By @kakashi_copiador

Aula 06

CNU - Concurso Nacional Unificado (Bloco Temático 1 - Infraestrutura, Exatas e Engenharia) Bizu Estratégico - 2024 (Pós-Edital)

Autor:

Vinícius Peron Fineto, Aline Calado Fernandes, Diogo Matias das Neves, Fernanda Harumi Amaral Jo, Elizabeth Menezes de Pinho Alves, Guilherme Carvalho, Arthur Fontes da Silva Jr, Leo Mandarino, Paulo Júnior, Leonardo Mathias

12 de Fevereiro de 2024

Bizu Estratégico de GEOPROCESSAMENTO E ANÁLISE DE DADOS (Concurso Nacional Unificado – PÓS EDITAL)

Olá, prezado aluno. Tudo certo?

Neste material, traremos uma seleção de *bizus* da disciplina de Geoprocessamento e Análise de Dados para o Concurso Nacional Unificado - CNU.

O objetivo é proporcionar uma revisão rápida e de alta qualidade aos alunos por meio de tópicos que possuem as maiores chances de incidência em prova.

Todos os *bizus* destinam-se a alunos que já estejam na fase bem final de revisão (que já estudaram bastante o conteúdo teórico da disciplina e, nos últimos dias, precisam revisar por algum material bem curto e objetivo).

Esse *bizu* foi elaborado com base nos cursos de Geoprocessamento e Análise de Dados dos professores Alexandre Vastella, Equipe Exatas Estratégia Concursos e Monik Begname de Castro.

Fernanda Harumi

@fernandaharu_

Leonardo Mathias

@profleomathias

ANÁLISE ESTATÍSTICA

Pessoal, segue abaixo uma análise estatística dos assuntos mais exigidos pela Cesgranrio no âmbito da disciplina de Geoprocessamento e Análise de Dados em concursos.

Geoprocessamento e Análise de Dados	
Assunto	**% de cobrança**
Estimação Pontual e Intervalar	47,90%
Apresentação de dados	30,25%
Variáveis Aleatórias Discretas	8,40%

Com essa análise, podemos verificar quais são os temas mais exigidos pela banca e, através disso, focaremos nos principais pontos em nossa revisão!

A disciplina de Geoprocessamento e Análise de Dados no Edital do Concurso Nacional Unificado - CNU abordou o seguinte conteúdo programático:

> 2 GEOPROCESSAMENTO E ANÁLISE DE DADOS 2.1 Noções Básicas de Cartografia, escala, sistemas de coordenadas, projeção cartográfica Sistema Global de Posicionamento Por Satélites Artificiais. 2.2 Noções básicas de Geografia Urbana, urbanismo, conceitos de território e estrutura territorial brasileira. 2.3 Armazenamento de informações geoespaciais em ambiente de banco de dados relacional e orientado a objeto. Infraestrutura de dados espaciais. Sensoriamento remoto. 2.4 Noções de Inferência Estatística de dados geoespaciais: População e Amostra, Seleção de amostra, Estatística e Parâmetro, Distribuições amostrais. 2.5 Noções de Amostragem de dados geoespaciais: Amostragem Probabilística e Não probabilística. 2.6 Noções de Estimação de dados geoespaciais: Estimação Pontual e Estimação Intervalar.

MAPA DO BIZU

Segue uma tabela contendo a numeração dos bizus referentes a cada tópico abordado e os respectivos cadernos de questões selecionados no nosso SQ.

Geoprocessamento e Análise de Dados – CNU		
Assunto	**Bizus**	**Caderno de Questões**
Apresentação de dados	1 a 3	http://questo.es/r6gksq
Variáveis Aleatórias Discretas	4 a 5	http://questo.es/bvu002
Estimação Pontual e Intervalar	6 a 9	http://questo.es/mdzgxb

Apresentação

Olá, futuro (a) aprovado(a)! Antes de darmos início aos nossos trabalhos, farei uma breve apresentação.

Meu nome é Fernanda Harumi Amaral Jo, sou natural de São Paulo, me formei em Contabilidade pela USP e hoje ocupo o cargo de Auditor de Controle Externo no Tribunal de Contas do Estado de São Paulo (TCE-SP), tendo sido aprovada no último certame, realizado em 2017.

Hoje também Integro a Equipe de Coaching do Estratégia Concursos junto com renomados profissionais e ex-concurseiros de todo o Brasil.

Como pode perceber, há pouco tempo eu estava justamente aí, onde você concurseiro está. Logo utilizarei as experiências e conhecimentos adquirodos ao longo da minha trajetória para auxiliá-lo na disciplina de Geoprocessamento e Análise de Dados. Fiz uma análise bem cautelosa dos pontos mais exigidos pela banca, e todos eles estão aqui! Cada questão vale ouro, então não podemos dar bobeira! Mãos à obra!

Fernanda Harumi

Apresentação de dados

1) Conceitos iniciais

- População: **Uma POPULAÇÃO é um conjunto que contém TODOS OS INDIVÍDUOS, OBJETOS OU ELEMENTOS a serem estudados, que apresentam uma ou mais características em comum.** A população pode ser finita, quando apresenta um número pequeno ou limitado de observações; ou infinita, quando apresenta um número muito grande ou ilimitado de observações.

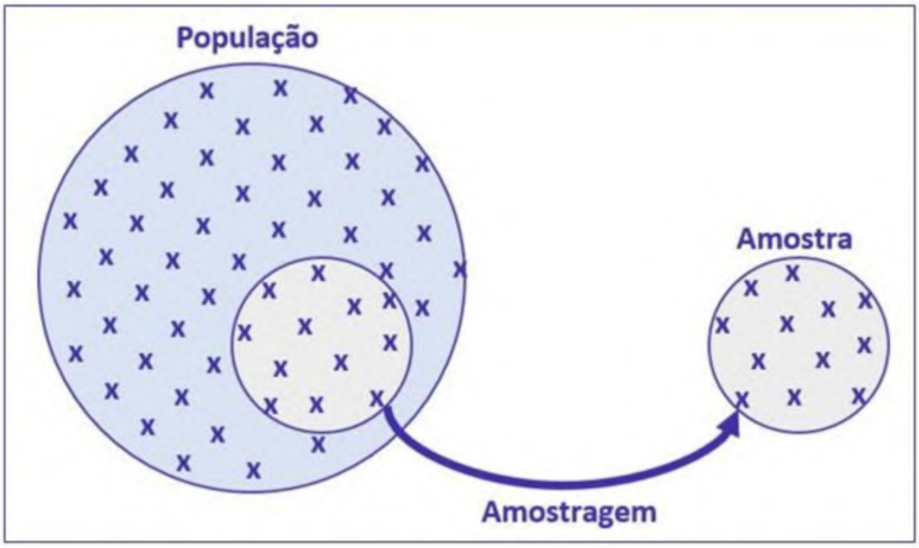
- Amostra: **Uma AMOSTRA é um SUBCONJUNTO EXTRAÍDO DA POPULAÇÃO para análise, devendo ser representativo daquele grupo.** A partir das informações colhidas da amostra, os resultados obtidos podem ser utilizados para generalizar, inferir ou tirar conclusões acerca da população. Como exemplo, podemos citar as pesquisas eleitorais, em que uma amostra de eleitores deve ser extraída de acordo com a proporcionalidade de gênero, idade, grau de instrução e classe social.

- Censo: **O CENSO, ou recenseamento, é um estudo dos dados relativos a TODOS os elementos de uma população.** O censo pode custar muito caro e demandar um tempo considerável, de forma que um estudo considerando apenas uma parcela da população pode ser uma alternativa mais simples, rápida e menos onerosa. Como exemplos, podemos citar a pesquisa sobre o grau de escolaridade dos habitantes brasileiros, o estudo sobre a renda dos brasileiros e a pesquisa de emprego.

- Amostragem: **A AMOSTRAGEM é um processo que consiste na SELEÇÃO CRITERIOSA dos elementos a serem submetidos à investigação.** Se forem cometidos erros no processo de seleção da amostra, muito provavelmente, o estudo ficará comprometido e os resultados serão tendenciosos. Portanto, devemos garantir que a amostra seja representativa da população. Isso significa que, com exceção de pequenas discrepâncias inerentes à aleatoriedade existente no processo de amostragem, uma amostra deve possuir as mesmas características básicas da população, no que diz respeito às variáveis que desejamos pesquisar.

- Os PARÂMETROS são DESCRIÇÕES NUMÉRICAS de CARACTERÍSTICAS POPULACIONAIS que raramente são **conhecidas**. Em geral, é muito caro ou demorado obter os dados da população inteira. Assim, **algumas medidas precisam ser estimadas a partir de critérios ou métodos definidos pelo pesquisador**, **para representar características desconhecidas de uma população (por exemplo, a proporção de homens e mulheres na população brasileira)**. Normalmente, os parâmetros populacionais são constantes para uma população.

- **As ESTATÍSTICAS são MEDIDAS NUMÉRICAS OBTIDAS DE AMOSTRAS representativas extraídas da população**. A partir das informações colhidas da amostra, as **estatísticas amostrais obtidas podem ser utilizadas para inferir ou tirar conclusões acerca dos parâmetros populacionais**, como a proporção de homens e mulheres na população brasileira. De forma resumida, as estatísticas (ou estimadores) são **descrições numéricas de características amostrais.** Normalmente, as estatísticas amostrais diferem de umaamostra para outra.



POPULAÇÃO

CONJUNTO de TODOS os elementos a serem estudados, que apresentam uma ou mais características em comum.

CENSO

ESTUDO dos dados relativos a TODOS os ELEMENTOS de uma população.

AMOSTRA

SUBCONJUNTO extraído DA POPULAÇÃO para análise, devendo ser representativo daquele grupo

AMOSTRAGEM

PROCESSO que consiste na seleção criteriosa dos elementos a serem submetidos à investigação.

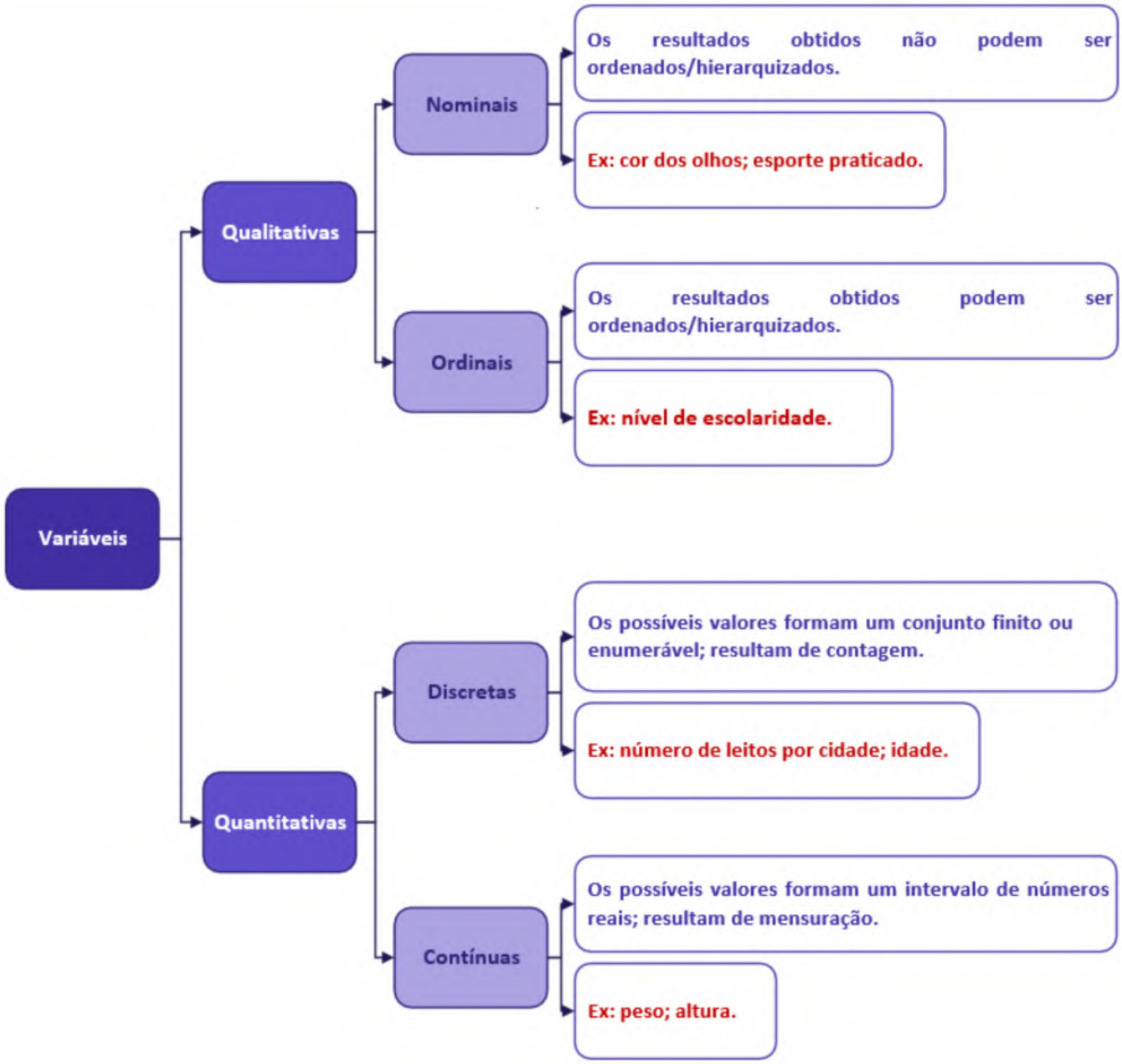
PARÂMETROS

Descrições numéricas de características da POPULAÇÃO, que normalmente precisam ser estimadas.

ESTATÍSTICAS

Medidas numéricas extraídas de AMOSTRAS representativas extraídas da população.

2) Variáveis


Variáveis
Qualitativas
Nominais
Os resultados obtidos não podem ser ordenados/hierarquizados.
Ex: cor dos olhos; esporte praticado.
Ordinais
Os resultados obtidos podem ser ordenados/hierarquizados.
Ex: nível de escolaridade.
Quantitativas
Discretas
Os possíveis valores formam um conjunto finito ou enumerável; resultam de contagem.
Ex: número de leitos por cidade; idade.
Contínuas
Os possíveis valores formam um intervalo de números reais; resultam de mensuração.
Ex: peso; altura.

3) Elementos de uma Distribuição de Frequências

Item	Definição	Símbolos e Fórmulas
Número de Classes	As classes são os intervalos nos quais o fenômeno é subdividido.	$k = 1 + 3{,}3 \times \log n$ $k = \sqrt{n}$
Limites de Classe	Correspondem aos valores extremos.	l_{inf} e l_{sup}
Amplitude de um Intervalo de Classe	Distância entre os limites inferiores (ou superiores) de classes consecutivas.	$h = l_{sup} - l_{inf}$
Amplitude total	Diferença entre o limite superior da última classe (limite superior máximo) e o limite inferior da primeira classe (limite inferior mínimo).	$AT = l_{máx} - l_{mín}$ $AT = h \times k$
Ponto Médio	Média aritmética simples dos valores extremos de uma classe.	$PM = \frac{(l_{inf} + l_{sup})}{2}$ $PM = l_{inf} + \frac{h}{2}$ $PM = l_{sup} - \frac{h}{2}$
Frequência Absoluta Simples	Número de observações correspondentes a uma determinada classe ou a um determinado valor.	f_i
Frequência Absoluta Acumulada	Total das frequências de todos os valores inferiores ao limite superior do intervalo de uma dada classe	$f_{ac_i} = f_1 + f_2 + f_3 + \cdots + f_i$

Frequência Relativa Simples	Proporção de dados existentes em uma determinada classe.	$F_i = \frac{f_i}{\sum f_i} = \frac{f_i}{n}$
Frequência Relativa Acumulada	Proporção de valores inferiores ao limite superior do intervalo de uma dada classe.	$F_{ac_i} = F_1 + F_2 + F_3 + \cdots + F_i$
Densidade de Frequência	Quociente entre a frequência da classe (absoluta ou relativa) e sua amplitude	$a = \frac{f}{h}$

Variáveis Aleatórias Discretas

4) Noções gerais

Variáveis Aleatórias Discretas	• Quantidade de valores possíveis é **enumerável** (finito ou não) • Atribuímos **probabilidades a resultados específicos**
Variáveis Aleatórias Contínuas	• Assumem qualquer valor dentro de um **intervalo** • Os resultados possíveis são **infinitos e não enumeráveis** • Não atribuímos probabilidade a resultados específicos, apenas a intervalos

5) Principais fórmulas

Função de Distribuição Acumulada: $F(x) = P(X \leq x)$

Esperança Matemática (média): $E(X) = \sum x.P(X = x)$

- $E(kX) = k.E(X)$
- $E(X + k) = E(X) + k$
- $E(X \pm Y) = E(X) \pm E(Y)$
- $E(k) = k$
- Se X e Y forem **independentes**, então $E(X \times Y) = E(X) \times E(Y)$

Moda: valor de **X** com maior probabilidade

Mediana: divide a distribuição em duas partes iguais, $F(x_{Med}) = 0,5$

Variância: $V(X) = \sum(x - \mu)^2 \times P(X = x)$

- $V(X + k) = V(X)$
- $V(k.X) = k^2.V(X)$
- $V(k) = 0$
- Se X e Y forem **independentes**, então $V(X \pm Y) = V(X) + V(Y)$

Desvio Padrão: $\sigma = \sqrt{V(X)}$

Covariância: $Cov(X, Y) = E(X.Y) - E(X).E(Y)$

Correlação: $\rho(X, Y) = \frac{Cov(X,Y)}{\sigma_X.\sigma_Y}$

- Se X e Y forem **independentes**, então $Cov(X, Y) = 0, \rho(X, Y) = 0$

Variância da Soma e da Diferença

$$\boldsymbol{V(X + Y) = V(X) + V(Y) + 2.Cov(X, Y)}$$

$$\boldsymbol{V(X - Y) = V(X) + V(Y) - 2.Cov(X, Y)}$$

Coeficiente de Variação: $C_V = \frac{\sigma}{\mu}$

Variância Relativa: $V_R = (C_V)^2 = \frac{V(X)}{\mu^2}$

Estimação Pontual e Intervalar

6) Distribuição amostral

- A distribuição dos elementos de uma amostra aleatória qualquer segue a mesma distribuição populacional.

Para estimarmos a **média da população μ**, utilizamos como estimador a **média amostral 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:

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

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:

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

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:

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

7) Distribuição Amostral da Proporção

- 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:

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

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:

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

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:

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

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

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

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:

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

8) 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:

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

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)

9) Estimação Intervalar

- 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 $(\boldsymbol{\hat{\theta} - E; \hat{\theta} + E})$, que também pode ser indicado como $\boldsymbol{\hat{\theta} \pm E}$. 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**.

- 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).

- 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.

Estimação Intervalar

Intervalo para a Média – Variância conhecida (distribuição normal):

Margem de Erro: $E = z.\frac{\sigma}{\sqrt{n}}$; Tamanho amostral: $n = \left(z.\frac{\sigma}{E}\right)^2$

Intervalo para a Média – Variância desconhecida (distribuição t-Student):

Margem de Erro: $E = t_{n-1}.\frac{s}{\sqrt{n}}$; Tamanho Amostral: $n = \left(t.\frac{s}{E}\right)^2$

Intervalo para a Proporção:

Margem de Erro: $E = z.\sqrt{\frac{\hat{p}.\hat{q}}{n}}$; Tamanho Amostral: $n = \left(\frac{z}{E}\right)^2 \hat{p}.\hat{q}$

Intervalo para a Variância: $\left(\frac{(n-1).s^2}{\chi^2_{SUP}}; \frac{(n-1).s^2}{\chi^2_{INF}}\right)$

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: $\bar{X} \pm t_{n-1}.\frac{\sigma}{\sqrt{n}}$

Vamos ficando por aqui.

Esperamos que tenha gostado do nosso Bizu!

Bons estudos!

"A diferença entre o sonho e a realidade é a quantidade certa de tempo e trabalho" (William Douglas)

Fernanda Harumi

*@fernandaharu_*

Leonardo Mathias

@profleomathias

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.