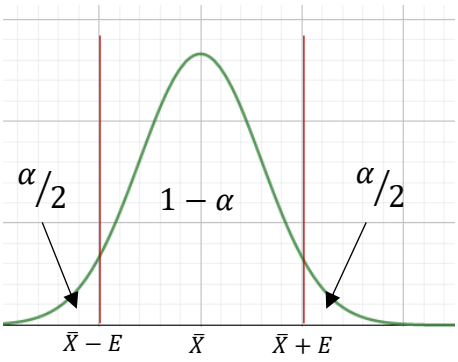
Supondo que a média amostral tenha sido $\bar{X} = 10$, o intervalo de confiança é:

$$\bar{X} + E = 10 + 0{,}98 = 10{,}98$$

$$\bar{X} - E = 10 - 0{,}98 = 9{,}02$$

$$(9{,}02;\ 10{,}98)$$

A probabilidade associada aos valores não incluídos no intervalo de confiança é **complementar** (igual a α). Por se tratar de uma distribuição **simétrica**, a probabilidade associada aos valores superiores ao intervalo é $\alpha/2$ e aos valores inferiores é também $\alpha/2$:



Por isso, é comum utilizar a notação $z_{\alpha/2}$ para indicar o limite do intervalo na distribuição normal padrão (no nosso exemplo, temos $z_{\alpha/2} = 1{,}96$).

Esse método de construção do intervalo de confiança pode ser chamado de **método da quantidade pivotal**.

Isso porque uma quantidade $Q(\boldsymbol{X}; \theta)$ é considerada pivotal quando a distribuição de Q não depende do parâmetro populacional θ, apenas dos valores observados na amostra $\boldsymbol{X}$.

Tamanho Amostral

A questão pode fornecer, além do nível de confiança $(1 - \alpha)$, o valor do erro máximo tolerado (E) e indague a respeito do **tamanho** necessário da **amostra** n. Para isso, podemos utilizar a mesma fórmula do erro que vimos há pouco:

$$E = z.\frac{\sigma}{\sqrt{n}}$$

Reorganizando essa fórmula, temos:

$$n = \left(z.\frac{\sigma}{E}\right)^2$$

Podemos observar que, quanto **maior** o **nível de confiança** desejado (z), **maior** será o **tamanho amostral**; e quanto **maior** o **erro máximo** (E), **menor** será o **tamanho amostral**.

Pontue-se que o **erro máximo** também pode ser chamado de **distância máxima** (ou **diferença máxima**) entre a estimativa e o parâmetro populacional com probabilidade $1 - \alpha$.

Ademais, em vez de fornecer o erro máximo, a questão pode fornecer a **amplitude** do intervalo de confiança, que corresponde ao **dobro do erro máximo**.

No exemplo que vimos anteriormente, obtivemos uma margem de erro de $E = 0{,}98$, com uma amostra de $n = 36$. Vamos supor que o erro máximo tolerável seja a metade, ou seja, $E = 0{,}49$. Considerando que os demais valores permaneçam constantes, ou seja, $z = 1{,}96$ e $\sigma = 3$, o tamanho da nova amostra n_2, necessário para obter a margem de erro desejada é:

$$n_2 = \left(z.\frac{\sigma}{E}\right)^2 = \left(1{,}96 \times \frac{3}{0{,}49}\right)^2 = (4 \times 3)^2 = (12)^2 = 144$$

Note que o **tamanho da amostra** **quadruplicou** para que a **margem de erro** fosse reduzida à **metade**, mantendo o nível de confiança, para a mesma população (portanto, o mesmo desvio padrão).

Na verdade, poderíamos verificar que isso aconteceria, sem conhecer os dados do problema. Vejamos: a amostra inicial é dada por:

$$n_1 = \left(z.\frac{\sigma}{E}\right)^2$$

E a nova amostra necessária para que o erro se reduza à **metade** é:

$$n_2 = \left(z.\frac{\sigma}{E/_2}\right)^2 = \left(z.\frac{\sigma}{E}.2\right)^2 = 4.\underbrace{\left(z.\frac{\sigma}{E}\right)^2}_{n_1} = 4.n_1$$

Pontue-se que tais fórmulas pressupõem uma população infinita ou amostras extraídas com reposição.

Fator de correção

Caso a população seja **finita** e as amostras extraídas **sem reposição**, será necessário aplicar o **fator de correção para população finita**. Para isso, devemos multiplicar a fórmula do erro pelo fator $\sqrt{\frac{N-n}{N-1}}$:

$$E = z.\frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}$$

Por exemplo, para uma amostra de tamanho $n = 36$ com o mesmo nível de confiança $(z = 1{,}96)$ e o mesmo desvio padrão populacional $(\sigma = 3)$, vamos calcular o novo valor da margem de erro, considerando que a população tem tamanho $N = 100$:

$$E = 1{,}96.\frac{3}{\sqrt{36}}.\sqrt{\frac{100-36}{100-1}} = 1{,}96.\frac{3}{6}\sqrt{\frac{64}{99}} \cong 0{,}98.\frac{8}{10} \cong 0{,}78$$

Enquanto a margem de erro para uma população infinita, considerando os mesmos parâmetros, é de 0,98, a margem de erro para a população finita é de 0,78.

Essa **redução** do erro ocorre sempre que aplicamos o fator de correção, mantendo os demais parâmetros iguais, porque o fator é **menor que 1**, uma vez que o tamanho da amostra n é maior que 1.

$$\sqrt{\frac{N-n}{N-1}} < 1$$

Porém, quando o tamanho da população N for muito maior que o tamanho da amostra n, esse fator se aproxima de 1, tornando-se dispensável.

Esse ajuste também **diminui** o **tamanho da amostra** necessário. Para calcular n, nessas condições, primeiro calculamos uma estimativa **inicial** para a amostra, como vimos anteriormente, **desconsiderando** o fator de correção:

$$n_0 = \left(z.\frac{\sigma}{E}\right)^2$$

Em seguida, **ajustamos** a amostra:

$$n = \frac{N.n_0}{N+n_0}$$

A amostra ajustada n é menor do que a amostra inicial n_0, porque o fator que multiplica n_0 é **menor que 1**, sempre que $n_0 > 1$:

$$\frac{N}{N+n_0} < 1$$

Para ilustrar, vamos supor os mesmos valores que vimos anteriormente, no cálculo do tamanho amostral, quais sejam $E = 0{,}49$, $z = 1{,}96$ e $\sigma = 3$. A **amostra inicial** será a mesma que calculamos anteriormente:

$$n_0 = \left(1{,}96 \times \frac{3}{0{,}49}\right)^2 = (4 \times 3)^2 = 144$$

E a amostra **ajustada**, considerando uma população de tamanho $N = 1000$, é:

$$n = \frac{1000 \times 144}{1000 + 144} = \frac{144.000}{1.144} \cong 126$$

(CESPE 2020/TJ-PA) Uma equipe de engenheiros da qualidade, com vistas a estimar vida útil de determinado equipamento, utilizou uma amostra contendo 225 unidades e obteve uma média de 1.200 horas de duração, com desvio padrão de 150 horas.

Considerando-se, para um nível de confiança de 95%, z = 1,96, é correto afirmar que a verdadeira duração média do equipamento, em horas, estará em um intervalo entre

a) 1.190,00 e 1.210,00.

b) 1.185,20 e 1.214,80.

c) 1.177,50 e 1.222,50.

d) 1.180,40 e 1.219,60.

e) 1.174,20 e 1.225,80.

Comentários:

Vamos listar as informações do enunciado:

- Média amostral: $\bar{X} = 1.200$
- Nível de confiança: $z = 1{,}96$
- Desvio padrão da população: $\sigma = 150$
- Tamanho da amostra: $n = 225$

Agora, substituímos esses dados na fórmula do intervalo de confiança para a média:

$$\bar{X} \pm z \times \frac{\sigma}{\sqrt{n}}$$

$$1200 \pm 1{,}96 \times \frac{150}{\sqrt{225}}$$

$$1200 \pm 1{,}96 \times \frac{150}{15}$$

$$1200 \pm 1{,}96 \times 10$$

$$1200 \pm 19{,}6$$

Logo, o intervalo é $(1200 - 19{,}6 = 1180{,}4;\ 1200 + 19{,}6 = 1219{,}6)$

Gabarito: D.

(FGV/2022 – TJDFT) Uma grande amostra foi selecionada para estimar o tempo médio de tramitação de um tipo particular de ação em uma comarca. Essa amostra demonstrou que o intervalo bilateral de 95% de confiança para o tempo médio de tramitação estava entre 8 e 10 anos.

Com o objetivo de aumentar a precisão dessa estimativa, um estatístico resolveu diminuir a confiança para 85%. O novo intervalo de confiança passou a ser, aproximadamente, igual a:

a) $9 \pm 0{,}26$

b) $9 \pm 0{,}34$

c) $9 \pm 0{,}72$

d) $9 \pm 0{,}88$

e) $9 \pm 1{,}44$

Nessa prova, foi fornecida a tabela normal padrão da forma P(Z > Z_0), parcialmente replicada a seguir.

Z_0	Segunda decimal de Z_0									
	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,00	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,10	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,20	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,30	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,40	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,50	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,60	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,70	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,80	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,90	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,00	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183

Comentários:

A questão informa que, para um nível de 95% de confiança, o intervalo para estimar a média foi (8; 10), ou seja, 9 ± 1. O erro portanto é dado por:

$$E = z_{95\%}.\frac{\sigma}{\sqrt{n}} = 1$$

E a questão pede o novo intervalo, quando **reduzimos o nível de confiança** para 85%, sabendo que os demais parâmetros continuarão os mesmos.

Assim, precisamos comparar o valor de z para 95% de confiança e o valor de z para 85% de confiança.

A prova apresenta a tabela normal da forma P(Z > Z_0). Para um nível de $1 - \alpha = 95\%$ de confiança, resta $\frac{\alpha}{2} = 2{,}5\%$ abaixo do limite inferior do intervalo e $\frac{\alpha}{2} = 2{,}5\%$ acima do limite superior do intervalo.

Logo, precisamos do valor de Z_0 associado a uma probabilidade P(Z > Z_0) = 2,5% = 0,025. Pela tabela fornecida, temos que Z_0 = 1,96 ≅ 2.

Para um nível de $1 - \alpha = 85\%$ de confiança, temos $\frac{\alpha}{2} = 7{,}5\%$ abaixo do limite inferior e $\frac{\alpha}{2} = 7{,}5\%$ acima do limite superior do intervalo.

Logo, precisamos do valor de Z_0 associado a uma probabilidade P(Z > Z_0) = 7,5% = 0,075. Pela tabela fornecida, temos que Z_0 = 1,44.

Agora, vamos calcular a razão entre os valores de z:

$$\frac{z_{85\%}}{z_{95\%}} \cong \frac{1{,}44}{2} = 0{,}72$$

Considerando que o erro é diretamente proporcional ao valor de z, a razão entre os erros segue essa mesma proporção:

$$\frac{E_{85\%}}{E_{95\%}} \cong 0{,}72$$

Sabendo que o erro a 95% de confiança é igual a 1, então o erro a 85% de confiança é:

$$E_{85\%} \cong 0{,}72 \times E_{95\%} = 0{,}72 \times 1 = 0{,}72$$

Logo, o intervalo de confiança é da forma $9 \pm 0{,}72$.

Gabarito: C

(FCC/2014 - TRT/MA) Para responder às questões use, dentre as informações dadas abaixo, as que julgar apropriadas.

Se Z tem distribuição normal padrão, então: P(Z<0,25)=0,599, P(Z<0,80)=0,84, P(Z<1)=0,841, P(Z<1,96)=0,975, P(Z<3,09)=0,999

Considere X_1, X_2, ...X_n uma amostra aleatória simples, com reposição, da distribuição da variável X, que tem distribuição normal com média μ e variância 36. Seja X a média amostral dessa amostra.

O valor de n para que a distância entre X e μ seja, no máximo, igual a 0,49, com probabilidade de 95% é igual a:

a) 256

b) 225

c) 400

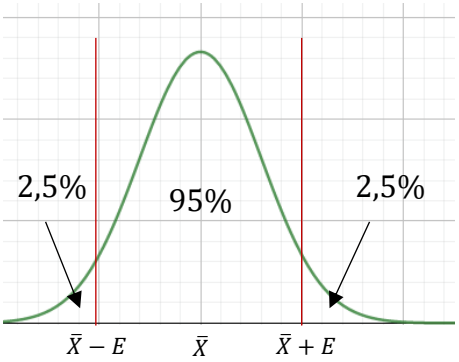
d) 144

e) 576

Comentários:

Ao dizer que a **distância** entre a média amostral (estimativa) e a média populacional (parâmetro populacional) seja no máximo igual a 0,49 com probabilidade de 95%, a questão forneceu o erro máximo $E = 0{,}49$ e o nível de confiança de 95%.

Para que 95% da distribuição esteja no intervalo $(\bar{X} - E; \bar{X} + E)$, 2,5% da distribuição estará acima desse intervalo e 2,5% abaixo:



Assim, precisamos do valor de z cuja probabilidade P(Z < z) seja igual a 2,5% + 95% = 97,5%.

Pelos valores fornecidas observamos que z = 1,96.

Sabendo que a variância é $\sigma^2 = 36$, ou seja, o desvio padrão é $\sigma = \sqrt{\sigma^2} = \sqrt{36} = 6$, podemos calcular o tamanho amostral:

$$n = \left(z.\frac{\sigma}{E}\right)^2$$

$$n = \left(1{,}96.\frac{6}{0{,}49}\right)^2 = (24)^2 = 576$$

Gabarito: E.

(FGV/2022 – MPE/SC) O tempo, em horas diárias, que homens com idades entre os 40 e 50 anos acessam redes sociais segue uma distribuição Normal com média 2,5 e desvio padrão 1,5. Para o mesmo grupo etário de mulheres, esse tempo segue também uma distribuição Normal com média 3 e desvio padrão 1. Serão retiradas duas amostras casuais e independentes, uma de homens e outra de mulheres.

O tamanho mínimo da amostra da população das mulheres que se pretende com probabilidade pelo menos 0,95 e cuja diferença em valor absoluto entre a média amostral e a média populacional não exceda 0,1 é, aproximadamente:

a) 20;

b) 100;

c) 250;

d) 385;

e) 500.

Comentários:

Essa questão também trabalha com o **tamanho amostral**, para um intervalo de confiança para a média, com variância conhecida:

$$n = \left(\frac{z}{E}.\sigma\right)^2$$

Em que z é o valor da tabela normal padrão associado ao nível de confiança desejado; E é a margem de erro e σ é o desvio padrão.

A questão pede o tamanho mínimo para o tempo das mulheres, em que o **desvio padrão** é $\sigma = 1$.

O enunciado informa, ainda, que a probabilidade do intervalo (nível de confiança) é de 95%, logo, **z = 1,96**.

Ademais, a questão informa que a **diferença máxima** entre a média amostral e a média populacional, que corresponde à **margem de erro**, é E = 0,1. Substituindo esses dados na fórmula, temos:

$$n = \left(\frac{1{,}96}{0{,}1}.1\right)^2 = (19{,}6)^2 = 384{,}16$$

Logo, o menor tamanho amostral, isto é, o menor número inteiro maior que o valor calculado, é 385.

Gabarito: D

(CESPE 2016/TCE-PA) Considerando uma população finita em que a média da variável de interesse seja desconhecida, julgue o item a seguir.

Se uma amostra aleatória simples, sem reposição, for obtida de uma população finita constituída por N = 45 indivíduos, o fator de correção para população finita não será considerado na definição do tamanho da amostra para a estimação da média.

Comentários:

Quando uma população é **finita** e a amostra é extraída **sem reposição**, utilizamos, sim, um **ajuste** na equação, chamado **fator de correção**, que influencia o cálculo do tamanho amostral.

Gabarito: Errado.

População com Variância Desconhecida

Sendo a **variância populacional desconhecida**, precisamos estimá-la, a partir da **variância amostral** (**estimador não tendencioso para a variância**):

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Esse estimador vale para populações **infinitas** **OU** amostras extraídas **com** reposição.

Caso a população seja **finita** de tamanho N **E** a amostra seja extraída **sem** reposição, é necessário aplicar o **fator de correção**, multiplicando a variância amostral por $\frac{N-n}{N-1}$:

$$s_*^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \times \frac{N-n}{N-1}$$

Com a estimativa para a variância populacional, calculamos a **variância da média amostral** $\bar{X}$, de forma análoga à que vimos anteriormente, substituindo a variância populacional σ^2 pela variância amostral s^2:

$$V(\bar{X}) = \frac{s^2}{n}$$

E o desvio padrão (ou erro padrão) da média amostral será a raiz quadrada:

$$EP(\bar{X}) = \sqrt{V(\bar{X})} = \sqrt{\frac{s^2}{n}}$$

$$EP(\bar{X}) = \frac{s}{\sqrt{n}}$$

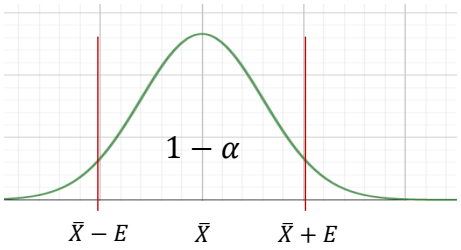
O método para a construção do intervalo de confiança será similar ao que vimos antes, porém utilizaremos a distribuição de **t-Student**, quando a população seguir distribuição **normal** (ou quando o tamanho da amostra permitir essa aproximação) com variância **desconhecida**.

Essa distribuição é similar à normal, também com formato de sino, porém mais achatada no centro e com caudas mais largas, ou seja, apresenta **maior variabilidade**.

ESCLARECENDO!

Precisamos utilizar uma outra distribuição, que não a distribuição normal, porque agora a **variância** deixou de ser um valor fixo e passou a ser também uma **estimativa**, o que justifica o uso de uma distribuição com **maior variabilidade** do que a normal.

Para isso, também será necessário encontrar os valores de $\bar{X} + E$ e $\bar{X} - E$ que delimitam uma probabilidade de $\mathbf{1 - \alpha}$, considerando a distribuição de t-Student.



A distribuição de t-Student também possui uma tabela para a distribuição padrão, com média igual a 0 e desvio padrão igual a 1. Assim, para encontrar os limites $\bar{X} + E$ e $\bar{X} - E$, utilizamos a seguinte transformação, similar à transformação para a normal padrão:

$$t = \frac{valor\ procurado - média\ da\ distribuição}{desvio\ padrão\ da\ distribuição}$$

No caso, os **valores procurados** são $X = \bar{X} + E$ e $X = \bar{X} - E$; a **média** da distribuição é $\bar{X}$; e o **desvio padrão** da distribuição é $EP(\bar{X}) = \frac{s}{\sqrt{n}}$:

$$t = \frac{\bar{X} + E\ - \bar{X}}{\frac{s}{\sqrt{n}}} = \frac{E}{\frac{s}{\sqrt{n}}}$$

Reorganizando essa expressão, obtemos a fórmula do erro:

$$E = t.\frac{s}{\sqrt{n}}$$

Ou seja, o erro é calculado assim como fizemos para a população com variância conhecida, apenas substituindo a variável da normal padrão z pela variável de t-Student t e o desvio padrão da população σ pela sua estimativa s.

O intervalo de confiança é da forma $\bar{X} \pm E = \left(\bar{X} - t.\frac{s}{\sqrt{n}}; \bar{X} + t.\frac{s}{\sqrt{n}}\right)$.

O valor de t depende do **nível de confiança**, assim como o valor de z, mas também do **número de graus de liberdade** da distribuição, igual ao **tamanho da amostra menos 1**:

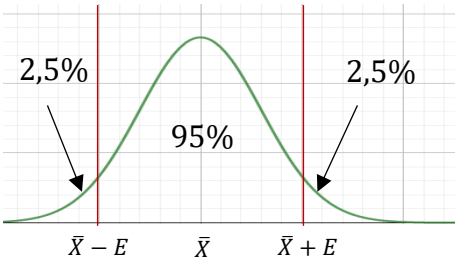
$$nº\ de\ graus\ de\ liberdade = \boldsymbol{n - 1}$$

Por exemplo, para uma amostra de tamanho $n = 5$, precisamos buscar o valor de t considerando $n - 1 = 4$ graus de liberdade.

Abaixo, inserimos parte da tabela de t-Student, que apresenta os valores de t para os quais as probabilidades $P(T < t)$ constam na primeira linha, considerando os graus de liberdade indicados na primeira coluna:

n	$t_{0,55}$	$t_{0,60}$	$t_{0,70}$	$t_{0,80}$	$t_{0,90}$	$t_{0,95}$	$t_{0,975}$	$t_{0,99}$	$t_{0,995}$
1	0,1584	0,3249	0,7265	1,3764	3,0777	6,3138	12,7062	31,8205	63,6567
2	0,1421	0,2887	0,6172	1,0607	1,8856	2,9200	4,3027	6,9646	9,9248
3	0,1366	0,2767	0,5844	0,9785	1,6377	2,3534	3,1824	4,5407	5,8409
4	0,1338	0,2707	0,5686	0,9410	1,5332	2,1318	2,7764	3,7469	4,6041
5	0,1322	0,2672	0,5594	0,9195	1,4759	2,0150	2,5706	3,3649	4,0321

Supondo que o nível de confiança seja $1 - \alpha = 95\%$, então a probabilidade associada aos valores acima e abaixo desse intervalo é $\frac{\alpha}{2} = 2{,}5\%$.



Ou seja, a probabilidade associada aos valores abaixo do valor crítico superior é:

$$P(X < \bar{X} + E) = 95\% + 2{,}5\% = 97{,}5\%$$

Assim, devemos buscar o valor de t para o qual $P(T < t) = 0{,}975$, considerando $n - 1 = 4$ graus de liberdade. Pela tabela acima, temos $t = 2{,}7764 \cong 2{,}78$. Logo, o intervalo será da seguinte forma:

$$\bar{X} \pm 2{,}78.\frac{s}{\sqrt{5}}$$

Vamos supor que a variância amostral observada seja $s^2 = 0{,}0125$. Nesse caso, o desvio padrão é dado por:

$$s = \sqrt{s^2} = \sqrt{0{,}0125}$$

Portanto, a margem de erro para esse exemplo é:

$$E = 2{,}78.\frac{s}{\sqrt{5}} = 2{,}78.\frac{\sqrt{0{,}0125}}{\sqrt{5}} = 2{,}78 \times \sqrt{0{,}0025} = 2{,}78 \times 0{,}05 = 0{,}139$$

Vamos supor que a média observada dessa amostra tenha sido $\bar{X} = 1{,}8$. Assim, o intervalo de confiança é:

$$\bar{X} + E = 1{,}8 + 0{,}139 = 1{,}939$$

$$\bar{X} - E = 1{,}8 - 0{,}139 = 1{,}661$$

$$(1{,}661;\ 1{,}939)$$

(FCC/2019 - SEFAZ-BA) Para obter um intervalo de confiança de 90% para a média p de uma população normalmente distribuída, de tamanho infinito e variância desconhecida, extraiu-se uma amostra aleatória de tamanho 9 dessa população, obtendo-se uma média amostral igual a 15 e variância igual a 16.

Considerou-se a distribuição t de *Student* para o teste unicaudal tal que a probabilidade P(t - t_0) = 0,05, com n graus de liberdade.

Com base nos dados da amostra, esse intervalo é igual a

Dados:

n	7	8	9	10	11
$t_{0,05}$	1,90	1,86	1,83	1,81	1,80

a) (12,56; 17,44)

b) (13,76; 16,24)

c) (12,47; 17,53)

d) (12,59; 17,41)

e) (12,52; 17,48)

Comentários:

Essa questão trabalha com um intervalo de confiança para a média de uma população de tamanho **infinito** (não precisamos do fator de correção) com **variância desconhecida**, dado por:

$$\bar{X} \pm E = \left(\bar{X} - t.\frac{s}{\sqrt{n}}; \bar{X} + t.\frac{s}{\sqrt{n}}\right)$$

Pelo enunciado, sabemos que o tamanho amostral é $n = 9$. Logo, temos $n - 1 = 8$ graus de liberdade.

Pela tabela fornecida, para 8 de graus de liberdade, observamos que $t = 1{,}86$.

Considerando que a variância amostral observada é $s^2 = 16$, logo $s = \sqrt{s^2} = \sqrt{16} = 4$, o erro é dado por:

$$E = t.\frac{s}{\sqrt{N}} = 1{,}86 \times \frac{4}{3} = 2{,}48$$

Assim, o intervalo de confiança para a média $\bar{X} = 15$ é:

$$\bar{X} - E = 15 - 2{,}48 = 12{,}52$$
$$\bar{X} + E = 15 + 2{,}48 = 17{,}48$$
$$(12{,}52;\ 17{,}48)$$

Gabarito: E.

(FGV/2022 - TJDFT) Em um modelo de simulação de uma fila com apenas um servidor para atendimento, foram realizadas 9 replicações para determinar o número médio de pessoas em fila. Os resultados obtidos para cada replicação estão no quadro a seguir.

Replicação	1	2	3	4	5	6	7	8	9	Média	Desvio padrão	Variância
Média de pessoas na fila	3,8	3,5	4,5	1,4	2	1,5	1,1	3,2	2,5	2,61	1,20	1,44

O intervalo bilateral de confiança de 95% para a média é, aproximadamente:

a) (1,83; 3,39)

b) (1,73; 3,50)

c) (1,69; 3,53)

d) (0,33; 4,83)

e) (0,04; 5,12)

Nessa prova, foi fornecida a tabela de t-Student da forma P(T > t_0), parcialmente replicada a seguir.

Grau de liberdade	Área da cauda superior				
	0,1	0,05	0,025	0,01	0,005
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169

Comentários:

Essa questão trabalha com um intervalo de confiança para a média, com variância desconhecida:

$$\bar{X} \pm t.\frac{s}{\sqrt{n}}$$

Em que s é a estimativa para o desvio padrão. Pela tabela fornecida na questão, temos $\bar{X} = 2{,}61$ e $s = 1{,}20$.

Ademais, o enunciado informa que $n = 9$ (logo, $\sqrt{n} = \sqrt{9} = 3$).

Um intervalo com 95% de confiança deixa 2,5% abaixo do limite inferior e 2,5% acima do limite superior.

Logo, precisamos do valor de t_0 associada a uma probabilidade $P(T > t_0) = 2,5\% = 0,025$.

Pela tabela de t-Student fornecida, observamos que, para n - 1 = **8 graus de liberdade** e **0,025 de área da cauda superior**, temos $t_0 = 2,306 \cong 2,3$. Substituindo esses dados na fórmula do intervalo de confiança:

$$2,61 \pm 2,3.\frac{1,2}{3} = 2,61 \pm 2,3 \times 0,4 = 2,61 \pm 0,92 = (1,69; 3,53)$$

Gabarito: C

Diferença entre populações

Para **comparar** os resultados provenientes de **duas populações independentes**, podemos construir um intervalo de confiança para a **diferença entre as médias** de duas populações.

Suponha que desejamos comparar o tempo médio de estudo entre os concurseiros solteiros e os concurseiros casados. Para isso, selecionamos uma amostra de cada um dos grupos e calculamos as médias de cada uma delas, $\overline{x_S}$ e $\overline{x_C}$. Em seguida, construímos um intervalo para a diferença $\overline{x_S} - \overline{x_C}$.

Analogamente ao intervalo de confiança para a média de uma população, somamos e subtraímos um **erro** da diferença observada nas amostras:

$$IC = (\overline{x_S} - \overline{x_C}) \pm E$$

E o erro corresponde ao produto do **valor crítico** tabelado, associado ao nível de confiança desejado, z, multiplicado pelo **erro padrão** (ou desvio padrão) do estimador, assim como vimos anteriormente:

$$E = z.\sigma_{\overline{x_S}-\overline{x_C}}$$

A diferença está no cálculo do erro padrão $\sigma_{\overline{x_S}-\overline{x_C}}$. Vejamos.

A variância da diferença entre as médias corresponde à **soma** das variâncias, uma vez que as médias são variáveis independentes:

$$Var(\overline{x_S} - \overline{x_C}) = Var(\overline{x_S}) + Var(\overline{x_C})$$

Como vimos antes, a variância da média amostral de cada população é a **razão** entre a variância da população e o tamanho da amostra:

$$Var(\overline{x_S} - \overline{x_C}) = \frac{Var(X_S)}{n_S} + \frac{Var(X_C)}{n_C}$$

Sabendo que o erro padrão é a **raiz quadrada** da variância, temos:

$$\sigma_{\overline{x_S}-\overline{x_C}} = \sqrt{\frac{Var(X_S)}{n_S} + \frac{Var(X_C)}{n_C}}$$

Vamos supor que a variância da população de solteiros seja $Var(X_S) = 1$ e a de casados seja $Var(X_C) = 3$. Supondo que tenha sido extraída uma amostra de tamanho 100 de cada população, ou seja, $n_S = 100$ e $n_C = 100$, o erro padrão da diferença entre médias é:

$$\sigma_{\overline{x_S}-\overline{x_C}} = \sqrt{\frac{1}{100} + \frac{3}{100}} = \sqrt{\frac{4}{100}} = \frac{2}{10} = 0{,}2$$

Para $z = 2$, que corresponde a um nível de aproximadamente 95% de confiança, a margem de erro é:

$$E = 2 \times 0{,}2 = 0{,}4$$

Supondo que a média observada na amostra dos solteiros tenha sido $\overline{x_S} = 7$ horas e a dos casados tenha sido $\overline{x_C} = 6$ horas, o intervalo de confiança para a diferença entre as médias é:

$$IC = (7 - 6) \pm 0{,}4 = 1 \pm 0{,}4 = (0{,}6; 1{,}4)$$

Alternativamente, podemos calcular as probabilidades associadas às diferenças entre as médias, utilizando a **transformação** para a normal padrão:

$$z = \frac{valor\ observado - média\ populacional}{desvio\ padrão}$$

$$z = \frac{(\overline{x_S} - \overline{x_C}) - (\mu_S - \mu_C)}{\sigma_{\overline{x_S}-\overline{x_C}}}$$

Para o nosso exemplo, supondo que as verdadeiras médias populacionais sejam $\mu_S = 6{,}8$ e $\mu_C = 6{,}2$, o valor de z associado à diferença observada, $\overline{x_S} - \overline{x_C} = 1$, é:

$$z = \frac{1 - (6{,}8 - 6{,}2)}{0{,}2} = \frac{0{,}4}{0{,}2} = 2$$

Sabendo que $P(-2 \leq Z \leq 2) \cong 95\%$, para esse exemplo, temos:

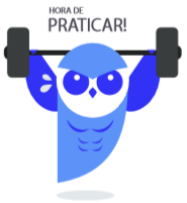
$$P(-1 \leq \overline{x_S} - \overline{x_C} \leq 1) \cong 95\%$$

Se a questão fornecer os **desvios padrão das populações**, é necessário elevá-los ao quadrado, para obter as **variâncias**, antes de aplicar a fórmula.

Não some os desvios padrão populacionais diretamente.

Caso as variâncias populacionais sejam desconhecidas e estimadas a partir das amostras, utilizamos a distribuição de t-Student, em vez da distribuição normal. Denotando as populações por X e Y, o erro padrão do estimador, nessa situação, é dado por:

$$s_{\bar{x}-\bar{y}} = \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$



(FGV/2023 - TCE/ES) Duas máquinas de empacotar, X e Y, estão reguladas de modo que cada pacote tenha média de 5 quilos e desvio padrão de 0,2 quilo.

Seja $\bar{X}$ o peso médio dos pacotes enchidos pela máquina X e $\bar{Y}$ o peso médio dos pacotes enchidos pela máquina Y. Suponha que as máquinas operem de forma independente e que os pesos dos pacotes enchidos por elas sigam uma distribuição normal.

Selecionou-se uma amostra aleatória de 128 pacotes de cada máquina. A probabilidade de que a diferença entre os pesos médios não ultrapasse 5%, isto é, $Prob(-0{,}05 < \bar{X} - \bar{Y} < 0{,}05)$, é:

a) $Prob(-2{,}5 < Z < 2{,}5)$, sendo $Z \sim N(0,1)$;

b) $Prob(-2 < Z < 2)$, sendo $Z \sim N(0,1)$;

a) $Prob(-1{,}25 < Z < 1{,}25)$, sendo $Z \sim N(0,1)$;

a) $Prob(-1 < Z < 1)$, sendo $Z \sim N(0,1)$;

a) $Prob(-0{,}25 < Z < 0{,}25)$, sendo $Z \sim N(0,1)$.

Comentários:

Para resolver essa questão, precisamos utilizar a fórmula da normal padrão para a diferença entre as médias:

$$z = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sigma_{\bar{x}-\bar{y}}}$$

Em que o erro padrão do estimador da diferença é dado por:

$$\sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{Var(X)}{n_X} + \frac{Var(Y)}{n_Y}}$$

O enunciado informa que o desvio padrão de ambas as populações é igual a 0,2, logo a variância é:

$$Var(X) = Var(Y) = 0{,}2^2 = 0{,}04$$

Sabendo que o tamanho da amostra de cada população é $n_X = n_Y = 128$, temos:

$$\sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{0{,}04}{128} + \frac{0{,}04}{128}} = \sqrt{2 \times \frac{0{,}04}{128}} = \sqrt{\frac{0{,}04}{64}} = \frac{0{,}2}{8} = 0{,}025$$

Sabendo que as médias são iguais, temos $\mu_X - \mu_Y = 0$. Assim, o valor de z associado a uma diferença $|\bar{x} - \bar{y}| = 0{,}05$ é:

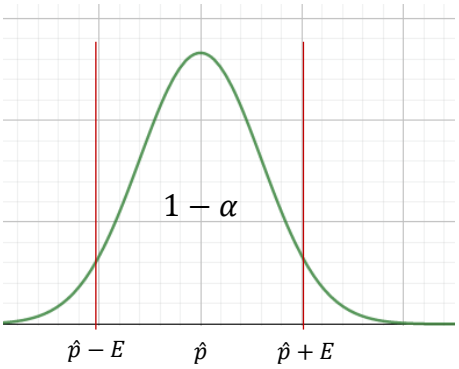
$$z = \frac{0{,}05 - 0}{0{,}025} = 2$$

Gabarito: B

Intervalo de Confiança para a Proporção

Para estimar a proporção populacional p, utilizamos a **proporção observada na amostra** $\hat{p}$.

Para viabilizar os cálculos, as questões consideram a aproximação da distribuição desse estimador a uma **normal**. Assim, devemos encontrar os valores limites que delimitam uma área sob a curva normal, correspondente a $1 - \alpha$:



A variância da proporção amostral é dada por:

$$V(\hat{p}) = \frac{\hat{p}.\hat{q}}{n}$$

Sendo $\hat{q} = 1 - \hat{p}$. O desvio padrão (erro padrão do estimador) é, portanto:

$$EP(\hat{p}) = \sqrt{V(\hat{p})} = \sqrt{\frac{\hat{p}.\hat{q}}{n}}$$

Sabendo que a média da distribuição é a proporção amostral $\hat{p}$, então a transformação para a normal padrão de $\hat{p} + E$ é:

$$z = \frac{valor\ procurado - média}{desvio\ padrão}$$

$$z = \frac{\hat{p} + E\ - \hat{p}}{\sqrt{\frac{\hat{p}.\hat{q}}{n}}} = \frac{E}{\sqrt{\frac{\hat{p}.\hat{q}}{n}}}$$

Reorganizando essa expressão, obtemos a fórmula do **erro do intervalo de confiança**:

$$E = z.\sqrt{\frac{\hat{p}.\hat{q}}{n}}$$

E o intervalo de confiança será $\hat{p} \pm E = \left(\hat{p} - z.\sqrt{\frac{\hat{p}.\hat{q}}{n}}; \hat{p} + z.\sqrt{\frac{\hat{p}.\hat{q}}{n}}\right)$.

Vamos supor que tenhamos estimado a proporção amostral de defeito em $\hat{p} = 0{,}2$ (logo, $\hat{q} = 1 - \hat{p} = 0{,}8$), considerando uma amostra de $n = 100$ peças.

Nesse caso, o desvio padrão (ou erro padrão) é dado por:

$$EP(\hat{p}) = \sqrt{\frac{\hat{p}.\hat{q}}{n}} = \sqrt{\frac{0{,}2 \times 0{,}8}{100}} = \frac{\sqrt{0{,}16}}{\sqrt{100}} = \frac{0{,}4}{10} = 0{,}04$$

Considerando um nível de confiança de 95% (z = 1,96), a margem de erro será:

$$E = z.\sqrt{\frac{\hat{p}.\hat{q}}{n}} = 1{,}96 \times 0{,}04 = 0{,}0784 \cong 0{,}08$$

Então, o intervalo de confiança para a proporção será:

$$\hat{p} + z.\sqrt{\frac{\hat{p}.\hat{q}}{n}} \cong 0{,}2 + 0{,}08 = 0{,}28$$

$$\hat{p} - z.\sqrt{\frac{\hat{p}.\hat{q}}{n}} \cong 0{,}2 - 0{,}08 = 0{,}12$$

$$(0{,}12; 0{,}28)$$

Tamanho Amostral

Também podemos determinar o tamanho amostral, para um dado nível de confiança $(1 - \alpha)$ e um valor de erro máximo E. Reorganizando a fórmula do erro que acabamos de ver, temos:

$$n = \left(\frac{z}{E}\right)^2 \hat{p}.\hat{q}$$

Vamos supor que o erro máximo tolerável seja E = 0,04, mantendo os demais parâmetros iguais aos do exemplo anterior ($z = 1{,}96$ e $\hat{p} = 0{,}2$). Nesse caso, o tamanho da amostra n_2 será:

$$n_2 = \left(\frac{1{,}96}{0{,}04}\right)^2 \times 0{,}2 \times 0{,}8 = (49)^2 \times 0{,}16 = 2401 \times 0{,}16 \cong 384$$

TOME NOTA!

Para calcular o **tamanho máximo da amostra n**, utilizamos $\hat{p} = \hat{q} = 0,5$, pois essa é a proporção que **maximiza a variância** da proporção amostral.

Assim, mesmo **sem** conhecer a proporção amostral, é possível determinar um **valor seguro** para o **tamanho amostral**, considerando $\hat{p} = \hat{q} = 0,5$ na fórmula do tamanho amostral:

Supondo $z = 1,96$ e um erro máximo tolerável $E = 0,04$, o **valor seguro** para o tamanho amostral n_*, sendo desconhecida a proporção amostral, é:

$$n_* = \left(\frac{1,96}{0,04}\right)^2 \times 0,5 \times 0,5 = (49)^2 \times 0,25 = 2401 \times 0,25 \cong 600$$

Quanto **mais próximo** de $\hat{p} = \hat{q} = 0,5$, **maior** será o tamanho amostral. Por exemplo, o tamanho amostral para $\hat{p} = 0,4$ (e $\hat{q} = 1 - \hat{p} = 0,6$) é **maior** do que para $\hat{p} = 0,3$ (e $\hat{q} = 1 - \hat{p} = 0,7$).

Ademais, como multiplicamos essas duas proporções no cálculo do tamanho amostral, não importa qual proporção assume o maior ou o menor valor. Assim, o tamanho da amostra para $\hat{p} = 0,4$ ($\hat{q} = 1 - \hat{p} = 0,6$) é **igual** ao tamanho da amostra para $\hat{p} = 0,6$ ($\hat{q} = 1 - \hat{p} = 0,4$).

Essas fórmulas valem para uma população **infinita** **OU** amostras extraídas **com reposição**.

Fator de correção

Caso a população seja **finita** e as amostras extraídas **sem reposição**, será necessário aplicar o fator de correção para população finita. Para isso, devemos multiplicar a fórmula do erro pelo fator $\sqrt{\frac{N-n}{N-1}}$:

$$E = z.\sqrt{\frac{\hat{p}.\hat{q}}{n}}.\sqrt{\frac{N-n}{N-1}}$$

Assim como vimos para a média, esse ajuste para populações finitas afeta no cálculo do **tamanho amostral**. Para calcular o tamanho da amostra necessário, primeiro calculamos uma **estimativa inicial** para a amostra, desconsiderando-se o fator de correção:

$$n_0 = \left(\frac{z}{E}\right)^2 \hat{p}.\hat{q}$$

Quando a questão não informa uma estimativa para as proporções e nem um nível de confiança, utilize a seguinte fórmula para a **amostra inicial**, que considera apenas a **margem de erro**:

$$n_0 = \frac{1}{E^2}$$

Em seguida, ajustamos a amostra inicial:

$$n = \frac{N \times n_0}{N + n_0}$$

Considerando um erro amostral $E = 0{,}05$, a estimativa inicial para o tamanho da amostra é:

$$n_0 = \frac{1}{0{,}05^2} = \frac{1}{0{,}0025} = 400$$

Supondo que a população tem tamanho $N = 1000$, o tamanho amostral ajustado será:

$$n = \frac{1.000 \times 400}{1.000 + 400} = \frac{400.000}{1.400} \cong 286$$

O ajuste da amostra inicial, quando ela é calculada com base apenas na margem de erro $n_0 = \frac{1}{E^2}$, corresponde à **fórmula de Slovin** para o tamanho amostral[1]:

$$n = \frac{N}{1+N.E^2}$$

[1] Vamos demonstrar isso! Considerando $n_0 = \frac{1}{E^2}$, a amostra ajustada pode ser calculada como:

$$n = \frac{N \times n_0}{N + n_0} = \frac{N \times \frac{1}{E^2}}{N + \frac{1}{E^2}} = \frac{N}{E^2 \times \left(N + \frac{1}{E^2}\right)} = \frac{N}{N.E^2 + 1}$$

Que é a fórmula de Slovin!

(CESPE 2018/PF) Determinado órgão governamental estimou que a probabilidade p de um ex-condenado voltar a ser condenado por algum crime no prazo de 5 anos, contados a partir da data da libertação, seja igual a 0,25. Essa estimativa foi obtida com base em um levantamento por amostragem aleatória simples de 1.875 processos judiciais, aplicando-se o método da máxima verossimilhança a partir da distribuição de Bernoulli. Sabendo que P(Z < 2) = 0,975, em que Z representa a distribuição normal padrão, julgue o item que se segue, em relação a essa situação hipotética.

A estimativa intervalar 0,25 ± 0,05 representa o intervalo de 95% de confiança do parâmetro populacional p.

Comentários:

O intervalo de confiança para a proporção é dado por:

$$\hat{p} \pm E = \hat{p} \pm z \times \sqrt{\frac{\hat{p} \times \hat{q}}{n}}$$

O enunciado informa que

- A proporção amostral é $\hat{p} = 0{,}25$, logo $\hat{q} = 1 - \hat{p} = 0{,}75$
- O tamanho da amostra é $n = 1.875$
- O valor de z para 95% de confiança é $z = 2$

Substituindo esses dados na fórmula do erro do intervalo, temos:

$$E = 2 \times \sqrt{\frac{0{,}25 \times 0{,}75}{1.875}} = 2 \times \sqrt{\frac{0{,}1875}{1.875}} = 2 \times 0{,}01 = 0{,}02$$

Assim, o intervalo de confiança é:

$$\hat{p} \pm E = 0{,}25 \pm 0{,}02$$

Gabarito: Errado.

(2019/FMS) Uma pesquisa tem como finalidade conhecer a proporção de pessoas em Teresina que teriam interesse em frequentar uma nova franquia de lanchonete vinda do exterior. O empreendedor diz que só vale a pena a instalação da franquia, se pelo menos 10% da população tivesse interesse em frequentar o estabelecimento.

Supondo que a proporção máxima da população não será maior que 30%, qual tamanho de amostra (aproximado) tal que a diferença entre a proporção populacional e proporção amostral não tenha um erro maior que três pontos percentuais, com uma confiança de 95%. Obs.: $z_\gamma = 1{,}96$.

a) 897

b) 683

c) 700

d) 300

e) 654

Comentários:

O tamanho amostral para a construção de um **intervalo de confiança para a proporção** é dado por:

$$n = \left(\frac{z}{E}\right)^2 \hat{p}.\hat{q}$$

O enunciado informa que $z = 1{,}96$ e $E = 0{,}03$.

Ademais, informa que a proporção **máxima** $\hat{p}$ é de 30%. Essa é a proporção que deve ser utilizada, por ser a **mais próxima de 50%**, para maximizar o tamanho amostral.

Sendo $\hat{p} = 0{,}3$, logo $\hat{q} = 1 - \hat{p} = 0{,}7$, temos:

$$n = \left(\frac{1{,}96}{0{,}03}\right)^2 \times 0{,}3 \times 0{,}7 \cong 897$$

Gabarito: A.

(FGV/2022 – MPE/SC) Uma empresa recebeu um lote muito grande, milhões de peças de refugo, e deseja saber quantas peças deverá examinar para estimar a proporção de itens defeituosos, de modo que o erro de estimação seja no máximo 2%. Será empregada uma seleção aleatória de itens onde cada um será classificado como defeituoso ou não defeituoso. Deseja-se extrair uma amostra aleatória de tamanho n.

Tendo como padrão um grau de confiança de 95%, o tamanho da amostra necessário para garantir o processo é:

a) 189;

b) 384;

c) 600;

d) 1681;

e) 2401.

Comentários:

Essa questão também trabalha com o tamanho amostral, para a estimação de proporções:

$$n = \left(\frac{z}{E}\right)^2 .\hat{p}.\hat{q}$$

Em que z é o valor da tabela normal padrão associado ao nível de confiança desejado; E é a margem de erro (ou erro máximo de estimação); $\hat{p}$ é a estimativa para a proporção de sucesso e $\hat{q}$ é a estimativa para a proporção de fracasso, sendo $\hat{q} = 1 - \hat{p}$.

O enunciado informa que o erro máximo é $E = 2\% = 0{,}02$; e que o grau de confiança é de 95%, logo, z=1,96.

A questão não informa a estimativa para a proporção de sucesso, logo, devemos considerar aquela que maximiza o tamanho da amostra, qual seja $\hat{p} = \hat{q} = 0{,}5$. Substituindo esses dados na fórmula, temos:

$$n = \left(\frac{1{,}96}{0{,}02}\right)^2 \times 0{,}5 \times 0{,}5 = (98)^2 \times 0{,}25 = 2401$$

Gabarito: E

Intervalo de Confiança para a Variância

O **estimador da variância** s^2 (variância amostral), multiplicado pelo fator $\left(\frac{n-1}{\sigma^2}\right)$, segue uma distribuição **qui-quadrado** com $n-1$ **graus de liberdade**:

$$\chi^2_{n-1} = \left(\frac{n-1}{\sigma^2}\right) s^2$$

Como essa distribuição é **assimétrica**, utilizaremos a tabela da distribuição qui-quadrado para encontrar tanto o limite superior χ^2_{SUP} quanto o limite inferior χ^2_{INF} do intervalo de confiança:

$$\chi^2_{INF} < \left(\frac{n-1}{\sigma^2}\right) s^2 < \chi^2_{SUP}$$

Isolando σ^2 nessa expressão, temos:

$$\frac{(n-1).s^2}{\chi^2_{SUP}} < \sigma^2 < \frac{(n-1).s^2}{\chi^2_{INF}}$$

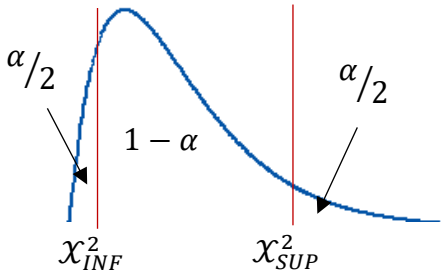
Ou seja, para calcular o limite **inferior**, dividimos pelo valor crítico **superior** da tabela $\boldsymbol{\chi^2_{SUP}}$, e para calcular o limite **superior** dividimos pelo valor crítico **inferior** da tabela $\boldsymbol{\chi^2_{INF}}$.

Isso porque χ^2_{SUP} é **maior** que χ^2_{INF}. Assim, quando **dividimos** por χ^2_{SUP}, obtemos um valor **menor** (limite inferior) do que quando dividimos por χ^2_{INF} (limite superior).

Assim, o **intervalo de confiança para a variância** é da forma:

$$\left(\frac{(n-1).s^2}{\chi^2_{SUP}}; \frac{(n-1).s^2}{\chi^2_{INF}}\right)$$

Os valores χ^2_{SUP} e χ^2_{INF} são obtidos a partir da **tabela** da distribuição qui-quadrado, dependendo do **nível de confiança** $1-\alpha$ desejado:



Observamos que o valor $\mathcal{X}^2_{INF}$ deixa $\frac{\alpha}{2}$ da distribuição abaixo e o valor $\mathcal{X}^2_{SUP}$ deixa $\frac{\alpha}{2}$ da distribuição acima, assim como vimos para os outros intervalos de confiança. A diferença é que agora não estamos mais lidando com uma distribuição simétrica:

$$P(\mathcal{X}^2 < \mathcal{X}^2_{INF}) = \frac{\alpha}{2}$$

$$P(\mathcal{X}^2 < \mathcal{X}^2_{SUP}) = 1 - \frac{\alpha}{2}$$

Os valores da tabela da distribuição qui-quadrado também dependem do número de graus de liberdade, igual ao **tamanho da amostra menos 1**.

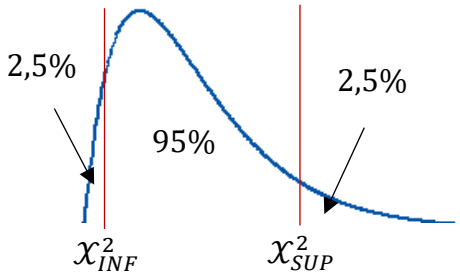
$$número\ de\ graus\ de\ liberdade = \boldsymbol{n - 1}$$

Para uma amostra de tamanho n = 5, por exemplo, devemos utilizar os valores da tabela da distribuição qui-quadrado com n - 1 = 4 graus de liberdade.

A tabela a seguir apresenta os valores de x que delimitam as probabilidades $P(\mathcal{X}^2 < x)$ da distribuição qui-quadrado com 4 graus de liberdade indicadas na primeira linha:

$P(\mathcal{X}^2_4 < x)$	**0,005**	**0,01**	**0,025**	**0,05**	**0,1**	**0,25**	**0,5**	**0,75**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
x	0,21	0,30	0,48	0,71	1,06	1,92	3,36	5,39	7,78	9,49	11,14	13,28	14,86

Para um nível de confiança de $1 - \alpha = 95\%$, temos:



Pela tabela acima, observamos que o valor de x que delimita uma probabilidade $P(\mathcal{X}^2 < x) = 0{,}025$, que corresponde ao limite inferior, é $x = \mathcal{X}^2_{INF} = 0{,}48$; e o valor de x que delimita uma probabilidade $P(\mathcal{X}^2 < x) = 0{,}975$, que corresponde ao limite superior, é $x = \mathcal{X}^2_{SUP} = 11{,}14$.

Considerando que a variância amostral é $s^2 = 0{,}0125$, os limites do intervalo de confiança são:

$$\frac{(n-1).s^2}{\mathcal{X}^2_{SUP}} = \frac{4 \times 0{,}0125}{11{,}14} = \frac{0{,}05}{11{,}14} \cong 0{,}004$$

$$\frac{(n-1).s^2}{\mathcal{X}^2_{INF}} = \frac{4 \times 0{,}0125}{0{,}48} = \frac{0{,}05}{0{,}48} \cong 0{,}104$$

E o intervalo é $(0{,}004; 0{,}104)$.

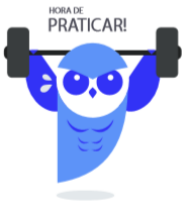
Caso a questão forneça a soma dos quadrados $\sum(X_i - \bar{X})^2$, em vez da estimativa para a variância s^2, podemos utilizá-lo diretamente no cálculo do intervalo de confiança.

$$IC = \left(\frac{(n-1).s^2}{\mathcal{X}^2_{SUP}}; \frac{(n-1).s^2}{\mathcal{X}^2_{INF}}\right)$$

Sabendo que $s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$, o intervalo de confiança pode ser calculado como:

$$IC = \left(\frac{(n-1)}{\mathcal{X}^2_{SUP}} \times \frac{\sum(X_i - \bar{X})^2}{n-1}; \frac{(n-1)}{\mathcal{X}^2_{INF}} \times \frac{\sum(X_i - \bar{X})^2}{n-1}\right)$$

$$IC = \left(\frac{\sum(X_i - \bar{X})^2}{\mathcal{X}^2_{SUP}}; \frac{\sum(X_i - \bar{X})^2}{\mathcal{X}^2_{INF}}\right)$$



(FGV/2019 – DPE-RJ) Com o objetivo de produzir uma estimativa por intervalo para a variância populacional, realiza-se uma amostra de tamanho n = 4, obtendo-se, após a extração, os seguintes resultados:

$X_1 = 6$, $X_2 = 3$, $X_3 = 11$ e $X_4 = 12$

Informações adicionais:

$P(X^2_4 < 0{,}75) = 0{,}05$ $P(X^2_3 < 0{,}40) = 0{,}05$

$P(X^2_4 < 10{,}8) = 0{,}95$ $P(X^2_3 < 9) = 0{,}95$

Então, sobre o resultado da estimação, e considerando-se um grau de confiança de 90%, tem-se que:

a) $5 < \sigma^2 < 72$;

b) $8 < \sigma^2 < 180$;

c) $6 < \sigma^2 < 135$;

d) $4 < \sigma^2 < 22$;

e) $6 < \sigma^2 < 24$.

Comentários:

O intervalo de confiança para a variância é dado por:

$$\left(\frac{(n-1).s^2}{\mathcal{X}^2_{SUP}}; \frac{(n-1).s^2}{\mathcal{X}^2_{INF}}\right)$$

Primeiro, precisamos calcular a estimativa para variância s^2 (variância amostral):

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Para isso, precisamos da média amostral:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{6+3+11+12}{4} = \frac{32}{4} = 8$$

Logo, o estimador da variância é:

$$s^2 = \frac{(6-8)^2 + (3-8)^2 + (11-8)^2 + (12-8)^2}{4-1} = \frac{(-2)^2 + (-5)^2 + (3)^2 + (4)^2}{3}$$

$$s^2 = \frac{4+25+9+16}{3} = \frac{54}{3} = 18$$

O enunciado informa que o tamanho da amostra é n = 4, logo, o número de graus de liberdade é n – 1 = 3.

Sabendo que o intervalo de confiança é de $1-\alpha = 90\% = 0{,}9$, precisamos dos valores da tabela que delimitam as probabilidades $P(\mathcal{X}^2 < \mathcal{X}^2_{INF}) = \frac{\alpha}{2} = 0{,}05$ e $P(\mathcal{X}^2 < \mathcal{X}^2_{SUP}) = 1 - \frac{\alpha}{2} = 0{,}95$.

O enunciado informa que $P(\mathcal{X}^2_3 < 0{,}40) = 0{,}05$, logo $\mathcal{X}^2_{INF} = 0{,}4$; e $P(\mathcal{X}^2_3 < 0{,}9) = 0{,}95$, logo $\mathcal{X}^2_{Sup} = 9$.

Assim, os limites inferior e superior são:

$$L_{Inf} = \frac{3 \times 18}{9} = 6$$

$$L_{Sup} = \frac{3 \times 18}{0{,}4} = 135$$

Gabarito: C.

ESQUEMATIZANDO

Estimação Intervalar

Intervalo para a Média – Variância conhecida (distribuição normal)**:**

Margem de Erro: $\boldsymbol{E = z.\frac{\sigma}{\sqrt{n}}}$; Tamanho amostral: $\boldsymbol{n = \left(z.\frac{\sigma}{E}\right)^2}$

Intervalo para a Média – Variância desconhecida (distribuição t-Student)**:**

Margem de Erro: $\boldsymbol{E = t_{n-1}.\frac{s}{\sqrt{n}}}$; Tamanho Amostral: $\boldsymbol{n = \left(t.\frac{s}{E}\right)^2}$

Intervalo para a Proporção:

Margem de Erro: $\boldsymbol{E = z.\sqrt{\frac{\hat{p}.\hat{q}}{n}}}$; Tamanho Amostral: $\boldsymbol{n = \left(\frac{z}{E}\right)^2 \hat{p}.\hat{q}}$

Intervalo para a Variância: $\boldsymbol{\left(\frac{(n-1).s^2}{x^2_{SUP}}; \frac{(n-1).s^2}{x^2_{INF}}\right)}$

RESUMO DA AULA

Estimadores

⇒ **Média amostral** $\bar{X}$**:** $E(\bar{X}) = \mu, V(\bar{X}) = \frac{\sigma^2}{n}, EP(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

⇒ **Proporção amostral** $\hat{p}$**:** $E(\hat{p}) = p, V(\hat{p}) = \frac{p.q}{n}, EP(\hat{p}) = \sqrt{\frac{p.q}{n}}$

⇒ Estimador da **variância amostral:** $s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$

Erro Padrão é o **desvio padrão** (raiz quadrada da variância) do estimador

Propriedades dos Estimadores

⇒ Suficiente (contempla todas as informações para estimar o parâmetro populacional)

⇒ Não Tendencioso (esperança do estimador é igual ao parâmetro populacional)

⇒ Eficiente (menor variância possível)

⇒ Consistente (estimativas convergem com o aumento do tamanho amostral)

Métodos de Estimação

⇒ Método dos Momentos: Resulta em $\bar{X}$ como estimador para a média; e em $\widehat{\sigma^2} = \frac{\sum(X_i - \bar{X})^2}{n}$ como estimador para a variância ($\widehat{\sigma^2}$ é tendencioso)

⇒ Método da Máxima Verossimilhança: Resulta em $\hat{p}$ como estimador para a proporção; para uma população normal, em $\bar{X}$ como estimador da média e $\widehat{\sigma^2} = \frac{\sum(X_i - \bar{X})^2}{n}$ como estimador para a variância ($\widehat{\sigma^2}$ é tendencioso)

⇒ Método dos Mínimos Quadrados: Resulta em $\bar{X}$ como estimador para a média; e em $\hat{p}$ como estimador para a proporção

Estimação Intervalar

Erro do Intervalo E (**metade** da **amplitude**)

⇒ Erro depende do nível de confiança desejado e do tamanho da amostra

⇒ Intervalo de confiança para população com variância conhecida: $\bar{X} \pm z.\frac{\sigma}{\sqrt{n}}$

⇒ Intervalo de confiança para população com variância desconhecida: $\bar{X} \pm t_{n-1}.\frac{\sigma}{\sqrt{n}}$

INFERÊNCIA BAYESIANA

Introdução

A inferência Bayesiana (ou estatística Bayesiana) considera **premissas diferentes** da inferência clássica (ou frequentista), que é a inferência que estudamos até agora.

Por um lado, a **inferência clássica** (ou **frequentista**) considera que **toda** a informação disponível está representada na **amostra**; e que o parâmetro populacional é **fixo**.

Já, na **inferência Bayesiana**, o parâmetro populacional a ser estimado é uma **variável aleatória** (e não fixo como na inferência clássica), que depende dos resultados observados na **amostra**. Essa distribuição do parâmetro populacional condicionada aos resultados da amostra é chamada de **distribuição a posteriori**.

A distribuição a posteriori depende não só das informações disponíveis na amostra (chamada de função de **verossimilhança** da amostra), mas também de outras informações que **conhecemos** e que não estarão representadas na amostra (**não observáveis**). Por exemplo, no contexto de eleições democráticas, sabemos que um candidato não terá 100% dos votos.

Essas informações previamente conhecidas podem ser consideradas **subjetivas** e compõem a chamada **distribuição a priori**. Elas são incorporadas no modelo Bayesiano, de modo que as estimativas variam de acordo com a distribuição a priori utilizada.

Por outro lado, a inserção de informações que não são verdadeiras torna o modelo **inapropriado**. Por exemplo, se for considerado que nenhum candidato terá 70% dos votos, mas se isso for de fato possível, a estimativa estará **enviesada**.

Ademais, o que chamávamos de **intervalo de confiança** na inferência clássica passa a ser chamado de **intervalo de credibilidade**, baseado na distribuição a posteriori. Enquanto, na inferência clássica, o nível de confiança é a probabilidade de o intervalo de confiança englobar o verdadeiro parâmetro populacional; na inferência Bayesiana, como o parâmetro populacional é uma variável aleatória, o nível de credibilidade é a probabilidade é de o parâmetro populacional pertencer ao intervalo de credibilidade.

Inferência Clássica ou Frequentista	Inferência Bayesiana
• Parâmetro populacional é **fixo** • Toda a informação está **disponível** na amostra • Apenas distribuição a **posteriori** • Intervalo de **confiança**	• Parâmetro populacional é **variável aleatória** • Depende de informações **não observáveis** (além das informações da amostra) • Depende de uma **distribuição a priori** (além da distribuição a posteriori) • Intervalo de **credibilidade**

(2017 – TRF – 2ª Região – *Adaptada*) Sobre a abordagem bayesiana para estimar um parâmetro θ, julgue as afirmativas a seguir.

I – Uma distribuição de probabilidade é atribuída para esse parâmetro.

II – A definição da distribuição priori pode ser totalmente subjetiva.

Comentários:

Na inferência bayesiana, considera-se que o parâmetro de interesse não é fixo, mas sim, uma variável aleatória. Ou seja, de fato, atribui-se uma distribuição ao parâmetro de interesse e a afirmativa I está correta.

Ademais, considera-se uma distribuição a priori, com informações subjetivas, que não podem ser observada na amostra. Logo, a afirmativa II também está correta.

Resposta: I e II certas.

(Cebraspe/2019 – TJ-AM) A respeito dos diferentes métodos de estimação de parâmetros, julgue o item a seguir.

A estimação de parâmetros pelo método bayesiano independe da distribuição a priori utilizada.

Comentários:

Na inferência bayesiana, são consideradas informações não observáveis na amostra (subjetivas) que compõem a chamada distribuição a priori. Logo, o item está incorreto.

Gabarito: Errado.

(VUNESP/2021 – EsFCEx) Assinale a alternativa correta sobre a comparação entre estatística clássica e estatística bayesiana.

a) A estatística bayesiana utiliza somente o conhecimento prévio do pesquisador (informação a priori) para a estimação dos parâmetros.

b) Somente a estatística clássica utiliza a verossimilhança na realização de suas inferências.

c) Na estatística clássica, as inferências são feitas com base na verossimilhança, e tratam os parâmetros como aleatórios e desconhecidos e os dados como aleatórios e conhecidos.

d) As inferências obtidas na estatística clássica e na estatística bayesiana para amostras pequenas são sempre coincidentes.

e) Na estatística clássica, é considerado que há apenas um valor para o parâmetro analisado e não uma distribuição de probabilidade.

Comentários:

Essa questão compara os conceitos da estatística clássica com os da estatística bayesiana.

Em relação à alternativa A, a estatística bayesiana considera, não apenas o conhecimento prévio (a priori), mas também as informações disponíveis na **amostra** (função de verossimilhança). Logo, a alternativa A está incorreta.

Em relação à alternativa B, a estatística clássica não é a única que considera as informações disponíveis na amostra (verossimilhança) - a estatística bayesiana também considera essa fonte de informações. Logo, a alternativa B está incorreta.

Em relação à alternativa C, a estatística clássica considera os parâmetros desconhecidos como **fixos**, e não como aleatórios. Logo, a alternativa C está incorreta.

Em relação à alternativa D, a inferência clássica não coincide necessariamente com a inferência Bayesiana, principalmente para amostras pequenas, isto é, quando há poucas informações disponíveis na amostra. Assim, a alternativas D está incorreta.

Em relação à alternativa E, na estatística clássica considera-se que o parâmetro desconhecido é fixo, ou seja, que há um único valor para ele, e não uma distribuição de probabilidade. Logo, a alternativa E está correta.

Gabarito: E.

Teorema de Bayes

Do ponto de vista Bayesiano, dispomos de alguma **informação** a respeito do parâmetro populacional θ desconhecido e buscamos **aumentar** esse nível de informação, observando os resultados das **amostras**.

O quanto o nível de informação aumenta é calculado pelo **Teorema de Bayes** (da Teoria de Probabilidade):

$$p(\theta|x) = \frac{p(\theta, x)}{p(x)} = \frac{p(x|\theta) \times p(\theta)}{p(x)}$$

A probabilidade $p(\theta)$ corresponde à distribuição **a priori** do parâmetro θ e representa o nível de informação antes da observação da amostra.

A probabilidade $p(x|\theta)$ dos resultados da amostra, condicionada a θ, corresponde à função de plausibilidade ou **verossimilhança** do resultado, em relação ao parâmetro θ. Ela pode ser indicada como $l(\theta; p)$.

Combinando essas duas fontes de informação (a priori e de verossimilhança), obtemos a distribuição **a posteriori** $p(\theta|x)$, representa o nível de informação após a observação da amostra.

A probabilidade $p(x)$, no denominador da fórmula, chamada **preditiva**, é a probabilidade associada ao resultado da amostra, para **todos** os possíveis valores de θ.

Se o parâmetro θ for uma variável discreta, ou seja, se houver uma quantidade contável de valores possíveis que ele pode assumir, a probabilidade preditiva pode ser calculada pelo Teorema da Probabilidade Total:

$$p(x) = \sum_i p(x, \theta_i) = \sum_i p(x|\theta_i) \times p(\theta_i)$$

Caso contrário, substituímos o somatório pela integral:

$$p(x) = \int p(x, \theta).d\theta = \int p(x|\theta).p(\theta).d\theta$$

Em ambos os casos, à direita, temos a fórmula da **esperança** da distribuição da amostra condicionada a θ, para todos os valores possíveis de θ:

$$p(x) = E_\theta[p(X|\theta)]$$

A probabilidade preditiva não é uma função do parâmetro θ e funciona como uma **constante normalizadora**, isto é, uma constante pela qual você divide o produto da probabilidade a priori pela função de verossimilhança (numerador da fórmula de Bayes), para que a probabilidade associada a todo o Espaço Amostral seja igual a 1.

Assim, a probabilidade a **posteriori** é **proporcional** ao produto da probabilidade a **priori** pela função de **verossimilhança**:

$$p(\theta|x) \propto p(x|\theta) \times p(\theta)$$

Que também pode ser representado como:

$$p(\theta|x) \propto l(\theta; x) \times p(\theta)$$

Para ilustrar, vamos considerar o exemplo[1] de um médico que acredita que um paciente esteja doente e vai aplicar um teste para obter mais informações. Nesse caso, o parâmetro desconhecido é o indicador da doença:

$$\theta = \begin{cases} 1, & se\ o\ paciente\ tem\ a\ doença \\ 0, & se\ o\ paciente\ não\ tem\ a\ doença \end{cases}$$

Na inferência clássica, as conclusões a respeito do parâmetro desconhecido seriam totalmente baseadas no resultado do teste. Porém, na inferência Bayesiana, a informação **prévia** que o médico detém, com base em seu contato com o paciente e em sua experiência, é incorporada ao modelo.

Vamos supor que o médico acredite que a probabilidade de o paciente ter a doença seja de 70%. Essa é a probabilidade a **priori** do parâmetro:

$$p(\theta) = P(\theta = 1) = 0{,}7$$

[1] Gamerman e Migon (1993)

E vamos supor que o teste indique um falso positivo com probabilidade de 40% e um falso negativo com probabilidade de 5%.

Assim, a probabilidade de o teste resultar em $X = 1$, dado que o paciente **não** está doente (falso positivo), é igual a 40%; e a probabilidade de o teste resultar em $X = 0$, dado que o paciente **não** está doente (resultado negativo adequado) é complementar:

$$P(X = 1|\theta = 0) = 0{,}4$$

$$P(X = 0|\theta = 0) = 1 - 0{,}4 = 0{,}6$$

Ademais, a probabilidade de o teste resultar em $X = 0$, dado que o paciente está **doente** (falso negativo), é igual a 5%; e a probabilidade de o teste resultar em $X = 1$, dado que o paciente está **doente** (resultado positivo adequado) é complementar:

$$P(X = 0|\theta = 1) = 0{,}05$$

$$P(X = 1|\theta = 1) = 1 - 0{,}05 = 0{,}95$$

Vamos considerar que o resultado do teste foi **positivo** $X = 1$.

Com base no resultado do teste, imaginamos que a probabilidade de o paciente ter a doença tenha **aumentado**.

Para calcular essa **nova probabilidade** de o paciente ter a doença, dado que o resultado do teste foi positivo (distribuição a posteriori), utilizamos o **Teorema de Bayes**:

$$P(\theta = 1|X = 1) = \frac{P(X = 1|\theta = 1) \times P(\theta = 1)}{P(X = 1)}$$

A função de **verossimilhança** $p(x|\theta) = P(X = 1|\theta = 1)$ corresponde à probabilidade de o resultado ser positivo (resultado observado), dado que o paciente está doente:

$$p(x|\theta) = P(X = 1|\theta = 1) = 0{,}95$$

A distribuição à **priori** $p(\theta) = P(\theta = 1)$ corresponde à probabilidade indicada inicialmente pelo médico:

$$P(\theta = 1) = 0{,}70$$

E a distribuição **preditiva** $p(x) = P(X = 1)$ corresponde à probabilidade de obter o resultado obtido $X = 1$, independentemente do valor de θ. Ela pode ser calculada pelo **Teorema da Probabilidade Total**:

$$P(X = 1) = P(X = 1|\theta = 1) \times P(\theta = 1) + P(X = 1|\theta = 0) \times P(\theta = 0)$$

Sabendo que a probabilidade a priori de o paciente não estar doente é $P(\theta = 0) = 1 - 0{,}7 = 0{,}3$, temos:

$$P(X = 1) = 0{,}95 \times 0{,}7 + 0{,}4 \times 0{,}3 = 0{,}665 + 0{,}12 = 0{,}785$$

Assim, a probabilidade à **posteriori**, que corresponde à probabilidade de o paciente estar doente, após o resultado do teste, é dada por:

$$P(\theta = 1|X = 1) = \frac{P(X = 1|\theta = 1) \times P(\theta = 1)}{P(X = 1)} = \frac{0{,}95 \times 0{,}70}{0{,}785} = \frac{0{,}665}{0{,}785} \cong 0{,}847$$

Que, de fato, é **maior** do que a probabilidade a priori. O aumento não foi tão grande porque a probabilidade do **falso positivo** $P(X = 1|\theta = 0) = 40\%$ é considerável.

Múltiplas Observações

É possível realizar um **novo procedimento** Y, após o primeiro X, para aumentar ainda mais o nível de informação a respeito do parâmetro θ.

Na hipótese de independência condicional de X e Y dado θ, a nova distribuição a posteriori, qual seja, a distribuição de θ após os resultados dos dois procedimentos $p(\theta|x, y)$ é dada por:

$$p(\theta|x, y) = \frac{p(y|\theta) \times p(\theta|x)}{p(y|x)}$$

Em que $p(y|\theta)$ é a função de **verossimilhança** do segundo procedimento, em relação a θ; e $p(\theta|x)$ corresponde à nova distribuição a **priori**, após o resultado do primeiro procedimento X.

Ou seja, a distribuição considerada a posteriori, após o teste X, agora é considerada a priori, antes de Y.

No denominador, temos a distribuição **preditiva** de y dado x, isto é, a probabilidade do resultado de Y, dado o resultado de X, para **todos** os possíveis valores de θ.

Na hipótese de independência condicional de X e Y dado θ, a probabilidade preditiva condicional de Y dado X, para o caso discreto, pode ser calculada como:

$$p(y|x) = \sum_i p(y, \theta_i|x) = \sum_i p(y|\theta_i) \times p(\theta_i|x)$$

Para o caso contínuo, substituímos o somatório pela integral:

$$p(y|x) = \int p(y, \theta|x).d\theta = \int p(y|\theta).p(\theta|x).d\theta$$

Em ambos os casos, à direita, temos a fórmula da esperança da distribuição de Y condicionada a θ, para todos os valores possíveis de θ dado x:

$$p(y|x) = E_{\theta|x}[p(Y|\theta)]$$

A probabilidade preditiva condicionada também não é uma função do parâmetro θ, funcionando como uma **constante normalizadora**.

Assim, temos a seguinte relação de proporcionalidade:

$$p(\theta|x,y) \propto p(y|\theta) \times p(\theta|x)$$

Sabendo que $p(\theta|x) \propto p(x|\theta) \times p(\theta)$, então:

$$p(\theta|x,y) \propto p(y|\theta) \times p(x|\theta) \times p(\theta)$$

Também podemos representar essas expressões, utilizando as funções de verossimilhança:

$$p(\theta|x,y) \propto l_y(\theta;y) \times p(\theta|x)$$

$$p(\theta|x,y) \propto l_y(\theta;y) \times l_x(\theta;x) \times p(\theta)$$

Por se tratar de um produto, podemos concluir que a **ordem** dos procedimentos **não** interfere no resultado.

Generalizando a proporcionalidade para n procedimentos $x_1, \dots x_n$, temos:

$$p(\theta|x_1, \dots, x_n) \propto p(x_n|\theta) \times p(\theta|x_{n-1}, \dots, x_1)$$

Sabendo que $p(\theta|x_{n-1}, \dots, x_1) \propto p(x_{n-1}|\theta) \times \dots \times p(x_1|\theta) \times p(\theta)$, temos:

$$p(\theta|x_1, \dots, x_n) \propto [\prod_{i=1}^{n} p(x_i|\theta)].p(\theta)$$

Também podemos representar essas expressões, utilizando as funções de verossimilhança:

$$p(\theta|x_1, \dots, x_n) \propto l_n(\theta;x_n) \times p(\theta|x_{n-1}, \dots, x_1)$$

$$p(\theta|x_1, \dots, x_n) \propto [\prod_{i=1}^{n} l_i(\theta;x_i)].p(\theta)$$

Vamos voltar ao nosso exemplo do médico e supor que seja aplicado um segundo teste Y, cuja probabilidade de um falso positivo seja de 4% e a probabilidade de um falso negativo seja de 1%

Assim, a probabilidade de o teste resultar em $Y = 1$, dado que o paciente **não** está doente (falso positivo), é igual a 4%; e a probabilidade de o teste resultar em $Y = 0$, dado que o paciente **não** está doente (resultado negativo adequado) é complementar:

$$P(Y = 1|\theta = 0) = 0{,}04$$

$$P(Y = 0|\theta = 0) = 1 - 0{,}04 = 0{,}96$$

Ademais, a probabilidade de o teste resultar em $Y = 0$, dado que o paciente está **doente** (falso negativo), é igual a 1%; e a probabilidade de o teste resultar em $Y = 1$, dado que o paciente está **doente** (resultado positivo adequado) é complementar:

$$P(Y = 0|\theta = 1) = 0{,}01$$

$$P(Y = 1|\theta = 1) = 1 - 0{,}01 = 0{,}99$$

Vamos supor que o resultado tenha sido **negativo** $Y = 0$.

Com base no resultado do teste, imaginamos que a probabilidade de o paciente ter a doença tenha diminuído. Para calcular essa probabilidade, dado que o resultado do teste X foi positivo e que o resultado do teste Y foi negativo (distribuição a posteriori), utilizamos o **Teorema de Bayes**:

$$P(\theta = 1|X = 1, Y = 0) = \frac{P(Y = 0|\theta = 1) \times P(\theta = 1|X = 1)}{P(Y = 0|X = 1)}$$

A função de **verossimilhança** $p(y|\theta) = P(Y = 0|\theta = 1)$ corresponde à probabilidade de o resultado ser negativo (resultado observado), dado que o paciente está doente:

$$p(y|\theta) = P(Y = 0|\theta = 1) = 0{,}01$$

A distribuição a **priori** $p(\theta|x) = P(\theta = 1|X = 1)$ agora corresponde à probabilidade de o paciente estar doente, após o resultado do teste X, que calculamos anteriormente (como a probabilidade a posteriori para o teste X):

$$P(\theta = 1|X = 1) = 0{,}847$$

E a distribuição **preditiva** $p(y|x) = P(Y = 0|X = 1)$ corresponde à probabilidade de obter o resultado do teste Y ser negativo, sabendo que o resultado do teste X foi positivo, independentemente do valor de θ. Ela pode ser calculada pelo **Teorema da Probabilidade Total**:

$$P(Y = 0|X = 1) = P(Y = 0|\theta = 0) \times P(\theta = 0|X = 1) + P(Y = 0|\theta = 1) \times P(\theta = 1|X = 1)$$

Sabendo que a probabilidade a priori de o paciente não estar doente, dado $X = 1$, é complementar $P(\theta = 0|X = 1) = 1 - 0{,}847 = 0{,}153$, temos:

$$P(Y = 0|X = 1) = 0{,}96 \times 0{,}153 + 0{,}01 \times 0{,}847 = 0{,}14688 + 0{,}00847 = 0{,}15535$$

Assim, a probabilidade a **posteriori**, que corresponde à probabilidade de o paciente estar doente, após o resultado do teste Y, é dada por:

$$P(\theta = 1|X = 1, Y = 0) = \frac{P(Y = 0|\theta = 1) \times P(\theta = 1|X = 1)}{P(Y = 0|X = 1)} = \frac{0{,}01 \times 0{,}847}{0{,}15535} = \frac{0{,}00847}{0{,}15535} \cong 0{,}055$$

A probabilidade a posteriori foi tão inferior à probabilidade a priori porque a probabilidade do falso negativo é muito pequena.

Distribuição Normal

Vamos supor que uma única observação X (teste único) tenha distribuição normal, com **média desconhecida** θ e **variância conhecida** σ^2, que corresponde à função de **verossimilhança**:

$$X|\theta \sim N(\theta, \sigma^2)$$

E, ainda, que a distribuição a **priori** da média θ siga distribuição **normal**, com parâmetros μ_0 e τ_0^2:

$$\theta \sim N(\mu_0, \tau_0^2)$$

Nessa situação, a distribuição a **posteriori** $\theta|X$ será **normal** com parâmetros μ_1 e τ_1^2, dados por:

$$\mu_1 = \frac{\frac{\mu_0}{\tau_0^2} + \frac{x}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}, \qquad \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}$$

Para entender melhor esse resultado, vamos considerar que o **inverso da variância**, que é necessariamente positiva, corresponde à **precisão**, a qual representa o nível de informação.

Assim, a **precisão a posteriori** $p_1 = \frac{1}{\tau_1^2}$ corresponde à **soma** das precisões a priori $p_0 = \frac{1}{\tau_0^2}$ e da verossimilhança $p_v = \frac{1}{\sigma^2}$:

$$p_1 = p_0 + p_v$$

Note que ela **não** depende do resultado observado x.

Além disso, vamos chamar de w a **razão** entre a precisão a **priori** e a precisão a **posteriori** (total), que pode variar no intervalo (0, 1):

$$w = \frac{p_0}{p_1} = \frac{p_0}{p_0 + p_v}$$

Com isso, a **média** a posteriori pode ser calculada como:

$$\mu_1 = \frac{\mu_0 . p_0 + x . p_v}{p_0 + p_v} = \mu_0 . w + x . (1 - w)$$

Ou seja, a média a posteriori é uma **combinação linear** entre a média a priori e o resultado observado, assumindo um valor intermediário:

$$\min\{\mu_0, x\} \leq \mu_1 \leq \max\{\mu_0, x\}$$

Ademais, quanto **maior a precisão a priori** em relação à precisão a posteriori (total), mais **próxima da média a priori** será a média a posteriori.

Vale acrescentar que a distribuição **preditiva** de X (não condicionada) segue distribuição **normal** com média igual à média a priori e variância igual à **soma** das variâncias a priori e de verossimilhança:

$$X \sim N(\mu_0, \tau_0^2 + \sigma^2)$$

Por exemplo[2], vamos supor dois físicos A e B, que queiram estimar uma constante física desconhecida θ.

O físico A, mais experiente, acredita que a constante tenha distribuição normal com média $\mu_{0A} = 900$ e variância $\tau_{0A}^2 = 400$. Essa é a distribuição a **priori** para o físico **A**.

O físico A, menos experiente, acredita que a constante tenha distribuição normal com média $\mu_{0B} = 800$ e variância $\tau_{0B}^2 = 6400$. Essa é a distribuição a **priori** para o físico **B**. A menor experiência do físico B é traduzida pela maior variância (menor precisão) da distribuição a priori.

Agora vamos supor um aparelho para **testar** a referida constante, cujos resultados seguem distribuição normal com média θ e variância $\sigma^2 = 1600$. Essa é a **função de verossimilhança**.

Vamos considerar que o resultado obtido seja $x = 850$ e calcular a distribuição a posteriori para cada físico.

A **precisão a posteriori** é a **soma** das precisões a priori e da verossimilhança. Para o físico A, temos:

$$p_{1A} = p_{0A} + p_v = \frac{1}{\tau_{0A}^2} + \frac{1}{\sigma^2}$$

$$p_{1A} = \frac{1}{400} + \frac{1}{1600} = 0{,}0025 + 0{,}000625 = 0{,}003125$$

A **variância a posteriori** é o **inverso** desse resultado:

$$\tau_{1A}^2 = \frac{1}{p_{1A}} = \frac{1}{0{,}003125} = 320$$

Para calcular a média a posteriori, vamos calcular a proporção w entre a precisão a **priori** e a precisão **total**:

$$w_A = \frac{p_{0A}}{p_{1A}} = \frac{0{,}0025}{0{,}003125} = 0{,}8$$

Ou seja, a precisão a priori corresponde a 80% da precisão a posteriori.

Por fim, a **média a posteriori** para o físico A é:

$$\mu_{1A} = \mu_{0A}.w_A + x.(1 - w_A)$$

$$\mu_{1A} = 900 \times 0{,}8 + 850 \times 0{,}2 = 720 + 170 = 890$$

Observamos que a média a posteriori está **entre** a média a priori e o resultado (como vimos antes). Como a precisão a priori corresponde a uma parcela **significativa** da precisão total, a média a posteriori é bem **próxima** da média a **priori**.

Também podemos calcular o **aumento** da precisão, em relação à precisão a priori:

$$\frac{p_{1A} - p_{0A}}{p_{0A}} = \frac{1}{w_A} - 1 = \frac{1}{0{,}8} - 1 = 1{,}25 - 1 = 0{,}25$$

Ou seja, a precisão aumentou em 25% para o físico A.

[2] Box & Tiao (1992)

Agora, vamos fazer o mesmo para o físico B. A **precisão a posteriori** é dada por:

$$p_{1B} = p_{0B} + p_v = \frac{1}{\tau_{0B}^2} + \frac{1}{\sigma^2}$$

$$p_{1A} = \frac{1}{6400} + \frac{1}{1600} = 0{,}00015625 + 0{,}000625 = 0{,}00078125$$

A **variância a posteriori** é o **inverso** desse resultado:

$$\tau_{1B}^2 = \frac{1}{p_{1B}} = \frac{1}{0{,}00078125} = 1280$$

Agora, vamos calcular a **proporção** w entre a precisão a priori e a precisão total:

$$w_B = \frac{p_{0B}}{p_{1B}} = \frac{0{,}00015625}{0{,}00078125} = 0{,}2$$

Ou seja, a precisão a priori corresponde a 20% da precisão a posteriori.

Por fim, a **média a posteriori** para o físico B é:

$$\mu_{1B} = \mu_{0B}.w_B + x.(1 - w_B)$$

$$\mu_{1B} = 800 \times 0{,}2 + 850 \times 0{,}8 = 160 + 680 = 840$$

Novamente, a média a posteriori está **entre** a média a priori e o resultado, mas, para o físico B, a nova média está muito mais **próxima** do **resultado** do que da média a posteriori, uma vez que a precisão a priori é bem **pequena** em relação ao total.

E o **aumento** da precisão foi na proporção de:

$$\frac{p_{1B} - p_{0B}}{p_{0B}} = \frac{1}{w_B} - 1 = \frac{1}{0{,}2} - 1 = 5 - 1 = 4$$

Ou seja, a precisão aumentou em 400%.

Embora esse aumento seja muito maior do que o aumento na precisão para o físico A (de 25%), a precisão a posteriori do físico B $\left(\frac{1}{1280} = 0{,}00078125\right)$ continua sendo bem inferior à do físico A $\left(\frac{1}{320} = 0{,}003125\right)$.

Distribuições Conjugadas

Quando a **conjugação** entre distribuições está presente, as distribuições a **priori** e a **posteriori** pertencem à **mesma classe** de distribuições.

Nessa situação, o aprendizado com as observações (ou testes) se resume em **atualizar** os parâmetros da distribuição à priori.

Isso ocorreu no exemplo que acabamos de ver - as distribuições tanto a priori quanto a posteriori são normais; e o aprendizado se resumo em atualizar os parâmetros da distribuição.

Para não confundir com o parâmetro desconhecido que desejamos estimar, chamamos os parâmetros da distribuição a priori (que são atualizados) de **hiperparâmetros**.

A **definição** de distribuições conjugadas é a seguinte:

> *Sendo* $F = \{p(x|\theta)\}$ *uma classe de distribuições* ***amostrais****, então uma classe de distribuições* P *é* ***conjugada*** *a* F *se:*
>
> $$\forall p(x|\theta) \in F \; e \; p(\theta) \in P \rightarrow p(\theta|x) \in P$$

Ou seja, as distribuições a **priori** $p(\theta)$ e a **posteriori** $p(\theta|x)$ pertencerão à **mesma classe** P, se a função de **verossimilhança** $p(x|\theta)$ pertencer a uma classe F **conjugada** a P.

Distribuição Beta

Vamos supor que uma população siga distribuição de **Bernoulli** com parâmetro $0 < \theta < 1$ e que seja extraída uma amostra $X_1, X_2, \dots, X_n$ de tamanho n.

Nessa situação, cada elemento da amostra assume o valor 1 com probabilidade θ e o valor 0 com probabilidade $(1 - \theta)$.

Assim, a distribuição de probabilidade de **cada elemento** da amostra pode ser representada como:

$$p(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}, \qquad x_i = 0, 1$$

E a função de **verossimilhança** $p(x|\theta)$, que corresponde à distribuição **conjunta** dos resultados da amostra, condicionada ao parâmetro θ, é dada por:

$$p(x|\theta) = p(x_1|\theta) \times p(x_2|\theta) \times \dots \times p(x_n|\theta)$$

$$p(x|\theta) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Fazendo $t = \sum_{i=1}^{n} x_i$, podemos representar a função de **verossimilhança** como:

$$p(x|\theta) = \theta^{t}(1-\theta)^{n-t}$$

Essa distribuição é proporcional a uma distribuição **Beta**.

A distribuição **Beta** depende dos parâmetros $\alpha > 0$ e $\beta > 0$ e a sua f.d.p. é dada por:

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha).\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1$$

Em que Γ é a função **gama**. Para valores inteiros e positivos, temos:

$$\Gamma(n+1) = n!$$

Na prática, a razão $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha).\Gamma(\beta)}$ funciona como uma **constante de normalização**.

A **média** da distribuição Beta é:

$$E(X) = \frac{\alpha}{\alpha+\beta}$$

Pontue-se que a distribuição **uniforme** no intervalo $(0,1)$ é um caso **particular** da distribuição Beta com parâmetros $\boldsymbol{\alpha = 1}$ e $\boldsymbol{\beta = 1}$.

Essa distribuição uniforme pode ser utilizada como distribuição a priori quando não há informações a respeito da proporção desconhecida, chamada **priori não informativa**.

Comparando os expoentes de θ e $(1-\theta)$ da função de verossimilhança e os expoentes de x e $(1-x)$ da distribuição Beta, podemos observar que:

$$t = \alpha - 1 \rightarrow \alpha = 1 + t$$

$$n - t = \beta - 1 \rightarrow \beta = 1 + n - t$$

Portanto, a função de verossimilhança de uma população de Bernoulli segue uma distribuição **Beta** com parâmetros $\alpha = 1 + t$ e $\beta = 1 + n - t$.

Pelo Teorema de Bayes, a distribuição a **posteriori** é proporcional ao produto da função de verossimilhança pela distribuição a priori:

$$p(\theta|x) \propto p(x|\theta) \times p(\theta) = \theta^t(1-\theta)^{n-t} \times p(\theta)$$

Vamos supor que a distribuição a **priori** de θ seja **Beta** com parâmetros α_0 e β_0. Podemos indicar a distribuição a **posteriori** como:

$$p(\theta|x) \propto p(x|\theta) \times p(\theta) \propto \theta^t(1-\theta)^{n-t} \times \theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}$$

$$p(\theta|x) \propto \theta^{\alpha_0+t-1}(1-\theta)^{\beta_0+n-t-1}$$

Comparando os expoentes de θ e $(1-\theta)$ dessa função e os expoentes de x e $(1-x)$ da distribuição Beta, podemos observar que:

$$\alpha_0 + t - 1 = \alpha - 1 \rightarrow \boldsymbol{\alpha = \alpha_0 + t}$$

$$\beta_0 + n - t - 1 = \beta - 1 \rightarrow \boldsymbol{\beta = \beta_0 + n - t}$$

Ou seja, se a função de **verossimilhança** seguir distribuição de **Bernoulli** e a distribuição a **priori** seguir distribuição **Beta** com parâmetros α_0 e β_0, a distribuição a **posteriori** seguirá distribuição **Beta** com parâmetros $\boldsymbol{\alpha_0 + t}$ e $\boldsymbol{\beta_0 + n - t}$, em que $t = \sum_{i=1}^{n} x_i$.

Assim, concluímos que a distribuição **Beta** é **conjugada** à distribuição de **Bernoulli**.

Embora tenhamos verificado para as distribuições Beta com parâmetros inteiros, podemos ampliar para todas as distribuições Beta (com parâmetros positivos).

Por exemplo, vamos supor que, de uma amostra de tamanho $n = 5$, extraída de uma população de **Bernoulli** com parâmetro θ desconhecido, foram observadas $t = 3$ sucessos.

Vamos considerar que a distribuição a **priori** para o parâmetro desconhecido é uniforme no intervalo (0, 1), o que corresponde a uma distribuição **Beta** com parâmetros $\alpha_0 = 1$ e $\beta_0 = 1$.

Nessa situação, a distribuição a **posteriori** será **Beta** com os seguintes parâmetros:

$$\alpha = \alpha_0 + t = 1 + 3 = 4$$

$$\beta = \beta_0 + n - t = 1 + 5 - 3 = 3$$

Agora, vamos verificar o que ocorre quando a função de **verossimilhança** segue distribuição **binomial** com parâmetros n (conhecido) e θ (desconhecido), dada por:

$$p(x|\theta) = C_{n,x}.\theta^x(1-\theta)^{n-x}$$

Que é proporcional a uma distribuição **Beta** com parâmetros $\alpha = 1 + x$ e $\beta = 1 + n - x$.

Pelo **Teorema de Bayes**, a distribuição a **posteriori** é proporcional a:

$$p(\theta|x) \propto p(x|\theta) \times p(\theta) = C_{n,x}.\theta^x(1-\theta)^{n-x} \times p(\theta)$$

Considerando que $C_{n,x}$ é uma constante nessa situação, temos:

$$p(\theta|x) \propto \theta^x(1-\theta)^{n-x} \times p(\theta)$$

Sendo a distribuição a **priori** uma distribuição **Beta** com parâmetros α_0 e β_0, temos:

$$p(\theta|x) \propto \theta^x(1-\theta)^{n-x} \times \theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}$$

$$p(\theta|x) \propto \theta^{\alpha_0+x-1}(1-\theta)^{\beta_0+n-x-1}$$

Ou seja, a distribuição a **posteriori** segue distribuição **Beta** com parâmetros $\boldsymbol{\alpha = \alpha_0 + x}$ e $\boldsymbol{\beta = \beta_0 + n - x}$.

Assim, a distribuição **Beta** também é **conjugada** à distribuição **binomial**.

Por exemplo, vamos supor a distribuição Beta a priori com parâmetros $\alpha_0 = 0{,}5$ e $\beta_0 = 0{,}5$. Assim, supondo uma amostra de tamanho $n = 10$, com $x = 2$ observações favoráveis, a distribuição a posteriori terá os seguintes parâmetros:

$$\alpha = 0{,}5 + 2 = 2{,}5$$

$$\beta = 0{,}5 + 10 - 2 = 8{,}5$$

A distribuição **Beta** com parâmetros $\boldsymbol{\alpha_0 = \frac{1}{2}}$ e $\boldsymbol{\beta_0 = \frac{1}{2}}$ também é uma distribuição a **priori não informativa**, utilizada quando a informação a priori é vaga. Ela resulta da aplicação do **Método de Jeffreys** para o modelo binomial (incluindo, a distribuilção de Bernoulli).

De maneira geral, a distribuição a priori não informativa obtida por esse método é proporcional à **raiz quadrada** da **informação esperada de Fisher**:

$$p_J(\theta) \propto \sqrt{I(\theta)}$$

Para distribuições da família exponencial, a informação esperada de Fisher pode ser calculada como:

$$I(\theta) = -E_X\left(\frac{\partial^2}{\partial\theta^2}\ln p(x|\theta)\right)$$

Sempre que a distribuição a posteriori possuir normalidade assintótica, como é o caso das funções que estudamos aqui, a distribuição a priori obtida pelo **Método de Jeffreys** corresponde à distribuição a priori não informativa de **referência**, qual seja, aquela que **maximiza** a falta de informação.

A distribuição Beta também é **conjugada** às distribuições amostrais **geométrica** e **binomial negativa**.

(VUNESP/2021 – EsFCEx) Na abordagem bayesiana, com base no conhecimento que se tem sobre um parâmetro θ, pode-se definir uma família paramétrica de densidades. Nesse caso, a distribuição a priori é representada por uma forma funcional, cujos parâmetros devem ser especificados de acordo com esse conhecimento. Essa abordagem, em geral, facilita a análise e o caso mais importante é o de prioris conjugadas. A ideia é que as distribuições a priori e a posteriori pertençam à mesma classe de distribuições e, assim, a atualização do conhecimento que se tem do parâmetro θ envolve apenas uma mudança nos hiperparâmetros.

Nesse caso, assinale a alternativa em que é correto afirmar que a priori é conjugada.

a) Distribuição a priori Beta de parâmetros inteiros é conjugada à família Bernoulli.

b) Distribuição a priori Uniforme é conjugada à família Normal.

c) Distribuição a priori Poisson é conjugada à família Normal.

d) Distribuição a priori Normal é conjugada à família Bernoulli.

e) Distribuição a priori Beta de parâmetros inteiros é conjugada à família Gama.

Comentários:

A distribuição **Beta** é conjugada à distribuição de **Bernoulli**, ou seja, quando a população segue distribuição de Bernoulli e a distribuição a priori pertence à família Beta, então a distribuição a posteriori também pertencerá à família Beta.

A distribuição Beta também é conjugada à distribuição binomial, geométrica e binomial negativa.

Gabarito: A

(FGV/2021 – FunSaúde/CE) Suponha que $X_1, X_2, \dots, X_n$ seja uma amostra aleatória de uma distribuição Bernoulli (θ), θ desconhecido.

Se usarmos uma distribuição a priori Beta $(\alpha = 1, \beta = 1)$ para θ, se $n = 10$ e se as observações são 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, então a distribuição a posteriori de θ, dada a amostra observada tem distribuição Beta com parâmetros iguais a

a) 6 e 4

b) 7 e 5

c) 6 e 10

d) 8 e 6

e) 9 e 5

Comentários:

Quando a população segue distribuição de **Bernoulli** e a distribuição a priori segue distribuição **Beta** com parâmetros α e β, então a distribuição a posteriori será **Beta**, com parâmetros $\boldsymbol{\alpha + t}$ e $\boldsymbol{\beta + n - t}$, em que $t = \sum_{i=1}^{n} x_i$.

O enunciado informa que $\alpha = 1$, $\beta = 1$ e $n = 10$. A partir da amostra fornecida, observamos que $t = 6$. Assim, os parâmetros da distribuição a posteriori são:

$$\alpha + t = 1 + 6 = 7$$

$$\beta + n - t = 1 + 10 - 6 = 5$$

Gabarito: B

(2022 – PROGEP-FURG) Seja $\pi(\theta) \sim Beta(\alpha, \beta), \alpha, \beta > 0$ uma distribuição a priori e $f(x|\theta) = Bin(n, \theta)$ uma função de verossimilhança. Portanto, visto que as distribuições Beta e Binomial pertencem à mesma família, ou seja, são conjugadas, a distribuição posterior que se procura é dada por

a) $\pi(x|\theta) \sim Beta(x + \alpha, n - x + \beta)$

b) $\pi(x|\theta) \sim Beta(x + \alpha - 1, n - x + \beta)$

c) $\pi(x|\theta) \sim Bin(x + \alpha, n - x + \beta)$

d) $\pi(\theta|x) \sim Bin(x, n - x)$

e) $\pi(\theta|x) \sim Beta(x + \alpha, n - x + \beta)$

Comentários:

A distribuição **Beta** é **conjugada** à distribuição **binomial**, ou seja, quando a função de verossimilhança segue distribuição binomial e a distribuição a priori pertence à família Beta, então a distribuição a posteriori também pertencerá à família Beta.

Sendo α e β os parâmetros da distribuição a priori, então os parâmetros da distribuição a posteriori serão $\boldsymbol{\alpha + x}$ e $\boldsymbol{\beta + n - x}$.

Acrescenta-se que a distribuição a **posteriori**, que representa a distribuição do parâmetro desconhecido θ após a observação da amostra x é indicada como $\pi(\theta|x)$, e não como $\pi(x|\theta)$.

Gabarito: E

(2019 – EsFCEx) Seja X urna variável aleatória com distribuição de Bernoulli de parâmetro $0 < \theta < 1$, a priori de Jeffreys para este modelo é dada por:

a) $Beta\left(\frac{1}{4}, \frac{1}{4}\right)$

b) $Beta\left(\frac{1}{2}, \frac{1}{2}\right)$

c) $Beta\left(1,\frac{1}{2}\right)$

d) $Beta\left(\frac{1}{2},1\right)$

e) $Beta\left(\frac{1}{4},1\right)$

Comentários:

A distribuição a priori de **Jeffreys** para o modelo binomial (incluindo a distribuição de Bernoulli) para o parâmetro θ, que corresponde à distribuição a priori **não informativa** que maximiza a falta de informação, é a distribuição:

$$Beta\left(\frac{1}{2},\frac{1}{2}\right)$$

Gabarito: B

A distribuição **Dirichlet** é uma generalização da distribuição Beta para $k = 2$. A f.d.p. da distribuição Dirichlet é dada por:

$$p(x_1,\dots,x_k|\alpha_1,\dots\alpha_k) = \frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}x_1^{\alpha_1-1}\dots x_k^{\alpha_k-1}, \quad \sum_{i=1}^{k}x_i = 1, \ \ \alpha_1,\dots\alpha_k > 0$$

Essa distribuição é **conjugada** à distribuição **multinomial**, que é uma generalização da distribuição binomial para $k = 2$. A f.d.p. da distribuição multinomial é dada por:

$$p(x_1,x_2,\dots,x_k|\theta) = \frac{n!}{\prod_{i=1}^{k}x_i!}.\prod_{i=1}^{k}\theta_i^{x_i}$$

Distribuição Normal

Anteriormente, vimos que, para uma **única** observação, a família de distribuições normais é **conjugada** à distribuição normal, com média desconhecida e variância conhecida.

Sendo μ_0 e τ_0^2 os parâmetros da distribuição a **priori**; e σ^2 a variância da função de **verossimilhança**, então os parâmetros da distribuição a **posteriori** μ_1 e τ_1^2, após uma única observação, são dados por:

$$\mu_1 = \frac{\frac{\mu_0}{\tau_0^2}+\frac{x}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}, \qquad \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2}+\frac{1}{\sigma^2}$$

Para uma amostra de **tamanho n**, substituímos x por $\bar{x}$ (média amostral) e σ^2 por $\frac{\sigma^2}{n}$ (variância da média amostral), de modo que os parâmetros da distribuição a **posteriori** são dados por:

$$\mu_1 = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n.\bar{x}}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}, \qquad \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$$

Assim, como vimos anteriormente, o **inverso da variância** corresponde à **precisão**, que representa o nível de informação. Assim, a **precisão a posteriori** $p_1 = \frac{1}{\tau_1^2}$ corresponde à **soma** das precisões a priori $p_0 = \frac{1}{\tau_0^2}$ e da verossimilhança $p_v = \frac{n}{\sigma^2}$:

$$p_1 = p_0 + p_v$$

Também podemos definir a **razão** entre a **precisão** a priori $p_0 = \frac{1}{\tau_0^2}$ e a **precisão** a posteriori $p_1 = \frac{1}{\tau_1^2}$:

$$w = \frac{p_0}{p_1} = \frac{p_0}{p_0 + p_v}$$

De modo que a média a posteriori pode ser reescrita como:

$$\mu_1 = \frac{\mu_0.p_0 + \bar{x}.p_v}{p_0 + p_v} = w.\mu_0 + (1 - w).\bar{x}$$

Que corresponde a uma **combinação linear** entre a média a priori e o resultado observado, assumindo um valor intermediário:

$$\min\{\mu_0, \bar{x}\} \leq \mu_1 \leq \max\{\mu_0, \bar{x}\}$$

Por exemplo, vamos considerar o exemplo anterior do físico A, cuja distribuição a priori para a constante física desconhecida é normal com média $\mu_0 = 900$ e variância $\tau_0^2 = 400$; com o mesmo aparelho de teste, cujos resultados seguem distribuição normal com média θ e variância $\sigma^2 = 1600$.

Agora, vamos supor que, em uma amostra de tamanho $n = 100$, tenha sido verificada uma média $\bar{x} = 870$.

A **precisão a posteriori** é a **soma** das precisões a priori e da verossimilhança:

$$p_1 = p_0 + p_v = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} = \frac{1}{400} + \frac{100}{1600} = 0{,}0025 + 0{,}0625 = 0{,}065$$

A **variância a posteriori** é o **inverso** desse resultado:

$$\tau_1^2 = \frac{1}{p_1} = \frac{1}{0{,}065} \cong 15{,}4$$

A **proporção** w entre a precisão a **priori** e a precisão **total é**:

$$w = \frac{p_0}{p_1} = \frac{0{,}0025}{0{,}065} \cong 0{,}04$$

E a **média a posteriori** é:

$$\mu_1 = \mu_0.w + x.(1-w) = 900 \times 0{,}04 + 870 \times 0{,}96 = 36 + 835{,}2 = 871{,}2$$

Que está **entre** a média a priori e a média amostral, mas muito mais próximo da **média amostral**, porque a sua precisão $p_v = \frac{n}{\sigma^2} = 0{,}0625$ é muito superior à precisão do físico $p_0 = \frac{1}{\tau_0^2} = 0{,}0025$ e assim a razão entre a precisão a priori e a precisão total é muito pequena $w = \frac{p_0}{p_1} \cong 0{,}04$.

A **priori não informativa** para esse caso pode ser obtida, fazendo a **variância** da distribuição a **priori** tender ao **infinito** $\tau_0^2 \to \infty$.

Nessa situação, os parâmetros da **distribuição a posteriori** serão:

$$\mu_1 = \frac{0+\frac{n.\bar{x}}{\sigma^2}}{0+\frac{n}{\sigma^2}} = \bar{x}, \qquad \frac{1}{\tau_1^2} = 0 + \frac{n}{\sigma^2}$$

Ou seja, os parâmetros da distribuição a posteriori **coincidem** com os parâmetros da inferência **clássica**.

(Cebraspe/2014 – ANATEL - *Adaptada*) Considere uma amostra aleatória simples X_1, X_2, ..., X_n retirada de uma distribuição normal apresenta média μ e desvio padrão 1. Com base nessas hipóteses, julgue o item seguinte.

A distribuição a priori conjugada da média μ é normal.

Comentários:

O enunciado informa que a amostra é extraída de uma população **normal** com média desconhecida e **variância conhecida** (no caso, o desvio padrão foi informado). A distribuição conjugada a essa distribuição amostral é a distribuição **normal**.

Isso significa que se a função de verossimilhança é normal com variância conhecida; e se a distribuição a priori é normal, então a distribuição a posteriori também é normal.

Resposta: Certo.

Distribuição Gama

Agora, vamos supor que uma população siga distribuição de **Poisson** com parâmetro θ e que seja extraída uma amostra $X_1, X_2, \dots, X_n$ de tamanho n.

Nessa situação, a distribuição de probabilidade de **cada elemento** da amostra pode ser representada como:

$$p(x_i|\theta) = \frac{e^{-\theta}.\theta^{x_i}}{x_i!}$$

E a função de **verossimilhança** $p(x|\theta)$, que corresponde à distribuição **conjunta** dos resultados da amostra, condicionada ao parâmetro θ, é dada por:

$$p(x|\theta) = p(x_1|\theta) \times p(x_2|\theta) \times \dots \times p(x_n|\theta) = \frac{e^{-n.\theta}.\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

Fazendo $t = \sum_{i=1}^{n} x_i$ e considerando que $\prod_{i=1}^{n} x_i!$ é uma constante nessa situação, podemos representar a função de verossimilhança como:

$$p(x|\theta) \propto e^{-n.\theta}.\theta^{t}$$

Essa distribuição é proporcional a uma distribuição **Gama**.

A distribuição **Gama** depende dos parâmetros $\alpha > 0$ e $\beta > 0$ e a sua f.d.p. é dada por:

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \ 0 < x < 1$$

Sendo $\frac{\beta^{\alpha}}{\Gamma(\alpha)}$ uma constante de **normalização**. A **média** e a **variância** dessa distribuição são:

$$E(X) = \frac{\alpha}{\beta}, \quad V(X) = \frac{\alpha}{\beta^2}$$

Pontue-se que a distribuição **exponencial** é um caso particular da distribuição Gama com $\alpha = 1$; e a distribuição de **Erlang** (hipoexponencial), formada pela **soma** de n variáveis exponenciais independentes, também é um caso particular com $\alpha = n$.

Ademais, a distribuição **qui-quadrado** com k graus de liberdade também é um caso particular da distribuição Gama, com $\alpha = \frac{k}{2}$ e $\beta = \frac{1}{2}$.

Pelo Teorema de Bayes, a distribuição a **posteriori** é proporcional ao produto da função de verossimilhança pela distribuição a priori:

$$p(\theta|x) \propto p(x|\theta) \times p(\theta) = e^{-n.\theta}.\theta^t \times p(\theta)$$

Supondo que a distribuição a **priori** de θ seja **Gama** com parâmetros α_0 e β_0, podemos indicar a distribuição a posteriori como:

$$p(\theta|x) \propto p(x|\theta) \times p(\theta) \propto e^{-n.\theta}.\theta^t \times \theta^{\alpha_0-1}e^{-\beta_0\theta}$$

$$p(\theta|x) \propto \theta^{\alpha_0+t-1}e^{-(\beta_0+n)\theta}$$

Comparando os expoentes dessa função com os da distribuição **Gama**, podemos observar que:

$$\alpha - 1 = \alpha_0 + t - 1 \rightarrow \boldsymbol{\alpha = \alpha_0 + t}$$

$$\boldsymbol{\beta = \beta_0 + n}$$

Ou seja, se a função de **verossimilhança** seguir distribuição de **Poisson** e a distribuição a **priori** seguir distribuição **Gama** com parâmetros α_0 e β_0, a distribuição a **posteriori** seguirá distribuição **Gama** com parâmetros $\alpha_0 + t$ e $\beta_0 + n$, em que $t = \sum_{i=1}^{n} x_i$.

Assim, concluímos que a distribuição **Gama** é **conjugada** à distribuição de **Poisson**.

Acrescenta-se que a **média** da distribuição a **posteriori** $\frac{\alpha_0+t}{\beta_0+n}$ é uma **combinação linear** entre a **média a priori** $\frac{\alpha_0}{\beta_0}$ e a **média amostral** $\frac{t}{n}$, assumindo um valor intermediário entre esses dois extremos.

Por exemplo, vamos supor que, de uma amostra de tamanho $n = 5$, extraída de uma população de **Poisson** com média λ_P desconhecida, foram observadas $t = 3$ ocorrências em $n = 5$ horas.

Vamos considerar que a distribuição a **priori** desse parâmetro seja exponencial com parâmetro $\lambda_E = 2$, o que corresponde a uma distribuição **Gama** com parâmetros $\alpha_0 = 1$ e $\beta_0 = 2$.

Nessa situação, a distribuição a **posteriori** será **Gama** com os seguintes parâmetros:

$$\alpha = \alpha_0 + t = 1 + 3 = 4$$

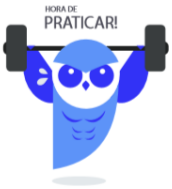
$$\beta = \beta_0 + n = 2 + 5 = 7$$

A média dessa distribuição é dada por:

$$E(X_{Pos}) = \frac{\alpha_0 + t}{\beta_0 + n} = \frac{4}{7} \cong 0{,}57$$

Que é **intermediário** entre a média a priori $\frac{\alpha_0}{\beta_0} = \frac{1}{2} = 0{,}5$ e a média amostral $\frac{t}{n} = \frac{3}{5} = 0{,}6$.

A distribuição **Gama** também é **conjugada** à distribuição normal com média conhecida e **variância desconhecida**; bem como à distribuição normal com **média e variância desconhecidas**.



(2019 – EsFCEx) Em Estimação Bayesiana, prioris conjugadas são utilizadas no sentido que as distribuições a priori e a posteriori pertençam a mesma classe de distribuições. Assinale a alternativa que NÃO É um exemplo de família conjugada natural as distribuições:

a) Bernoulli e Beta.

b) Poisson e Gama.

c) Multinomial e Dirichlet.

d) Binomial com n conhecido e normal com σ conhecido.

e) Binomial com n conhecido e Beta.

Comentários:

Precisamos identificar a alternativa que **não** indica distribuições **conjugadas**.

A distribuição **Beta** é conjugada tanto à distribuição de **Bernoulli** e quanto à distribuição **binomial** (com n conhecido), logo as alternativas A e E estão corretas.

A distribuição **Dirichlet** (generalização da distribuição Beta) é conjugada à distribuição **multinomial** (generalização da distribuição binomial), logo a alternativa C está correta.

A distribuição **Gama** é conjugada à distribuição de **Poisson**, logo a alternativa B está correta.

No entanto, as distribuições **binomial** e **normal** **não** são conjugadas. A distribuição normal com variância conhecida é conjugada à própria distribuição normal.

Gabarito: D

(Cebraspe/2016 – FUNPRESP) A amostra aleatória simples $X_1, X_2, \dots, X_n$ foi retirada de uma distribuição de Poisson, em que a média é *M* e a variância é *V* e a média amostral é $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$. Com relação a essa amostra, julgue o item a seguir.

Em inferência bayesiana, a distribuição *a priori* conjugada para o parâmetro *M* segue a distribuição normal, e a distribuição preditiva *a posteriori* segue a distribuição binomial.

Comentários:

A distribuição conjugada à distribuição de **Poisson** é a distribuição **Gama**.

Isso significa que quando a amostra é extraída de uma distribuição de Poisson e a distribuição a priori segue distribuição Gama, a distribuição a posteriori também segue distribuição Gama.

Gabarito: Errado

(FGV/2014 – SUSAM) Suponha que $X_1, X_2, \dots, X_n$ seja uma amostra aleatória de uma distribuição Poisson com parâmetro λ desconhecido ($\lambda > 0$) e que a distribuição a priori de λ seja uma distribuição gama (α, β).

Assinale a opção que indica a distribuição a posteriori de λ, dado que $X_i = x_i, i = 1, \dots, n$

a) $gama(\alpha + \sum_{i=1}^{n} x_i\,, n\beta)$

b) $gama(\alpha + \sum_{i=1}^{n} x_i\,, \beta + n)$

c) $gama(\alpha + \sum_{i=1}^{n} x_i\,, \beta)$

d) $normal(\alpha + (\sum_{i=1}^{n} x_i)^2)$

e) $Beta(\alpha - \sum_{i=1}^{n} x_i\,, \beta/n\,)$

Comentários:

Quando a população segue distribuição de **Poisson** com parâmetro λ e a distribuição a priori segue distribuição **Gama** com parâmetros α e β, então a distribuição a posteriori seguirá distribuição **Gama** com os parâmetros α' e β' dados por:

$$\alpha' = \alpha + \sum_{i=1}^{n} x_i$$

$$\beta' = \beta + n$$

Gabarito: B

Estimadores de Bayes

Um **bom** estimador $\delta(x)$, calculado a partir dos dados da amostra, para o parâmetro θ, é aquele cujo **erro** $\delta(x) - \theta$ se aproxima de **zero**, com alta probabilidade.

Assim, para cada estimativa a, podemos associar uma **perda** $L(a, \theta)$, de modo que quanto maior a diferença entre a e θ, maior a perda.

A **perda esperada a posteriori** é dada por:

$$E[L(a, \theta|x) = \int L(a, \theta) p(\theta|x) dx$$

O estimador de Bayes é aquele que **minimiza** a perda esperada.

Agora, veremos funções **simétricas** de perda, em que o peso de uma superestimação é igual ao de uma subestimação, ou seja, a gravidade desses erros é a mesma.

A função de perda mais utilizada é a função de **perda quadrática**:

$$L(a, \theta) = (a - \theta)^2$$

Considerando essa função de perda, o estimador de Bayes para o parâmetro θ corresponde à **média** da distribuição a **posteriori**.

Por exemplo, vamos supor uma amostra aleatória extraída de uma população de **Bernoulli** com parâmetro θ desconhecido. Vamos considerar a distribuição a priori conjugada, qual seja **Beta** com parâmetros α e β.

Nessa situação, a distribuição a posteriori será **Beta** com parâmetros $\alpha + t$ e $\beta + n - t$, em que $t = \sum_{i=1}^{n} x_i$.

Pela função de **perda quadrática**, o estimador de Bayes para θ será a **média** da distribuição a posteriori:

$$E(X) = \frac{\alpha'}{\alpha' + \beta'} = \frac{\alpha + t}{\alpha + t + \beta + n - t} = \frac{\alpha + t}{\alpha + \beta + n}$$

Vamos considerar o exemplo da amostra de tamanho $n = 5$, extraída de uma população de Bernoulli com parâmetro θ desconhecido, em que foram observadas $t = 3$ sucessos; e que a distribuição a priori é uniforme no intervalo (0, 1), a qual corresponde a uma distribuição Beta com parâmetros $\alpha = 1$ e $\beta = 1$.

A **média** da distribuição a **posteriori** é dada por:

$$E(X) = \frac{\alpha + t}{\alpha + \beta + n} = \frac{1 + 3}{1 + 1 + 5} = \frac{4}{7}$$

Essa é a **estimativa de Bayes**, considerando a função de perda quadrática, para a proporção desconhecida θ de sucesso da população de Bernoulli.

Agora, vamos supor uma amostra extraída de uma população que segue distribuição de **Poisson** com parâmetro θ desconhecido. Vamos considerar a distribuição a priori conjugada, qual seja **Gama** com parâmetros α e β.

Nessa situação, a distribuição a posteriori seguirá distribuição Gama com parâmetros $\alpha + t$ e $\beta + n$, em que $t = \sum_{i=1}^{n} x_i$.

Pela função de **perda quadrática**, o estimador de Bayes para θ será a **média** da distribuição a posteriori:

$$E(X) = \frac{\alpha'}{\beta'} = \frac{\alpha + t}{\beta + n}$$

Vamos considerar o exemplo da amostra de tamanho $n = 5$, extraída de uma população de **Poisson** com média λ_P desconhecida, em que foram observadas $t = 3$; e a distribuição a **priori** exponencial com parâmetro $\lambda_E = 2$, o que corresponde a uma distribuição **Gama** com parâmetros $\alpha = 1$ e $\beta = 2$.

Calculamos a **média** dessa distribuição, dada por:

$$E(X_{Pos}) = \frac{\alpha + t}{\beta + n} = \frac{4}{7} \cong 0{,}57$$

Essa é a **estimativa de Bayes**, considerando a função de perda quadrática, para a média desconhecida λ_P da população de Poisson.

Entende-se que a função da perda quadrática penaliza excessivamente o erro de estimação. Uma função de perda cuja penalidade cresce linearmente com o erro de estimação é a função de **perda absoluta**:

$$L(a, \theta) = |a - \theta|$$

Considerando essa função de perda, o estimador de Bayes para o parâmetro θ corresponde à **mediana** da distribuição a **posteriori**.

Outra função de perda que penaliza ainda menos os erros de estimação é a chamada função de **perda 0-1**, que associa uma perda fixa ao erro, independentemente de sua magnitude, atribuindo uma perda de 1 para valores incorretos e 0 para valores corretos:

$$L(a, \theta) = \begin{cases} 1, & se\ |a - \theta| > \epsilon \\ 0, & se\ |a - \theta| < \epsilon \end{cases}$$

Para todo $\epsilon > 0$.

Considerando essa função de perda, o estimador de Bayes para o parâmetro θ corresponde à **moda** da distribuição a **posteriori**, o qual é denominado Estimador de Máxima Verossimilhança Generalizado (EMVG).

No caso contínuo, a moda da distribuição a posteriori corresponde ao valor do parâmetro para o qual a derivada da função é nula:

$$\frac{d[p(\theta|x)]}{dx} = 0$$

Sabendo que $p(\theta|x) \propto p(x|\theta) \times p(\theta)$, podemos calcular a moda como:

$$\frac{d[p(x|\theta) \times p(\theta)]}{dx} = 0$$

Assim, para esse estimador, **não** é necessário conhecer a distribuição a posteriori exata.

Por exemplo, vamos supor uma amostra aleatória de tamanho n extraída de uma população com distribuição **normal** com média θ desconhecida e variância conhecida σ^2; e a distribuição a priori conjugada, qual seja **normal** com média μ_0 e variância σ_0^2.

Nessa situação, a distribuição a posteriori será **normal**, em que a média, a mediana e a moda **coincidem**.

Assim, o estimador de Bayes para θ, considerando qualquer uma das três funções de perda que vimos, é a **média** da distribuição a posteriori, dada por:

$$\mu_1 = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n.\bar{x}}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$

No exemplo do físico A, em que a média da amostra de tamanho $n = 100$ foi $\bar{x} = 870$, calculamos que a média da distribuição a posteriori foi $\mu_1 = 871{,}2$. Esse é a estimativa de **Bayes**, considerando qualquer função de perda que vimos.

(FGV/2022 – TRT/PB) Deseja-se estimar a proporção θ de itens defeituosos numa grande produção de itens. Suponha que não se tenha informação prévia sobre o valor de θ, de modo que uma densidade a priori Uniforme no intervalo (0,1) seja usada para θ.

Suponha ainda que uma amostra aleatória simples de tamanho 15 seja obtida (note: são 15 ensaios Bernoulli(θ)) e que 2 itens defeituosos e 13 não defeituosos sejam constatados. Se a função de perda de erro quadrático for usada, a estimativa de Bayes a posteriori para q é igual a

a) 2/15.

b) 3/17.

c) 4/15.

d) 5/17.

e) 2/13

Comentários:

Utilizando a função de perda quadrática $L(a, \theta) = (a - \theta)^2$, o estimador de Bayes para o parâmetro θ desconhecido corresponde à média da distribuição a posteriori.

Quando a amostra segue distribuição de Bernoulli e a distribuição a priori é Beta com parâmetros α e β, a distribuição a posteriori será Beta com parâmetros $\alpha + t$ e $\beta + n - t$, em que $t = \sum_{i=1}^{n} x_i$. A média dessa distribuição é dada por:

$$E(X) = \frac{\alpha'}{\alpha' + \beta'} = \frac{\alpha + t}{\alpha + t + \beta + n - t} = \frac{\alpha + t}{\alpha + \beta + n}$$

A distribuição uniforme no intervalo (0, 1) corresponde a uma distribuição Beta com parâmetros $\alpha = 1$ e $\beta = 1$. Além disso, o enunciado informa que a amostra tem tamanho $n = 15$, sendo verificados $t = 2$ itens defeituosos. Assim, o estimador de Bayes para o parâmetro desconhecido é:

$$E(X) = \frac{\alpha + t}{\alpha + \beta + n} = \frac{1 + 2}{1 + 1 + 15} = \frac{3}{17}$$

Gabarito: B

Estimação por Intervalos

Enquanto na estimação pontual resumimos toda a informação presente na distribuição a posteriori em um único valor; na estimação por intervalos, construímos um **intervalo** em torno desse ponto, para indicar o a **precisão** da estimativa.

Esse intervalo, construído a partir da distribuição a **posteriori**, é chamado de **intervalo de credibilidade** (ou intervalo de confiança Bayesiano), cuja definição é:

C é um intervalo de credibilidade ao nível $1 - \alpha$ para θ se $P(\theta \in C) \geq 1 - \alpha$.

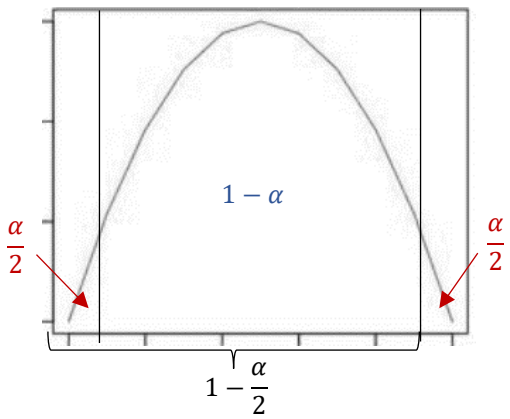
Na estatística Bayesiana, podemos dizer que $1 - \alpha$ é a probabilidade de o verdadeiro parâmetro (que é uma variável aleatória) **pertencer** ao intervalo de credibilidade.

Embora a probabilidade de $\theta \in C$ possa ser maior que $1 - \alpha$, pela definição, em geral, utilizamos a igualdade, pois desejamos que o intervalo seja o **menor possível**, pois quanto menor o intervalo, mais precisa é a estimativa.

O sinal de maior ou igual é considerado quando não for possível satisfazer a igualdade, o que pode ocorrer com variáveis discretas.

Os **extremos** do intervalo de confiança são os quantis $Q_{\frac{\alpha}{2}}$ e $Q_{1-\frac{\alpha}{2}}$ da distribuição a posteriori, tais que:

$$P\left(\theta \middle| x < Q_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}, \qquad P\left(\theta \middle| x < Q_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$$



Por exemplo, para um nível $1-\alpha=95\%$, ou seja, $\alpha=5\%$ de credibilidade, os extremos do intervalo serão $Q_{2,5\%}$ e $Q_{97,5\%}$ da distribuição a posteriori.

Assim como os intervalos de confiança da estatística clássica, os intervalos de credibilidade são **invariantes** a transformações 1 a 1, o que significa que podemos aplicar uma **função** ao parâmetro θ e aos limites do intervalo de credibilidade $[a,b]$, de modo que o novo intervalo $[\varphi(a),\varphi(b)]$ será um intervalo com o **mesmo nível de credibilidade** para o novo parâmetro $\varphi(\theta)$.

Já, os **intervalos de máxima densidade a posteriori** (MDP ou HPD - *Highest Posterior Density*) **minimizam o tamanho do intervalo**, mantendo o nível de confiança $1-\alpha$.

Para isso, considera-se o intervalo de valores de θ com **maior densidade** a posteriori.

A definição de intervalo de máxima densidade a posteriori é:

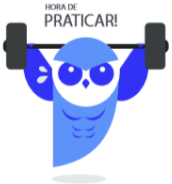
> *Um intervalo de credibilidade C ao nível $1-\alpha$ para θ é de máxima densidade a posteriori se $C=\{\theta: p(\theta|x)\geq k_\alpha\}$, sendo k_α a maior constante tal que $P(\theta\in C)\geq 1-\alpha$*

Em outras palavras, as densidades da distribuição a posteriori delimitada pelo intervalo MDP serão as **maiores possíveis**, desde que a probabilidade de o parâmetro pertencer ao intervalo seja $1-\alpha$.

Consequentemente, os intervalos MDP apresentam as seguintes propriedades:

- a **densidade** para qualquer ponto pertencente ao intervalo é **maior** que a densidade para qualquer ponto não pertencente ao intervalo;
- para um nível de credibilidade $1-\alpha$, o **intervalo** é o de **menor** comprimento.

Se a distribuição a posteriori for unimodal e simétrica (com duas caudas), o intervalo MDP **coincide** com o **intervalo de credibilidade**. Caso contrário, é possível encontrar intervalos MDP que consistem na **união** de intervalos.



(Cebraspe/2019 – TJ-AM) Acerca de métodos usuais de estimação intervalar, julgue o item subsecutivo.

Intervalos de credibilidade independem da distribuição a priori utilizada.

Comentários:

Embora o intervalo de credibilidade seja construído em torno da distribuição a posteriori, esta depende da distribuição a priori. Assim, os intervalos de credibilidade não independem da distribuição a priori.

Gabarito: Errado.

(VUNESP/2021 – EsFCEx) Uma alternativa bayesiana em relação ao intervalo de confiança empregado na estatística clássica é o intervalo de credibilidade.

Assinale a alternativa que define corretamente o que representa o intervalo de credibilidade de 95%.

a) O intervalo de credibilidade de 95% para um parâmetro é o intervalo delimitado pelos percentis 5% e 95% da distribuição a posteriori para o parâmetro.

b) O intervalo de credibilidade de 95% para um parâmetro é o intervalo delimitado pelos percentis 2,5% e 97,5% da distribuição a priori para o parâmetro.

c) O intervalo de credibilidade de 95% para um parâmetro é o intervalo delimitado pelos percentis 2,5% e 97,5% da distribuição a posteriori para o parâmetro

d) O intervalo de credibilidade de 95% para um parâmetro é o intervalo delimitado pelos percentis 5% e 95% da distribuição a priori para o parâmetro.

e) O intervalo de credibilidade de 95% para um parâmetro é o intervalo delimitado pelos percentis 5% e 95% da função de verossimilhança dos dados.

Comentários:

Um intervalo de credibilidade a um nível $1 - \alpha = 95\%$ ($\alpha = 5\%$) é delimitado pelos quantis $Q_{\alpha/2} = Q_{2,5\%}$ e $Q_{1-\alpha/2} = Q_{97,5\%}$ da distribuição a **posteriori**. Assim, a única alternativa correta é a C.

Gabarito: C

(Cebraspe/2013 – CNJ) Com relação a inferência estatística, julgue o item a seguir.

Entre todos os intervalos que possuem o mesmo nível de credibilidade, o intervalo HPD (highest probability density) é o que proporciona a maior amplitude possível.

Comentários:

O intervalo HPD é aquele que concentra as **maiores densidades** da distribuição a posteriori do parâmetro desconhecido, o que torna o intervalo com a **menor amplitude** possível, dentre aqueles com o mesmo nível de credibilidade.

Gabarito: Errado

(2022 – FURG) O intervalo de credibilidade delimita, com probabilidade especificada, um conjunto de parâmetros plausíveis para o parâmetro de acordo com sua distribuição posterior marginal.

É correto afirmar que:

a) O intervalo de credibilidade percentil 95% para θ simbolizado por $ICr_{95\%}$ é o intervalo delimitado pelos percentis 0% e 95% da distribuição posterior marginal $p(\theta \mid x)$ para θ.

b) O intervalo de credibilidade percentil 95% para θ simbolizado por $ICr_{95\%}$ inclui os valores mais extremos em cada cauda da distribuição posterior marginal.

c) O intervalo de credibilidade percentil 95% para θ simbolizado por $ICr_{95\%}$ é o intervalo delimitado pelos percentis 2,5% e 97,5% da distribuição posterior marginal $p(\theta \mid x)$ para θ.

d) O intervalo de maiores densidades de 95% para θ simbolizando por $HDIr_{95\%}$ inclui os valores que representam as 5% menores densidades na distribuição posterior.

e) O intervalo de maiores densidades de 95% para θ simbolizando por $HDIr_{95\%}$ exclui os valores que representam as 5% menores densidades na distribuição posterior de forma igualitária, independente da simetria da distribuição posterior.

Comentários:

Essa questão cobra conceitos relativos ao intervalo de credibilidade e ao intervalo de máxima densidade, que a questão chama de maiores densidades.

Um intervalo de credibilidade a um nível de 95% ($\alpha = 5\%$) é delimitado pelos quantis $Q_{\alpha/2} = Q_{2,5\%}$ e $Q_{1-\alpha/2} = Q_{97,5\%}$, como indicado na alternativa C; e **não** $Q_{0\%}$ e $Q_{95\%}$, como indicado na alternativa A.

A alternativa B afirma que os limites do intervalo são os extremos da distribuição $Q_{0\%}$ e $Q_{100\%}$, o que também **não** está correta.

Já o intervalo de máxima densidade concentra as **maiores** densidades da distribuição a posteriori, desde que a probabilidade de o parâmetro pertencer ao intervalo. Assim, a alternativa D está incorreta, pois afirma que o intervalo inclui as 5% menores densidades. Ademais, a alternativa E está incorreta, pois afirma que o intervalo exclui as menores densidades de forma igualitária, independentemente da simetria da distribuição.

Gabarito: C

Distribuição Normal

Se a função de verossimilhança for normal com média desconhecida θ variância conhecida σ^2; e a distribuição a priori for normal com média μ_0 e variância τ_0^2, então a distribuição a posteriori será **normal** com média μ_1 e variância τ_1^2, os quais são obtidos a partir dos parâmetros da distribuição a priori e dos dados amostrais. Assim, podemos dizer que a seguinte expressão segue a **normal padrão**:

$$Z = \frac{\theta - \mu_1}{\tau_1}$$

Em que θ representa a distribuição posteriori do parâmetro desconhecido, condicionado à amostra obtida.

A partir de um nível de credibilidade $1 - \alpha$, podemos obter o valor de z pela tabela normal padrão tal que $P(-z \leq Z \leq z) = 1 - \alpha$ e, portanto:

$$P\left(-z \leq \frac{\theta - \mu_1}{\tau_1} \leq z\right) = 1 - \alpha$$

Isolando o parâmetro desconhecido, temos:

$$P(\mu_1 - z.\tau_1 \leq \theta \leq \mu_1 + z.\tau_1) = 1 - \alpha$$

Em outras palavras, o seguinte intervalo corresponde ao **intervalo de credibilidade** para o parâmetro desconhecido, condicionado à amostra obtida:

$$[\mu_1 - z.\tau_1;\ \mu_1 + z.\tau_1]$$

Em razão da simetria da normal, esse intervalo é **MDP**.

No exemplo do físico A e a amostra de tamanho $n = 100$, calculamos a variância a posteriori $\tau_1^2 \cong 15{,}4$, o que corresponde a um desvio padrão $\tau_1 \cong 3{,}9$, e a média a posteriori $\mu_1 \cong 871{,}2$. O intervalo de credibilidade de 95%, associado a $z = 1{,}96$, é dado por:

$$IC = \mu_1 \pm z.\tau_1 = 871{,}2 \pm 1{,}96 \times 3{,}9 = 871{,}2 \pm 7{,}6$$

$$IC = [863{,}6; 878{,}8]$$

Utilizando a **priori não informativa**, em que $\tau_0^2 \to \infty$, os parâmetros da distribuição a posteriori coincidem com os da inferência clássica $\mu_1 = \bar{x}$ e $\tau_1^2 = \sigma^2/n$.

Nessa situação, o **intervalo de credibilidade** se torna:

$$\left[\bar{x} - z.\frac{\sigma}{\sqrt{n}};\ \bar{x} + z.\frac{\sigma}{\sqrt{n}}\right]$$

Que também **coincide** numericamente com o intervalo de confiança da inferência clássica. No entanto, a sua interpretação continua sendo distinta.

Para o mesmo exemplo, em que $\bar{x} = 870$, $\sigma^2 = 1600$ (o que corresponde a um desvio padrão $\sigma = 40$) e $n = 100$, o intervalo de credibilidade ao nível de 95%, em que $z = 1{,}96$, é:

$$IC = \bar{x} \pm z.\frac{\sigma}{\sqrt{n}} = 870 \pm 1{,}96 \times \frac{40}{\sqrt{100}} = 870 \pm 1{,}96 \times 4 = 870 \pm 7{,}84$$

$$IC = [862{,}16; 877{,}84]$$

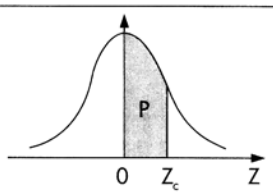
A inferência Bayesiana como um todo envolve o cálculo de diversas integrais; e algumas delas podem ser extremamente complexas. Nessa situação, utilizamos métodos baseados em **simulação**, que fornecem resultados **aproximados** para esses cálculos.

QUESTÕES COMENTADAS - VUNESP

Estimação Intervalar

Utilize a tabela a seguir para resolver as próximas questões, de 1 a 5.

Tabela I — Distribuição Normal Padrão
$Z \sim N(0, 1)$
Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

1. (VUNESP/2015 – TJ-SP) Leia o texto a seguir para responder à questão.

O Sr. Manoel comprou uma padaria, e foi garantido o faturamento médio de R$ 1.000,00 por dia de funcionamento. Durante os primeiros 16 dias, considerados como uma amostra de 16 valores da população, obteve-se o faturamento médio de R$ 910,00 e desvio padrão de R$ 80,00. Sentindo-se enganado pelo vendedor, o Sr. Manoel entrou com ação de perdas e danos. O juiz sugeriu, então, efetuar o teste de hipótese, indicado ao nível de significância de 5% para confirmar ou refutar a ação.

Supondo-se que a distribuição seja normal com desvio padrão de R$ 120,00 e que a amostra dos 16 dias tenha acusado o valor de R$ 910,00, então o intervalo de confiança para a verdadeira média com 95% de confiança é de, aproximadamente,

a) R$ 850,00 < x < R$ 970,00

b) R$ 800,00 < x < R$ 1.020,00

c) R$ 790,00 < x < R$ 1.030,00

d) R$ 900,00 < x < R$ 920,00

e) R$ 890,00 < x < R$ 930,00

Comentários:

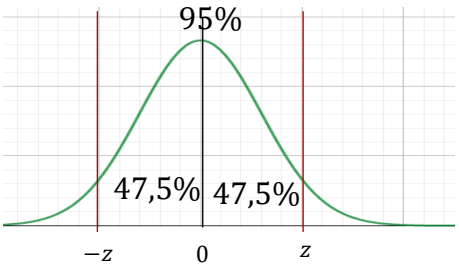
Essa questão trabalha com a estimação intervalar para a média de uma população com distribuição normal e variância conhecida, em que o intervalo de confiança é dado por:

$$\bar{X} \pm z.\frac{\sigma}{\sqrt{n}}$$

O enunciado informa que:

- A média é $\bar{X} = 910$
- O desvio padrão é $\sigma = 120$
- O tamanho da amostra é $n = 16$

O enunciado informa, ainda, que o nível de confiança é de 95%, conforme ilustrado abaixo:



Ou seja, precisamos do valor de z para o qual P(0 < Z < z) = 47,5% = 0,475. Pela tabela fornecida acima, observamos que z = 1,96. Substituindo esse e os demais dados na fórmula do intervalo de confiança, temos:

$$L_{SUP} = 910 + 1{,}96 \times \frac{120}{\sqrt{16}} = 910 + 1{,}96 \times \frac{120}{4} = 910 + 1{,}96 \times 30 \cong 910 + 60 = 970$$

$$L_{INF} = 910 - 1{,}96 \times \frac{120}{\sqrt{16}} = 910 - 1{,}96 \times \frac{120}{4} = 910 - 1{,}96 \times 30 \cong 910 - 60 = 850$$

Gabarito: A

2. (VUNESP/2015 – TJ-SP) Para avaliar o tempo médio de viagem entre o ponto inicial e o ponto final de uma linha de ônibus, retira-se uma amostra de 36 observações (viagens), encontrando-se, para essa amostra, o tempo médio de 50 minutos e o desvio padrão de 6 minutos, com distribuição normal. Considerando-se um intervalo de confiança de 95% para o tempo médio populacional, é correto afirmar que o valor mais próximo para limite inferior desse intervalo é o tempo de

a) 40 min.

b) 44 min.

c) 48 min.

d) 52 min.

e) 56 min.

Comentários:

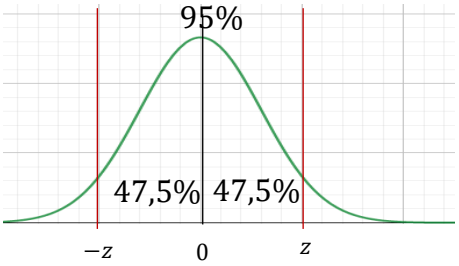
Essa questão trabalha com a estimação intervalar para a média de uma população com distribuição normal e variância conhecida, em que o intervalo de confiança é dado por:

$$\bar{X} \pm z.\frac{\sigma}{\sqrt{n}}$$

O enunciado informa que:

- A média é $\bar{X} = 50$ min
- O desvio padrão é $\sigma = 6$ min
- O tamanho da amostra é $n = 36$

O enunciado informa, ainda, que o nível de confiança é de 95%, conforme ilustrado abaixo:



Ou seja, precisamos do valor de z para o qual P(0 < Z < z) = 47,5% = 0,475. Pela tabela fornecida acima, observamos que z = 1,96. Substituindo esse e os demais dados na fórmula do limite inferior do intervalo de confiança, temos:

$$L_{INF} = 50 - 1{,}96 \times \frac{6}{\sqrt{36}} = 50 - 1{,}96 \times \frac{6}{6} = 50 - 1{,}96 \cong 48$$

Gabarito: C

3. (VUNESP/2014 – TJ-PA) Supondo que em uma amostra de 4 baterias automotivas tenha-se calculado o tempo de vida média de 4 anos. Sabe-se que o tempo de vida da bateria é uma distribuição normal com desvio padrão de 1 ano e meio.

Então, o intervalo de 90% de confiança para a média de todas as baterias é de, aproximadamente:

a) 4 anos $\pm$ 15 meses

b) 4 anos $\pm$ 12 meses

c) 4 anos $\pm$ 10 meses

d) 4 anos $\pm$ 8 meses

e) 4 anos $\pm$ 5 meses

Comentários:

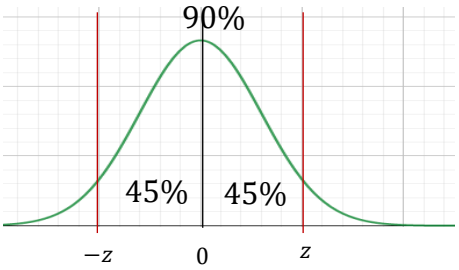
Essa questão trabalha com a estimação intervalar para a média de uma população com distribuição normal e variância conhecida, em que o intervalo de confiança é dado por:

$$\bar{X} \pm z.\frac{\sigma}{\sqrt{n}}$$

O enunciado informa que:

- A média é $\bar{X} = 4$ anos
- O desvio padrão é $\sigma = 1{,}5$ ano
- O tamanho da amostra é $n = 4$

O enunciado informa, ainda, que o nível de confiança é de 90%, conforme ilustrado abaixo:



Ou seja, precisamos do valor de z para o qual P(0 < Z < z) = 45% = 0,45. Pela tabela fornecida acima, observamos que z = 1,64. Substituindo esse e os demais dados na fórmula do intervalo de confiança, temos:

$$4 \pm 1{,}64 \times \frac{1{,}5}{\sqrt{4}} = 4 \pm 1{,}64 \times \frac{1{,}5}{2} = 4 \pm 0{,}82 \times 1{,}5 = 4 \pm 1{,}23$$

1,23 ano corresponde ao seguinte número de meses:

$$1{,}23 \times 12 = 14{,}76 \cong 15$$

Ou seja, o intervalo de confiança é 4 anos $\pm$ 15 meses

Gabarito: A

4. (VUNESP/2014 – TJ-PA) O sindicato dos médicos de uma região divulgou uma nota na qual afirma que os médicos plantonistas daquela região trabalham em média, mais de 40 horas por semana, o que, supostamente, contraria acordos anteriores firmados com a categoria. Para verificar essa afirmativa realizou-se uma nova pesquisa na qual 16 médicos escolhidos de modo aleatório declararam seu tempo de trabalho semanal, obtendo-se para tempo médio e para desvio padrão, respectivamente, os valores 42 horas e 3 horas. Considere-se que a distribuição de frequência das horas trabalhadas pelos médicos é normal. Com esses dados o intervalo de confiança $\mu_x = \overline{x} \pm t_{\alpha/2} \times \frac{s_x}{\sqrt{n}}$ de 95% para a média populacional, é:

a) $\mu_x = 42 \pm 6$

b) $\mu_x = 42 \pm 3$

c) $\mu_x = 42 \pm 2{,}55$

d) $\mu_x = 42 \pm 1{,}6$

e) $\mu_x = 42 \pm 0{,}75$

Comentários:

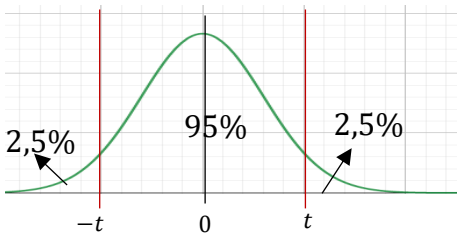
Essa questão trabalha com a estimação intervalar para a média de uma população com distribuição normal e variância desconhecida, em que o intervalo de confiança é dado por:

$$\bar{X} \pm t.\frac{s}{\sqrt{n}}$$

O enunciado informa que:

- A média é $\bar{X} = 42$ horas
- O desvio padrão amostral é $s = 3$ horas
- O tamanho da amostra é $n = 16$

O enunciado informa, ainda, que o nível de confiança é de 1 – p = 95%, logo p = 5%:



Ademais, sabemos que n – 1= 15. Pela tabela, observamos que o valor de t associado a p = 5% e 15 graus de liberdade é t = 2,131. Substituindo esse e os demais dados na fórmula do intervalo de confiança, temos:

$$42 \pm 2{,}131 \times \frac{3}{\sqrt{16}} = 42 \pm 2{,}131 \times \frac{3}{4} \cong 42 \pm 1{,}6$$

Gabarito: D

5. (VUNESP/2013 – MPE-ES) Pesquisa recente sobre o tempo total para que os ônibus de determinada linha urbana percorram todo o trajeto entre o ponto inicial e o ponto final, programados para essa viagem, detectou que os tempos de viagem são normalmente distribuídos com tempo médio gasto de 53 minutos e com desvio-padrão amostral de 9 minutos. Nessa pesquisa, foram observados e computados os dados de 16 viagens escolhidas aleatoriamente.

Com um intervalo de confiança de 98%, utilizando-se a tabela t de Student para estimar o erro amostral, e arredondando para cima o valor desse erro, é correto afirmar que o tempo médio dessa viagem varia entre

a) 45 min e 61 min

b) 47 min e 59 min

c) 50 min e 56 min

d) 51 min e 55 min

e) 52 min e 54 min

Comentários:

Essa questão trabalha com a estimação intervalar para a média de uma população com distribuição normal e variância desconhecida, em que o intervalo de confiança é dado por:

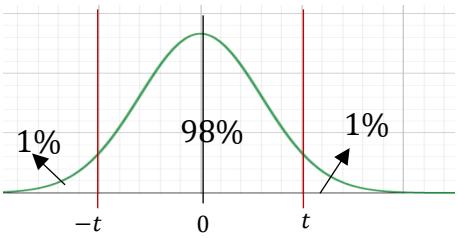
$$\bar{X} \pm t.\frac{s}{\sqrt{n}}$$

O enunciado informa que:

- A média é $\bar{X} = 53$ minutos

- O desvio padrão amostral é $s = 9$ minutos
- O tamanho da amostra é $n = 16$

O enunciado informa, ainda, que o nível de confiança é de 1 – p = 98%, logo p = 2%:



Ademais, sabemos que n – 1= 15. Pela tabela, observamos que o valor de t associado a p = 2% e 15 graus de liberdade é t = 2,602. Substituindo esse e os demais dados na fórmula do intervalo de confiança, temos:

$$L_{SUP} = 53 + 2{,}602 \times \frac{9}{\sqrt{16}} = 53 + 2{,}602 \times \frac{9}{4} \cong 53 + 6 = 59$$

$$L_{INF} = 53 - 2{,}602 \times \frac{9}{\sqrt{16}} = 53 - 2{,}602 \times \frac{9}{4} \cong 53 - 6 = 47$$

Gabarito: B

6. (VUNESP/2014 – EMPLASA) Um jornal deseja estimar a proporção de jornais impressos com não conformidades. Em uma amostra aleatória de 100 jornais dentre todos os jornais impressos durante um dia, observou-se que 20 têm algum tipo de não conformidade. Para um nível de confiança de 90%, Z = 1,64. Então pode-se concluir que apresentam não conformidades.

a) 21,64%, no máximo

b) 26,56%, no máximo

c) 26,56%, no mínimo

d) 21,64%, no mínimo

e) 20%, no mínimo

Comentários:

Essa questão trabalha com a estimação intervalar para proporções, em que o intervalo de confiança é dado por:

$$\hat{p} \pm z. \sqrt{\frac{\hat{p}.\hat{q}}{n}}$$

O enunciado informa que dos n = 100 jornais, 20 apresentam não conformidade. Logo, a proporção amostral de não conformidade é:

$$\hat{p} = \frac{20}{100} = 0{,}2$$

E o complementar é $\hat{q} = 1 - \hat{p} = 1 - 0{,}2 = 0{,}8$. Sabendo que z = 1,64, temos:

$$L_{SUP} = 0{,}2 + 1{,}64.\sqrt{\frac{0{,}2.0{,}8}{100}} = 0{,}2 + 1{,}64.\frac{\sqrt{0{,}16}}{10} = 0{,}2 + 0{,}164 \times 0{,}4 = 0{,}2656 = 26{,}56\%$$

Gabarito: B

QUESTÕES COMENTADAS - MULTIBANCAS

Distribuição Amostral

CEBRASPE

1. (CESPE 2019/TJ-AM) Em determinado município brasileiro, realizou-se um levantamento para estimar o percentual P de pessoas que conhecem o programa justiça itinerante. Para esse propósito, foram selecionados 1.000 domicílios por amostragem aleatória simples de um conjunto de 10 mil domicílios. Nos domicílios selecionados, foram entrevistados todos os residentes maiores de idade, que totalizaram 3.000 pessoas entrevistadas, entre as quais 2.250 afirmaram conhecer o programa justiça itinerante. De acordo com essa situação hipotética, julgue o seguinte item.

O número médio de pessoas maiores de idade por domicílio foi igual a 3 pessoas por domicílio; e o erro padrão do estimador do percentual P é inversamente proporcional a $3\sqrt{1.000}$.

Comentários:

Como foram entrevistadas 3000 pessoas em 1000 domicílios, a média de pessoas maiores de idade por domicílio é igual a:

$$\frac{3000}{1000} = 3\ pessoas\ por\ domicílio.$$

Por outro lado, o erro padrão da proporção é igual a:

$$EP = \sqrt{\frac{p(1-p)}{n}},$$

em que p é a proporção amostral e n o tamanho da amostra.

Sendo a amostra do número de pessoas de tamanho igual a 3000, o número de pessoas que conhecem o programa igual a 2250, então $p = 2250/3000$ e

$$EP = \sqrt{\frac{\frac{2250}{3000} \times \frac{750}{3000}}{3000}}$$

$$EP = \sqrt{\frac{750}{1000} \times \frac{750}{3000} \times \frac{1}{3000}} = \sqrt{\frac{750}{1000} \times \frac{750}{1000} \times \frac{1}{9000}} = \frac{750}{1000} \times \frac{1}{\sqrt{9000}} = 0{,}75 \times \frac{1}{3\sqrt{1000}}$$

o que mostra que o erro é inversamente proporcional a $3\sqrt{1000}$.

Gabarito: Certo.

2. (CESPE 2019/TJ-AM) Para avaliar a satisfação dos servidores públicos de certo tribunal no ambiente de trabalho, realizou-se uma pesquisa. Os servidores foram classificados em três grupos, de acordo com o nível do cargo ocupado. Na tabela seguinte, k é um índice que se refere ao grupo de servidores, e N_k denota o tamanho populacional de servidores pertencentes ao grupo k.

Nível do Cargo	k	N_k	n_k	p_k
I	1	500	50	0,7
II	2	300	20	0,8
III	3	200	10	0,9

De cada grupo k foi retirada uma amostra aleatória simples sem reposição de tamanho n_k; p_k representa a proporção de servidores amostrados do grupo k que se mostraram satisfeitos no ambiente de trabalho.

A partir das informações e da tabela apresentadas, julgue o próximo item.

Com relação ao grupo $k = 2$, o erro padrão da estimativa da proporção dos servidores satisfeitos no ambiente de trabalho foi inferior a $0{,}1$.

Comentários:

O erro padrão para a proporção é dada pela relação:

$$EP = \sqrt{\frac{p(1-p)}{n}}.$$

Para $k = 2$, temos $p_2 = 0{,}8$ e $n_2 = 20$, de modo que:

$$EP = \sqrt{\frac{0{,}8 \times 0{,}2}{20}} = \sqrt{0{,}008} \cong 0{,}089$$

Gabarito: Certo.

3. (CESPE 2019/TJ-AM) Em determinado município brasileiro, realizou-se um levantamento para estimar o percentual P de pessoas que conhecem o programa justiça itinerante. Para esse propósito, foram selecionados 1.000 domicílios por amostragem aleatória simples de um conjunto de 10 mil domicílios. Nos domicílios selecionados, foram entrevistados todos os residentes maiores de idade, que totalizaram 3.000 pessoas entrevistadas, entre as quais 2.250 afirmaram conhecer o programa justiça itinerante.

De acordo com essa situação hipotética, julgue o seguinte item.

A estimativa do percentual de pessoas que conhecem o programa justiça itinerante foi inferior a 60%.

Comentários:

Após selecionados os domicílios, foram entrevistados todos os residentes maiores de idade, que totalizaram 3.000 pessoas. Nesse grupo, o número de participantes que conhecia o programa justiça itinerante foi de 2.250 pessoas.

Dessa forma, a estimativa de pessoas que conhecem o programa é de:

$$\hat{p} = \frac{2250}{3000} = 0{,}75 = 75\%$$

Gabarito: Errado.

4. (CESPE 2019/TJ-AM) Para estimar a proporção de menores infratores reincidentes em determinado município, foi realizado um levantamento estatístico. Da população-alvo desse estudo, constituída por 10.050 menores infratores, foi retirada uma amostra aleatória simples sem reposição, composta por 201 indivíduos. Nessa amostra foram encontrados 67 reincidentes.

Com relação a essa situação hipotética, julgue o seguinte item.

Se a amostragem fosse com reposição, a estimativa da variância da proporção amostral teria sido superior a 0,001.

Comentários:

O Teorema Central do Limite nos garante que a variância da proporção amostral, na amostragem com reposição, tem distribuição aproximadamente normal, sendo calculada pela relação:

$$Var(\hat{p}) = \frac{p(1-p)}{n}$$

Em que p é a proporção amostral e n é o tamanho da amostra. Como a amostra tem tamanho $n = 201$ e $p = 67/201 = 1/3$, então:

$$Var(\hat{p}) = \frac{p(1-p)}{n}$$

$$Var(\hat{p}) = \frac{\frac{1}{3} \times \frac{2}{3}}{201}$$

$$Var(\hat{p}) = \frac{2}{1809}$$

$$Var(\hat{p}) \cong 0{,}0011$$

Gabarito: Certo.

5. (CESPE 2018/PF) Determinado órgão governamental estimou que a probabilidade p de um ex-condenado voltar a ser condenado por algum crime no prazo de 5 anos, contados a partir da data da libertação, seja igual a 0,25. Essa estimativa foi obtida com base em um levantamento por amostragem aleatória simples de 1.875 processos judiciais, aplicando-se o método da máxima verossimilhança a partir da distribuição de Bernoulli. Sabendo que $P(Z < 2) = 0,975$, em que Z representa a distribuição normal padrão, julgue o item que se segue, em relação a essa situação hipotética.

O erro padrão da estimativa da probabilidade p foi igual a 0,01.

Comentários:

O estimador de máxima verossimilhança para a proporção populacional é justamente a proporção amostral. A variância da estimativa de p é dada por:

$$s^2 = \frac{\hat{p}\hat{q}}{n}$$

Em que $\hat{p}$ é a proporção amostral de casos de interesse (0,25) e $\hat{q} = 1 - \hat{p} = 0,75$. Além disso, n é o tamanho da amostra, que é de 1.875 processos.

$$s^2 = \frac{25 \times 0,75}{1.875}$$

Dividindo numerador e denominador por 75:

$$s^2 = \frac{0,25 \times 0,01}{25}$$

Dividindo numerador e denominador por 25:

$$s^2 = 0,01 \times 0,01$$

O desvio padrão é a raiz quadrada da variância:

$$s = 0,01$$

Gabarito: Certo.

6. (CESPE 2018/PF) Uma pesquisa realizada com passageiros estrangeiros que se encontravam em determinado aeroporto durante um grande evento esportivo no país teve como finalidade investigar a sensação de segurança nos voos internacionais. Foram entrevistados 1.000 passageiros, alocando-se a amostra de acordo com o continente de origem de cada um — África, América do Norte (AN), América do Sul (AS), Ásia/Oceania (A/O) ou Europa. Na tabela seguinte, N é o tamanho populacional de passageiros em voos internacionais no período de interesse da pesquisa; n é o tamanho da amostra por origem; P é o percentual dos passageiros entrevistados que se manifestaram satisfeitos no que se refere à sensação de segurança.

Origem	N	n	P
África	100.000	100	80
AN	300.000	300	70
AS	100.000	100	90
A/O	300.000	300	80
Europa	200.000	200	80
Total	1.000.000	1.000	P_{pop}

Em cada grupo de origem, os passageiros entrevistados foram selecionados por amostragem aleatória simples. A última linha da tabela mostra o total populacional no período da pesquisa, o tamanho total da amostra e P_{pop} representa o percentual populacional de passageiros satisfeitos.

A partir dessas informações, julgue o item.

Considerando o referido desenho amostral, estima-se que o percentual populacional P_{pop} seja inferior a 79%.

Comentários:

A proporção amostral ($\hat{p}$) nos dá a estimativa da proporção populacional (p).

Origem	N	n	P(%)	$n \times P$
África	100.000	100	80%	80
AN	300.000	300	70%	210
AS	100.000	100	90%	90
A/O	300.000	300	80%	240
Europa	200.000	200	80%	160
total	**1.000.000**	**1.000**		**780**

Logo:

$$\hat{p} = \frac{780}{1.000} = 78\%.$$

De fato, a estimativa é inferior a 79%.

Gabarito: Certo.

7. (CESPE 2018/STM) Em um tribunal, entre os processos que aguardam julgamento, foi selecionada aleatoriamente uma amostra contendo 30 processos. Para cada processo da amostra que estivesse há mais de 5 anos aguardando julgamento, foi atribuído o valor 1; para cada um dos outros, foi atribuído o valor 0. Os dados da amostra são os seguintes:

1 1 0 0 1 0 1 1 0 1 1 1 1 0 1 1 1 0 1 0 1 0 1 0 1 1 1 0 1 1

A proporção populacional de processos que aguardam julgamento há mais de 5 anos foi denotada por p; a proporção amostral de processos que aguardam julgamento há mais de 5 anos foi representada por $\hat{p}$

A variância da proporção amostral $\hat{p}$ sob a hipótese nula H_0: $p = 0{,}5$ é menor que 0,1.

Comentários:

A variância da proporção amostral é dada por:

$$\sigma_{\hat{p}} = \frac{pq}{n}$$

Na fórmula acima, temos:

i) p é a proporção populacional de casos favoráveis. A questão disse que $p = 0{,}5$;

ii) q é a proporção populacional de casos desfavoráveis. $q = 1 - p = 0{,}5$;

iii) n é o tamanho da amostra ($n = 30$).

$$= \frac{0{,}5 \times 0{,}5}{30}$$

$$= \frac{0{,}25}{30}$$

$$\cong 0{,}00833$$

Gabarito: Certo.

8. (CESPE 2018/STM) Em um tribunal, entre os processos que aguardam julgamento, foi selecionada aleatoriamente uma amostra contendo 30 processos. Para cada processo da amostra que estivesse há mais de 5 anos aguardando julgamento, foi atribuído o valor 1; para cada um dos outros, foi atribuído o valor 0. Os dados da amostra são os seguintes:

1 1 0 0 1 0 1 1 0 1 1 1 1 0 1 1 1 0 1 0 1 0 1 0 1 1 1 0 1 1

A proporção populacional de processos que aguardam julgamento há mais de 5 anos foi denotada por p; a proporção amostral de processos que aguardam julgamento há mais de 5 anos foi representada por $\hat{p}$

Estima-se que, nesse tribunal, $p > 60\%$.

Comentários:

Na amostra de tamanho 30, há 20 casos favoráveis, ou seja, 20 valores iguais a "1".

Portanto, a proporção amostral fica:

$$\hat{p} = \frac{20}{30} \cong 66{,}67\%$$

Como é comum usarmos a proporção amostral para estimar a proporção populacional, pode-se dizer que a estimativa é superior a 60%.

Gabarito: Certo.

9. (CESPE 2018/EBSERH) $X_1, X_2, \ldots, X_{10}$ representa uma amostra aleatória simples retirada de uma distribuição normal com média μ e variância σ^2, ambas desconhecidas. Considerando que $\hat{\mu}$ e $\hat{\sigma}$ representam os respectivos estimadores de máxima verossimilhança desses parâmetros populacionais, julgue o item subsecutivo.

A razão $\frac{\hat{\mu}-\mu}{\sigma}$ segue uma distribuição normal padrão.

Comentários:

Pode-se mostrar que o estimador de máxima verossimilhança para a média de uma distribuição normal $N(\mu, \sigma^2)$ é igual a

$$\hat{\mu} = \bar{X} = \frac{X_1 + X_2 + \cdots + X_{10}}{10} = Média\ amostral$$

Para concluir a questão, precisamos de duas informações adicionais:

1. a distribuição da média amostral de uma distribuição normal com média μ e variância σ^2 é dada por

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

2. para uma variável aleatória X com distribuição normal com média μ e variância σ^2, a transformação:

$$Z = \frac{X - \mu}{\sigma}$$

é tal que $Z \sim N(0,1)$. Juntando as duas informações, chegamos à transformação:

$$\frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

e não a $\frac{\hat{\mu}-\mu}{\sigma}$.

Gabarito: Errado.

10. (CESPE 2016/TCE-PA) Uma amostra aleatória simples $X_1, X_2, ..., X_n$ foi retirada de uma população normal com média e desvio padrão iguais a 10. Julgue o próximo item, a respeito da média amostral $\bar{X} = [X_1 + X_2 + \ldots + X_n]/n$.

A variância de $\bar{X}$ é igual a 100.

Comentários:

O desvio padrão populacional vale 10, isso é:

$$\sigma = 10$$

A variância é igual ao quadrado do desvio padrão:

$$\sigma^2 = 100$$

A média amostral tem variância igual à populacional, dividida por n, em que n é o tamanho da amostra.

$$V(\bar{X}) = \frac{100}{n}$$

Logo, esse valor **não** é igual a 100.

Gabarito: Errado.

11. (CESPE 2016/TCE-PA) Uma amostra aleatória simples $X_1, X_2, ..., X_n$ foi retirada de uma população normal com média e desvio padrão iguais a 10. Julgue o próximo item, a respeito da média amostral $\bar{X} = [X_1 + X_2 + \ldots + X_n]/n$.

A média amostral segue uma distribuição t de *Student* com $n - 1$ graus de liberdade.

Comentários:

O enunciado garantiu que a distribuição original é normal. Deste modo, a média aritmética $\bar{X}$ dependerá de uma soma de n variáveis normais, independentes e identicamente distribuídas. Será, também, uma variável **normal** (e não t de Student).

Gabarito: Errado.

12. (CESPE 2016/TCE-PA) Uma amostra aleatória simples $X_1, X_2, ..., X_n$ foi retirada de uma população normal com média e desvio padrão iguais a 10. Julgue o próximo item, a respeito da média amostral $\bar{X} = [X_1 + X_2 + \ldots + X_n]/n$.

$P(\bar{X} - 10 > 0) \leq 0{,}5$

Comentários:

O exercício pediu a seguinte probabilidade:

$$P(\bar{X} - 10 > 0)$$

$$= P(\bar{X} > 10)$$

Sabemos que X é normal com média 10.

A média amostral $\bar{X}$, por sua vez, também é normal com média 10. Ou seja, seu comportamento é simétrico ao redor de 10. Isto significa que a chance de termos valores maiores que 10 é igual à chance de valores menores que 10, ambas valendo 50%.

$$P(\bar{X} > 10) = 0{,}5$$

O item afirmou que a probabilidade seria menor ou igual a 0,5. Como vimos, ela é justamente igual, o que torna o item correto.

Gabarito: Certo.

13. (CESPE 2016/TCE-PA) Considere um processo de amostragem de uma população finita cuja variável de interesse seja binária e assuma valor 0 ou 1, sendo a proporção de indivíduos com valor 1 igual a p = 0,3. Considere, ainda, que a probabilidade de cada indivíduo ser sorteado seja a mesma para todos os indivíduos da amostragem e que, após cada sorteio, haja reposição do indivíduo selecionado na amostragem. A partir dessas informações, julgue o item subsequente.

Se, dessa população, for coletada uma amostra aleatória de tamanho $n = 1$, a probabilidade de um indivíduo apresentar valor 1 é igual a 0,5.

Comentários:

Como na população temos 30% de casos iguais a 1 (ou seja, a população tem 30% de casos favoráveis), então a chance de a extração resultar em caso favorável é justamente de 30%.

Gabarito: Errado.

FCC

14. (FCC 2015/DPE-SP) Uma amostra aleatória simples, com reposição, de n observações $X_1, X_2, \ldots X_n$ foi selecionada de uma população com distribuição uniforme contínua no intervalo $[-2, b]$, $b > -2$. Sabe-se que:

I. a média dessa distribuição uniforme é igual a 10.

II. o desvio padrão de $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ é igual a 0,4.

Nessas condições, o valor de n é igual a,

a) 100.

b) 400.

c) 225.

d) 300.

e) 324.

Comentário

O desvio padrão da média amostral é dado por:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Então n pode ser calculado como:

$$\sqrt{n} = \frac{\sigma}{\sigma_{\bar{X}}}$$

$$n = \frac{\sigma^2}{\sigma_{\bar{X}}^2}$$

Ou seja, n é a razão entre a variância populacional e o quadrado do desvio padrão da média amostral (variância da média amostral). O enunciado informa que $\sigma_{\bar{X}} = 0{,}4$, portanto, $\sigma_{\bar{X}}^2 = 0{,}4^2 = 0{,}16$. Agora, para calcular o valor de n, precisamos da variância populacional, σ^2.

Para isso, o enunciado informa que a população segue distribuição uniforme contínua no intervalo [-2, b]. Para podermos encontrar o valor de b, o enunciado informa que a média da distribuição é E(X) = 10. Sabendo que a média dessa distribuição é igual à média aritmética dos extremos do intervalo, temos:

$$E(X) = \frac{a + b}{2}$$

$$10 = \frac{-2 + b}{2}$$

$$-2 + b = 20$$

$$b = 22$$

E a variância dessa distribuição é dada por:

$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{[22-(-2)]^2}{12} = \frac{[24]^2}{12} = 2 \times 24 = 48$$

Então, o tamanho amostral é:

$$n = \frac{\sigma^2}{\sigma_{\bar{X}}{}^2} = \frac{48}{0{,}16} = 300$$

Gabarito: D

15. (FCC 2015/TRT 3ª Região) Se Z tem distribuição normal padrão, então:

P(Z < 0,5) = 0,591; P(Z < 1) = 0,841; P(Z < 1,15) = 0,8951; P(Z < 1,17) = 0,879; P(Z < 1,2) = 0,885; P(Z<1,4)=0,919;P(Z<1,64)=0,95;P(Z<2)=0,977;P(Z<2,06)=0,98;P(Z<2,4)= 0,997.

Considere que X é a variável aleatória, que representa as idades, em anos, dos trabalhadores de certa indústria. Suponha que X tem distribuição normal com média de μ anos e desvio padrão de 5 anos.

Uma amostra aleatória, com reposição, de 16 trabalhadores será selecionada e sejam X_1, X_2, ..., X_{16} as idades observadas e $\overline{X} = \frac{\sum_1^{16} X_i}{n}$ a média desta amostra. Sabendo-se que a probabilidade de $\overline{X}$ ser superior a 30 anos é igual a 0,919, o valor de μ, em anos, é igual a:

a) 28,25

b) 31,75

c) 30,50

d) 32,50

e) 30,85

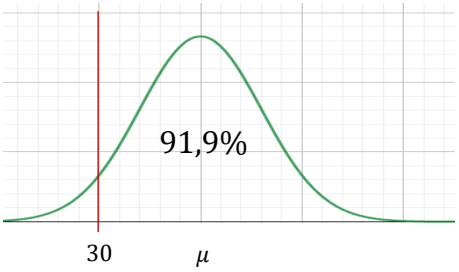
Comentários:

O enunciado informa que a população segue distribuição normal com média μ e desvio padrão $\sigma = 5$, então a média amostral, com tamanho $n = 16$, segue distribuição normal, com a seguinte transformação:

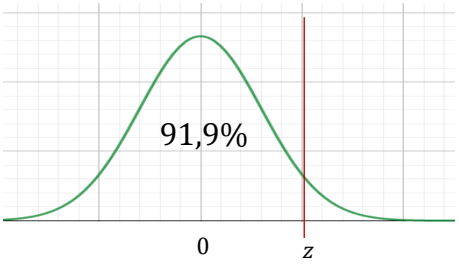
$$z_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\frac{5}{4}}$$

$$\bar{X} - \mu = \frac{5}{4}.z_{\bar{X}}$$

O enunciado informa que a probabilidade de $\bar{X}$ ser superior a 30 é $P(\bar{X} > 30) = 0{,}919$, conforme ilustrado a seguir:



Ou seja, X = 30 é um valor menor do que a média. Por isso, vamos associá-lo ao valor de -z. Pela simetria da curva normal padrão, o valor de z está representado a seguir:



Ou seja, temos P(Z < z) = 0,919. Pelos dados do enunciado, vemos que z = 1,4, pois P(Z<1,4) = 0,919. Portanto, o valor de X = 30 está associado a z = -1,4:

$$30 - \mu = \frac{5}{4} \times (-1{,}4)$$

$$\mu = 30 + 1{,}75 = 31{,}75$$

Gabarito: B

QUESTÕES COMENTADAS - MULTIBANCAS

Estimação Pontual

CEBRASPE

1. (CESPE 2018/STM) Diversos processos buscam reparação financeira por danos morais. A tabela seguinte mostra os valores, em reais, buscados em 10 processos — numerados de 1 a 10 — de reparação por danos morais, selecionados aleatoriamente em um tribunal.

Processo	Valor
1	3.700
2	3.200
3	2.500
4	2.100
5	3.000
6	5.200
7	5.000
8	4.000
9	3.200
10	3.100

A partir dessas informações e sabendo que os dados seguem uma distribuição normal, julgue o item subsequente.

Se μ = estimativa pontual para a média dos valores buscados como reparação por danos morais no referido tribunal, então $3.000 < \mu < 3.300$.

Comentários:

A média amostral fornece-nos uma estimativa pontual da média:

Processo	Valor
1	3.700
2	3.200
3	2.500
4	2.100
5	3.000
6	5.200
7	5.000
8	4.000
9	3.200
10	3.100
Total	**35.000**

O total amostral é de R$ 35.000,00.

Agora, basta dividirmos por 10, já que são 10 elementos na amostra.

$$\bar{X} = \frac{35.000}{10} = 3.500$$

Logo, este valor não está entre 3.000 e 3.300.

Gabarito: Errado.

2. (CESPE 2016/TCE-PA) Considerando uma população finita em que a média da variável de interesse seja desconhecida, julgue o item a seguir.

Considere uma amostragem com três estratos, cujos pesos populacionais sejam 0,2, 0,3 e 0,5. Considere, ainda, que os tamanhos das amostras em cada estrato correspondam, respectivamente, a n_1= 20, n_2= 30 e n_3= 50, e que as médias amostrais sejam 12 kg, 6 kg e 8 kg, respectivamente. Nessa situação, a estimativa pontual da média populacional, com base nessa amostra, é igual a 8,2 kg.

Comentários:

A estimativa pontual da média populacional corresponde às médias amostrais, ponderadas pelos tamanhos dos respectivos estratos:

$$\bar{X} = 0{,}2 \times 12 + 0{,}3 \times 6 + 0{,}5 \times 8 = 2{,}4 + 1{,}8 + 4 = 8{,}2$$

Gabarito: Certo.

3. (CESPE 2016/TCE-PA) Considere um processo de amostragem de uma população finita cuja variável de interesse seja binária e assuma valor 0 ou 1, sendo a proporção de indivíduos com valor 1 igual a p = 0,3. Considere, ainda, que a probabilidade de cada indivíduo ser sorteado seja a mesma para todos os indivíduos da amostragem e que, após cada sorteio, haja reposição do indivíduo selecionado na amostragem. A partir dessas informações, julgue o item subsequente.

Caso, em uma amostra de tamanho n = 10, os valores observados sejam A = {1, 0, 1, 0, 1, 0, 0, 1, 0, 0}, a estimativa via estimador de máxima verossimilhança para a média populacional será igual a 0,4.

Comentários:

A média de uma população com distribuição de Bernoulli corresponde à proporção populacional p. O estimador para a proporção pelo método de máxima verossimilhança é a proporção amostral:

$$\hat{p} = \frac{\sum X}{n} = \frac{1+0+1+0+1+0+0+1+0+0}{10} = \frac{4}{10} = 0{,}4$$

Gabarito: Certo.

4. (CESPE 2018/ABIN) A quantidade diária de emails indesejados recebidos por um atendente é uma variável aleatória X que segue distribuição de Poisson com média e variância desconhecidas. Para estimá-las, retirou-se dessa distribuição uma amostra aleatória simples de tamanho quatro, cujos valores observados foram 10, 4, 2 e 4.

Com relação a essa situação hipotética, julgue o seguinte item.

No que se refere à média amostral $\bar{X} = \frac{(X_1+X_2+X_3+X_4)}{4}$, na qual X_1, X_2, X_3, X_4 representa uma amostra aleatória simples retirada dessa distribuição X, é correto afirmar que a estimativa da variância do estimador $\bar{X}$ seja igual a 1,25.

Comentários:

A estimativa da variância da média amostral é dada por:

$$V(\bar{X}) = \frac{V(X)}{n}$$

Para uma distribuição de Poisson, temos:

$$E(X) = V(X) = \lambda$$

Logo, podemos estimar λ a partir da média amostral:

$$\hat{\lambda} = \bar{X} = \frac{(X_1 + X_2 + X_3 + X_4)}{4} = \frac{(10+4+2+4)}{4} = \frac{20}{4} = 5$$

Logo, temos:

$$V(\bar{X}) = \frac{5}{4} = 1{,}25$$

Gabarito: Certo.

5. (CESPE 2015/Telebras) Um analista da TELEBRAS, a fim de verificar o tempo durante o qual um grupo de consumidores ficou sem o serviço de Internet do qual eram usuários, selecionou uma amostra de 10 consumidores críticos. Os dados coletados, em minutos, referentes a esses consumidores foram listados na tabela seguinte.

consumidor	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
tempo (em minutos)	8	2	3	5	7	7	10	9	4	5

Com base nessa situação hipotética, julgue o item subsequente.

Se os dados seguissem uma distribuição normal, a expressão matemática que permite calcular a variância estimada pelo método de máxima verossimilhança teria denominador igual a 9.

Comentários:

O estimador de máxima verossimilhança para a variância de uma população normal é:

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$$

Sabendo que a amostra tem tamanho n = 10, então o denominador é igual a 10

Gabarito: Errado.

6. (CESPE 2018/ABIN) Considerando que os principais métodos para a estimação pontual são o método dos momentos e o da máxima verossimilhança, julgue o item a seguir.

Para a distribuição normal, o método dos momentos e o da máxima verossimilhança fornecem os mesmos estimadores aos parâmetros μ e σ.

Comentários:

Os estimadores para a média e para a variância obtidos tanto pelo método dos momentos, quanto pelo método da máxima verossimilhança são:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$$

Gabarito: Certo.

7. (CESPE 2020/TJ-PA) Considerando que a inferência estatística é um processo que consiste na utilização de observações feitas em uma amostra com o objetivo de estimar as propriedades de uma população, assinale a opção correta.

a) Na estatística inferencial, um parâmetro é um valor conhecido, extraído de uma amostra, utilizado para a estimação de uma grandeza populacional.

b) Independentemente do tamanho da amostra, um estimador consistente sempre irá convergir para o verdadeiro valor da grandeza populacional.

c) A amplitude de uma amostra definirá se a média amostral poderá ser um estimador de máxima verossimilhança da média populacional.

d) Sendo â um estimador de máxima verossimilhança de um parâmetro a, então $(a)^{1/2} = (â)^{1/2}$.

e) Sendo a e b estimadores de um mesmo parâmetro cujas variâncias são simbolizadas por Var(a) e Var(b). Se Var(a) > Var(b), então é correto afirmar que a é um melhor estimador que b.

Comentários:

Em relação à alternativa A, um parâmetro (populacional) é **desconhecido**, sendo estimado a **partir** de uma amostra. Logo, a alternativa A está incorreta. A definição apresentada é de estatística.

Em relação à alternativa B, um estimador consistente converge para a grandeza populacional quando o tamanho da amostra tende ao infinito. Logo, a consistência **depende** do tamanho e a alternativa B está incorreta.

Em relação à alternativa C, o estimador de máxima verossimilhança da média amostral depende da **distribuição da população**. Logo, a alternativa C está incorreta.

Em relação à alternativa D, para um estimador de máxima verossimilhança â, que corresponda a um parâmetro a, a raiz do estimador corresponderá à raiz do parâmetro. Logo, a alternativa D está correta.

Em relação à alternativa E, se Var(a) > Var(b), nas condições indicadas na alternativa, então podemos concluir que b é um estimador melhor (mais eficiente).

Gabarito: D.

8. (CESPE 2018/PF – Papiloscopista) Em determinado município, o número diário X de registros de novos armamentos segue uma distribuição de Poisson, cuja função de probabilidade é expressa por $P(X = k) = \frac{e^{-M}.M^k}{k!}$ em que k = 0, 1, 2, ..., e M é um parâmetro.

	dia				
	1	2	3	4	5
realização da variável X	6	8	0	4	2

Considerando que a tabela precedente mostra as realizações da variável aleatória X em uma amostra aleatória simples constituída por cinco dias, julgue o item que segue.

A estimativa de máxima verossimilhança do desvio padrão da distribuição da variável X é igual a 2 registros por dia.

Comentários:

A estimativa de máxima verossimilhança para o parâmetro λ de uma distribuição de Poisson é dada por:

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{6 + 8 + 0 + 4 + 2}{5} = \frac{20}{5} = 4$$

Em uma distribuição de Poisson, a variância é igual a esse parâmetro, $V(X) = \lambda$. Logo, o desvio padrão é dado por:

$$\sigma = \sqrt{V(X)} = \sqrt{\lambda} = \sqrt{4} = 2$$

Gabarito: Certo.

FGV

9. (FGV/2014 – Prefeitura de Recife/PE) Avalie se as seguintes propriedades de um estimador de um certo parâmetro são desejáveis:

I. Ser não tendencioso para esse parâmetro.

II. Ter variância grande.

III. Ter erro quadrático médio grande.

Assinale:

a) se apenas a propriedade I estiver correta.

b) se apenas as propriedades I e II estiverem corretas.

c) se apenas as propriedades I e III estiverem corretas.

d) se apenas as propriedades II e III estiverem corretas.

e) se todas as propriedades estiverem corretas.

Comentários:

Em relação ao item I, uma das características desejáveis do estimador é ele ser não tendencioso. Logo, o item I está correto.

Em relação ao item II, o estimador deve ter a **menor** variância possível (eficiência). Logo, o item II está incorreto.

Em relação ao item III, o Erro Quadrático Médio, que é igual à variância do estimador, se ele for não tendencioso, deve ser o menor possível. Logo o item III está incorreto.

Gabarito: A

10. (FGV/2017 – IBGE) Dois estimadores, $\widehat{\theta}_1$ e $\widehat{\theta}_2$, para um parâmetro populacional θ, têm seus Erros Quadráticos Médios (EQM) dados por:

$$EQM(\widehat{\theta}_1) = \frac{3.\sigma^2}{n} + \left(\frac{\theta-1}{n}\right)^2 \text{ e } EQM(\widehat{\theta}_2) = \frac{\sqrt{n}.\sigma^2}{s} + \left(1 + \frac{\theta}{n}\right)^2$$

Com base apenas nas expressões, onde as primeiras parcelas são as variâncias, é correto concluir que:

a) $\widehat{\theta}_1$ é não tendencioso;

b) $\widehat{\theta}_1$ é um estimador consistente;

c) $\widehat{\theta}_2$ é assintoticamente não tendencioso;

d) $\widehat{\theta}_2$ não é um estimador consistente;

e) ambos são assintoticamente eficientes.

Comentários:

O Erro Quadrático Médio de um estimador é a soma da variância com o quadrado do viés:

$$EQM(\hat{\theta}) = V(\hat{\theta}) + [b(\hat{\theta})]^2$$

Como as primeiras parcelas das expressões são as variâncias, então as segundas parcelas são os vieses. Portanto, os estimadores são tendenciosos (alternativa A errada).

Para um estimador consistente, quando n é grande, a sua esperança tende ao parâmetro populacional e a sua variância tende a zero:

$$\lim_{n\to\infty} E(\hat{\theta}) = \theta$$

$$\lim_{n\to\infty} V(\hat{\theta}) = 0$$

O viés do primeiro estimador, $\left(\frac{\theta-1}{n}\right)^2$, é dividido por n e, por isso, diminui conforme n aumenta:

$$\lim_{n\to\infty} E(\widehat{\theta_1}) = \theta$$

A variância do primeiro estimador, $\frac{3.\sigma^2}{n}$, também é dividida por n, logo:

$$\lim_{n\to\infty} V(\widehat{\theta_1}) = 0$$

Portanto, o primeiro estimador é consistente (alternativa B correta).

Já em relação ao viés do segundo estimador, $\left(1+\frac{\theta}{n}\right)^2$, quando n aumenta, o viés tende a **1**, pois a expressão $\frac{\theta}{n}$ tende a 0. Por isso, o estimador é tendencioso, mesmo quando n tende a infinito (alternativa C errada).

Apesar de o estimador não apresentar as propriedades assintóticas associadas à consistência, não podemos concluir que o estimador é inconsistente, pois essas propriedades são condições suficientes, porém não necessárias (alternativa D errada).

A variância do segundo estimador, $\frac{\sqrt{n}.\sigma^2}{s}$, também não tende a zero quando n tende a infinito (pelo contrário, essa razão aumento com o crescimento de n). Logo, esse estimador não é assintoticamente eficiente (alternativa E errada).

Gabarito: B

11. (FGV/2019 – DPE-RJ) Sejam θ_1, θ_2 e θ_3 estimadores de um parâmetro populacional θ gerados a partir de uma amostra do tipo AAS de tamanho n.

Sabe-se ainda que θ_1 é eficiente quando comparada com uma certa classe de estimadores, que θ_2 e θ_3 são tendenciosos, mas θ_2 não é assintoticamente tendencioso. Então:

a) Se $\lim_{n\to\infty} V(\theta_3) \neq 0$, o estimador não é consistente;

b) Se $\theta^* = \frac{\theta_2+\theta_3}{2}$, então θ^* é um estimador inconsistente de θ;

c) Se $\theta^{**} = \theta_1 + \theta_2 - \theta_3$, então θ^{**} é não tendencioso;

d) Se $\lim_{n\to\infty} V(\theta_1) = 0$, então θ_1 é um estimador consistente;

e) Se $\lim_{n\to\infty} EQM(\theta_2) \neq 0$, então θ_2 não é consistente.

Comentários:

Essa questão pode ser mais rapidamente resolvida ao atentarmos para o fato de que **não** podemos concluir que um estimador seja **inconsistente** (ou **não consistente**), com base nas propriedades assintóticas de não tendência e eficiência, apenas que ele seja consistente, se apresentar tais propriedades.

Assim, eliminamos as alternativas A, B e E, pois elas concluem que o estimador é inconsistente. Em particular, a alternativa A afirma que o estimador θ_3 será inconsistente se não for assintoticamente eficiente. Já a alternativa B define o estimador θ^* como a média dos estimadores θ_2 e θ_3. Como não podemos concluir que θ_2 e θ_3 são inconsistentes, então também não podemos afirmar a sua média, θ^*, será inconsistente. Por fim, a alternativa E afirma que o estimador θ_3 será inconsistente se o seu erro quadrático médio não tender a zero quando n tende a infinito. Se o EQM desse estimador não tende a zero, a sua variância não tende a zero e/ou o seu viés não tende a zero. De qualquer forma, não podemos afirmar que o estimador é consistente, nem inconsistente.

Em relação à alternativa C, para saber se o estimador $\theta^{**} = \theta_1 + \theta_2 - \theta_3$ é tendencioso ou não, precisamos calcular o seu viés:

$$b(\theta^{**}) = E(e) = E(\theta^{**}) - \theta = E(\theta_1 + \theta_2 - \theta_3) - \theta = E(\theta_1) + E(\theta_2) - E(\theta_3) - \theta$$

O primeiro estimador, θ_1, é eficiente, sendo, portanto, não tendencioso, ou seja, $E(\theta_1) = \theta$:

$$b(\theta^{**}) = \theta + E(\theta_2) - E(\theta_3) - \theta = E(\theta_2) - E(\theta_3)$$

Ou seja, para que o estimador θ^{**} fosse não tendencioso seria necessário que os vieses dos estimadores θ_2 e θ_3 fossem iguais. Como isso não ocorre, então **não** podemos dizer que θ^{**} é não tendencioso. Logo, a alternativa C está incorreta.

Em relação à alternativa D, o enunciado afirma que θ_1 é eficiente, o que significa que esse estimador é não tendencioso. Então, se a sua variância tender a zero quando n tende a infinito (assintoticamente eficiente), o estimador θ_1 será consistente (alternativa certa).

Gabarito: D

12. (FGV/2017 – IBGE) Para estimar a média de certa população μ, desconhecida, partindo apenas de duas observações amostrais, cogita-se o emprego de um dos seguintes estimadores, onde X_1 e X_2 representam os indivíduos da amostra ex ante.

$$\widehat{\theta} = \frac{2}{3}X_1 + \frac{1}{3}X_2 \text{ e } \tilde{\theta} = \frac{2}{7}X_1 + \frac{4}{7}X_2$$

Sobre os estimadores, é correto afirmar que:

a) o estimador $\tilde{\theta}$ é mais eficiente do que $\hat{\theta}$;

b) o estimador $\tilde{\theta}$ subestima, em média, o valor verdadeiro de μ;

c) a tendenciosidade de $\hat{\theta}$, T ($\hat{\theta}$), é igual a μ/4;

d) o Erro Quadrático Médio de $\tilde{\theta}$, EQM($\tilde{\theta}$), é $\frac{20}{49}\cdot\sigma^2$;

e) a tendenciosidade de $\tilde{\theta}$, T($\tilde{\theta}$), é igual a $\frac{-\mu}{7}$.

Comentários:

Pelas expressões dos estimadores fornecidas, observamos que o primeiro estimador, $\hat{\theta}$, é uma espécie de média ponderada, em que o valor do primeiro elemento da amostra, X_1, tem peso 2, enquanto o primeiro elemento da amostra, X_2, tem peso 1. Já o segundo estimador, $\tilde{\theta}$, não faz uma média ponderada, deixando de fora uma parte dos valores (mais especificamente, 1/7).

O viés (ou tendenciosidade) do segundo estimador pode ser calculado como (lembre-se das propriedades da esperança):

$$b(\tilde{\theta}) = E(e) = E(\tilde{\theta}) - \mu = E\left(\frac{2}{7}X_1 + \frac{4}{7}X_2\right) = \frac{2}{7}E(X_1) + \frac{4}{7}E(X_2)$$

Como a amostra segue a mesma distribuição da população, então $E(X_1) = E(X) = \mu$:

$$b(\tilde{\theta}) = \frac{2}{7}\mu + \frac{4}{7}\mu - \mu = \frac{6}{7}\mu - \mu = -\frac{\mu}{7}$$

Portanto, a alternativa E está correta.

Em relação a alternativa B, não podemos dizer que o estimador subestima o valor verdadeiro, pois o valor de μ é fixo. O que ocorre é que, em média, o estimador é inferior ao parâmetro (viés negativo).

Para calcular o EQM desse estimador, precisamos da sua variância (lembre-se das propriedades da variância). Supondo as amostras independentes, temos:

$$V(\tilde{\theta}) = V\left(\frac{2}{7}X_1 + \frac{4}{7}X_2\right) = \left(\frac{2}{7}\right)^2 V(X_1) + \left(\frac{4}{7}\right)^2 V(X_2) = \frac{4}{49}V(X_1) + \frac{16}{49}V(X_2)$$

Como a amostra segue a mesma distribuição da população, então $V(X_1) = V(X) = \sigma^2$:

$$V(\tilde{\theta}) = \frac{4}{49}\sigma^2 + \frac{16}{49}\sigma^2 = \frac{20}{49}\sigma^2$$

O EQM do estimador é a soma da sua variância com o quadrado do seu viés:

$$EQM(\tilde{\theta}) = V(\tilde{\theta}) + [b(\tilde{\theta})]^2 = \frac{20}{49}\sigma^2 + \left(-\frac{\mu}{7}\right)^2 = \frac{20}{49}\sigma^2 + \frac{\mu^2}{49} = \frac{20\sigma^2 + \mu^2}{49}$$

Portanto, a alternativa D está errada (o valor fornecido é o valor da variância do estimador).

Com relação ao estimador $\hat{\theta}$, o seu viés é dado por:

$$b(\hat{\theta}) = E(e) = E(\hat{\theta}) - \mu = E\left(\frac{2}{3}X_1 + \frac{1}{3}X_2\right) - \mu = \frac{2}{3}E(X_1) + \frac{1}{3}E(X_2) - \mu$$

$$b(\hat{\theta}) = \frac{2}{3}\mu + \frac{1}{3}\mu - \mu = \mu - \mu = 0$$

Portanto, o estimador $\hat{\theta}$ é não tendencioso e a alternativa C está errada. Ademais, a alternativa A está errada porque esse estimador é não tendencioso, enquanto $\tilde{\theta}$ é tendencioso. Por esse motivo, $\tilde{\theta}$ não pode ser mais eficiente que $\hat{\theta}$.

Gabarito: E

13. (FGV/2015 – TJ-BA) Para estimar a média populacional de uma distribuição, com base em uma amostra de tamanho n = 3, são propostos os seguintes estimadores:

$$\hat{\mu} = \frac{3}{7}X_1 + \frac{5}{7}X_2 - \frac{1}{7}X_3$$

$$\tilde{\mu} = \frac{1}{3}X_1 + \frac{2}{5}X_2 + \frac{1}{4}X_3$$

$$\bar{\bar{\mu}} = \frac{1}{12}X_1 + \frac{1}{4}X_2 + \frac{2}{3}X_3$$

Sobre esses estimadores é correto afirmar que:

a) os estimadores são todos não tendenciosos;

b) apenas o estimador $\hat{\mu}$ é não viesado e eficiente com relação aos demais;

c) exceto $\bar{\bar{\mu}}$, os outros estimadores são não tendenciosos e consistentes;

d) dentre os três estimadores sugeridos, o que apresenta a menor variância é $\tilde{\mu}$, mas não é o mais eficiente;

e) o erro quadrático médio de $\bar{\bar{\mu}}$ difere da sua variância a razão de 1/3 da variância populacional

Comentários:

Vamos primeiro calcular o viés dos 3 estimadores. Para o primeiro estimador, temos:

$$b(\hat{\mu}) = E(e) = E(\hat{\mu}) - \mu = E\left(\frac{3}{7}X_1 + \frac{5}{7}X_2 - \frac{1}{7}X_3\right) - \mu$$

$$b(\hat{\mu}) = \frac{3}{7}E(X_1) + \frac{5}{7}E(X_2) - \frac{1}{7}E(X_3) - \mu = \frac{3}{7}\mu + \frac{5}{7}\mu - \frac{1}{7}\mu - \mu = \mu - \mu = 0$$

Ou seja, o primeiro estimador é não tendencioso. Para o segundo estimador, temos:

$$b(\tilde{\mu}) = E(e) = E(\tilde{\mu}) - \mu = E\left(\frac{1}{3}X_1 + \frac{2}{5}X_2 + \frac{1}{4}X_3\right) - \mu$$

$$b(\tilde{\mu}) = \frac{1}{3}E(X_1) + \frac{2}{5}E(X_2) + \frac{1}{4}E(X_3) - \mu = \frac{1}{3}\mu + \frac{2}{5}\mu + \frac{1}{4}\mu - \mu = \frac{20 + 24 + 15 - 60}{60}\mu = \frac{-\mu}{60}$$

Portanto, o segundo estimador é tendencioso. Para o terceiro estimador, temos:

$$b(\bar{\bar{\mu}}) = E(e) = E(\bar{\bar{\mu}}) - \mu = E\left(\frac{1}{12}X_1 + \frac{1}{4}X_2 + \frac{2}{3}X_3\right) - \mu$$

$$b(\bar{\bar{\mu}}) = \frac{1}{12}E(X_1) + \frac{1}{4}E(X_2) + \frac{2}{3}E(X_3) - \mu = \frac{1}{12}\mu + \frac{1}{4}\mu + \frac{2}{3}\mu - \mu = \frac{1 + 3 + 8 - 12}{12}\mu = 0$$

Logo, o terceiro estimador é não tendencioso.

Com isso, vemos que a alternativa A, B e C estão incorretas. Além disso, como estimador $\bar{\bar{\mu}}$ é não tendencioso, o seu EQM é **igual** à sua variância. Portanto, a alternativa E também está incorreta.

Para verificar a alternativa D, vamos calcular as variâncias dos estimadores, supondo as amostras independentes:

$$V(\hat{\mu}) = V\left(\frac{3}{7}X_1 + \frac{5}{7}X_2 - \frac{1}{7}X_3\right) = \frac{9}{49}V(X_1) + \frac{25}{49}V(X_2) + \frac{1}{49}V(X_3) = \frac{9}{49}\sigma^2 + \frac{25}{49}\sigma^2 + \frac{1}{49}\sigma^2$$

$$V(\hat{\mu}) = \frac{35}{49}\sigma^2 \cong 0{,}7\sigma^2$$

$$V(\tilde{\mu}) = V\left(\frac{1}{3}X_1 + \frac{2}{5}X_2 + \frac{1}{4}X_3\right) = \frac{1}{9}V(X_1) + \frac{4}{25}V(X_2) + \frac{1}{16}V(X_3) = \frac{1}{9}\sigma^2 + \frac{4}{25}\sigma^2 + \frac{1}{16}\sigma^2 =$$

$$V(\tilde{\mu}) = \frac{400 + 144 + 225}{3600}\sigma^2 = \frac{378}{3600}\sigma^2 = \frac{63}{600}\sigma^2 \cong 0{,}1\sigma^2$$

$$V(\bar{\mu}) = V\left(\frac{1}{12}X_1 + \frac{1}{4}X_2 + \frac{2}{3}X_3\right) = \frac{1}{144}V(X_1) + \frac{1}{16}V(X_2) + \frac{4}{9}V(X_3) = \frac{1}{144}\sigma^2 + \frac{1}{16}\sigma^2 + \frac{4}{9}\sigma^2$$

$$V(\bar{\mu}) = \frac{1 + 9 + 64}{144}\sigma^2 = \frac{74}{144}\sigma^2 \cong 0{,}5\sigma^2$$

Logo, o estimador de menor variância é $\tilde{\mu}$, porém ele não é o mais eficiente, por ser tendencioso. Assim, a alternativa D está correta.

Gabarito: D

14. (FGV 2018/AL-RO) Suponha que X_1, X_2, ..., X_n seja uma amostra aleatória simples de uma variável aleatória populacional qualquer com média μ e variância finita. Considere os seguintes estimadores de μ:

$T_1 = X_1$

$T_2 = X_1 + X_2 + X_3 - X_4 - X_5$.

$T_3 = (X_1 + X_2 + X_3)/3$.

$T_4 = X_1 - X_2$.

$T_5 = (X_1 + X_2 + X_3 + X_4 + X_5)/5$.

São estimadores não tendenciosos de μ:

a) T_1 e T_2, somente.

b) T_1, T_2 e T_3, somente.

c) T_1, T_2, T_3 e T_5, somente.

d) T_2, T_3, T_4 e T_5, somente.

e) T_1, T_2, T_3, T_4 e T_5.

Comentários:

Para estimadores não tendenciosos, a esperança do estimador é igual ao parâmetro a ser estimado:

$$E(T) = \mu$$

A esperança do primeiro estimador, $T_1 = X_1$, é:

$$E(T_1) = E(X_1)$$

Sabemos que a amostra segue a mesma distribuição da população, logo, $E(X_1) = E(X) = \mu$:

$$E(T_1) = \mu$$

Portanto, o primeiro estimador é não tendencioso. A esperança do segundo estimador, $T_2 = X_1 + X_2 + X_3 - X_4 - X_5$, é (considere as propriedades da esperança):

$$E(T_2) = E(X_1 + X_2 + X_3 - X_4 - X_5) = E(X_1) + E(X_2) + E(X_3) - E(X_4) - E(X_5)$$

Como $E(X_i) = E(X) = \mu$, temos:

$$E(T_2) = \mu + \mu + \mu - \mu - \mu = \mu$$

Logo, o segundo estimador é não tendencioso. Para o terceiro estimador, $T_3 = \frac{X_1+X_2+X_3}{3}$, temos:

$$E(T_3) = E\left(\frac{X_1 + X_2 + X_3}{3}\right) = \frac{1}{3}[E(X_1) + E(X_2) + E(X_3)] = \frac{1}{3}[\mu + \mu + \mu] = \frac{3\mu}{3} = \mu$$

Ou seja, o terceiro estimador também é não tendencioso. Para o quarto estimador, $T_4 = X_1 - X_2$:

$$E(T_4) = E(X_1 - X_2) = E(X_1) - E(X_2) = \mu - \mu = 0$$

Como a esperança do estimador é diferente do parâmetro populacional, μ, o quarto estimador é tendencioso. Para o quinto estimador, $T_5 = \frac{X_1+X_2+X_3+X_4+X_5}{5}$:

$$E(T_5) = E\left(\frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}\right) = \frac{1}{5}[E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5)]$$

$$E(T_5) = \frac{1}{5}[\mu + \mu + \mu + \mu + \mu] = \frac{5\mu}{5} = \mu$$

Assim, concluímos que o quinto estimador é não tendencioso.

Gabarito: C

15. (FGV 2018/AL-RO) Suponha que X_1, X_2, ..., X_n seja uma amostra aleatória simples de uma variável aleatória populacional qualquer com média μ e variância finita. Considere os seguintes estimadores de μ:

$T_1 = X_1$
$T_2 = X_1 + X_2 + X_3 – X_4 – X_5$.
$T_3 = (X_1 + X_2 + X_3)/3$.
$T_4 = X_1 – X_2$.
$T_5 = (X_1 + X_2 + X_3 + X_4 + X_5)/5$.

O estimador não tendencioso de variância uniformemente mínima de μ é:

a) T_1

b) T_2

c) T_3

d) T_4

e) T_5

Comentários:

Na questão anterior, vimos que os estimadores não tendenciosos são T_1, T_2, T_3, T_5. Agora, vamos calcular as suas variâncias, lembrando que as amostras seguem a mesma distribuição da população, logo $V(X_i) = V(X)$ e supondo as amostras independentes. Para o primeiro estimador, $T_1 = X_1$:

$$V(T_1) = V(X_1) = V(X)$$

Para o segundo estimador, $T_2 = X_1 + X_2 + X_3 - X_4 - X_5$, temos:

$$V(T_2) = V(X_1 + X_2 + X_3 - X_4 - X_5) = V(X_1) + V(X_2) + V(X_3) + V(X_4) + V(X_5) = 5.V(X)$$

Para o terceiro estimador, $T_3 = \frac{X_1+X_2+X_3}{3}$:

$$V(T_3) = V\left(\frac{X_1 + X_2 + X_3}{3}\right) = \frac{1}{9}[V(X_1) + V(X_2) + V(X_3)] = \frac{3.V(X)}{9} = \frac{V(X)}{3}$$

E para o quinto estimador, $T_5 = \frac{X_1+X_2+X_3+X_4+X_5}{5}$:

$$V(T_5) = V\left(\frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}\right) = \frac{1}{25}[V(X_1) + V(X_2) + V(X_3) + V(X_4) + V(X_5)]$$

$$V(T_5) = \frac{5.V(X)}{25} = \frac{V(X)}{5}$$

Assim, o estimador de menor variância é o quinto estimador, T_5.

Gabarito: E

16. (FGV 2014/DPE-RJ) Sejam $\widehat{\theta_1}$ e $\widehat{\theta_2}$ dois estimadores pontuais, ambos não tendenciosos e igualmente eficientes do parâmetro θ (Var($\widehat{\theta_1}$) = Var($\widehat{\theta_2}$)) , sendo que a covariância entre eles é igual a ½.Var($\widehat{\theta_1}$). Então, também é não tendencioso e mais eficiente o estimador

a) $\frac{2}{3}\widehat{\theta_1} + \frac{1}{3}\widehat{\theta_2}$

b) $\frac{1}{5}\widehat{\theta_1} + \frac{3}{4}\widehat{\theta_2}$

c) $\frac{3}{5}\widehat{\theta_1} + \frac{4}{7}\widehat{\theta_2}$

d) $\frac{\widehat{\theta_1}+\widehat{\theta_2}}{2}$

e) $\frac{3}{8}\widehat{\theta_1} + \frac{7}{8}\widehat{\theta_2}$

Comentários:

Esta questão envolve a combinação de estimadores e pergunta pelo estimador não tendencioso e mais eficiente (ou seja, com menor variância).

Em relação à alternativa A, vamos calcular a esperança do estimador $\frac{2}{3}\widehat{\theta_1} + \frac{1}{3}\widehat{\theta_2}$ para sabermos se o estimador é tendencioso ou não:

$$E\left(\frac{2}{3}\widehat{\theta_1} + \frac{1}{3}\widehat{\theta_2}\right) = \frac{2}{3}E(\widehat{\theta_1}) + \frac{1}{3}E(\widehat{\theta_2})$$

Como $\widehat{\theta_1}$ e $\widehat{\theta_2}$ são não tendenciosos, então a sua esperança é $E(\widehat{\theta_1}) = E(\widehat{\theta_2}) = \theta$:

$$E\left(\frac{2}{3}\widehat{\theta_1} + \frac{1}{3}\widehat{\theta_2}\right) = \frac{2}{3}\theta + \frac{1}{3}\theta = \theta$$

Portanto, o estimador é não tendencioso. Agora, vamos calcular a sua variância, utilizando a fórmula da variância da soma:

$$V(X,Y) = V(X) + V(Y) + 2.Cov(X,Y)$$

$$V\left(\frac{2}{3}\widehat{\theta_1} + \frac{1}{3}\widehat{\theta_2}\right) = V\left(\frac{2}{3}\widehat{\theta_1}\right) + V\left(\frac{1}{3}\widehat{\theta_2}\right) + 2.Cov\left(\frac{2}{3}\widehat{\theta_1}, \frac{1}{3}\widehat{\theta_2}\right)$$

Agora, aplicamos as propriedades da variância, $V(k.X) = k^2V(X)$, e da covariância:

$$Cov(l.X, k.Y) = l \times k \times Cov(X,Y)$$

$$V\left(\frac{2}{3}\widehat{\theta_1} + \frac{1}{3}\widehat{\theta_2}\right) = \frac{4}{9}V(\widehat{\theta_1}) + \frac{1}{9}V(\widehat{\theta_2}) + 2 \times \frac{2}{3} \times \frac{1}{3}.Cov(\widehat{\theta_1}, \widehat{\theta_2})$$

O enunciado informa que as variâncias são iguais, $V(\widehat{\theta_1}) = V(\widehat{\theta_2})$, e que a covariância é $Cov(\widehat{\theta_1}, \widehat{\theta_2}) = V(\widehat{\theta_1})/2$, então:

$$V\left(\frac{2}{3}\widehat{\theta_1} + \frac{1}{3}\widehat{\theta_2}\right) = \frac{5}{9}V(\widehat{\theta_1}) + \frac{4}{9} \times \frac{V(\widehat{\theta_1})}{2} = \boldsymbol{\frac{7}{9}V(\widehat{\theta_1}) \cong 0,78V(\widehat{\theta_1})}$$

Vejamos o estimador da alternativa B, $\frac{1}{5}\widehat{\theta_1} + \frac{3}{4}\widehat{\theta_2}$. A sua esperança é:

$$E\left(\frac{1}{5}\widehat{\theta_1} + \frac{3}{4}\widehat{\theta_2}\right) = \frac{1}{5}E(\widehat{\theta_1}) + \frac{3}{4}E(\widehat{\theta_2}) = \frac{1}{5}\theta + \frac{3}{4}\theta = \frac{4+15}{20}\theta = \frac{19}{20}\theta$$

Como a esperança do estimador é diferente do parâmetro θ, então o estimador é tendencioso e, por isso, a resposta não pode ser a alternativa B.

Em relação ao estimador na alternativa C, $\frac{3}{5}\widehat{\theta_1} + \frac{4}{7}\widehat{\theta_2}$, a esperança é:

$$E\left(\frac{3}{5}\widehat{\theta_1} + \frac{4}{7}\widehat{\theta_2}\right) = \frac{3}{5}E(\widehat{\theta_1}) + \frac{4}{7}E(\widehat{\theta_2}) = \frac{3}{5}\theta + \frac{4}{7}\theta = \frac{21+20}{35}\theta = \frac{41}{35}\theta$$

Como a esperança do estimador é diferente do parâmetro θ, então o estimador é tendencioso e a resposta não pode ser a alternativa C.

A esperança do estimador da alternativa D, $\frac{\widehat{\theta_1}+\widehat{\theta_2}}{2}$, é:

$$E\left(\frac{\widehat{\theta_1}+\widehat{\theta_2}}{2}\right)=\frac{E(\widehat{\theta_1}+\widehat{\theta_2})}{2}=\frac{\theta+\theta}{2}=\theta$$

Portanto, o estimador é não tendencioso. Agora, vamos calcular a sua variância:

$$V\left(\frac{\widehat{\theta_1}+\widehat{\theta_2}}{2}\right)=V\left(\frac{\widehat{\theta_1}}{2}\right)+V\left(\frac{\widehat{\theta_2}}{2}\right)+2.Cov\left(\frac{\widehat{\theta_1}}{2},\frac{\widehat{\theta_2}}{2}\right)$$

$$V\left(\frac{\widehat{\theta_1}+\widehat{\theta_2}}{2}\right)=\frac{1}{4}V(\widehat{\theta_1})+\frac{1}{4}V(\widehat{\theta_2})+2\times\frac{1}{2}\times\frac{1}{2}Cov(\widehat{\theta_1},\widehat{\theta_2})$$

Sabendo que $V(\widehat{\theta_1})=V(\widehat{\theta_2})$ e $Cov(\widehat{\theta_1},\widehat{\theta_2})={}^{V(\widehat{\theta_1})}/_{2}$, então:

$$V\left(\frac{\widehat{\theta_1}+\widehat{\theta_2}}{2}\right)=\frac{1}{4}V(\widehat{\theta_1})+\frac{1}{4}V(\widehat{\theta_1})+\frac{1}{2}\times\frac{1}{2}V(\widehat{\theta_1})=\mathbf{\frac{3}{4}V(\widehat{\theta_1})=0,75V(\widehat{\theta_1})}$$

Em relação à alternativa E, $\frac{3}{8}\widehat{\theta_1}+\frac{7}{8}\widehat{\theta_2}$, a esperança é:

$$E\left(\frac{3}{8}\widehat{\theta_1}+\frac{7}{8}\widehat{\theta_2}\right)=\frac{3}{8}E(\widehat{\theta_1})+\frac{7}{8}E(\widehat{\theta_2})=\frac{3}{8}\theta+\frac{7}{8}\theta=\frac{10}{8}\theta=\frac{5}{4}\theta$$

Como a esperança do estimador é diferente do parâmetro θ, então o estimador é tendencioso e a resposta não pode ser a alternativa E.

Portanto, os estimadores não tendenciosos são os da alternativa A e D. Como a variância do estimador da alternativa A, $\mathbf{\frac{7}{9}V(\widehat{\theta_1})\cong 0,78V(\widehat{\theta_1})}$, é MAIOR que a variância do estimador da alternativa D, $\mathbf{\frac{3}{4}V(\widehat{\theta_1})=0,75V(\widehat{\theta_1})}$, então o estimador mais eficiente é o da alternativa D.

Gabarito: D

17. (FGV 2014/DPE-RJ) Seja o estimador $\widehat{\theta}$ de um parâmetro populacional θ tal que $EQM(\widehat{\theta})-Var(\widehat{\theta})=\left(k-\frac{1}{n}\right)^2$, onde k (≠ zero) é uma constante que depende do verdadeiro valor de θ e n é o tamanho da amostra. Então, o estimador será

a) assintoticamente eficiente se $\lim_{n\to\infty} Var(\hat{\theta})=0$

b) assintoticamente tendencioso.

c) assintoticamente tendencioso, subestimando o parâmetro θ.

d) assintoticamente tendencioso, superestimando o parâmetro θ.

e) consistente, desde que $\lim_{n\to\infty} EQM(\hat{\theta})=(k)^2$

Comentários:

O erro quadrático médio (EQM) é a soma da variância do estimador e o quadrado do seu viés:

$$EQM(\hat{\theta}) = Var(\hat{\theta}) + [b(\hat{\theta})]^2$$

Então a diferença entre o EQM e a variância fornecida no enunciado é justamente o quadrado do viés do estimador:

$$[b(\hat{\theta})]^2 = \left(k - \frac{1}{n}\right)^2$$

$$|b(\hat{\theta})| = \left|k - \frac{1}{n}\right|$$

Podemos observar que quando n fica muito grande (tende a infinito), o viés tende a k, que é diferente de zero, conforme o enunciado. Portanto, o estimador é assintoticamente tendencioso (alternativa B correta).

Sendo assintoticamente tendencioso, ele não será eficiente, mesmo que a variância seja nula para $n \to \infty$, e também não pode ser consistente (alternativas A e E incorretas). Também não sabemos se o viés é positivo ou negativo (alternativas C e D incorretas).

Gabarito: B

18. (FGV 2018/AL-RO) Se X_1, X_2, ..., X_n é uma amostra aleatória simples de uma distribuição Bernoulli (p), então o estimador de máxima verossimilhança da variância populacional é

a) $\bar{X}$

b) $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$

c) $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$

d) $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{2n}$

e) $\bar{X}(1-\bar{X})$

Comentários:

O estimador de máxima verossimilhança da proporção populacional é a proporção amostral, que é calculada da mesma forma que a média amostral:

$$\hat{p} = \frac{\sum X_i}{n} = \bar{X}$$

Sabendo que a variância da distribuição de Bernoulli é $V(X) = p.q = p.(1-p)$, então a estimativa da variância é:

$$\hat{p}.(1-\hat{p}) = \bar{X}.(1-\bar{X})$$

Gabarito: E

19. (FGV 2018/AL-RO) Se $X_1, X_2, ..., X_n$ é uma amostra aleatória simples de uma distribuição exponencial com parâmetro θ, ou seja,

$f(x|\theta) = \theta e^{-\theta x}$, $\theta > 0$,

então, o estimador de θ pelo método dos momentos é

a) $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

b) $\frac{1}{\bar{X}} = \frac{n}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$

c) $\bar{X}^2$

d) $\sum_{i=1}^{n} X_i$

e) $\frac{\sum_{i=1}^{n} X_i}{2}$

Comentários:

Para uma distribuição exponencial, a **média** é o inverso do parâmetro (que normalmente chamamos de λ, mas o enunciado está chamando de θ):

$$E(X) = \frac{1}{\theta}$$

O estimador de máxima verossimilhança para essa distribuição também considera a média amostral, sendo calculado pelo inverso da média:

$$\hat{\theta} = \frac{1}{\bar{X}} = \frac{n}{\sum X_i}$$

Gabarito: B

FCC

20. (FCC 2015/DPE-SP) Dois estimadores não viesados $E_1 = mX + (m - 1)Y - (2m - 2)Z$ e $E_2 = 1,5X - Y + 0,5Z$ são utilizados para estimar a média μ de uma população normal e variância σ^2 diferente de zero. O parâmetro m é um número real e (X, Y, Z) corresponde a uma amostra aleatória, com reposição, da população. Se E_1 é mais eficiente que E_2, então

a) m < 1/6 ou m > 2

b) 3/2 < m < 2

c) 1/6 < m < 2

d) 1/6 < m < 3/2

e) 1 < m < 2

Comentário

Sendo dois estimadores não viesados, o estimador mais eficiente é aquele com a **menor** variância. A variância de E_1 é:

$$V(E_1) = V[m.X + (m-1).Y - (2m-2).Z]$$

Sendo X, Y e Z uma amostra aleatória extraída com reposição, então essas variáveis são independentes. Então, pelas propriedades da variância, temos:

$$V(E_1) = m^2.V(X) + (m-1)^2.V(Y) + (2m-2)^2.V(Z)$$

Sabendo que a distribuição da amostra é igual à distribuição da população, temos $V(X) = V(Y) = V(Z) = \sigma^2$:

$$V(E_1) = m^2.\sigma^2 + (m-1)^2.\sigma^2 + (2m-2)^2.\sigma^2$$

$$V(E_1) = \sigma^2[m^2 + m^2 - 2m + 1 + 4.m^2 - 8m + 4)$$

$$V(E_1) = \sigma^2(6.m^2 - 10m + 5)$$

Segundo o mesmo raciocínio, a variância de E_2 é dada por:

$$V(E_2) = V[1{,}5.X - Y + 0{,}5.Z] = 2{,}25.V(X) + V(Y) + 0{,}25.V(Z)$$

$$V(E_2) = 2{,}25.\sigma^2 + \sigma^2 + 0{,}25.\sigma^2 = 3{,}5.\sigma^2$$

Para que E_1 seja mais eficiente que E_2, é necessário ter:

$$V(E_1) < V(E_2)$$

$$\sigma^2.(6.m^2 - 10m + 5) < 3{,}5.\sigma^2$$

$$6.m^2 - 10m + 5 < 3{,}5$$

$$6.m^2 - 10m + 1{,}5 < 0$$

Agora, buscamos as raízes da equação $6.m^2 - 10m + 1{,}5$:

$$\Delta = b^2 - 4.a.c = (-10)^2 - 4 \times 6 \times 1{,}5 = 100 - 36 = 64$$

$$m = \frac{-b \pm \sqrt{\Delta}}{2.a} = \frac{10 \pm 8}{2 \times 6}$$

$$m_1 = \frac{2}{12} = \frac{1}{6}; \qquad m_2 = \frac{18}{12} = \frac{3}{2}$$

Como a equação $6.m^2 - 10m + 1{,}5$ tem a sua concavidade voltada para cima, ela será negativa para valores entre as duas raízes:

$$\frac{1}{6} < m < \frac{3}{2}$$

Gabarito: D

21. (FCC 2013/TRT 5ª Região) Em 20 experiências de 4 provas cada uma, obteve-se a seguinte distribuição:

x_i	0	1	2	3
n_i	10	8	1	1

Observação: n_i é o número de experiências nas quais um determinado acontecimento ocorreu x_i vezes.

Admitindo que este acontecimento trata de uma variável aleatória X obedecendo a uma distribuição binomial $P(x) = C_m^x . p^x (1-p)^{m-x}$, em que x é o número de ocorrências de um certo acontecimento em m provas, tem-se, com base nas 20 experiências, que a estimativa pontual de p pelo método da máxima verossimilhança é

a) 65,00%

b) 50,00%

c) 48,75%

d) 32,50%

e) 16,25%

Comentário

O estimador de máxima verossimilhança para a proporção populacional é a proporção amostral. Pela tabela, verificamos que dos 20 x 4 = 80 ensaios (de Bernoulli), 10 resultaram em 0 sucesso, 8 resultaram em 1 sucesso, 1 resultou em 2 sucessos e 1 resultou em 3 sucessos. Assim, a proporção de sucessos dessa amostra foi de:

$$\hat{p} = \frac{10 \times 0 + 8 \times 1 + 1 \times 2 + 1 \times 3}{20 \times 4} = \frac{8 + 2 + 3}{80} = \frac{13}{80} = 0{,}1624 = 16{,}25\%$$

Gabarito: E

22. (FCC 2016/TRT 20ª Região) Em uma sala estão presentes algumas pessoas e somente duas delas têm nível superior, sendo que o número de pessoas sem nível superior é desconhecido e sabendo-se apenas que é um número par. Foram selecionadas, desta sala, aleatoriamente, com reposição, 4 pessoas verificando-se que 3 delas não têm nível superior. Com base nesta seleção e utilizando o método da máxima verossimilhança encontra-se a estimativa do número de pessoas sem nível superior. Com isto, o número estimado total de pessoas presentes na sala é igual a

a) 12

b) 6

c) 10

d) 14

e) 8

Comentário

O estimador de máxima verossimilhança para a proporção populacional é a proporção amostral. O enunciado informa que 3 pessoas, dentre 4 selecionadas na amostra, não têm nível superior, ou seja, apenas 1 pessoa da amostra possui nível superior. Logo, a proporção de pessoas com nível superior é:

$$\hat{p} = \frac{1}{4}$$

Sabendo que 2 pessoas na sala possuem nível superior e que esse quantitativo corresponde a 1/4 do total, então o número total de pessoas é:

$$T \times \frac{1}{4} = 2$$

$$T = 2 \times 4 = 8$$

Gabarito: E

QUESTÕES COMENTADAS - MULTIBANCAS

Estimação Intervalar

CEBRASPE

1. (CESPE 2020/TJ-PA) Para determinado experimento, uma equipe de pesquisadores gerou 20 amostras de tamanho $n = 25$ de uma distribuição normal, com média $\mu = 5$ e desvio padrão $\sigma = 3$. Para cada amostra, foi montado um intervalo de confiança com coeficiente de 0,95 (ou 95%). Com base nessas informações, julgue os itens que se seguem.

I. Os intervalos de confiança terão a forma $\beta_i \pm 1,176$, em que β_i é a média da amostra i.
II. Para todos os intervalos de confiança, $\beta_i + \varepsilon \geq \mu \geq \beta_i - \varepsilon$, sendo ε a margem de erro do estimador.
III. Se o tamanho da amostra fosse maior, mantendo-se fixos os valores do desvio padrão e do nível de confiança, haveria uma redução da margem de erro ε.

Assinale a opção correta.

a) Apenas o item II está certo.

b) Apenas os itens I e II estão certos.

c) Apenas os itens I e III estão certos.

d) Apenas os itens II e III estão certos.

e) Todos os itens estão certos.

Comentários:

Vamos analisar as afirmativas: Item I – Certo. Vamos usar a fórmula do intervalo de confiança (IC) para uma amostra menor ou igual a 30:

$$\bar{X} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

$\bar{X} \rightarrow média\ amostral$
$Z_{\alpha/2} \rightarrow Z_{0,025} = 1,96 \rightarrow escore\ associado\ ao\ nível\ de\ confiança\ (1-\alpha)$
$\sigma = 3 \rightarrow desvio\ padrão$
$n = 25 \rightarrow tamanho\ da\ amostra$

$$\beta_i \pm 1,96 \times \frac{3}{5}$$

$$\beta_i \pm 1,176$$

Item II – Errado. O enunciado nos diz que o intervalo de confiança é de 95%, sendo assim 5% estará fora desse intervalo, logo, não é verdadeiro o que se afirma em II.

Item III – Certo. Quanto maior o tamanho da amostra, menor a margem de erro.

Gabarito: C.

2. (CESPE 2020/TJ-PA). Ao analisar uma amostra aleatória simples composta de 324 elementos, um pesquisador obteve, para os parâmetros média amostral e variância amostral, os valores 175 e 81, respectivamente.

Nesse caso, um intervalo de 95% de confiança de μ é dado por

a) (166,18; 183,82).

b) (174,02; 175,98).

c) (174,51; 175,49).

d) (163,35; 186,65).

e) (174,1775; 175,8225).

Comentários:

Vamos listar as informações do enunciado:

$$Média\ amostral\ \rightarrow \bar{X} = 175$$
$$Escore\ associado\ ao\ nível\ de\ confiança \rightarrow z = 1{,}96$$
$$Desvio\ padrão \rightarrow \sigma = \sqrt{81} = 9$$
$$Tamanho\ da\ amostra \rightarrow n = 324$$

Agora, podemos apenas aplicar na fórmula do intervalo de confiança para a média:

$$\bar{X} \pm z \times \frac{\sigma}{\sqrt{n}}$$

$$175 \pm 1{,}96 \times \frac{9}{\sqrt{324}}$$

$$175 \pm 1{,}96 \times \frac{9}{18}$$

$$175 \pm 1{,}96 \times 0{,}5$$

$$175 \pm 0{,}98$$

$$175 \pm 0{,}98$$

Logo, o intervalo é $(175 - 0{,}98 = 174{,}02;\ 175 + 0{,}98 = 175{,}98)$.

Gabarito: B.

3. (CESPE 2020/TJ-PA). A respeito dos intervalos de confiança, julgue os próximos itens.

I. Um intervalo de confiança tem mais valor do que uma estimativa pontual única, pois uma estimativa pontual não fornece nenhuma informação sobre o grau de precisão da estimativa.
II. Um intervalo de confiança poderá ser reduzido se o nível de confiança for menor e o valor da variância populacional for maior.
III. No cálculo de um intervalo de confiança para a média, deve-se utilizar a distribuição t em lugar da distribuição normal quando a variância populacional é desconhecida e o número de observações é inferior a 30.

Assinale a opção correta.

a) Apenas o item II está certo.

b) Apenas os itens I e II estão certos.

c) Apenas os itens I e III estão certos.

d) Apenas os itens II e III estão certos.

e) Todos os itens estão certos.

Comentários:

Item I – Correto. Uma estimativa pontual não nos dá um parâmetro para definir o quão precisa é a estimativa. Quando temos um intervalo de confiança, quanto maior a amostra, menor é o erro e maior é a precisão.

Item II – Errado. Quanto maior o nível de confiança, menor será a margem de erro e, consequentemente, menor será o intervalo de confiança.

Item III – Certo. Para $n < 30$ e com variância populacional desconhecida utiliza-se t-Student. Se aumentar a amostra com variância desconhecida, ou dada a variância independentemente do tamanho da amostra utiliza-se a distribuição normal.

Gabarito: C.

4. (CESPE 2020/TJ PA) Na construção de um intervalo de confiança para a média, conhecida a variância, considerando o intervalo na forma $[x + \varepsilon;\ x - \varepsilon]$, sendo x o valor do estimador da média e ε a semi-amplitude do intervalo de confiança ou, como é mais popularmente conhecida, a margem de erro do intervalo de confiança. Considere que, para uma determinada peça automotiva, um lote de 100 peças tenha apresentado espessura média de 4,561 polegada, com desvio padrão de 1,125 polegada. Um intervalo de confiança de 95% para a média apresentou limite superior de 4,7815 e limite inferior de 4,3405. Nessa situação, a margem de erro do intervalo é de, aproximadamente,

a) $\varepsilon = 0{,}4410$.

b) $\varepsilon = 0{,}3436$.

c) $\varepsilon = 0{,}2205$.

d) $\varepsilon = 0{,}1125$.

e) $\varepsilon = 0{,}1103$.

Comentários:

Vamos listar as informações do enunciado:

$Z_{\alpha/2} \rightarrow escore\ associado\ ao\ nível\ de\ confiança\ (1-\alpha)$
$nível\ de\ confiança\ de\ 95\%\ na\ tabela\ normal \rightarrow Z_{0{,}025} = 1{,}96$
$Desvio\ padrão \rightarrow \sigma = 1{,}125$
$Tamanho\ da\ amostra \rightarrow n = 100$

Agora, podemos apenas aplicar na fórmula da margem do erro no intervalo de confiança para a média:

$$\varepsilon = Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

$$\varepsilon = 1{,}96 \times \frac{1{,}125}{\sqrt{100}}$$

$$\varepsilon = 1{,}96 \times \frac{1{,}125}{10}$$

$$\varepsilon = 0{,}2205$$

Gabarito: C.

5. (CESPE 2020/TJ-PA) Tabela da distribuição T fornecida.

Em uma amostra aleatória de 20 municípios Paraenses, considerando-se os dados da Secretaria de Estado de Segurança Pública e Defesa Social relativos ao crime de lesão corporal, a média é igual a 87 e o desvio padrão igual a 101,9419.

Considerando-se, para 19 graus de liberdade, o coeficiente a = 2,093 e utilizando-se o valor aproximado 4,4721 para a raiz quadrada de 20, com o auxílio da distribuição t, um intervalo de 95% de confiança para a média deverá ter

a) Limite inferior de, aproximadamente, 38,78.

b) Limite superior de, aproximadamente, 143,12.

c) Amplitude 2c = 93,45.

d) Limite inferior de 39,29 e limite superior de 142,18.

e) Limite superior de, aproximadamente, 134,71.

Comentários:

Vamos listar as informações do enunciado:

$Média\ amostral\ \rightarrow \bar{X} = 87$
$t \rightarrow valor\ crítico\ associado\ ao\ nível\ de\ confiança\ da\ distribuição\ t - student$
$t = 2{,}093 \rightarrow para\ 95\%\ de\ confiança$
$Desvio\ padrão \rightarrow \sigma = 101{,}9419$
$Tamanho\ da\ amostra \rightarrow n = 20 \Rightarrow \sqrt{20} = 4{,}4721$

Agora, podemos apenas aplicar na fórmula do intervalo de confiança para a média:

$$\bar{X} \pm t \times \frac{\sigma}{\sqrt{n}}$$

$$87 \pm 2{,}093 \times \frac{101{,}9419}{\sqrt{20}}$$

$$87 \pm 2{,}093 \times \frac{101{,}9419}{4{,}4721}$$

$$87 \pm 2{,}093 \times 22{,}7950$$

$$87 \pm 47{,}71$$

Logo, o intervalo é $(87 - 47{,}71 = 39{,}29;\ 87 + 47{,}71 = 134{,}71)$

Gabarito: E.

6. (CESPE 2018/EBSERH) Uma amostra aleatória simples Y1, Y2, ..., Y25 foi retirada de uma distribuição normal com média nula e variância σ², desconhecida. Considerando que P(x² < 13) = P(X² > 41) = 0,025, em que x² representa a distribuição qui-quadrado com 25 graus de liberdade, e que $S^2 = \sum_{t=1}^{25} Y_i^2$, julgue o item a seguir.

[S²/41;S²/13] representa um intervalo de 95% de confiança para a variância σ².

Comentários:

Vamos calcular o intervalo de confiança para σ² com nível $100(1-\alpha)\%$. Assim, determinaremos se a afirmativa é verdadeira:

$$IC = (\sigma^2, 1-\alpha) = \left(\frac{S^2}{Q_{1-\alpha/2}}, \frac{S^2}{Q_{\alpha/2}}\right)$$

O enunciado nos informou que $n = 25$.

Sendo $\alpha = 0{,}05$, temos:

$$IC = (\sigma^2, 1 - 0{,}95) = \left(\frac{S^2}{Q_{0{,}975}}, \frac{S^2}{Q_{0{,}025}}\right)$$

$Q_{0{,}975} = 41$

$Q_{0{,}025} = 13$

Então:

$$P(X_{24}^2 \leq Q_{0{,}975}) = 0{,}975, \text{ ou seja, } P(X_{24}^2 \leq Q_{0{,}975}) = 0{,}025$$

Logo,

$$IC = (\sigma^2, 1 - 0{,}95) = \left(\frac{S^2}{41}, \frac{S^2}{13}\right)$$

Gabarito: Certo.

7. (CESPE 2018/PF) O tempo gasto (em dias) na preparação para determinada operação policial é uma variável aleatória X que segue distribuição normal com média M, desconhecida, e desvio padrão igual a 3 dias. A observação de uma amostra aleatória de 100 outras operações policiais semelhantes a essa produziu uma média amostral igual a 10 dias.

Com referência a essas informações, julgue o item que se segue, sabendo que P(Z > 2) = 0,025, em que Z denota uma variável aleatória normal padrão.

A expressão 10 dias ± 6 dias corresponde a um intervalo de 95% de confiança para a média populacional M.

Comentários:

Vamos listar as informações do enunciado:

$Média\ amostral \rightarrow \bar{X} = 10$
$Escore\ associado\ ao\ nível\ de\ confiança \rightarrow z = 2$
$Desvio\ padrão \rightarrow \sigma = 3$
$Tamanho\ da\ amostra \rightarrow n = 100$

Agora, podemos apenas aplicar na fórmula do intervalo de confiança para a média:

$$\bar{X} \pm z \times \frac{\sigma}{\sqrt{n}}$$

$$10 \pm 2 \times \frac{3}{\sqrt{100}}$$

$$10 \pm 2 \times \frac{3}{10}$$

$$10 \pm 2 \times 0{,}3$$

$$10 \pm 0{,}6$$

Logo, o intervalo é $(10 - 0{,}6 = 9{,}4;\ 10 + 0{,}6 = 10{,}6)$

Gabarito: Errado.

8. (CESPE 2016/TCE-PA) Em estudo acerca da situação do CNPJ das empresas de determinado município, as empresas que estavam com o CNPJ regular foram representadas por 1, ao passo que as com CNPJ irregular foram representadas por 0.

Considerando que a amostra

{0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1}

Foi extraída para realizar um teste de hipóteses, julgue o item subsequente.

Uma vez que a amostra é menor que 30, a estatística do teste utilizada segue uma distribuição t de Student.

Comentários:

Pelo enunciado, sabemos que a população segue distribuição de Bernoulli (0 ou 1). Nessa situação, utilizamos a proporção amostral, $\hat{p}$, para estimar a proporção populacional, e consideramos a distribuição normal para estimar o seu intervalo.

Gabarito: Errado.

9. (CESPE 2016/TCE-PA) Em estudo acerca da situação do CNPJ das empresas de determinado município, as empresas que estavam com o CNPJ regular foram representadas por 1, ao passo que as com CNPJ irregular foram representadas por 0.

Considerando que a amostra

{0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1}

Foi extraída para realizar um teste de hipóteses, julgue o item subsequente.

Sendo P(Z > 1,96) = 0,025 e P(Z > 1,645) = 0,05, em que Z representa a variável normal padronizada, e P(t_{20} > 2,086) = 0,025 e P(t_{19} > 1,729) = 0,05, em que t_{20} e t_{19} possuem distribuição t de Student com, respectivamente, 20 e 19 graus de liberdade, o erro utilizado para a construção do intervalo de confiança é menor que 15%, se considerado um nível de significância de 5%.

Comentários:

O erro máximo no intervalo de confiança de uma distribuição de proporção é dado por:

$$Z_0 \times \sqrt{\frac{\hat{p} \times \hat{q}}{n}}$$

Vamos aos dados do problema:

$\hat{p} = 0{,}6 \rightarrow proporção\ amostral\ de\ 12\ empresas\ regulares\ em\ 20$
$\hat{q} = 1 - \hat{p} = 0{,}4$
$n = 20 \rightarrow tamanho\ da\ amostra$
$Z_0 = 1{,}96 \rightarrow nível\ de\ confiança\ de\ 95\%\ na\ tabela\ normal$

$$1{,}96 \times \sqrt{\frac{0{,}6 \times 0{,}4}{20}}$$

$$1{,}96 \times \sqrt{\frac{0{,}24}{20}}$$

$$1{,}96 \times \sqrt{0{,}012}$$

$$1{,}96 \times 0{,}109$$

$$0{,}213\ ou\ 21{,}3\%$$

Gabarito: Errado.

10. (CESPE 2016/TCE-PR) A partir de um levantamento estatístico por amostragem aleatória simples em que se entrevistaram 2.400 trabalhadores, uma seguradora constatou que 60% deles acreditam que poderão manter seu atual padrão de vida na aposentadoria.

Considerando que P(|Z|≤3)=0,99, em que Z representa a distribuição normal padrão, assinale a opção correspondente ao intervalo de 99% de confiança para o percentual populacional de trabalhadores que acreditam que poderão manter seu atual padrão de vida na aposentadoria.

a) 60,0% ± 1,0%

b) 60,0% ± 1,5%

c) 60,0% ± 3,0%

d) 60,0% ± 0,2%

e) 60,0% ± 0,4%

Comentários:

O intervalo de confiança de uma distribuição de proporção é dado por:

$$\hat{p} \pm Z_0 \times \sqrt{\frac{\hat{p} \times \hat{q}}{n}}$$

Vamos aos dados do problema:

$\hat{p} = 0{,}6 \rightarrow proporção\ amostral$
$\hat{q} = 1 - \hat{p} = 0{,}4$
$n = 2.400 \rightarrow tamanho\ da\ amostra$
$Z_0 = 3 \rightarrow nível\ de\ confiança\ de\ 99\%\ na\ tabela\ normal$

$$0{,}6 \pm 3 \times \sqrt{\frac{0{,}6 \times 0{,}4}{2.400}}$$

$$0{,}6 \pm 3 \times \sqrt{\frac{0{,}24}{2.400}} = 0{,}6 \pm 3 \times \sqrt{0{,}0001}$$

$$0{,}6 \pm 3 \times 0{,}01 = 0{,}6 \pm 0{,}03$$

$$60\% \pm 3\%$$

Gabarito: C.

11. (CESPE 2016/TCE-PA) Suponha que o tribunal de contas de determinado estado disponha de 30 dias para analisar as contas de 800 contratos firmados pela administração. Considerando que essa análise é necessária para que a administração pública possa programar o orçamento do próximo ano e que o resultado da análise deve ser a aprovação ou rejeição das contas, julgue o item a seguir. Sempre que necessário, utilize que P(Z > 1,96) = 0,025 e P(Z > 1,645) = 0,05, em que Z representa a variável normal padronizada.

Se forem aprovados 90% dos contratos de uma amostra composta de 100 contratos, o erro amostral será superior a 10%.

Comentários:

A questão trata sobre o erro amostral da proporção:

$$e = Z_0 \times \sqrt{\frac{\hat{p} \times \hat{q}}{n}}$$

O enunciado não nos dá o nível de confiança, portanto, não é possível calcular o erro sem ele. Porém, para deixar evidente que a questão está errada, vamos adotar o nível de confiança de 95%, o escore normal padrão seria de 1,96, valor informado na questão. Então, fica:

$$e = 1{,}96 \times \sqrt{\frac{0{,}9 \times 0{,}1}{100}}$$

$$e = 1{,}96 \times \sqrt{0{,}0009}$$

$$e = 1{,}96 \times 0{,}03$$

$$e = 0{,}0579 \; ou \; 5{,}79\%$$

Logo, temos um erro amostral inferior a 10%.

Gabarito: Errado.

12. (CESPE 2016/TCE-PA) Suponha que o tribunal de contas de determinado estado disponha de 30 dias para analisar as contas de 800 contratos firmados pela administração. Considerando que essa análise é necessária para que a administração pública possa programar o orçamento do próximo ano e que o resultado da análise deve ser a aprovação ou rejeição das contas, julgue o item a seguir. Sempre que necessário, utilize que P(Z > 1,96) = 0,025 e P(Z > 1,645) = 0,05, em que Z representa a variável normal padronizada.

Considerando-se que, no ano anterior ao da análise em questão, 80% dos contratos tenham sido aprovados e que 0,615 seja o valor aproximado de 1,962×0,8×0,2, é correto afirmar que a quantidade de contratos de uma amostra com nível de 95% de confiança para a média populacional e erro amostral de 5% é inferior a 160.

Comentários:

Vamos listar os dados da questão:

$\hat{p} = 0{,}8 \rightarrow proporção\ amostral\ de\ contas\ aprovadas\ no\ ano\ anterior$
$\hat{q} = 1 - \hat{p} = 0{,}2$
$n \rightarrow tamanho\ da\ amostra$
$Z_0 \rightarrow escore\ normal\ associado\ ao\ nível\ de\ significância\ \ 5\% \rightarrow Z_0 = 1{,}96$

Temos que o erro amostral é dado por:

$$e = Z_0 \times \sqrt{\frac{\hat{p} \times \hat{q}}{n}}$$

$$0{,}05 = 1{,}96 \times \sqrt{\frac{0{,}8 \times 0{,}2}{n}}$$

$$\sqrt{n} = \frac{1{,}96 \times 0{,}4}{0{,}05}$$

$$\sqrt{n} = 15{,}68$$

$$n = 15{,}68^2$$

$$n = 245{,}8624$$

Gabarito: Errado.

13. (CESPE 2016/TCE-PA) A respeito de uma amostra de tamanho n = 10, com os valores amostrados {0,10, 0,06, 0,10, 0,12, 0,08, 0,10, 0,05, 0,15, 0,14, 0,11}, extraídos de determinada população, julgue o item seguinte.

Por um intervalo de confiança frequentista igual a (–0,11, 0,32), entende-se que a probabilidade de o parâmetro médio ser superior a –0,11 e inferior a 0,32 é igual ao nível de confiança γ.

Comentários:

A probabilidade de que o parâmetro dado no enunciado esteja no intervalo construído (–0,11, 0,32) é ou 0 ou 1. Pela abordagem frequentista, 95% dos intervalos de confiança conteriam o verdadeiro valor do parâmetro (a média populacional).

Gabarito: Errado.

14. (CESPE 2016/TCE-PA) Uma amostra aleatória, com n = 16 observações independentes e identicamente distribuídas (IID), foi obtida a partir de uma população infinita, com média e desvio padrão desconhecidos e distribuição normal.

Tendo essa informação como referência inicial, julgue o seguinte item.

Se a média amostral for igual a 3,2 e a variância amostral, igual a 4,0, o estimador de máxima verossimilhança para a média populacional será igual a 1,6.

Comentários:

Sabendo que a população segue distribuição normal, então o estimador de máxima verossimilhança para a média populacional é igual à média amostral. Sabendo que a média amostral é igual a 3,2, então o estimador de máxima verossimilhança para a média populacional é 3,2.

Gabarito: Errado.

15. (CESPE 2016/TCE-PA) Uma amostra aleatória, com n = 16 observações independentes e identicamente distribuídas (IID), foi obtida a partir de uma população infinita, com média e desvio padrão desconhecidos e distribuição normal.

Tendo essa informação como referência inicial, julgue o seguinte item.

Em um intervalo de 95% de confiança para a média populacional em questão, caso se aumente o tamanho da amostra em 100 vezes (passando a 1.600 observações), a largura total do intervalo de confiança será reduzida à metade.

Comentários:

A questão trata da amplitude do intervalo de confiança. Pela fórmula, podemos determinar se a afirmativa da questão está correta. Vamos analisar:

$$A = 2 \times t_0 \times \frac{\sigma}{\sqrt{n}}$$

$t_0 \rightarrow$ é $o\ escore\ de\ distribuição\ T\ associado\ a\ 95\%\ de\ confiança.$
$\sigma \rightarrow desvio\ padrão\ da\ população$
$n \rightarrow tamanho\ da\ amostra$

Aumentando o tamanho da amostra em 100 vezes, na fórmula ficará $\sqrt{100} = 10$. Logo, concluímos que A será dividido por 10 e não por 2.

Gabarito: Errado.

FGV

16. (FGV/2011 – AFRE/RJ) Um processo X segue uma distribuição normal com média populacional desconhecida, mas com desvio-padrão conhecido e igual a 4. Uma amostra com 64 observações dessa população é feita, com média amostral 45. Dada essa média amostral, a estimativa da média populacional, a um intervalo de confiança de 95%, é

a) (41;49).

b) (37;54).

c) (44,875;45,125).

d) (42,5;46,5).

e) (44;46).

Comentários:

Sabendo que o processo segue distribuição normal com média desconhecida, mas com desvio padrão conhecido ($\sigma = 4$), então a média amostral seguirá distribuição normal e o intervalo será da forma:

$$\bar{X} \pm z.\frac{\sigma}{\sqrt{n}}$$

O enunciado informa que a média amostral é $\bar{X} = 45$ e o tamanho amostral é $n = 64$. Para um intervalo de confiança de 95%, o valor crítico da normal padrão é $z = 1{,}96$. Assim, o intervalo é:

$$\bar{X} + z.\frac{\sigma}{\sqrt{n}} = 45 + 1{,}96\frac{4}{\sqrt{64}} = 45 + 1{,}96\frac{4}{8} = 45 + 0{,}98 \cong 46$$

$$\bar{X} - z.\frac{\sigma}{\sqrt{n}} = 45 - 1{,}96\frac{4}{\sqrt{64}} = 45 - 1{,}96\frac{4}{8} = 45 - 0{,}98 \cong 44$$

Gabarito: E

17. (FGV/2016 – IBGE) 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 $\bar{X}$ = 8 de grau de confiança pretendido é de 95%. Somam-se a todas essas informações os valores tabulados:

Φ(1,65) ≅ 0,95 Φ(1,96) ≅ 0,975

T_{25}(1,71) ≅ 0,95 T_{26}(1,70) ≅ 0,95 T_{25}(2,06) ≅ 0,975 T_{26}(2,05) ≅ 0,975

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

Φ(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;

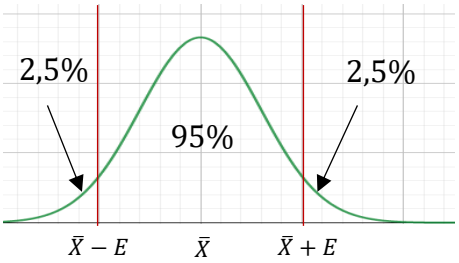
e) 4,60 e 11,40.

Comentários:

O enunciado informa que a variância da população é desconhecida, tendo apresentado a variância amostral, $\hat{s}^2 = 100$, e a média amostral, $\bar{X}$ = 8, e o tamanho da amostra, $n = 26$. Dessa forma, precisamos utilizar a distribuição de t-Student, em que o intervalo de confiança é da forma:

$$\bar{X} \pm t_{n-1}.\frac{s}{\sqrt{n}}$$

Para um intervalo de confiança de 95%, temos 2,5% na região crítica inferior:



Assim, devemos buscar o valor para o qual a função acumulada é T = 2,5% + 95% = 97,5% = 0,975. Para n – 1 = 25 graus de liberdade, o enunciado informa que $T_{25}(2{,}06) \cong 0{,}975$, ou seja, t_{n-1}= 2,06.

Considerando, ainda, que o desvio padrão é $\hat{s} = \sqrt{\hat{s}^2} = \sqrt{100} = 10$, o intervalo de confiança é:

$$\bar{X} + t_{n-1}.\frac{s}{\sqrt{n}} = 8 + 2{,}06 \times \frac{10}{\sqrt{26}} \cong 8 + 2{,}06 \times 2 = 8 + 4{,}12 = 12{,}12$$

$$\bar{X} - t_{n-1}.\frac{s}{\sqrt{n}} = 8 - 2{,}06 \times \frac{10}{\sqrt{26}} \cong 8 - 2{,}06 \times 2 = 8 - 4{,}12 = 3{,}88$$

Gabarito: B

18. (FGV/2010 – Fiscal de Rendas/RJ) Para estimar a proporção p de pessoas acometidas por uma certa gripe numa população, uma amostra aleatória simples de 1600 pessoas foi observada e constatou-se que, dessas pessoas, 160 estavam com a gripe.

Um intervalo aproximado de 95% de confiança para p será dado por:

a) (0,066, 0,134).

b) (0,085, 0,115).

c) (0,058, 0,142).

d) (0,091, 0,109).

e) (0,034, 0,166).

Comentários:

A partir da proporção amostral, o intervalo de confiança é calculado como:

$$\hat{p} \pm z.\sqrt{\frac{\hat{p}.\hat{q}}{n}}$$

A proporção encontrada na amostra foi de:

$$\hat{p} = \frac{160}{1600} = 0{,}1$$

Logo, $\hat{q} = 1 - \hat{p} = 0{,}9$. Também sabemos que o tamanho amostral é $n = 1600$ e o valor de z para um intervalo de confiança de 95% é $z = 1{,}96$. Portanto, a margem de erro do intervalo, $z.\sqrt{\frac{\hat{p}.\hat{q}}{n}}$, é:

$$z.\sqrt{\frac{\hat{p}.\hat{q}}{n}} = 1{,}96.\sqrt{\frac{0{,}1 \times 0{,}9}{1600}} = 1{,}96.\sqrt{\frac{0{,}09}{1600}} = 1{,}96.\frac{0{,}3}{40} = 0{,}0147 \cong 0{,}015$$

Assim, o intervalo é:

$$\hat{p} + z.\sqrt{\frac{\hat{p}.\hat{q}}{n}} = 0{,}1 + 0{,}015 = 0{,}115$$

$$\hat{p} - z.\sqrt{\frac{\hat{p}.\hat{q}}{n}} = 0{,}1 - 0{,}015 = 0{,}085$$

Gabarito: B

19. (FGV/2016 – IBGE) 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 ε e com um grau de confiança de (1-α). Assim, é correto afirmar que:

a) uma redução de α pode ser compensado por uma redução de ε, 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 ε e (1-α);

c) se a variância de $\hat{p}$ for máxima, o erro ε for 5% e a amostra tiver tamanho n = 1000, o nível de significância será de 5%;

d) fixos α e p, quanto maior a amostra menores serão os ganhos de precisão (redução de ε) gerados por seus incrementos;

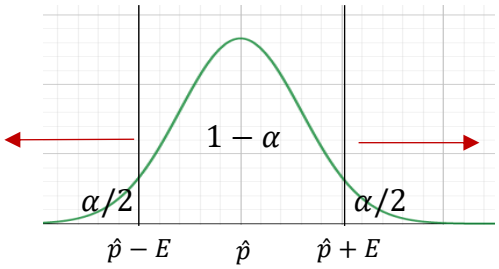
e) fixo ε, quanto menor a proporção populacional (p) e maior o nível de significância (α), maior deverá ser a amostra (n).

Comentários:

A questão trata do intervalo para a proporção amostral, em que a margem de erro é dada por:

$$\varepsilon = z.\sqrt{\frac{\hat{p}.\hat{q}}{n}}$$

Em relação à alternativa A, uma **redução do nível de significância α** implicaria no aumento do nível de confiança, $1-\alpha$ (isto é, **maior valor de z**), o que **aumentaria a margem de erro ε**, conforme ilustrado abaixo.



Por isso, a alternativa A está incorreta.

Analogamente, em relação à alternativa E, quanto maior o nível de significância α, menor o nível de confiança $1-\alpha$ (isto é, **menor valor de z**). Sabendo que o tamanho amostral é proporcional ao quadrado de z, $n = \left(\frac{z}{\varepsilon}\right)^2 \hat{p}.\hat{q}$, então para uma mesma margem de erro ε, teríamos uma amostra n **menor**. Por esse motivo, sabemos que a alternativa E está errada.

Em relação à alternativa B, o tamanho amostral é calculado como:

$$n = \left(\frac{z}{\varepsilon}\right)^2 \hat{p}.\hat{q}$$

E a variância é populacional é $V(X) = n.p.q$. Ou seja, a variância aumenta com o aumento do produto $p.q$. Quanto maior esse produto, **maior** será o tamanho amostral $n = \left(\frac{z}{\varepsilon}\right)^2 \hat{p}.\hat{q}$, para um mesmo erro ε e mesmo nível de confiança $1-\alpha$. Por isso, a alternativa B está incorreta.

Em relação à alternativa C, a variância é máxima para $p = q = 0{,}5$. Para um erro $\varepsilon = 0{,}05$ e para uma amostra de tamanho $n = 1000$, teremos:

$$\varepsilon = z.\sqrt{\frac{\hat{p}.\hat{q}}{n}}$$

$$z = \frac{\varepsilon}{\sqrt{\frac{\hat{p}.\hat{q}}{n}}} = \varepsilon \times \frac{\sqrt{n}}{\sqrt{\hat{p}.\hat{q}}}$$

$$z = 0{,}05 \times \frac{\sqrt{1000}}{\sqrt{0{,}5 \times 0{,}5}} = 0{,}05 \times \frac{10\sqrt{10}}{0{,}5} = \sqrt{10} \cong 3{,}16$$

Para um nível de significância $\alpha = 5\%$, descrito na alternativa, o valor crítico é $z_{\alpha/2} = 1{,}96$. Logo, a alternativa C está incorreta.

Em relação à alternativa D, para um mesmo α (ou seja, mesmo z) e mesmo $\hat{p}$, sabemos que o aumento da amostra gera um aumento da precisão, isto é, uma redução do erro ε:

$$\varepsilon = z.\sqrt{\frac{\hat{p}.\hat{q}}{n}}$$

Porém, quanto maior a amostra, menor será o efeito de determinado incremento. Por exemplo, se tivermos uma amostra de tamanho n = 10 e aumentarmos para n = 40, esse incremento quadruplica a amostra e, com isso, reduzimos o erro à metade:

$$\varepsilon_2 = z.\sqrt{\frac{\hat{p}.\hat{q}}{4.n}} = z.\sqrt{\frac{\hat{p}.\hat{q}}{n}} \times \frac{1}{2} = \frac{1}{2}\varepsilon$$

Por outro lado, o aumento de n = 100 para n = 130 (mesmo incremento de 30 unidades) representa um aumento de 1,3 e o novo erro terá uma redução aproximada de 12%. Ou seja, o ganho de precisão é bem menor e a alternativa D está correta.

Gabarito: D

FCC

20. (FCC 2017/TRT 11ª Região) Instruções: Considere as informações abaixo para responder à questão. Se Z tem distribuição normal padrão, então:

P(Z < 0,4) = 0,655; P(Z < 0,67) = 0,75; P(Z < 1,4) = 0,919; P(Z < 1,6) = 0,945;

P(Z < 1,64) = 0,95; P(Z < 1,75) = 0,96; P(Z < 2) = 0,977; P(Z < 2,05) = 0,98

A porcentagem do orçamento gasto com educação nos municípios de certo estado é uma variável aleatória X com distribuição normal com média μ(%) e variância 4(%)².

Uma amostra aleatória, com reposição, de tamanho n, X_1, X_2, ..., X_n, é selecionada da distribuição de X. Sendo X, a média amostral dessa amostra, o valor de n para que X não se distancie de sua média por mais do que 0,41% com probabilidade de 96% é igual a

a) 64

b) 100

c) 121

d) 81

e) 225

Comentário

Essa questão trabalha com o tamanho da amostra necessário para construir um intervalo de confiança para a media de uma população com distribuição normal e variância conhecida. O tamanho amostral pode ser obtido a partir da fórmula do erro:

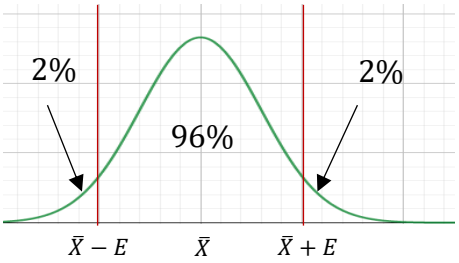
$$E = z.\frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = \frac{z.\sigma}{E}$$

$$n = \left(\frac{z.\sigma}{E}\right)^2$$

O enunciado informa que X não deve se distanciar da média $\bar{X}$ (medida em %), em mais do 0,41(%), o que corresponde à margem de erro, E = 0,41. Sabendo que a variância é $\sigma^2 = 4$, então o desvio padrão é $\sigma = \sqrt{\sigma^2} = 2$.

Para que o nível de confiança seja $1 - \alpha = 96\%$, então, $\alpha = 100\% - 96\% = 4\%$ e haverá $^{\alpha}/_{2} = 2\%$ acima e abaixo do intervalo de confiança:



Logo, o valor de z deve ser tal que P(Z < z) = 2% + 96% = 98% = 0,98. Pelos dados do enunciado observamos que esse valor é z = 2,05, pois P(Z < 2,05) = 0,98.

Substituindo os valores de $z = 2{,}05$, $\sigma = 2$ e $E = 0{,}41$, na fórmula do tamanho amostral, temos:

$$n = \left(z.\frac{\sigma}{E}\right)^2 = \left(2{,}05.\frac{2}{0{,}41}\right)^2 = (10)^2 = 100$$

Gabarito: B

21. (FCC 2015/SEFAZ-PI) Instrução: Para responder à questão utilize, dentre as informações dadas a seguir, as que julgar apropriadas. Se Z tem distribuição normal padrão, então:

P(Z < 0,4) = 0,655; P(Z < 1,2) = 0,885; P(Z < 1,6) = 0,945; P(Z < 1,8) = 0,964; P(Z < 2) = 0,977.

O efeito do medicamento A é o de baixar a pressão arterial de indivíduos hipertensos. O tempo, em minutos, decorrido entre a tomada do remédio e a diminuição da pressão é uma variável aleatória X com distribuição normal, tendo média μ e desvio padrão σ.

Uma amostra aleatória de n indivíduos hipertensos foi selecionada com o objetivo de se estimar μ. Supondo que o valor de σ é 10 min, o valor de n para que o estimador não se afaste de μ por mais do que 2 min, com probabilidade de 89%, é igual a

a) 36

b) 100

c) 81

d) 49

e) 64

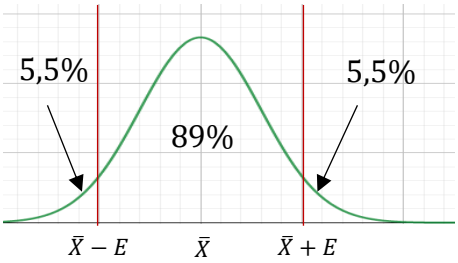
Comentário

Trata-se de estimação para o intervalo da média, sendo conhecido o desvio padrão populacional. Nesse caso, utilizamos a distribuição normal. O tamanho amostral é dado por:

$$n = \left(z.\frac{\sigma}{E}\right)^2$$

O enunciado informa que X não deve se distanciar da média em mais do 2 min, o que corresponde à margem de erro, E = 2. O enunciado também informa que o desvio padrão é $\sigma = 10$ min.

Para que o nível de confiança seja $1 - \alpha = 89\%$, então, $\alpha = 100\% - 89\% = 11\%$ e haverá $^{\alpha}/_{2} = 5{,}5\%$ acima e abaixo do intervalo de confiança:



Logo, o valor de z é tal que P(Z < z) = 5,5% + 89% = 94,5%. Pelos dados do enunciado observamos que esse valor é z = 1,6, pois P(Z < 1,6) = 0,945.

Substituindo os valores de $z = 1{,}6, \sigma = 10$ e $E = 2$, na fórmula do tamanho amostral, temos:

$$n = \left(z.\frac{\sigma}{E}\right)^2 = \left(1{,}6.\frac{10}{2}\right)^2 = (8)^2 = 64$$

Gabarito: E

22. (FCC 2015/TRT 3ª Região) Se Z tem distribuição normal padrão, então:

P(Z < 0,5) = 0,591; P(Z < 1) = 0,841; P(Z < 1,15) = 0,8951; P(Z < 1,17) = 0,879; P(Z < 1,2) = 0,885; P(Z<1,4)=0,919;P(Z<1,64)=0,95;P(Z<2)=0,977;P(Z<2,06)=0,98;P(Z<2,4)= 0,997.

Considere que X é a variável aleatória, que representa as idades, em anos, dos trabalhadores de certa indústria. Suponha que X tem distribuição normal com média de μ anos e desvio padrão de 5 anos.

Uma amostra aleatória, com reposição, de n trabalhadores será selecionada e sejam sejam X_1, X_2, ..., X_n as idades observadas e $\overline{X} = \frac{\sum_1^n X_i}{n}$ a média desta amostra.

Desejando-se que o valor absoluto da diferença entre $\overline{X}$ e sua média seja menor do que 6 meses, com probabilidade de 95,4%, o valor de n deverá ser igual a

a) 225.

b) 100.

c) 256.

d) 196.

e) 400.

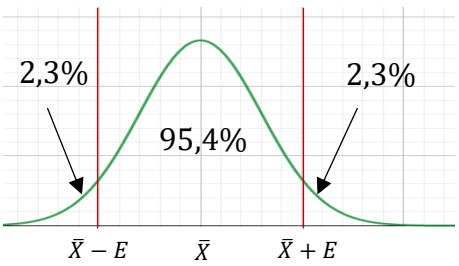
Comentário

Trata-se de estimação para o intervalo da média, sendo conhecido o desvio padrão populacional. Nesse caso, utilizamos a distribuição normal. O tamanho amostral é dado por:

$$n = \left(z.\frac{\sigma}{E}\right)^2$$

O enunciado informa que o valor absoluto da diferença entre $\overline{X}$ e a média μ não deve ser maior que 6 meses, ou seja, **0,5 ano**, o que corresponde à margem de erro, E = 0,5. O enunciado também informa que o desvio padrão é $\sigma = 5$ anos.

Para que o nível de confiança seja $1 - \alpha = 95{,}4\%$, então, $\alpha = 100\% - 95{,}4\% = 4{,}6\%$ e haverá $^{\alpha}/_{2} = 2{,}3\%$ acima e abaixo do intervalo de confiança:



Logo, o valor de z é tal que P(Z < z) = 2,3% + 95,4% = 97,7% = 0,977. Pelos dados do enunciado observamos que esse valor é z = 2, pois P(Z < 2) = 0,977.

Substituindo os valores de $z = 2, \sigma = 5$ e $E = 0{,}5$, na fórmula do tamanho amostral, temos:

$$n = \left(z.\frac{\sigma}{E}\right)^2 = \left(2.\frac{5}{0{,}5}\right)^2 = (20)^2 = 400$$

Gabarito: E

23. (FCC 2014/TRT 16ª Região) Se Z tem distribuição normal padrão, então:

P(Z<0,25) = 0,599, P(Z<0,80) = 0,84, P(Z<1) = 0,841, P(Z<1,96) = 0,975, P(Z<3,09) = 0,999.

Considere $X_1, X2, \dots, X_n$ uma amostra aleatória simples, com reposição, da distribuição da variável X, que tem distribuição normal com média μ e variância 36. Seja $\bar{X}$ a média amostral dessa amostra. O valor de n para que a distância entre $\bar{X}$ e μ seja, no máximo, igual a 0,49, com probabilidade de 95% é igual a:

a) 256

b) 225

c) 400

d) 144

e) 576

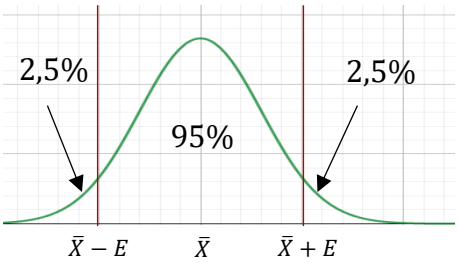
Comentários:

Trata-se de estimação para o intervalo da média, sendo conhecida a variância populacional. Nesse caso, utilizamos a distribuição normal. O tamanho amostral é dado por:

$$n = \left(z.\frac{\sigma}{E}\right)^2$$

O enunciado informa que a distância entre $\bar{X}$ e a média μ deve ser no máximo igual a 0,49, o que corresponde à margem de erro, E = 0,49. O enunciado também informa que a variância é $\sigma^2 = 36$, logo o desvio padrão é $\sigma = \sqrt{36} = 6$.

Para que o nível de confiança seja $1 - \alpha = 95\%$, então, $\alpha = 100\% - 95\% = 5\%$ e haverá $^{\alpha}/_{2} = 2{,}5\%$ acima e abaixo do intervalo de confiança:



Logo, o valor de z é tal que P(Z < z) = 2,5% + 95% = 97,5% = 0,975. Pelos dados do enunciado observamos que esse valor é z = 1,96, pois P(Z < 1,96) = 0,975.

Substituindo os valores de $z = 1{,}96, \sigma = 6$ e $E = 0{,}49$, na fórmula do tamanho amostral, temos:

$$n = \left(z.\frac{\sigma}{E}\right)^2 = \left(1{,}96.\frac{6}{0{,}49}\right)^2 = (24)^2 = 576$$

Gabarito: E

24. (FCC 2018/TRT 14ª Região) Uma pesquisa piloto realizada no setor de embalagens, referente aos motivos de demissão de funcionários, mostra que 34% dos casos de demissão, p*, tem como motivo a situação financeira da empresa. Utilizando um nível de confiança de 95%, a proporção p* obtida na pesquisa piloto, com uma margem de erro amostral e ≤ 3% e que P(Z ≥ 1,96) = 2,5%, o tamanho mínimo necessário da amostra para estimar a proporção de demissões causadas por motivos financeiros, no setor de embalagens, nas condições estipuladas é

a) 635

b) 1020

c) 2115

d) 854

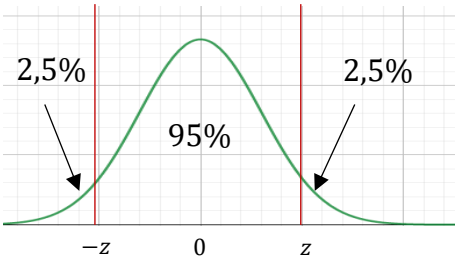
e) 958

Comentários:

Essa questão trabalha com a estimação intervalar para proporções. Nesse caso, o tamanho amostral é dado por:

$$n = \left(\frac{z}{E}\right)^2 \times \hat{p} \times \hat{q}$$

O nível de confiança deve ser de 95%, conforme ilustrado a seguir:



O enunciado informa que P(Z ≥ 1,96) = 2,5%. Ou seja, o valor de z que delimita um intervalo de confiança de 95% é z = 1,96.

Além disso, o enunciado informa que:

- A margem de erro amostral tolerada é E = 3% = 0,03;
- A proporção estimada é $\hat{p} = 34\% = 0{,}34$. Assim, o seu complemento é $\hat{q} = 1 - \hat{p} = 1 - 0{,}34 = 0{,}66$.

Substituindo esses dados na fórmula acima, temos:

$$n = \left(\frac{1{,}96}{0{,}03}\right)^2 \times 0{,}34 \times 0{,}66 \cong 957{,}8$$

Gabarito: E

25. (FCC 2016/TRT 20ª Região) Sejam duas variáveis aleatórias X e Y, normalmente distribuídas, com as populações de tamanho infinito e médias μ_X e μ_Y, respectivamente. Uma amostra aleatória de tamanho 64 foi extraída da população de X, apresentando um intervalo de confiança [1, 5] para μ_X, ao nível de confiança (1 − α). Uma outra amostra aleatória de tamanho 144 foi extraída da população de Y, independente da primeira, apresentando um intervalo de confiança [4, 10] para μ_Y, também ao nível de confiança de (1 − α). Se σ_X e σ_Y são os desvios padrões populacionais de X e Y, respectivamente, então σ_Y/σ_X apresenta um valor igual a

a) 2,000

b) 3,375

c) 1,500

d) 2,250

e) 2,500

Comentários:

Essa questão trabalha com o intervalo de confiança para a média de uma população com distribuição normal e variância conhecida, dado por:

$$(\bar{X} - E; \bar{X} + E)$$

Em que $E = z.\frac{\sigma}{\sqrt{n}}$. Observe que a amplitude do intervalo de confiança (diferença entre o limite superior e o limite inferior) é igual ao dobro do erro amostral:

$$A = 2 \times E = 2.z.\frac{\sigma}{\sqrt{n}}$$

Em relação à população X, o enunciado informa que:

- O tamanho da amostra é n = 64
- O intervalo de confiança é [1, 5], ou seja, a amplitude é A = 5 – 1 = 4.

Substituindo esses dados na fórmula acima, temos:

$$4 = 2.z.\frac{\sigma_X}{\sqrt{64}} = 2.z.\frac{\sigma_X}{8} = z.\frac{\sigma_X}{4}$$

$$\sigma_X = \frac{16}{z}$$

Em relação à população Y, o enunciado informa que:

- O tamanho da amostra é n = 144
- O intervalo de confiança é [4, 10], ou seja, a amplitude é A = 10 – 4 = 6.

Substituindo esses dados na fórmula anterior e sabendo que o nível de confiança é o mesmo (mesmo z), temos:

$$6 = 2.z.\frac{\sigma_Y}{\sqrt{144}} = 2.z.\frac{\sigma_Y}{12} = z.\frac{\sigma_Y}{6}$$

$$\sigma_Y = \frac{36}{z}$$

Assim, a razão entre os desvios padrão é:

$$\frac{\sigma_Y}{\sigma_X} = \frac{\frac{36}{z}}{\frac{16}{z}} = \frac{36}{16} = 2{,}25$$

Gabarito: D

26. (FCC 2020/AL-AP) De uma amostra aleatória de tamanho 64 extraída, com reposição, de uma população normalmente distribuída e variância conhecida σ^2, obteve-se um intervalo de confiança de 95% igual a [23, 27] para a média μ desta população. Desejando-se obter um intervalo de confiança de 95% para μ, porém com amplitude igual à metade da obtida anteriormente, é necessário extrair da população uma amostra aleatória, com reposição, de tamanho

a) 400

b) 1.024

c) 512

d) 256

e) 128

Comentários:

A amplitude do intervalo de confiança é igual ao dobro do erro amostral. O erro para uma população com distribuição normal e variância conhecida é dado por:

$$E = z.\frac{\sigma}{\sqrt{n}}$$

Para que a amplitude caia pela metade, é necessário que o erro amostral caia pela metade:

$$E_2 = \frac{1}{2}.E_1$$

Sendo o intervalo de confiança (z) e o desvio padrão (σ) constantes, conforme enunciado, então o novo tamanho amostral n deve ser:

$$z.\frac{\sigma}{\sqrt{n_2}} = \frac{1}{2} \times z.\frac{\sigma}{\sqrt{n_1}}$$

$$\frac{1}{\sqrt{n_2}} = \frac{1}{2\sqrt{n_1}}$$

$$\sqrt{n_2} = 2.\sqrt{n_1}$$

$$n_2 = (2.\sqrt{n_1})^2 = 4.n_1$$

Ou seja, a amostra precisa quadruplicar. Sabendo que $n_1 = 64$, o tamanho da nova amostra deve ser:

$$n_2 = 4 \times 64 = 256$$

Gabarito: D

27. (FCC 2019/Prefeitura de Recife/PE) Considere que na curva normal padrão (Z) a probabilidade P(-2 ≤ Z ≤ 2) = 95%. Uma amostra aleatória de tamanho 400 é extraída de uma população normalmente distribuída e de tamanho infinito. Dado que a variância desta população é igual a 64, obtém-se, com base na amostra, um intervalo de confiança de 95% para a média da população. A amplitude deste intervalo é igual a

a) 0,8

b) 6,4

c) 1,6

d) 12,8

e) 3,2

Comentários:

A amplitude do intervalo de confiança é igual ao dobro do erro amostral. Sendo a população normalmente distribuída com variância conhecida, temos:

$$A = 2 \times E = 2.z.\frac{\sigma}{\sqrt{n}}$$

O enunciado informa que:

- O tamanho da amostra é n = 400;
- P(-2 ≤ Z ≤ 2) = 95%, ou seja, o valor de z para um intervalo de confiança de 95% é z = 2;
- A variância da população é $\sigma^2 = 64$. Logo, o desvio padrão (raiz quadrada) é $\sigma = \sqrt{64} = 8$.

Substituindo esses dados na fórmula acima, temos:

$$A = 2 \times 2 \times \frac{8}{\sqrt{400}} = \frac{32}{20} = 1{,}6$$

Gabarito: C

28. (FCC 2017/TRT 11ª Região) Uma amostra aleatória de tamanho 64 é extraída de uma população de tamanho infinito, normalmente distribuída, média μ e variância conhecida σ². Obtiveram-se com base nos dados desta amostra, além de uma determinada média amostral x, 2 intervalos de confiança para μ aos níveis de 95% e 99%, sendo os limites superiores destes intervalos iguais a 20,98 e 21,29, respectivamente. Considerando que na curva normal padrão (Z) as probabilidades P(|Z| > 1,96) = 0,05 e P(|Z| > 2,58) = 0,01, encontra-se que σ² é igual a

a) 16,00

b) 6,25

c) 4,00

d) 12,25

e) 9,00

Comentários:

O limite superior do intervalo de confiança para a média de uma população com variância conhecida é dado por:

$$L_{SUP} = \bar{X} + z.\frac{\sigma}{\sqrt{n}}$$

Para encontrarmos a variância populacional σ^2, devemos utilizar os limites superiores informados considerando os dois níveis de confiança informados.

Para um nível de confiança de 95%, temos z = 1,96 (dado no enunciado) e $L_{SUP_1} = 20{,}98$. Sabendo que o tamanho da amostra é n = 64 (logo, $\sqrt{n} = \sqrt{64} = 8$), temos:

$$20{,}98 = \bar{X} + 1{,}96.\frac{\sigma}{8}$$

$$20{,}98 = \bar{X} + 0{,}245.\sigma$$

Para um nível de confiança de 99%, temos z = 2,58 (dado no enunciado) e $L_{SUP_2} = 21{,}29$. Considerando a mesma amostra de tamanho n = 64, temos:

$$21{,}29 = \bar{X} + 2{,}58.\frac{\sigma}{8}$$

$$21{,}29 = \bar{X} + 0{,}3225.\sigma$$

Subtraindo a primeira equação da segunda, temos:

$$21{,}29 - 20{,}98 = \bar{X} - \bar{X} + 0{,}3225.\sigma - 0{,}245.\sigma$$

$$0{,}31 = 0{,}0775.\sigma$$

$$\sigma = 4$$

Logo, a variância é o quadrado desse resultado:

$$\sigma^2 = 4^2 = 16$$

Gabarito: A

29. (FCC 2018/TRT – 14ª Região) Um intervalo de confiança com um nível de (1 – α) foi construído para a média μ_1 de uma população P_1, normalmente distribuída, de tamanho infinito e variância populacional igual a 144. Por meio de uma amostra aleatória de tamanho 36 obteve-se esse intervalo igual a [25,3; 34,7]. Seja uma outra população P_2, também normalmente distribuída, de tamanho infinito e independente da primeira. Sabe-se que a variância de P_2 é conhecida e que por meio de uma amostra aleatória de tamanho 64 de P_2 obteve-se um intervalo de confiança com um nível de (1 – α) para a média μ_2 de P_2 igual a [91,54; 108,46]. O desvio padrão de P_2 é igual a.

a) 28,80

b) 19,20

c) 23,04

d) 38,40

e) 14,40

Comentários:

Essa questão trabalha com a estimação intervalar para uma população com distribuição normal e variância conhecida. Nessa situação, o intervalo é dado por:

$$\bar{X} \pm z.\frac{\sigma}{\sqrt{n}}$$

Observe que o enunciado não forneceu um nível de confiança, pois o valor de z será obtido pelos dados da primeira população.

Para a primeira população, o enunciado informa que:

- O tamanho da amostra é $n_1 = 36$;
- A variância é $\sigma_1^2 = 144$. Logo, o desvio padrão (raiz quadrada) é $\sigma_1 = \sqrt{144} = 12$.

A média amostral para essa primeira população pode ser calculada pela média aritmética dos extremos do respectivo intervalo:

$$\bar{X}_1 = \frac{34{,}7 + 25{,}3}{2} = \frac{60}{2} = 30$$

A diferença para qualquer um dos extremos fornece o erro amostral:

$$E_1 = z.\frac{\sigma_1}{\sqrt{n_1}} = 34{,}7 - 30 = 4{,}7$$

Substituindo os valores do desvio padrão e tamanho amostral que conhecemos, temos:

$$z.\frac{12}{\sqrt{36}} = 4{,}7$$

$$z.\frac{12}{6} = 4{,}7$$

$$2.z = 4{,}7$$

$$z = 2{,}35$$

Em relação à segunda população, sabemos que o tamanho da amostra é $n_2 = 64$. A média amostral para essa população pode ser calculada pela média aritmética dos extremos do respectivo intervalo:

$$\bar{X}_2 = \frac{91{,}54 + 108{,}46}{2} = \frac{200}{2} = 100$$

A diferença para qualquer um dos extremos fornece o erro amostral:

$$E_2 = z.\frac{\sigma_2}{\sqrt{n_2}} = 108{,}46 - 100 = 8{,}46$$

Substituindo os valores de z = 2,35 e do tamanho amostral $n_2 = 64$, temos:

$$2{,}35.\frac{\sigma_2}{\sqrt{64}} = 8{,}46$$

$$\frac{\sigma_2}{8} = \frac{8{,}46}{2{,}35}$$

$$\sigma_2 = 3{,}6 \times 8 = 28{,}8$$

Gabarito: A

30. (FCC 2016/TRT – 20ª Região) Uma variável aleatória X tem distribuição normal, variância desconhecida e com uma população de tamanho infinito. Deseja-se construir um intervalo de confiança de 95% para a média μ da população com base em uma amostra aleatória de tamanho 9 extraída dessa população e considerando a distribuição t de Student. Nessa amostra, observou-se que a média apresentou um valor igual a 5 e a soma dos quadrados dos 9 elementos da amostra foi igual a 243.

Dados: Valores críticos (t_α) da distribuição de Student com n graus de liberdade, tal que a probabilidade $P(t > t_\alpha) = \alpha$.

n	$\alpha = 0{,}025$	$\alpha = 0{,}05$	$\alpha = 0{,}10$
7	2,36	1,89	1,41
8	2,31	1,86	1,40
9	2,26	1,83	1,38

O intervalo de confiança encontrado foi igual a

a) [4,055; 5,945]

b) [4,070; 5,930]

c) [4,300; 5,700]

d) [3,845; 6,155]

e) [3,870; 6,130]

Comentários:

Essa questão trabalha com a estimação intervalar para uma população com distribuição normal e variância desconhecida. Nessa situação, o intervalo é dado por:

$$\bar{X} \pm t.\frac{s}{\sqrt{n}}$$

O enunciado informa que:

- O tamanho da amostra é n = 9;
- A média é $\bar{X} = 5$;
- A soma dos quadrados dos elementos da amostra é $\sum X^2 = 243$. Com essa informação, podemos calcular a variância amostral:

$$s^2 = \left(\frac{\sum X^2}{n} - \bar{X}^2\right) \times \frac{n}{n-1}$$

$$s^2 = \left(\frac{243}{9} - 5^2\right) \times \frac{9}{8} = (27 - 25) \times \frac{9}{8} = 2 \times \frac{9}{8} = \frac{9}{4}$$

Logo, o desvio padrão amostral é:

$$s = \sqrt{s^2} = \sqrt{\frac{9}{4}} = \frac{3}{2} = 1{,}5$$

Por fim, sendo o intervalo de confiança igual a 95%, precisamos do valor de t associado a uma probabilidade P(T > t) = 2,5% = 0,025. Pela tabela, observamos que para n – 1 = 8 graus de liberdade, temos t = 2,31.

Substituindo os dados que conhecemos na fórmula do erro acima, temos:

$$L_{SUP} = 5 + 2{,}31.\frac{1{,}5}{\sqrt{9}} = 5 + 2{,}31.\frac{1{,}5}{3} = 5 + 2{,}31 \times 0{,}5 = 5 + 1{,}155 = 6{,}155$$

$$L_{INF} = 5 - 2{,}31.\frac{1{,}5}{\sqrt{9}} = 5 - 2{,}31.\frac{1{,}5}{3} = 5 - 2{,}31 \times 0{,}5 = 5 - 1{,}155 = 3{,}845$$

Gabarito: D

31. (FCC 2016/TRT – 20ª Região) De uma população normalmente distribuída, de tamanho infinito e variância desconhecida, é extraída uma amostra aleatória de tamanho 16 fornecendo um intervalo de confiança de (1 − α) igual a [4,91; 11,30] para a média μ da população. A variância amostral apresentou um valor igual a 36 e considerou-se a distribuição t de Student para obtenção do intervalo de confiança. Consultando a tabela da distribuição t de Student com o respectivo número de graus de liberdade e verificando o valor crítico $t_{\alpha/2}$ tal que a probabilidade $P(|t| > t_{\alpha/2}) = \alpha$, obtém-se que $t_{\alpha/2}$ é igual a

a) 4,26

b) 6,39

c) 2,13

d) 1,65

e) 8,52

Comentários:

Essa questão trabalha com a estimação intervalar para uma população com distribuição normal e variância desconhecida. Nessa situação, o intervalo é dado por:

$$\bar{X} \pm E$$

Em que $E = t.\frac{s}{\sqrt{n}}$.

O enunciado informa que:

- O tamanho da amostra é n = 16, logo $\sqrt{n} = \sqrt{16} = 4$;
- A variância amostral é $s^2 = 36$. Logo, o desvio padrão amostral é $s = \sqrt{s^2} = \sqrt{36} = 6$.

Pela média aritmética do intervalo, podemos calcular o valor da média amostral:

$$\bar{X} = \frac{L_{SUP} + L_{INF}}{2} = \frac{4{,}91 + 11{,}30}{2} = \frac{16{,}21}{2} = 8{,}105$$

A diferença para qualquer um dos extremos fornece o erro amostral:

$$E = 8{,}105 - 4{,}91 = 3{,}195$$

Substituindo os dados que conhecemos na fórmula do erro acima, temos:

$$E = t.\frac{s}{\sqrt{n}}$$

$$3{,}195 = t.\frac{6}{4} = t.\frac{3}{2}$$

$$t = 3{,}195 \times \frac{2}{3} = 2{,}13$$

Gabarito: C

32. (FCC 2016/TRT – 20ª Região) Uma amostra aleatória de tamanho 100 é extraída de uma população P_1 de tamanho infinito, com média μ_1, normalmente distribuída e com desvio padrão populacional igual a 2. Uma outra amostra aleatória, independente da primeira, de tamanho 400 é extraída de uma outra população P_2 de tamanho infinito, com média μ_2, normalmente distribuída e com desvio padrão populacional igual a 3. Considerando que na curva normal padrão (Z) as probabilidades P(Z > 1,64) = 0,05 e P(Z > 1,28) = 0,10 e que as médias das amostras tomadas de P_1 e P_2 foram iguais a 10 e 8, respectivamente, obtém-se que o intervalo de confiança de 90% para ($\mu_1 - \mu_2$) é

a) [1,79; 2,21]

b) [1,64; 2,36]

c) [1,66; 2,34]

d) [1,59; 2,41]

e) [1,68; 2,32]

Comentários:

Essa questão trabalha com a estimação intervalar para a diferença das médias de duas populações. Nessa situação, o intervalo é dado por:

$$(\bar{X} - \bar{Y}) \pm E$$

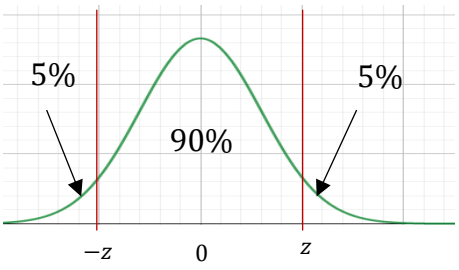
Em que o erro é dado por:

$$E = z.\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

O enunciado informa que:

- A média da primeira população é $\bar{X} = 10$;
- A média da segunda população é $\bar{Y} = 8$;
- O desvio padrão da primeira população é $\sigma_X = 2$; logo, a variância é o quadrado: $\sigma_X^2 = 4$;
- O tamanho da amostra da primeira população é $n_X = 100$;
- O desvio padrão da segunda população é $\sigma_Y = 3$; logo, a variância é o quadrado: $\sigma_Y^2 = 9$;
- O tamanho da amostra da primeira população é $n_Y = 400$;

Ademais, o nível de confiança é de 90%, conforme ilustrado abaixo:



O enunciado informa que P(Z > 1,64) = 0,05 = 5%. Ou seja, o valor de z que delimita um intervalo de confiança de 90% é z = 1,64.

Substituindo esses dados na fórmula do erro, temos:

$$E = 1{,}64.\sqrt{\frac{4}{100}+\frac{9}{400}} = 1{,}64.\sqrt{\frac{16+9}{400}} = 1{,}64.\sqrt{\frac{25}{400}} = 1{,}64.\frac{5}{20} = 0{,}41$$

Logo, o intervalo de confiança é dado por:

$$(10-8) \pm 0{,}41 = 2 \pm 0{,}41 = [1{,}59; 2{,}41]$$

Gabarito: D

VUNESP

33. (VUNESP/2015 – TJ-SP) Leia o texto a seguir para responder à questão.

O Sr. Manoel comprou uma padaria, e foi garantido o faturamento médio de R$ 1.000,00 por dia de funcionamento. Durante os primeiros 16 dias, considerados como uma amostra de 16 valores da população, obteve-se o faturamento médio de R$ 910,00 e desvio padrão de R$ 80,00. Sentindo-se enganado pelo vendedor, o Sr. Manoel entrou com ação de perdas e danos. O juiz sugeriu, então, efetuar o teste de hipótese, indicado ao nível de significância de 5% para confirmar ou refutar a ação.

Supondo-se que a distribuição seja normal com desvio padrão de R$ 120,00 e que a amostra dos 16 dias tenha acusado o valor de R$ 910,00, então o intervalo de confiança para a verdadeira média com 95% de confiança é de, aproximadamente,

a) R$ 850,00 < x < R$ 970,00

b) R$ 800,00 < x < R$ 1.020,00

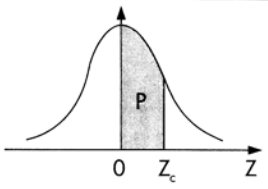
c) R$ 790,00 < x < R$ 1.030,00

d) R$ 900,00 < x < R$ 920,00

e) R$ 890,00 < x < R$ 930,00

Utilize a tabela a seguir para resolver esta e as próximas questões.

Tabela I — Distribuição Normal Padrão
$Z \sim N(0, 1)$
Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

Comentários:

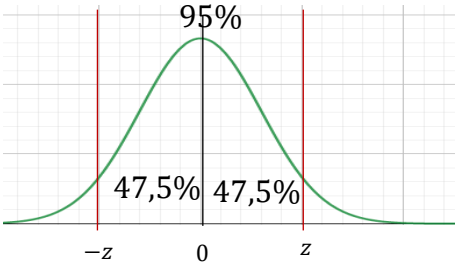
Essa questão trabalha com a estimação intervalar para a média de uma população com distribuição normal e variância conhecida, em que o intervalo de confiança é dado por:

$$\bar{X} \pm z.\frac{\sigma}{\sqrt{n}}$$

O enunciado informa que:

- A média é $\bar{X} = 910$
- O desvio padrão é $\sigma = 120$
- O tamanho da amostra é $n = 16$

O enunciado informa, ainda, que o nível de confiança é de 95%, conforme ilustrado abaixo:



Ou seja, precisamos do valor de z para o qual P(0 < Z < z) = 47,5% = 0,475. Pela tabela fornecida acima, observamos que z = 1,96. Substituindo esse e os demais dados na fórmula do intervalo de confiança, temos:

$$L_{SUP} = 910 + 1{,}96 \times \frac{120}{\sqrt{16}} = 910 + 1{,}96 \times \frac{120}{4} = 910 + 1{,}96 \times 30 \cong 910 + 60 = 970$$

$$L_{INF} = 910 - 1{,}96 \times \frac{120}{\sqrt{16}} = 910 - 1{,}96 \times \frac{120}{4} = 910 - 1{,}96 \times 30 \cong 910 - 60 = 850$$

Gabarito: A

34. (VUNESP/2015 – TJ-SP) Para avaliar o tempo médio de viagem entre o ponto inicial e o ponto final de uma linha de ônibus, retira-se uma amostra de 36 observações (viagens), encontrando-se, para essa amostra, o tempo médio de 50 minutos e o desvio padrão de 6 minutos, com distribuição normal. Considerando-se um intervalo de confiança de 95% para o tempo médio populacional, é correto afirmar que o valor mais próximo para limite inferior desse intervalo é o tempo de

a) 40 min.

b) 44 min.

c) 48 min.

d) 52 min.

e) 56 min.

Comentários:

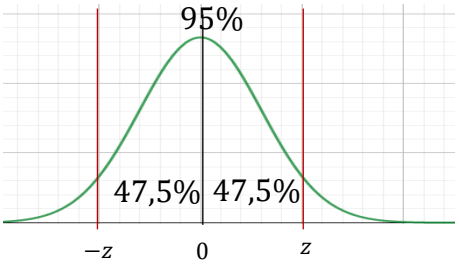
Essa questão trabalha com a estimação intervalar para a média de uma população com distribuição normal e variância conhecida, em que o intervalo de confiança é dado por:

$$\bar{X} \pm z.\frac{\sigma}{\sqrt{n}}$$

O enunciado informa que:

- A média é $\bar{X} = 50$ min
- O desvio padrão é $\sigma = 6$ min
- O tamanho da amostra é $n = 36$

O enunciado informa, ainda, que o nível de confiança é de 95%, conforme ilustrado abaixo:



Ou seja, precisamos do valor de z para o qual P(0 < Z < z) = 47,5% = 0,475. Pela tabela fornecida acima, observamos que z = 1,96. Substituindo esse e os demais dados na fórmula do limite inferior do intervalo de confiança, temos:

$$L_{INF} = 50 - 1{,}96 \times \frac{6}{\sqrt{36}} = 50 - 1{,}96 \times \frac{6}{6} = 50 - 1{,}96 \cong 48$$

Gabarito: C

35. (VUNESP/2014 – TJ-PA) Supondo que em uma amostra de 4 baterias automotivas tenha-se calculado o tempo de vida média de 4 anos. Sabe-se que o tempo de vida da bateria é uma distribuição normal com desvio padrão de 1 ano e meio.

Então, o intervalo de 90% de confiança para a média de todas as baterias é de, aproximadamente:

a) 4 anos $\pm$ 15 meses

b) 4 anos $\pm$ 12 meses

c) 4 anos $\pm$ 10 meses

d) 4 anos $\pm$ 8 meses

e) 4 anos $\pm$ 5 meses

Comentários:

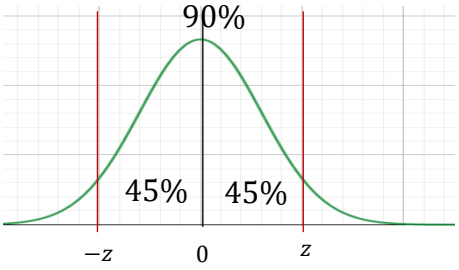
Essa questão trabalha com a estimação intervalar para a média de uma população com distribuição normal e variância conhecida, em que o intervalo de confiança é dado por:

$$\bar{X} \pm z.\frac{\sigma}{\sqrt{n}}$$

O enunciado informa que:

- A média é $\bar{X} = 4$ anos
- O desvio padrão é $\sigma = 1{,}5$ ano
- O tamanho da amostra é $n = 4$

O enunciado informa, ainda, que o nível de confiança é de 90%, conforme ilustrado abaixo:



Ou seja, precisamos do valor de z para o qual P(0 < Z < z) = 45% = 0,45. Pela tabela fornecida acima, observamos que z = 1,64. Substituindo esse e os demais dados na fórmula do intervalo de confiança, temos:

$$4 \pm 1{,}64 \times \frac{1{,}5}{\sqrt{4}} = 4 \pm 1{,}64 \times \frac{1{,}5}{2} = 4 \pm 0{,}82 \times 1{,}5 = 4 \pm 1{,}23$$

1,23 ano corresponde ao seguinte número de meses:

$$1{,}23 \times 12 = 14{,}76 \cong 15$$

Ou seja, o intervalo de confiança é 4 anos $\pm$ 15 meses

Gabarito: A

36. (VUNESP/2014 – TJ-PA) O sindicato dos médicos de uma região divulgou uma nota na qual afirma que os médicos plantonistas daquela região trabalham em média, mais de 40 horas por semana, o que, supostamente, contraria acordos anteriores firmados com a categoria. Para verificar essa afirmativa realizou-se uma nova pesquisa na qual 16 médicos escolhidos de modo aleatório declararam seu tempo de trabalho semanal, obtendo-se para tempo médio e para desvio padrão, respectivamente, os valores 42

horas e 3 horas. Considere-se que a distribuição de frequência das horas trabalhadas pelos médicos é normal. Com esses dados o intervalo de confiança $\mu_x = \overline{x} \pm t_{\alpha/2} \times \frac{s_x}{\sqrt{n}}$ de 95% para a média populacional, é:

a) $\mu_x = 42 \pm 6$

b) $\mu_x = 42 \pm 3$

c) $\mu_x = 42 \pm 2{,}55$

d) $\mu_x = 42 \pm 1{,}6$

e) $\mu_x = 42 \pm 0{,}75$

Comentários:

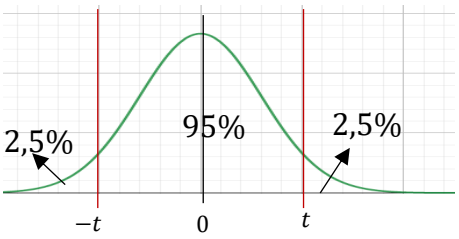
Essa questão trabalha com a estimação intervalar para a média de uma população com distribuição normal e variância desconhecida, em que o intervalo de confiança é dado por:

$$\bar{X} \pm t.\frac{s}{\sqrt{n}}$$

O enunciado informa que:

- A média é $\bar{X} = 42$ horas
- O desvio padrão amostral é $s = 3$ horas
- O tamanho da amostra é $n = 16$

O enunciado informa, ainda, que o nível de confiança é de 1 – p = 95%, logo p = 5%:



Ademais, sabemos que n – 1= 15. Pela tabela, observamos que o valor de t associado a p = 5% e 15 graus de liberdade é t = 2,131. Substituindo esse e os demais dados na fórmula do intervalo de confiança, temos:

$$42 \pm 2{,}131 \times \frac{3}{\sqrt{16}} = 42 \pm 2{,}131 \times \frac{3}{4} \cong 42 \pm 1{,}6$$

Gabarito: D

37. (VUNESP/2013 – MPE-ES) Pesquisa recente sobre o tempo total para que os ônibus de determinada linha urbana percorram todo o trajeto entre o ponto inicial e o ponto final, programados para essa viagem, detectou que os tempos de viagem são normalmente distribuídos com tempo médio gasto de 53 minutos e com desvio-padrão amostral de 9 minutos. Nessa pesquisa, foram observados e computados os dados de 16 viagens escolhidas aleatoriamente.

Com um intervalo de confiança de 98%, utilizando-se a tabela t de Student para estimar o erro amostral, e arredondando para cima o valor desse erro, é correto afirmar que o tempo médio dessa viagem varia entre

a) 45 min e 61 min

b) 47 min e 59 min

c) 50 min e 56 min

d) 51 min e 55 min

e) 52 min e 54 min

Comentários:

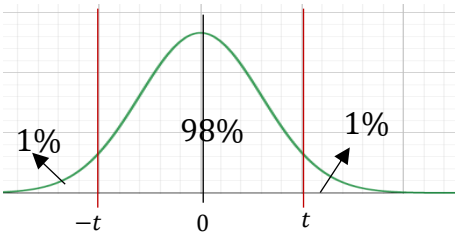
Essa questão trabalha com a estimação intervalar para a média de uma população com distribuição normal e variância desconhecida, em que o intervalo de confiança é dado por:

$$\bar{X} \pm t.\frac{s}{\sqrt{n}}$$

O enunciado informa que:

- A média é $\bar{X} = 53$ minutos
- O desvio padrão amostral é $s = 9$ minutos
- O tamanho da amostra é $n = 16$

O enunciado informa, ainda, que o nível de confiança é de 1 – p = 98%, logo p = 2%:



Ademais, sabemos que n – 1= 15. Pela tabela, observamos que o valor de t associado a p = 2% e 15 graus de liberdade é t = 2,602. Substituindo esse e os demais dados na fórmula do intervalo de confiança, temos:

$$L_{SUP} = 53 + 2{,}602 \times \frac{9}{\sqrt{16}} = 53 + 2{,}602 \times \frac{9}{4} \cong 53 + 6 = 59$$

$$L_{INF} = 53 - 2{,}602 \times \frac{9}{\sqrt{16}} = 53 - 2{,}602 \times \frac{9}{4} \cong 53 - 6 = 47$$

Gabarito: B

38. (VUNESP/2014 – EMPLASA) Um jornal deseja estimar a proporção de jornais impressos com não conformidades. Em uma amostra aleatória de 100 jornais dentre todos os jornais impressos durante um dia, observou-se que 20 têm algum tipo de não conformidade. Para um nível de confiança de 90%, Z = 1,64. Então pode-se concluir que apresentam não conformidades.

a) 21,64%, no máximo

b) 26,56%, no máximo

c) 26,56%, no mínimo

d) 21,64%, no mínimo

e) 20%, no mínimo

Comentários:

Essa questão trabalha com a estimação intervalar para proporções, em que o intervalo de confiança é dado por:

$$\hat{p} \pm z. \sqrt{\frac{\hat{p}.\hat{q}}{n}}$$

O enunciado informa que dos n = 100 jornais, 20 apresentam não conformidade. Logo, a proporção amostral de não conformidade é:

$$\hat{p} = \frac{20}{100} = 0{,}2$$

E o complementar é $\hat{q} = 1 - \hat{p} = 1 - 0{,}2 = 0{,}8$. Sabendo que z = 1,64, temos:

$$L_{SUP} = 0{,}2 + 1{,}64. \sqrt{\frac{0{,}2.0{,}8}{100}} = 0{,}2 + 1{,}64. \frac{\sqrt{0{,}16}}{10} = 0{,}2 + 0{,}164 \times 0{,}4 = 0{,}2656 = 26{,}56\%$$

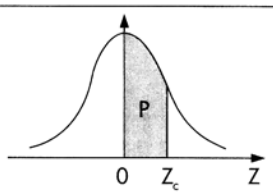
Gabarito: B

LISTA DE QUESTÕES - VUNESP

Estimação Intervalar

Utilize a tabela a seguir para resolver as próximas questões, de 1 a 5.

Tabela I – Distribuição Normal Padrão
$Z \sim N(0, 1)$
Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

1. (VUNESP/2015 – TJ-SP) Leia o texto a seguir para responder à questão.

O Sr. Manoel comprou uma padaria, e foi garantido o faturamento médio de R$ 1.000,00 por dia de funcionamento. Durante os primeiros 16 dias, considerados como uma amostra de 16 valores da população, obteve-se o faturamento médio de R$ 910,00 e desvio padrão de R$ 80,00. Sentindo-se enganado pelo vendedor, o Sr. Manoel entrou com ação de perdas e danos. O juiz sugeriu, então, efetuar o teste de hipótese, indicado ao nível de significância de 5% para confirmar ou refutar a ação.

Supondo-se que a distribuição seja normal com desvio padrão de R$ 120,00 e que a amostra dos 16 dias tenha acusado o valor de R$ 910,00, então o intervalo de confiança para a verdadeira média com 95% de confiança é de, aproximadamente,

a) R$ 850,00 < x < R$ 970,00

b) R$ 800,00 < x < R$ 1.020,00

c) R$ 790,00 < x < R$ 1.030,00

d) R$ 900,00 < x < R$ 920,00

e) R$ 890,00 < x < R$ 930,00

2. (VUNESP/2015 – TJ-SP) Para avaliar o tempo médio de viagem entre o ponto inicial e o ponto final de uma linha de ônibus, retira-se uma amostra de 36 observações (viagens), encontrando-se, para essa amostra, o tempo médio de 50 minutos e o desvio padrão de 6 minutos, com distribuição normal. Considerando-se um intervalo de confiança de 95% para o tempo médio populacional, é correto afirmar que o valor mais próximo para limite inferior desse intervalo é o tempo de

a) 40 min.

b) 44 min.

c) 48 min.

d) 52 min.

e) 56 min.

3. (VUNESP/2014 – TJ-PA) Supondo que em uma amostra de 4 baterias automotivas tenha-se calculado o tempo de vida média de 4 anos. Sabe-se que o tempo de vida da bateria é uma distribuição normal com desvio padrão de 1 ano e meio.

Então, o intervalo de 90% de confiança para a média de todas as baterias é de, aproximadamente:

a) 4 anos $\pm$ 15 meses

b) 4 anos $\pm$ 12 meses

c) 4 anos $\pm$ 10 meses

d) 4 anos $\pm$ 8 meses

e) 4 anos $\pm$ 5 meses

4. (VUNESP/2014 – TJ-PA) O sindicato dos médicos de uma região divulgou uma nota na qual afirma que os médicos plantonistas daquela região trabalham em média, mais de 40 horas por semana, o que, supostamente, contraria acordos anteriores firmados com a categoria. Para verificar essa afirmativa realizou-se uma nova pesquisa na qual 16 médicos escolhidos de modo aleatório declararam seu tempo de trabalho semanal, obtendo-se para tempo médio e para desvio padrão, respectivamente, os valores 42 horas e 3 horas. Considere-se que a distribuição de frequência das horas trabalhadas pelos médicos é normal. Com esses dados o intervalo de confiança $\mu_x = \overline{x} \pm t_{\alpha/2} \times \frac{s_x}{\sqrt{n}}$ de 95% para a média populacional, é:

a) $\mu_x = 42 \pm 6$

b) $\mu_x = 42 \pm 3$

c) $\mu_x = 42 \pm 2{,}55$

d) $\mu_x = 42 \pm 1{,}6$

e) $\mu_x = 42 \pm 0{,}75$

5. (VUNESP/2013 – MPE-ES) Pesquisa recente sobre o tempo total para que os ônibus de determinada linha urbana percorram todo o trajeto entre o ponto inicial e o ponto final, programados para essa viagem, detectou que os tempos de viagem são normalmente distribuídos com tempo médio gasto de 53 minutos e com desvio-padrão amostral de 9 minutos. Nessa pesquisa, foram observados e computados os dados de 16 viagens escolhidas aleatoriamente.

Com um intervalo de confiança de 98%, utilizando-se a tabela t de Student para estimar o erro amostral, e arredondando para cima o valor desse erro, é correto afirmar que o tempo médio dessa viagem varia entre

a) 45 min e 61 min

b) 47 min e 59 min

c) 50 min e 56 min

d) 51 min e 55 min

e) 52 min e 54 min

6. (VUNESP/2014 – EMPLASA) Um jornal deseja estimar a proporção de jornais impressos com não conformidades. Em uma amostra aleatória de 100 jornais dentre todos os jornais impressos durante um

dia, observou-se que 20 têm algum tipo de não conformidade. Para um nível de confiança de 90%, Z = 1,64. Então pode-se concluir que apresentam não conformidades.

a) 21,64%, no máximo

b) 26,56%, no máximo

c) 26,56%, no mínimo

d) 21,64%, no mínimo

e) 20%, no mínimo

GABARITO

1. LETRA A
2. LETRA C
3. LETRA D
4. LETRA D
5. LETRA B
6. LETRA B

LISTA DE QUESTÕES - MULTIBANCAS

Distribuição Amostral

CEBRASPE

1. (CESPE 2019/TJ-AM) Em determinado município brasileiro, realizou-se um levantamento para estimar o percentual P de pessoas que conhecem o programa justiça itinerante. Para esse propósito, foram selecionados 1.000 domicílios por amostragem aleatória simples de um conjunto de 10 mil domicílios. Nos domicílios selecionados, foram entrevistados todos os residentes maiores de idade, que totalizaram 3.000 pessoas entrevistadas, entre as quais 2.250 afirmaram conhecer o programa justiça itinerante.

De acordo com essa situação hipotética, julgue o seguinte item.

O número médio de pessoas maiores de idade por domicílio foi igual a 3 pessoas por domicílio; e o erro padrão do estimador do percentual P é inversamente proporcional a $3\sqrt{1.000}$.

2. (CESPE 2019/TJ-AM) Para avaliar a satisfação dos servidores públicos de certo tribunal no ambiente de trabalho, realizou-se uma pesquisa. Os servidores foram classificados em três grupos, de acordo com o nível do cargo ocupado. Na tabela seguinte, k é um índice que se refere ao grupo de servidores, e N_k denota o tamanho populacional de servidores pertencentes ao grupo k.

Nível do Cargo	k	N_k	n_k	p_k
I	1	500	50	0,7
II	2	300	20	0,8
III	3	200	10	0,9

De cada grupo k foi retirada uma amostra aleatória simples sem reposição de tamanho n_k; p_k representa a proporção de servidores amostrados do grupo k que se mostraram satisfeitos no ambiente de trabalho.

A partir das informações e da tabela apresentadas, julgue o próximo item.

Com relação ao grupo $k = 2$, o erro padrão da estimativa da proporção dos servidores satisfeitos no ambiente de trabalho foi inferior a $0{,}1$.

3. (CESPE 2019/TJ-AM) Em determinado município brasileiro, realizou-se um levantamento para estimar o percentual P de pessoas que conhecem o programa justiça itinerante. Para esse propósito, foram selecionados 1.000 domicílios por amostragem aleatória simples de um conjunto de 10 mil domicílios. Nos

domicílios selecionados, foram entrevistados todos os residentes maiores de idade, que totalizaram 3.000 pessoas entrevistadas, entre as quais 2.250 afirmaram conhecer o programa justiça itinerante.

De acordo com essa situação hipotética, julgue o seguinte item.

A estimativa do percentual de pessoas que conhecem o programa justiça itinerante foi inferior a 60%.

4. (CESPE 2019/TJ-AM) Para estimar a proporção de menores infratores reincidentes em determinado município, foi realizado um levantamento estatístico. Da população-alvo desse estudo, constituída por 10.050 menores infratores, foi retirada uma amostra aleatória simples sem reposição, composta por 201 indivíduos. Nessa amostra foram encontrados 67 reincidentes.

Com relação a essa situação hipotética, julgue o seguinte item.

Se a amostragem fosse com reposição, a estimativa da variância da proporção amostral teria sido superior a 0,001.

5. (CESPE 2018/PF) Determinado órgão governamental estimou que a probabilidade p de um ex-condenado voltar a ser condenado por algum crime no prazo de 5 anos, contados a partir da data da libertação, seja igual a 0,25. Essa estimativa foi obtida com base em um levantamento por amostragem aleatória simples de 1.875 processos judiciais, aplicando-se o método da máxima verossimilhança a partir da distribuição de Bernoulli.

Sabendo que $P(Z < 2) = 0,975$, em que Z representa a distribuição normal padrão, julgue o item que se segue, em relação a essa situação hipotética.

O erro padrão da estimativa da probabilidade p foi igual a 0,01.

6. (CESPE 2018/PF) Uma pesquisa realizada com passageiros estrangeiros que se encontravam em determinado aeroporto durante um grande evento esportivo no país teve como finalidade investigar a sensação de segurança nos voos internacionais. Foram entrevistados 1.000 passageiros, alocando-se a amostra de acordo com o continente de origem de cada um — África, América do Norte (AN), América do Sul (AS), Ásia/Oceania (A/O) ou Europa. Na tabela seguinte, N é o tamanho populacional de passageiros em voos internacionais no período de interesse da pesquisa; n é o tamanho da amostra por origem; P é o percentual dos passageiros entrevistados que se manifestaram satisfeitos no que se refere à sensação de segurança.

Origem	N	n	P
África	100.000	100	80

AN	300.000	300	70
AS	100.000	100	90
A/O	300.000	300	80
Europa	200.000	200	80
Total	1.000.000	1.000	P_{pop}

Em cada grupo de origem, os passageiros entrevistados foram selecionados por amostragem aleatória simples. A última linha da tabela mostra o total populacional no período da pesquisa, o tamanho total da amostra e P_{pop} representa o percentual populacional de passageiros satisfeitos.

A partir dessas informações, julgue o item.

Considerando o referido desenho amostral, estima-se que o percentual populacional P_{pop} seja inferior a 79%.

7. (CESPE 2018/STM) Em um tribunal, entre os processos que aguardam julgamento, foi selecionada aleatoriamente uma amostra contendo 30 processos. Para cada processo da amostra que estivesse há mais de 5 anos aguardando julgamento, foi atribuído o valor 1; para cada um dos outros, foi atribuído o valor 0. Os dados da amostra são os seguintes:

1 1 0 0 1 0 1 1 0 1 1 1 1 0 1 1 1 0 1 0 1 0 1 0 1 1 1 0 1 1

A proporção populacional de processos que aguardam julgamento há mais de 5 anos foi denotada por p; a proporção amostral de processos que aguardam julgamento há mais de 5 anos foi representada por $\hat{p}$

A variância da proporção amostral $\hat{p}$ sob a hipótese nula H_0: $p = 0{,}5$ é menor que 0,1.

8. (CESPE 2018/STM) Em um tribunal, entre os processos que aguardam julgamento, foi selecionada aleatoriamente uma amostra contendo 30 processos. Para cada processo da amostra que estivesse há mais de 5 anos aguardando julgamento, foi atribuído o valor 1; para cada um dos outros, foi atribuído o valor 0. Os dados da amostra são os seguintes:

1 1 0 0 1 0 1 1 0 1 1 1 1 0 1 1 1 0 1 0 1 0 1 0 1 1 1 0 1 1

A proporção populacional de processos que aguardam julgamento há mais de 5 anos foi denotada por p; a proporção amostral de processos que aguardam julgamento há mais de 5 anos foi representada por $\hat{p}$

Estima-se que, nesse tribunal, $p > 60\%$.

9. (CESPE 2018/EBSERH) $X_1, X_2, ..., X_{10}$ representa uma amostra aleatória simples retirada de uma distribuição normal com média μ e variância σ^2, ambas desconhecidas. Considerando que $\hat{\mu}$ e $\hat{\sigma}$ representam os respectivos estimadores de máxima verossimilhança desses parâmetros populacionais, julgue o item subsecutivo.

A razão $\frac{\hat{\mu}-\mu}{\sigma}$ segue uma distribuição normal padrão.

10. (CESPE 2016/TCE-PA) Uma amostra aleatória simples $X_1, X_2, ..., X_n$ foi retirada de uma população normal com média e desvio padrão iguais a 10. Julgue o próximo item, a respeito da média amostral $\bar{X} = [X_1 + X_2 + \ldots + X_n]/n$.

A variância de $\bar{X}$ é igual a 100.

11. (CESPE 2016/TCE-PA) Uma amostra aleatória simples $X_1, X_2, ..., X_n$ foi retirada de uma população normal com média e desvio padrão iguais a 10. Julgue o próximo item, a respeito da média amostral $\bar{X} = [X_1 + X_2 + \ldots + X_n]/n$.

A média amostral segue uma distribuição t de *Student* com $n - 1$ graus de liberdade.

12. (CESPE 2016/TCE-PA) Uma amostra aleatória simples $X_1, X_2, ..., X_n$ foi retirada de uma população normal com média e desvio padrão iguais a 10. Julgue o próximo item, a respeito da média amostral $\bar{X} = [X_1 + X_2 + \ldots + X_n]/n$.

$P(\bar{X} - 10 > 0) \leq 0{,}5$

13. (CESPE 2016/TCE-PA) Considere um processo de amostragem de uma população finita cuja variável de interesse seja binária e assuma valor 0 ou 1, sendo a proporção de indivíduos com valor 1 igual a p = 0,3. Considere, ainda, que a probabilidade de cada indivíduo ser sorteado seja a mesma para todos os indivíduos da amostragem e que, após cada sorteio, haja reposição do indivíduo selecionado na amostragem. A partir dessas informações, julgue o item subsequente.

Se, dessa população, for coletada uma amostra aleatória de tamanho $n = 1$, a probabilidade de um indivíduo apresentar valor 1 é igual a 0,5.

FCC

14. (FCC 2015/DPE-SP) Uma amostra aleatória simples, com reposição, de n observações $X_1, X_2, \ldots X_n$ foi selecionada de uma população com distribuição uniforme contínua no intervalo $[-2, b]$, $b > -2$. Sabe-se que:

I. a média dessa distribuição uniforme é igual a 10.

II. o desvio padrão de $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ é igual a 0,4.

Nessas condições, o valor de n é igual a,

a) 100.

b) 400.

c) 225.

d) 300.

e) 324.

15. (FCC 2015/TRT 3ª Região) Se Z tem distribuição normal padrão, então:

P(Z < 0,5) = 0,591; P(Z < 1) = 0,841; P(Z < 1,15) = 0,8951; P(Z < 1,17) = 0,879; P(Z < 1,2) = 0,885; P(Z<1,4)=0,919;P(Z<1,64)=0,95;P(Z<2)=0,977;P(Z<2,06)=0,98;P(Z<2,4)= 0,997.

Considere que X é a variável aleatória, que representa as idades, em anos, dos trabalhadores de certa indústria. Suponha que X tem distribuição normal com média de μ anos e desvio padrão de 5 anos.

Uma amostra aleatória, com reposição, de 16 trabalhadores será selecionada e sejam $X_1, X_2, \ldots, X_{16}$ as idades observadas e $\overline{X} = \frac{\sum_{1}^{16} X_i}{n}$ a média desta amostra. Sabendo-se que a probabilidade de $\overline{X}$ ser superior a 30 anos é igual a 0,919, o valor de μ, em anos, é igual a:

a) 28,25

b) 31,75

c) 30,50

d) 32,50

e) 30,85

GABARITO

1. CERTO
2. CERTO
3. ERRADO
4. CERTO
5. CERTO
6. CERTO
7. CERTO
8. CERTO
9. ERRADO
10. ERRADO
11. ERRADO
12. CERTO
13. ERRADO
14. LETRA D
15. LETRA B

LISTA DE QUESTÕES - MULTIBANCAS

Estimação Pontual

CEBRASPE

1. (CESPE 2018/STM) Diversos processos buscam reparação financeira por danos morais. A tabela seguinte mostra os valores, em reais, buscados em 10 processos — numerados de 1 a 10 — de reparação por danos morais, selecionados aleatoriamente em um tribunal.

Processo	Valor
1	3.700
2	3.200
3	2.500
4	2.100
5	3.000
6	5.200
7	5.000
8	4.000
9	3.200
10	3.100

A partir dessas informações e sabendo que os dados seguem uma distribuição normal, julgue o item subsequente.

Se μ = estimativa pontual para a média dos valores buscados como reparação por danos morais no referido tribunal, então $3.000 < \mu < 3.300$.

2. (CESPE 2016/TCE-PA) Considerando uma população finita em que a média da variável de interesse seja desconhecida, julgue o item a seguir.

Considere uma amostragem com três estratos, cujos pesos populacionais sejam 0,2, 0,3 e 0,5. Considere, ainda, que os tamanhos das amostras em cada estrato correspondam, respectivamente, a n_1= 20, n_2= 30 e n_3= 50, e que as médias amostrais sejam 12 kg, 6 kg e 8 kg, respectivamente. Nessa situação, a estimativa pontual da média populacional, com base nessa amostra, é igual a 8,2 kg.

3. (CESPE 2016/TCE-PA) Considere um processo de amostragem de uma população finita cuja variável de interesse seja binária e assuma valor 0 ou 1, sendo a proporção de indivíduos com valor 1 igual a p = 0,3. Considere, ainda, que a probabilidade de cada indivíduo ser sorteado seja a mesma para todos os indivíduos da amostragem e que, após cada sorteio, haja reposição do indivíduo selecionado na amostragem. A partir dessas informações, julgue o item subsequente.

Caso, em uma amostra de tamanho n = 10, os valores observados sejam A = {1, 0, 1, 0, 1, 0, 0, 1, 0, 0}, a estimativa via estimador de máxima verossimilhança para a média populacional será igual a 0,4.

4. (CESPE 2018/ABIN) A quantidade diária de emails indesejados recebidos por um atendente é uma variável aleatória X que segue distribuição de Poisson com média e variância desconhecidas. Para estimá-las, retirou-se dessa distribuição uma amostra aleatória simples de tamanho quatro, cujos valores observados foram 10, 4, 2 e 4.

Com relação a essa situação hipotética, julgue o seguinte item.

No que se refere à média amostral $\bar{X} = \frac{(X_1+X_2+X_3+X_4)}{4}$, na qual X_1, X_2, X_3, X_4 representa uma amostra aleatória simples retirada dessa distribuição X, é correto afirmar que a estimativa da variância do estimador $\bar{X}$ seja igual a 1,25.

5. (CESPE 2015/Telebras) Um analista da TELEBRAS, a fim de verificar o tempo durante o qual um grupo de consumidores ficou sem o serviço de Internet do qual eram usuários, selecionou uma amostra de 10 consumidores críticos. Os dados coletados, em minutos, referentes a esses consumidores foram listados na tabela seguinte.

consumidor	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
tempo (em minutos)	8	2	3	5	7	7	10	9	4	5

Com base nessa situação hipotética, julgue o item subsequente.

Se os dados seguissem uma distribuição normal, a expressão matemática que permite calcular a variância estimada pelo método de máxima verossimilhança teria denominador igual a 9.

6. (CESPE 2018/ABIN) Considerando que os principais métodos para a estimação pontual são o método dos momentos e o da máxima verossimilhança, julgue o item a seguir.

Para a distribuição normal, o método dos momentos e o da máxima verossimilhança fornecem os mesmos estimadores aos parâmetros μ e σ.

7. (CESPE 2020/TJ-PA) Considerando que a inferência estatística é um processo que consiste na utilização de observações feitas em uma amostra com o objetivo de estimar as propriedades de uma população, assinale a opção correta.

a) Na estatística inferencial, um parâmetro é um valor conhecido, extraído de uma amostra, utilizado para a estimação de uma grandeza populacional.

b) Independentemente do tamanho da amostra, um estimador consistente sempre irá convergir para o verdadeiro valor da grandeza populacional.

c) A amplitude de uma amostra definirá se a média amostral poderá ser um estimador de máxima verossimilhança da média populacional.

d) Sendo â um estimador de máxima verossimilhança de um parâmetro a, então $(a)^{1/2} = (\hat{a})^{1/2}$.

e) Sendo a e b estimadores de um mesmo parâmetro cujas variâncias são simbolizadas por Var(a) e Var(b). Se Var(a) > Var(b), então é correto afirmar que a é um melhor estimador que b.

8. (CESPE 2018/PF – Papiloscopista) Em determinado município, o número diário X de registros de novos armamentos segue uma distribuição de Poisson, cuja função de probabilidade é expressa por $P(X = k) = \frac{e^{-M}.M^k}{k!}$ em que k = 0, 1, 2, ..., e M é um parâmetro.

	dia				
	1	2	3	4	5
realização da variável X	6	8	0	4	2

Considerando que a tabela precedente mostra as realizações da variável aleatória X em uma amostra aleatória simples constituída por cinco dias, julgue o item que segue.

A estimativa de máxima verossimilhança do desvio padrão da distribuição da variável X é igual a 2 registros por dia.

FGV

9. (FGV/2014 – Prefeitura de Recife/PE) Avalie se as seguintes propriedades de um estimador de um certo parâmetro são desejáveis:

I. Ser não tendencioso para esse parâmetro.

II. Ter variância grande.

III. Ter erro quadrático médio grande.

Assinale:

a) se apenas a propriedade I estiver correta.

b) se apenas as propriedades I e II estiverem corretas.

c) se apenas as propriedades I e III estiverem corretas.

d) se apenas as propriedades II e III estiverem corretas.

e) se todas as propriedades estiverem corretas.

10. (FGV/2017 – IBGE) Dois estimadores, $\widehat{\theta}_1$ e $\widehat{\theta}_2$, para um parâmetro populacional θ, têm seus Erros Quadráticos Médios (EQM) dados por:

$$EQM(\widehat{\theta}_1) = \frac{3.\sigma^2}{n} + \left(\frac{\theta-1}{n}\right)^2 \text{ e } EQM(\widehat{\theta}_2) = \frac{\sqrt{n}.\sigma^2}{s} + \left(1 + \frac{\theta}{n}\right)^2$$

Com base apenas nas expressões, onde as primeiras parcelas são as variâncias, é correto concluir que:

a) $\hat{\theta}_1$ é não tendencioso;

b) $\hat{\theta}_1$ é um estimador consistente;

c) $\hat{\theta}_2$ é assintoticamente não tendencioso;

d) $\hat{\theta}_2$ não é um estimador consistente;

e) ambos são assintoticamente eficientes.

11. (FGV/2019 – DPE-RJ) Sejam θ_1, θ_2 e θ_3 estimadores de um parâmetro populacional θ gerados a partir de uma amostra do tipo AAS de tamanho n.

Sabe-se ainda que θ_1 é eficiente quando comparada com uma certa classe de estimadores, que θ_2 e θ_3 são tendenciosos, mas θ_2 não é assintoticamente tendencioso. Então:

a) Se $\lim_{n\to\infty} V(\theta_3) \neq 0$, o estimador não é consistente;

b) Se $\theta^* = \frac{\theta_2+\theta_3}{2}$, então θ^* é um estimador inconsistente de θ;

c) Se $\theta^{**} = \theta_1 + \theta_2 - \theta_3$, então θ^{**} é não tendencioso;

d) Se $\lim_{n\to\infty} V(\theta_1) = 0$, então θ_1 é um estimador consistente;

e) Se $\lim_{n\to\infty} EQM(\theta_2) \neq 0$, então θ_2 não é consistente.

12. (FGV/2017 – IBGE) Para estimar a média de certa população μ, desconhecida, partindo apenas de duas observações amostrais, cogita-se o emprego de um dos seguintes estimadores, onde X_1 e X_2 representam os indivíduos da amostra ex ante.

$\widehat{\theta} = \frac{2}{3}X_1 + \frac{1}{3}X_2$ **e** $\widetilde{\theta} = \frac{2}{7}X_1 + \frac{4}{7}X_2$

Sobre os estimadores, é correto afirmar que:

a) o estimador $\tilde{\theta}$ é mais eficiente do que $\hat{\theta}$;

b) o estimador $\tilde{\theta}$ subestima, em média, o valor verdadeiro de μ;

c) a tendenciosidade de $\hat{\theta}$, T ($\hat{\theta}$), é igual a μ/4;

d) o Erro Quadrático Médio de $\tilde{\theta}$, EQM($\tilde{\theta}$), é $\frac{20}{49}\cdot\sigma^2$;

e) a tendenciosidade de $\tilde{\theta}$, T($\tilde{\theta}$), é igual a $\frac{-\mu}{7}$.

13. (FGV/2015 – TJ-BA) Para estimar a média populacional de uma distribuição, com base em uma amostra de tamanho n = 3, são propostos os seguintes estimadores:

$$\widehat{\mu} = \frac{3}{7}X_1 + \frac{5}{7}X_2 - \frac{1}{7}X_3$$

$$\widetilde{\mu} = \frac{1}{3}X_1 + \frac{2}{5}X_2 + \frac{1}{4}X_3$$

$$\bar{\bar{\mu}} = \frac{1}{12}X_1 + \frac{1}{4}X_2 + \frac{2}{3}X_3$$

Sobre esses estimadores é correto afirmar que:

a) os estimadores são todos não tendenciosos;

b) apenas o estimador $\hat{\mu}$ é não viesado e eficiente com relação aos demais;

c) exceto $\bar{\bar{\mu}}$, os outros estimadores são não tendenciosos e consistentes;

d) dentre os três estimadores sugeridos, o que apresenta a menor variância é $\tilde{\boldsymbol{\mu}}$, mas não é o mais eficiente;

e) o erro quadrático médio de $\bar{\bar{\boldsymbol{\mu}}}$ difere da sua variância a razão de 1/3 da variância populacional

14. (FGV 2018/AL-RO) Suponha que X_1, X_2, ..., X_n seja uma amostra aleatória simples de uma variável aleatória populacional qualquer com média μ e variância finita. Considere os seguintes estimadores de μ:

$T_1 = X_1$
$T_2 = X_1 + X_2 + X_3 – X_4 – X_5$.
$T_3 = (X_1 + X_2 + X_3)/3$.
$T_4 = X_1 – X_2$.
$T_5 = (X_1 + X_2 + X_3 + X_4 + X_5)/5$.

São estimadores não tendenciosos de μ:

a) T_1 e T_2, somente.

b) T_1, T_2 e T_3, somente.

c) T_1, T_2, T_3 e T_5, somente.

d) T_2, T_3, T_4 e T_5, somente.

e) T_1, T_2, T_3, T_4 e T_5.

15. (FGV 2018/AL-RO) Suponha que X_1, X_2, ..., X_n seja uma amostra aleatória simples de uma variável aleatória populacional qualquer com média μ e variância finita. Considere os seguintes estimadores de μ:

$T_1 = X_1$
$T_2 = X_1 + X_2 + X_3 – X_4 – X_5$.
$T_3 = (X_1 + X_2 + X_3)/3$.
$T_4 = X_1 – X_2$.
$T_5 = (X_1 + X_2 + X_3 + X_4 + X_5)/5$.

O estimador não tendencioso de variância uniformemente mínima de μ é:

a) T_1

b) T_2

c) T_3

d) T_4

e) T_5

16. (FGV 2014/DPE-RJ) Sejam $\widehat{\theta_1}$ e $\widehat{\theta_2}$ dois estimadores pontuais, ambos não tendenciosos e igualmente eficientes do parâmetro θ (Var($\widehat{\theta_1}$) = Var($\widehat{\theta_2}$)) , sendo que a covariância entre eles é igual a ½.Var($\widehat{\theta_1}$). Então, também é não tendencioso e mais eficiente o estimador

a) $\frac{2}{3}\widehat{\theta_1} + \frac{1}{3}\widehat{\theta_2}$

b) $\frac{1}{5}\widehat{\theta_1} + \frac{3}{4}\widehat{\theta_2}$

c) $\frac{3}{5}\widehat{\theta_1} + \frac{4}{7}\widehat{\theta_2}$

d) $\frac{\widehat{\theta_1}+\widehat{\theta_2}}{2}$

e) $\frac{3}{8}\widehat{\theta_1} + \frac{7}{8}\widehat{\theta_2}$

17. (FGV 2014/DPE-RJ) Seja o estimador $\widehat{\theta}$ de um parâmetro populacional θ tal que $EQM(\widehat{\theta}) - Var(\widehat{\theta}) = \left(k - \frac{1}{n}\right)^2$, onde k (≠ zero) é uma constante que depende do verdadeiro valor de θ e n é o tamanho da amostra. Então, o estimador será

a) assintoticamente eficiente se $\lim_{n\to\infty} Var(\hat{\theta}) = 0$

b) assintoticamente tendencioso.

c) assintoticamente tendencioso, subestimando o parâmetro θ.

d) assintoticamente tendencioso, superestimando o parâmetro θ.

e) consistente, desde que $\lim_{n\to\infty} EQM(\hat{\theta}) = (k)^2$

18. (FGV 2018/AL-RO) Se X_1, X_2, ..., X_n é uma amostra aleatória simples de uma distribuição Bernoulli (p), então o estimador de máxima verossimilhança da variância populacional é

a) $\bar{X}$

b) $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$

c) $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$

d) $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{2n}$

e) $\bar{X}(1-\bar{X})$

19. (FGV 2018/AL-RO) Se $X_1, X_2, ..., X_n$ é uma amostra aleatória simples de uma distribuição exponencial com parâmetro θ, ou seja,

$f(x|\theta) = \theta e^{-\theta x}$, $\theta > 0$,

então, o estimador de θ pelo método dos momentos é

a) $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

b) $\frac{1}{\bar{X}} = \frac{n}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$

c) $\bar{X}^2$

d) $\sum_{i=1}^{n} X_i$

e) $\frac{\sum_{i=1}^{n} X_i}{2}$

FCC

20. (FCC 2015/DPE-SP) Dois estimadores não viesados $E_1 = mX + (m - 1)Y - (2m - 2)Z$ e $E_2 = 1,5X - Y + 0,5Z$ são utilizados para estimar a média μ de uma população normal e variância σ^2 diferente de zero. O parâmetro m é um número real e (X, Y, Z) corresponde a uma amostra aleatória, com reposição, da população. Se E_1 é mais eficiente que E_2, então

a) $m < 1/6$ ou $m > 2$

b) $3/2 < m < 2$

c) $1/6 < m < 2$

d) $1/6 < m < 3/2$

e) $1 < m < 2$

21. (FCC 2013/TRT 5ª Região) Em 20 experiências de 4 provas cada uma, obteve-se a seguinte distribuição:

x_i	0	1	2	3
n_i	10	8	1	1

Observação: n_i é o número de experiências nas quais um determinado acontecimento ocorreu x_i vezes.

Admitindo que este acontecimento trata de uma variável aleatória X obedecendo a uma distribuição binomial $P(x) = C_m^x . p^x (1-p)^{m-x}$, em que x é o número de ocorrências de um certo acontecimento em m provas, tem-se, com base nas 20 experiências, que a estimativa pontual de p pelo método da máxima verossimilhança é

a) 65,00%

b) 50,00%

c) 48,75%

d) 32,50%

e) 16,25%

22. (FCC 2016/TRT 20ª Região) Em uma sala estão presentes algumas pessoas e somente duas delas têm nível superior, sendo que o número de pessoas sem nível superior é desconhecido e sabendo-se apenas que é um número par. Foram selecionadas, desta sala, aleatoriamente, com reposição, 4 pessoas verificando-se que 3 delas não têm nível superior. Com base nesta seleção e utilizando o método da máxima verossimilhança encontra-se a estimativa do número de pessoas sem nível superior. Com isto, o número estimado total de pessoas presentes na sala é igual a

a) 12

b) 6

c) 10

d) 14

e) 8

GABARITO

1. ERRADO
2. CERTO
3. CERTO
4. CERTO
5. ERRADO
6. CERTO
7. LETRA D
8. CERTO
9. LETRA A
10. LETRA B
11. LETRA D
12. LETRA E
13. LETRA D
14. LETRA C
15. LETRA E
16. LETRA D
17. LETRA B
18. LETRA E
19. LETRA B
20. LETRA D
21. LETRA E
22. LETRA E

LISTA DE QUESTÕES - MULTIBANCAS

Estimação Intervalar

CEBRASPE

1. (CESPE 2020/TJ-PA) Para determinado experimento, uma equipe de pesquisadores gerou 20 amostras de tamanho $n = 25$ de uma distribuição normal, com média $\mu = 5$ e desvio padrão $\sigma = 3$. Para cada amostra, foi montado um intervalo de confiança com coeficiente de 0,95 (ou 95%). Com base nessas informações, julgue os itens que se seguem.

I. Os intervalos de confiança terão a forma $\beta_i \pm 1,176$, em que β_i é a média da amostra i.
II. Para todos os intervalos de confiança, $\beta_i + \varepsilon \geq \mu \geq \beta_i - \varepsilon$, sendo ε a margem de erro do estimador.
III. Se o tamanho da amostra fosse maior, mantendo-se fixos os valores do desvio padrão e do nível de confiança, haveria uma redução da margem de erro ε.

Assinale a opção correta.

a) Apenas o item II está certo.

b) Apenas os itens I e II estão certos.

c) Apenas os itens I e III estão certos.

d) Apenas os itens II e III estão certos.

e) Todos os itens estão certos.

2. (CESPE 2020/TJ-PA). Ao analisar uma amostra aleatória simples composta de 324 elementos, um pesquisador obteve, para os parâmetros média amostral e variância amostral, os valores 175 e 81, respectivamente.

Nesse caso, um intervalo de 95% de confiança de μ é dado por

a) (166,18; 183,82).

b) (174,02; 175,98).

c) (174,51; 175,49).

d) (163,35; 186,65).

e) (174,1775; 175,8225).

3. (CESPE 2020/TJ-PA). A respeito dos intervalos de confiança, julgue os próximos itens.

I. **Um intervalo de confiança tem mais valor do que uma estimativa pontual única, pois uma estimativa pontual não fornece nenhuma informação sobre o grau de precisão da estimativa.**
II. **Um intervalo de confiança poderá ser reduzido se o nível de confiança for menor e o valor da variância populacional for maior.**
III. **No cálculo de um intervalo de confiança para a média, deve-se utilizar a distribuição t em lugar da distribuição normal quando a variância populacional é desconhecida e o número de observações é inferior a 30.**

Assinale a opção correta.

a) Apenas o item II está certo.

b) Apenas os itens I e II estão certos.

c) Apenas os itens I e III estão certos.

d) Apenas os itens II e III estão certos.

e) Todos os itens estão certos.

4. (CESPE 2020/TJ PA) Na construção de um intervalo de confiança para a média, conhecida a variância, considerando o intervalo na forma $[x + \varepsilon;\ x - \varepsilon]$, sendo x o valor do estimador da média e ε a semi-amplitude do intervalo de confiança ou, como é mais popularmente conhecida, a margem de erro do intervalo de confiança. Considere que, para uma determinada peça automotiva, um lote de 100 peças tenha apresentado espessura média de 4,561 polegada, com desvio padrão de 1,125 polegada. Um intervalo de confiança de 95% para a média apresentou limite superior de 4,7815 e limite inferior de 4,3405. Nessa situação, a margem de erro do intervalo é de, aproximadamente,

a) $\varepsilon = 0{,}4410$.

b) $\varepsilon = 0{,}3436$.

c) $\varepsilon = 0{,}2205$.

d) $\varepsilon = 0{,}1125$.

e) $\varepsilon = 0{,}1103$.

5. (CESPE 2020/TJ-PA) Tabela da distribuição T fornecida.

Em uma amostra aleatória de 20 municípios Paraenses, considerando-se os dados da Secretaria de Estado de Segurança Pública e Defesa Social relativos ao crime de lesão corporal, a média é igual a 87 e o desvio padrão igual a 101,9419.

Considerando-se, para 19 graus de liberdade, o coeficiente a = 2,093 e utilizando-se o valor aproximado 4,4721 para a raiz quadrada de 20, com o auxílio da distribuição t, um intervalo de 95% de confiança para a média deverá ter

a) Limite inferior de, aproximadamente, 38,78.

b) Limite superior de, aproximadamente, 143,12.

c) Amplitude 2c = 93,45.

d) Limite inferior de 39,29 e limite superior de 142,18.

e) Limite superior de, aproximadamente, 134,71.

6. (CESPE 2018/EBSERH) Uma amostra aleatória simples Y1, Y2, ..., Y25 foi retirada de uma distribuição normal com média nula e variância σ², desconhecida. Considerando que P(x² < 13) = P(X² > 41) = 0,025, em que x² representa a distribuição qui-quadrado com 25 graus de liberdade, e que $S^2 = \sum_{t=1}^{25} Y_i^2$, julgue o item a seguir.

[S²/41;S²/13] representa um intervalo de 95% de confiança para a variância σ².

7. (CESPE 2018/PF) O tempo gasto (em dias) na preparação para determinada operação policial é uma variável aleatória X que segue distribuição normal com média M, desconhecida, e desvio padrão igual a 3 dias. A observação de uma amostra aleatória de 100 outras operações policiais semelhantes a essa produziu uma média amostral igual a 10 dias.

Com referência a essas informações, julgue o item que se segue, sabendo que P(Z > 2) = 0,025, em que Z denota uma variável aleatória normal padrão.

A expressão 10 dias ± 6 dias corresponde a um intervalo de 95% de confiança para a média populacional M.

8. (CESPE 2016/TCE-PA) Em estudo acerca da situação do CNPJ das empresas de determinado município, as empresas que estavam com o CNPJ regular foram representadas por 1, ao passo que as com CNPJ irregular foram representadas por 0.

Considerando que a amostra

{0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1}

Foi extraída para realizar um teste de hipóteses, julgue o item subsequente.

Uma vez que a amostra é menor que 30, a estatística do teste utilizada segue uma distribuição t de Student.

9. (CESPE 2016/TCE-PA) Em estudo acerca da situação do CNPJ das empresas de determinado município, as empresas que estavam com o CNPJ regular foram representadas por 1, ao passo que as com CNPJ irregular foram representadas por 0.

Considerando que a amostra

{0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1}

Foi extraída para realizar um teste de hipóteses, julgue o item subsequente.

Sendo P(Z > 1,96) = 0,025 e P(Z > 1,645) = 0,05, em que Z representa a variável normal padronizada, e P(t_{20} > 2,086) = 0,025 e P(t_{19} > 1,729) = 0,05, em que t_{20} e t_{19} possuem distribuição t de Student com, respectivamente, 20 e 19 graus de liberdade, o erro utilizado para a construção do intervalo de confiança é menor que 15%, se considerado um nível de significância de 5%.

10. (CESPE 2016/TCE-PR) A partir de um levantamento estatístico por amostragem aleatória simples em que se entrevistaram 2.400 trabalhadores, uma seguradora constatou que 60% deles acreditam que poderão manter seu atual padrão de vida na aposentadoria.

Considerando que P(|Z|≤3)=0,99, em que Z representa a distribuição normal padrão, assinale a opção correspondente ao intervalo de 99% de confiança para o percentual populacional de trabalhadores que acreditam que poderão manter seu atual padrão de vida na aposentadoria.

a) 60,0% ± 1,0%

b) 60,0% ± 1,5%

c) 60,0% ± 3,0%

d) 60,0% ± 0,2%

e) 60,0% ± 0,4%

11. (CESPE 2016/TCE-PA) Suponha que o tribunal de contas de determinado estado disponha de 30 dias para analisar as contas de 800 contratos firmados pela administração. Considerando que essa análise é necessária para que a administração pública possa programar o orçamento do próximo ano e que o resultado da análise deve ser a aprovação ou rejeição das contas, julgue o item a seguir. Sempre que necessário, utilize que P(Z > 1,96) = 0,025 e P(Z > 1,645) = 0,05, em que Z representa a variável normal padronizada.

Se forem aprovados 90% dos contratos de uma amostra composta de 100 contratos, o erro amostral será superior a 10%.

12. (CESPE 2016/TCE-PA) Suponha que o tribunal de contas de determinado estado disponha de 30 dias para analisar as contas de 800 contratos firmados pela administração. Considerando que essa análise é necessária para que a administração pública possa programar o orçamento do próximo ano e que o resultado da análise deve ser a aprovação ou rejeição das contas, julgue o item a seguir. Sempre que necessário, utilize que P(Z > 1,96) = 0,025 e P(Z > 1,645) = 0,05, em que Z representa a variável normal padronizada.

Considerando-se que, no ano anterior ao da análise em questão, 80% dos contratos tenham sido aprovados e que 0,615 seja o valor aproximado de 1,962×0,8×0,2, é correto afirmar que a quantidade de contratos de uma amostra com nível de 95% de confiança para a média populacional e erro amostral de 5% é inferior a 160.

13. (CESPE 2016/TCE-PA) A respeito de uma amostra de tamanho n = 10, com os valores amostrados {0,10, 0,06, 0,10, 0,12, 0,08, 0,10, 0,05, 0,15, 0,14, 0,11}, extraídos de determinada população, julgue o item seguinte.

Por um intervalo de confiança frequentista igual a (–0,11, 0,32), entende-se que a probabilidade de o parâmetro médio ser superior a –0,11 e inferior a 0,32 é igual ao nível de confiança γ.

14. (CESPE 2016/TCE-PA) Uma amostra aleatória, com n = 16 observações independentes e identicamente distribuídas (IID), foi obtida a partir de uma população infinita, com média e desvio padrão desconhecidos e distribuição normal.

Tendo essa informação como referência inicial, julgue o seguinte item.

Se a média amostral for igual a 3,2 e a variância amostral, igual a 4,0, o estimador de máxima verossimilhança para a média populacional será igual a 1,6.

15. (CESPE 2016/TCE-PA) Uma amostra aleatória, com n = 16 observações independentes e identicamente distribuídas (IID), foi obtida a partir de uma população infinita, com média e desvio padrão desconhecidos e distribuição normal.

Tendo essa informação como referência inicial, julgue o seguinte item.

Em um intervalo de 95% de confiança para a média populacional em questão, caso se aumente o tamanho da amostra em 100 vezes (passando a 1.600 observações), a largura total do intervalo de confiança será reduzida à metade.

FGV

16. (FGV/2011 – AFRE/RJ) Um processo X segue uma distribuição normal com média populacional desconhecida, mas com desvio-padrão conhecido e igual a 4. Uma amostra com 64 observações dessa população é feita, com média amostral 45. Dada essa média amostral, a estimativa da média populacional, a um intervalo de confiança de 95%, é

a) (41;49).

b) (37;54).

c) (44,875;45,125).

d) (42,5;46,5).

e) (44;46).

17. (FGV/2016 – IBGE) 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 $\overline{X}$ = 8 de grau de confiança pretendido é de 95%. Somam-se a todas essas informações os valores tabulados:

Φ(1,65) ≅ 0,95 Φ(1,96) ≅ 0,975

T_{25}(1,71) ≅ 0,95 T_{26}(1,70) ≅ 0,95 T_{25}(2,06) ≅ 0,975 T_{26}(2,05) ≅ 0,975

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

Φ(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.

18. (FGV/2010 – Fiscal de Rendas/RJ) Para estimar a proporção p de pessoas acometidas por uma certa gripe numa população, uma amostra aleatória simples de 1600 pessoas foi observada e constatou-se que, dessas pessoas, 160 estavam com a gripe.

Um intervalo aproximado de 95% de confiança para p será dado por:

a) (0,066, 0,134).

b) (0,085, 0,115).

c) (0,058, 0,142).

d) (0,091, 0,109).

e) (0,034, 0,166).

19. (FGV/2016 – IBGE) 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 ε e com um grau de confiança de (1-α). Assim, é correto afirmar que:

a) uma redução de α pode ser compensado por uma redução de ε, 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 ε e (1-α);

c) se a variância de $\hat{p}$ for máxima, o erro ε for 5% e a amostra tiver tamanho n = 1000, o nível de significância será de 5%;

d) fixos α e p, quanto maior a amostra menores serão os ganhos de precisão (redução de ε) gerados por seus incrementos;

e) fixo ε, quanto menor a proporção populacional (p) e maior o nível de significância (α), maior deverá ser a amostra (n).

FCC

20. (FCC 2017/TRT 11ª Região) Instruções: Considere as informações abaixo para responder à questão. Se Z tem distribuição normal padrão, então:

P(Z < 0,4) = 0,655; P(Z < 0,67) = 0,75; P(Z < 1,4) = 0,919; P(Z < 1,6) = 0,945;

P(Z < 1,64) = 0,95; P(Z < 1,75) = 0,96; P(Z < 2) = 0,977; P(Z < 2,05) = 0,98

A porcentagem do orçamento gasto com educação nos municípios de certo estado é uma variável aleatória X com distribuição normal com média μ(%) e variância 4(%)².

Uma amostra aleatória, com reposição, de tamanho n, X_1, X_2, ..., X_n, é selecionada da distribuição de X. Sendo $\bar{X}$, a média amostral dessa amostra, o valor de n para que $\bar{X}$ não se distancie de sua média por mais do que 0,41% com probabilidade de 96% é igual a

a) 64

b) 100

c) 121

d) 81

e) 225

21. (FCC 2015/SEFAZ-PI) Instrução: Para responder à questão utilize, dentre as informações dadas a seguir, as que julgar apropriadas. Se Z tem distribuição normal padrão, então:

P(Z < 0,4) = 0,655; P(Z < 1,2) = 0,885; P(Z < 1,6) = 0,945; P(Z < 1,8) = 0,964; P(Z < 2) = 0,977.

O efeito do medicamento A é o de baixar a pressão arterial de indivíduos hipertensos. O tempo, em minutos, decorrido entre a tomada do remédio e a diminuição da pressão é uma variável aleatória X com distribuição normal, tendo média μ e desvio padrão σ. Uma amostra aleatória de n indivíduos hipertensos foi selecionada com o objetivo de se estimar μ. Supondo que o valor de σ é 10 min, o valor de n para que o estimador não se afaste de μ por mais do que 2 min, com probabilidade de 89%, é igual a

a) 36

b) 100

c) 81

d) 49

e) 64

22. (FCC 2015/TRT 3ª Região) Se Z tem distribuição normal padrão, então:

P(Z < 0,5) = 0,591; P(Z < 1) = 0,841; P(Z < 1,15) = 0,8951; P(Z < 1,17) = 0,879; P(Z < 1,2) = 0,885; P(Z<1,4)=0,919;P(Z<1,64)=0,95;P(Z<2)=0,977;P(Z<2,06)=0,98;P(Z<2,4)= 0,997.

Considere que X é a variável aleatória, que representa as idades, em anos, dos trabalhadores de certa indústria. Suponha que X tem distribuição normal com média de μ anos e desvio padrão de 5 anos. Uma amostra aleatória, com reposição, de n trabalhadores será selecionada e sejam sejam X_1, X_2, ..., X_n as idades observadas e $\overline{X} = \frac{\sum_1^n X_i}{n}$ a média desta amostra.

Desejando-se que o valor absoluto da diferença entre $\overline{X}$ e sua média seja menor do que 6 meses, com probabilidade de 95,4%, o valor de n deverá ser igual a

a) 225.

b) 100.

c) 256.

d) 196.

e) 400.

23. (FCC 2014/TRT 16ª Região) Se Z tem distribuição normal padrão, então:

P(Z<0,25) = 0,599, P(Z<0,80) = 0,84, P(Z<1) = 0,841, P(Z<1,96) = 0,975, P(Z<3,09) = 0,999.

Considere $X_1, X2, \dots, X_n$ uma amostra aleatória simples, com reposição, da distribuição da variável X, que tem distribuição normal com média μ e variância 36. Seja $\overline{X}$ a média amostral dessa amostra. O valor de n para que a distância entre $\overline{X}$ e μ seja, no máximo, igual a 0,49, com probabilidade de 95% é igual a:

a) 256

b) 225

c) 400

d) 144

e) 576

24. (FCC 2018/TRT 14ª Região) Uma pesquisa piloto realizada no setor de embalagens, referente aos motivos de demissão de funcionários, mostra que 34% dos casos de demissão, p*, tem como motivo a situação financeira da empresa. Utilizando um nível de confiança de 95%, a proporção p* obtida na pesquisa piloto, com uma margem de erro amostral e ≤ 3% e que P(Z ≥ 1,96) = 2,5%, o tamanho mínimo necessário da amostra para estimar a proporção de demissões causadas por motivos financeiros, no setor de embalagens, nas condições estipuladas é

a) 635

b) 1020

c) 2115

d) 854

e) 958

25. (FCC 2016/TRT 20ª Região) Sejam duas variáveis aleatórias X e Y, normalmente distribuídas, com as populações de tamanho infinito e médias μ_X e μ_Y, respectivamente. Uma amostra aleatória de tamanho 64 foi extraída da população de X, apresentando um intervalo de confiança [1, 5] para μ_X, ao nível de confiança $(1 - \alpha)$. Uma outra amostra aleatória de tamanho 144 foi extraída da população de Y, independente da primeira, apresentando um intervalo de confiança [4, 10] para μ_Y, também ao nível de confiança de $(1 - \alpha)$. Se σ_X e σ_Y são os desvios padrões populacionais de X e Y, respectivamente, então σ_Y/σ_X apresenta um valor igual a

a) 2,000

b) 3,375

c) 1,500

d) 2,250

e) 2,500

26. (FCC 2020/AL-AP) De uma amostra aleatória de tamanho 64 extraída, com reposição, de uma população normalmente distribuída e variância conhecida σ^2, obteve-se um intervalo de confiança de 95% igual a [23, 27] para a média μ desta população. Desejando-se obter um intervalo de confiança de 95% para μ, porém com amplitude igual à metade da obtida anteriormente, é necessário extrair da população uma amostra aleatória, com reposição, de tamanho

a) 400

b) 1.024

c) 512

d) 256

e) 128

27. (FCC 2019/Prefeitura de Recife/PE) Considere que na curva normal padrão (Z) a probabilidade $P(-2 \leq Z \leq 2) = 95\%$. Uma amostra aleatória de tamanho 400 é extraída de uma população normalmente distribuída e de tamanho infinito. Dado que a variância desta população é igual a 64, obtém-se, com base na amostra, um intervalo de confiança de 95% para a média da população. A amplitude deste intervalo é igual a

a) 0,8

b) 6,4

c) 1,6

d) 12,8

e) 3,2

28. (FCC 2017/TRT 11ª Região) Uma amostra aleatória de tamanho 64 é extraída de uma população de tamanho infinito, normalmente distribuída, média μ e variância conhecida σ^2. Obtiveram-se com base nos dados desta amostra, além de uma determinada média amostral x, 2 intervalos de confiança para μ aos níveis de 95% e 99%, sendo os limites superiores destes intervalos iguais a 20,98 e 21,29, respectivamente. Considerando que na curva normal padrão (Z) as probabilidades P(|Z| > 1,96) = 0,05 e P(|Z| > 2,58) = 0,01, encontra-se que σ^2 é igual a

a) 16,00

b) 6,25

c) 4,00

d) 12,25

e) 9,00

29. (FCC 2018/TRT – 14ª Região) Um intervalo de confiança com um nível de (1 – α) foi construído para a média μ_1 de uma população P_1, normalmente distribuída, de tamanho infinito e variância populacional igual a 144. Por meio de uma amostra aleatória de tamanho 36 obteve-se esse intervalo igual a [25,3; 34,7]. Seja uma outra população P_2, também normalmente distribuída, de tamanho infinito e independente da primeira. Sabe-se que a variância de P_2 é conhecida e que por meio de uma amostra aleatória de tamanho 64 de P_2 obteve-se um intervalo de confiança com um nível de (1 – α) para a média μ_2 de P_2 igual a [91,54; 108,46]. O desvio padrão de P_2 é igual a.

a) 28,80

b) 19,20

c) 23,04

d) 38,40

e) 14,40

30. (FCC 2016/TRT – 20ª Região) Uma variável aleatória X tem distribuição normal, variância desconhecida e com uma população de tamanho infinito. Deseja-se construir um intervalo de confiança de 95% para a média μ da população com base em uma amostra aleatória de tamanho 9 extraída dessa população e considerando a distribuição t de Student. Nessa amostra, observou-se que a média apresentou um valor igual a 5 e a soma dos quadrados dos 9 elementos da amostra foi igual a 243.

Dados: Valores críticos (t_α) da distribuição de Student com n graus de liberdade, tal que a probabilidade $P(t > t_\alpha) = \alpha$.

n	$\alpha = 0,025$	$\alpha = 0,05$	$\alpha = 0,10$
7	2,36	1,89	1,41
8	2,31	1,86	1,40
9	2,26	1,83	1,38

O intervalo de confiança encontrado foi igual a

a) [4,055; 5,945]

b) [4,070; 5,930]

c) [4,300; 5,700]

d) [3,845; 6,155]

e) [3,870; 6,130]

31. (FCC 2016/TRT – 20ª Região) De uma população normalmente distribuída, de tamanho infinito e variância desconhecida, é extraída uma amostra aleatória de tamanho 16 fornecendo um intervalo de confiança de (1 – α) igual a [4,91; 11,30] para a média μ da população. A variância amostral apresentou um valor igual a 36 e considerou-se a distribuição t de Student para obtenção do intervalo de confiança. Consultando a tabela da distribuição t de Student com o respectivo número de graus de liberdade e verificando o valor crítico $t_{\alpha/2}$ tal que a probabilidade $P(|t| > t_{\alpha/2}) = \alpha$, obtém-se que $t_{\alpha/2}$ é igual a

a) 4,26

b) 6,39

c) 2,13

d) 1,65

e) 8,52

32. (FCC 2016/TRT – 20ª Região) Uma amostra aleatória de tamanho 100 é extraída de uma população P_1 de tamanho infinito, com média μ_1, normalmente distribuída e com desvio padrão populacional igual a 2. Uma outra amostra aleatória, independente da primeira, de tamanho 400 é extraída de uma outra população P_2 de tamanho infinito, com média μ_2, normalmente distribuída e com desvio padrão populacional igual a 3. Considerando que na curva normal padrão (Z) as probabilidades $P(Z > 1,64) = 0,05$ e $P(Z > 1,28) = 0,10$ e que as médias das amostras tomadas de P_1 e P_2 foram iguais a 10 e 8, respectivamente, obtém-se que o intervalo de confiança de 90% para $(\mu_1 - \mu_2)$ é

a) [1,79; 2,21]

b) [1,64; 2,36]

c) [1,66; 2,34]

d) [1,59; 2,41]

e) [1,68; 2,32]

VUNESP

33. (VUNESP/2015 – TJ-SP) Leia o texto a seguir para responder à questão.

O Sr. Manoel comprou uma padaria, e foi garantido o faturamento médio de R$ 1.000,00 por dia de funcionamento. Durante os primeiros 16 dias, considerados como uma amostra de 16 valores da população, obteve-se o faturamento médio de R$ 910,00 e desvio padrão de R$ 80,00. Sentindo-se enganado pelo vendedor, o Sr. Manoel entrou com ação de perdas e danos. O juiz sugeriu, então, efetuar o teste de hipótese, indicado ao nível de significância de 5% para confirmar ou refutar a ação.

Supondo-se que a distribuição seja normal com desvio padrão de R$ 120,00 e que a amostra dos 16 dias tenha acusado o valor de R$ 910,00, então o intervalo de confiança para a verdadeira média com 95% de confiança é de, aproximadamente,

a) R$ 850,00 < x < R$ 970,00

b) R$ 800,00 < x < R$ 1.020,00

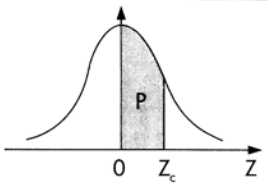
c) R$ 790,00 < x < R$ 1.030,00

d) R$ 900,00 < x < R$ 920,00

e) R$ 890,00 < x < R$ 930,00

Utilize a tabela a seguir para resolver esta e as próximas questões.

Tabela I — Distribuição Normal Padrão
$Z \sim N(0, 1)$
Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

34. (VUNESP/2015 – TJ-SP) Para avaliar o tempo médio de viagem entre o ponto inicial e o ponto final de uma linha de ônibus, retira-se uma amostra de 36 observações (viagens), encontrando-se, para essa amostra, o tempo médio de 50 minutos e o desvio padrão de 6 minutos, com distribuição normal. Considerando-se um intervalo de confiança de 95% para o tempo médio populacional, é correto afirmar que o valor mais próximo para limite inferior desse intervalo é o tempo de

a) 40 min.

b) 44 min.

c) 48 min.

d) 52 min.

e) 56 min.

35. (VUNESP/2014 – TJ-PA) Supondo que em uma amostra de 4 baterias automotivas tenha-se calculado o tempo de vida média de 4 anos. Sabe-se que o tempo de vida da bateria é uma distribuição normal com desvio padrão de 1 ano e meio.

Então, o intervalo de 90% de confiança para a média de todas as baterias é de, aproximadamente:

a) 4 anos $\pm$ 15 meses

b) 4 anos $\pm$ 12 meses

c) 4 anos $\pm$ 10 meses

d) 4 anos $\pm$ 8 meses

e) 4 anos $\pm$ 5 meses

36. (VUNESP/2014 – TJ-PA) O sindicato dos médicos de uma região divulgou uma nota na qual afirma que os médicos plantonistas daquela região trabalham em média, mais de 40 horas por semana, o que, supostamente, contraria acordos anteriores firmados com a categoria. Para verificar essa afirmativa realizou-se uma nova pesquisa na qual 16 médicos escolhidos de modo aleatório declararam seu tempo de trabalho semanal, obtendo-se para tempo médio e para desvio padrão, respectivamente, os valores 42 horas e 3 horas. Considere-se que a distribuição de frequência das horas trabalhadas pelos médicos é normal. Com esses dados o intervalo de confiança $\mu_x = \overline{x} \pm t_{\alpha/2} \times \frac{s_x}{\sqrt{n}}$ de 95% para a média populacional, é:

a) $\mu_x = 42 \pm 6$

b) $\mu_x = 42 \pm 3$

c) $\mu_x = 42 \pm 2{,}55$

d) $\mu_x = 42 \pm 1{,}6$

e) $\mu_x = 42 \pm 0{,}75$

37. (VUNESP/2013 – MPE-ES) Pesquisa recente sobre o tempo total para que os ônibus de determinada linha urbana percorram todo o trajeto entre o ponto inicial e o ponto final, programados para essa viagem, detectou que os tempos de viagem são normalmente distribuídos com tempo médio gasto de 53 minutos e com desvio-padrão amostral de 9 minutos. Nessa pesquisa, foram observados e computados os dados de 16 viagens escolhidas aleatoriamente.

Com um intervalo de confiança de 98%, utilizando-se a tabela t de Student para estimar o erro amostral, e arredondando para cima o valor desse erro, é correto afirmar que o tempo médio dessa viagem varia entre

a) 45 min e 61 min

b) 47 min e 59 min

c) 50 min e 56 min

d) 51 min e 55 min

e) 52 min e 54 min

38. (VUNESP/2014 – EMPLASA) Um jornal deseja estimar a proporção de jornais impressos com não conformidades. Em uma amostra aleatória de 100 jornais dentre todos os jornais impressos durante um dia, observou-se que 20 têm algum tipo de não conformidade. Para um nível de confiança de 90%, Z = 1,64. Então pode-se concluir que apresentam não conformidades.

a) 21,64%, no máximo

b) 26,56%, no máximo

c) 26,56%, no mínimo

d) 21,64%, no mínimo

e) 20%, no mínimo

GABARITO

1. LETRA C
2. LETRA B
3. LETRA C
4. LETRA C
5. LETRA E
6. CERTO
7. ERRADO
8. ERRADO
9. ERRADO
10. LETRA C
11. ERRADO
12. ERRADO
13. ERRADO
14. ERRADO
15. ERRADO
16. LETRA E
17. LETRA B
18. LETRA B
19. LETRA D
20. LETRA B
21. LETRA E
22. LETRA E
23. LETRA E
24. LETRA E
25. LETRA D
26. LETRA D
27. LETRA C
28. LETRA A
29. LETRA A
30. LETRA D
31. LETRA C
32. LETRA D
33. LETRA A
34. LETRA C
35. LETRA A
36. LETRA D
37. LETRA B
38. LETRA B

ESSA LEI TODO MUNDO CONHECE:
PIRATARIA É CRIME.

Mas é sempre bom revisar o porquê e como você pode ser prejudicado com essa prática.

1. 

Professor investe seu tempo para elaborar os cursos e o site os coloca à venda.

2. 

Pirata divulga ilicitamente (grupos de rateio), utilizando-se do anonimato, nomes falsos ou laranjas (geralmente o pirata se anuncia como formador de "grupos solidários" de rateio que não visam lucro).

3. 

Pirata cria alunos fake praticando falsidade ideológica, comprando cursos do site em nome de pessoas aleatórias (usando nome, CPF, endereço e telefone de terceiros sem autorização).

4. 

Pirata compra, muitas vezes, clonando cartões de crédito (por vezes o sistema anti-fraude não consegue identificar o golpe a tempo).

5. 

Pirata fere os Termos de Uso, adultera as aulas e retira a identificação dos arquivos PDF (justamente porque a atividade é ilegal e ele não quer que seus fakes sejam identificados).

6. 

Pirata revende as aulas protegidas por direitos autorais, praticando concorrência desleal e em flagrante desrespeito à Lei de Direitos Autorais (Lei 9.610/98).

7. 

Concurseiro(a) desinformado participa de rateio, achando que nada disso está acontecendo e esperando se tornar servidor público para exigir o cumprimento das leis.

8. 

O professor que elaborou o curso não ganha nada, o site não recebe nada, e a pessoa que praticou todos os ilícitos anteriores (pirata) fica com o lucro.

Deixando de lado esse mar de sujeira, aproveitamos para agradecer a todos que adquirem os cursos honestamente e permitem que o site continue existindo.