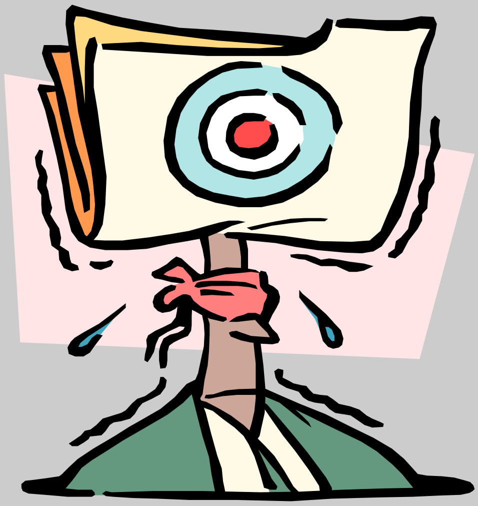
Material Didático

Série

Estatística Básica



Texto III

Estimação

Enfoque: Exatas

Prof. Lori Viali, Dr.

SUMÁRIO

INTRODUÇÃO

A inferência estatística tem por objetivo fazer generalizações sobre uma população com base em valores amostrais. A inferência pode ser feita estimando os parâmetros:

(a) Por ponto e

(b) Por intervalo.

A estimação por ponto é feita através de um único valor, enquanto que a estimação por intervalo fornece um intervalo de valores em torno do valor da estimativa pontual. Na estimação por ponto o objetivo é utilizar a informação amostral e apriorística para se calcular um valor que seria, em certo sentido, nossa melhor avaliação quanto ao valor, de fato, do parâmetro em questão.

Na estimativa por intervalo, usa-se a mesma informação com o propósito de se produzir um intervalo que contenha o valor verdadeiro do parâmetro com algum nível de probabilidade.

Exemplo:

Uma amostra aleatória simples de 400 pessoas de uma cidade é extraída e 300 respondem que acham a administração municipal boa ou ótima. Então o valor p = 300/400 = 75% é uma estimativa por ponto do percentual de pessoas da cidade que acham a administração boa ou ótima. Esta mesma estimativa poderia ser enunciado como de: 70% a 80% das pessoas da cidade acham a administração boa ou ótima. Neste caso, teríamos uma estimativa por intervalo da proporção. Note-se que o centro do intervalo é o valor “75%” da estimativa pontual.

1. ESTIMAÇÃO POR PONTO

O problema da estimação por ponto é o de produzir uma estimativa que realmente represente a melhor avaliação do valor do parâmetro. Deve-se especificar em primeiro lugar o que se entende por "melhor avaliação" e em segundo os estimadores que satisfaçam estas especificações. Como um estimador é uma variável aleatória cujo valor varia de amostra para amostra, suas propriedades são, de fato, as propriedades da distribuição amostral.

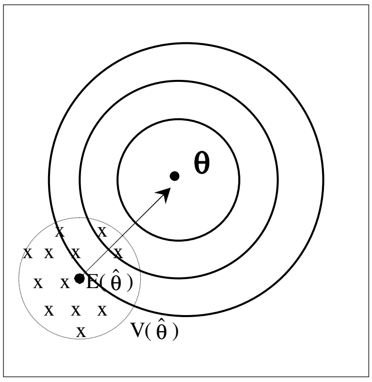
1.1. NOTAÇÃO

Vai-se considerar uma variável aleatória X (população) cuja distribuição é caracterizada, entre outras coisas, por um parâmetro θ, que gostaríamos de estimar.

Um estimador do parâmetro θ, que é obtido através de uma fórmula dos valores amostrais: X_1, X_2, ..., X_n, é anotado por $\hat{\theta}$. As características básicas da distribuição de $\hat{\theta}$ são sua média $\mu_{\hat{\theta}} = E(\hat{\theta})$ e sua variância $\sigma^2_{\hat{\theta}} = Var(\hat{\theta}) = E(\hat{\theta} - \mu_{\hat{\theta}})^2 = E(\hat{\theta}^2) - \mu^2_{\hat{\theta}}$.

O desvio padrão de $\hat{\theta}$, representado por $\sigma_{\hat{\theta}} = \sqrt{Var(\hat{\theta})}$ é denominado de "erro padrão de $\hat{\theta}$".

Figura 1.1 - Relação entre EQM e variância de um estimador



Além destes, os seguintes conceitos são de importância:

Erro amostral $\varepsilon = \hat{\theta} - \theta$, que é a diferença entre o valor do estimador $\hat{\theta}$ e o verdadeiro valor a ser estimado θ. O tamanho do erro amostral varia de amostra para amostra.

Viés ou tendenciosidade $Viés(\hat{\theta}) = E(\hat{\theta}) - \theta$ como sendo a diferença entre a média da distribuição amostral de $\hat{\theta}$ e o valor do parâmetro θ. Este valor é, para cada estimador, fixo, podendo ou não ser zero.

Erro quadrado (quadrático) médio EQM($\hat{\theta}$) = E = (ε) = E($\hat{\theta}$ - θ)^2 é uma variância que mede a dispersão do estimador em torno do verdadeiro parâmetro, ao invés de em torno de sua média.

Existe uma relação entre o EQM($\hat{\theta}$) e a Var($\hat{\theta}$), conforme, mostrado abaixo:

$\text{EQM}(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = E(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2 = E\{[\hat{\theta} - E(\hat{\theta})] + [E(\hat{\theta}) - \theta]\}^2 = E[\hat{\theta} - E(\hat{\theta})]^2 + 2.[E(\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] + E[E(\hat{\theta}) - \theta]^2 = E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta] = E[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta}) - \theta]^2 = \text{Var}(\hat{\theta}) + \text{Viés}(\hat{\theta})^2$, pois

$2.[E(\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] = 2.[E(\hat{\theta}) - \hat{\theta}][E(\hat{\theta}) - E(\hat{\theta})] = 0$.

Desta forma:

EQM($\hat{\theta}$) = Var($\hat{\theta}$) + Viés($\hat{\theta}$)^2, isto é, o EQM é a soma da variância do estimador com sua tendenciosidade elevada ao quadrado.

1.2. PROPRIEDADES DOS ESTIMADORES

As propriedades desejáveis para um estimador são: não-tendenciosidade, precisão ou eficiência, validade ou acurácia e consistência.

1.2.1. NÃO-TENDENCIOSIDADE

Um estimador $\hat{\theta}$ é dito não-tendencioso (Imparcial, justo, não-viciado, não-viezado, *unbiased*) de um parâmetro θ se E($\hat{\theta}$) = θ.

Se E($\hat{\theta}$) $\neq$ θ, então $\hat{\theta}$ é dito "viciado" e E($\hat{\theta}$) - θ é dito "viés" do estimador (*bias of the estimate*).

Exemplos:

(1) $\overline{\mathbf{X}}$ é um estimador não-tendencioso de μ.

Prova:

$$E(\overline{X}) = \mathbf{E}\left(\frac{\sum \mathbf{X}}{\mathbf{n}}\right) = \frac{1}{\mathbf{n}}\mathbf{E}(\sum \mathbf{X}) = \frac{1}{n}\sum E(X) = \frac{1}{n}\sum \mu = (n.\mu)/n = \mu.$$

(2) $\hat{\sigma}^2 = \dfrac{\sum(\mathbf{X} - \overline{\mathbf{X}})^2}{\mathbf{n}}$ é um estimador tendencioso de σ^2.

Prova:

Considere a soma $\sum(\mathbf{X} - \overline{\mathbf{X}})^2$ e observe que ela poderá ser escrita da seguinte maneira:

$$\sum\left(X-\bar{X}\right)^2 = \sum\left(X-\mu+\mu-\bar{X}\right)^2 = \sum\left(X-\mu\right)^2 + 2\sum(X-\mu)(\mu-\bar{X}) + \sum\left(\mu-\bar{X}\right)^2.$$

Como μ - $\bar{X}$ é constante e $\sum(X-\mu) = \sum X - n.\mu = n.\bar{X} - n.\mu = n(\bar{X}-\mu)$, segue:

$\sum\left(X-\bar{X}\right)^2 = \sum\left(X-\mu\right)^2 - n.(\bar{X}-\mu)^2$, pois

$2\sum(X-\mu)(\mu-\bar{X}) = 2(\mu-\bar{X}).n(\bar{X}-\mu) = -2n(\mu-\bar{X})$ e

$2\mu-\bar{X}) + \sum(\mu-\bar{X})^2 = -2n\,(\mu-\bar{X})^2 + n\,(\mu-\bar{X})^2 = -n\,(\mu-\bar{X})^2$

Portanto:

$$E(\hat{\sigma}^2) = E\left(\frac{\sum\left(X-\bar{X}\right)^2}{n}\right) = \frac{1}{n}E\left(\sum\left(X-\bar{X}\right)^2\right) = \frac{1}{n}E\left(\sum(X-\mu)^2 - n.(\bar{X}-\mu)^2\right) = \frac{1}{n}\{\sum E(X-\mu)^2 - nE(\bar{X}-\mu)^2\}$$

$$= \frac{1}{n}\{\sum Var(X) - nVar(\bar{X})\} = \frac{1}{n}\{n\sigma^2 - n\frac{\sigma^2}{n}\} = \frac{1}{n}(n\sigma^2 - \sigma^2\} = \sigma^2 - \sigma^2/n = (n\sigma^2 - \sigma^2)/n =$$

$$= (n-1)\sigma^2/n$$

(3) $S^2 = \dfrac{\sum\left(X-\bar{X}\right)^2}{n-1}$ é um estimador não-viciado de σ^2.

Prova:

$$E(S^2) = E\left(\frac{\sum\left(X-\bar{X}\right)^2}{n-1}\right) = \frac{1}{n-1}E\left[\sum\left(X-\bar{X}\right)^2\right] = \frac{1}{n-1}E\left[\sum\left(X-E(X)+E(X)-\bar{X}\right)^2\right] =$$

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

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

(4) $S^2 = \dfrac{\sum\left(X-\bar{X}\right)^2}{n-1}$ é um estimador tendencioso de σ^2, se a amostragem for realizada sem reposição de uma população finita.

Prova:

(5) $T = n.\bar{X}$, total amostral, é um estimador tendencioso de $\tau = \sum X$, total populacional.

Prova:

$E(T) = E(n.\bar{X}) = n.\ E(\bar{X}) = n.\mu \neq N.\mu = \tau$.

(6) $\bar{T} = N.\bar{X}$ é um estimador não-tendencioso de $\tau = \sum X$.

Prova:

$E(\overline{T}) = E(N.\overline{X}) = N.\ E(\overline{X}) = N.\mu = N.\mu = \tau.$

(7) P é um estimador não-tendencioso de π.

Prova:

É um caso particular de $\overline{X}$, onde $X = \begin{cases} 0, & \text{com probabilidade } 1 - \pi \\ 1, & \text{com prabalidade } \pi \end{cases}$

(8) $\hat{\sigma}^2_{\overline{X}} = \dfrac{S^2}{n}$ é um estimador não-tendencioso de $\sigma^2_{\overline{X}} = \dfrac{\sigma^2}{n}$

Prova:

$E(\hat{\sigma}^2_{\overline{X}}) = E(\dfrac{S^2}{n}) = \dfrac{E(S^2)}{n} = \sigma^2/n.$

Se a amostragem for sem reposição de população finita então:

$\hat{\sigma}^2_{\overline{X}} = \dfrac{\hat{S}^2}{n}\dfrac{N-n}{N-1}$ é um estimador não tendencioso de $\sigma^2_{\overline{X}} = \dfrac{\sigma^2}{n}\dfrac{N-n}{N-1}$, onde $\hat{S}^2 = \dfrac{N-1}{N}S^2$

A não-tendenciosidade ou ausência de viés é uma qualidade desejável para os estimadores. Entretanto, essa qualidade é insuficiente como critério para selecionar um estimador, pois, por exemplo, toda média ponderada dos valores amostrais é é um estimador não-tendencioso da média populacional.

Prova:

Seja $M = \sum\pi_i.X_i$, onde $\sum\pi_i = 1$, um estimador de μ. Então:

$E(M) = E(\sum\pi_i.X_i) = \sum E(\pi_i.X_i) = \sum\pi_i.E(X_i) = \mu.\sum\pi_i = \mu.$

Outra propriedade desejável para um estimador é a da precisão ou eficiência.

1.2.2. PRECISÃO OU EFICIÊNCIA

A precisão ou eficiência é a proximidade das observações (estimativas) do seu valor esperado.

Definição:

Dados dois estimadores não-tendenciosos $\hat{\theta}_1$ e $\hat{\theta}_2$ de um mesmo parâmetro θ, diremos que $\hat{\theta}_1$ é mais eficiente que $\hat{\theta}_2$ se $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$. A eficiência relativa de $\hat{\theta}_1$ em relação a $\hat{\theta}_2$ é definida como sendo $V(\hat{\theta}_1)/V(\hat{\theta}_2)$.

Exemplo:

Qual dos dois estimadores abaixo é mais eficiente para estimar a média da população?

$\overline{X}_1 = 0{,}3X_1 + 0{,}7X_2$ ou $\overline{X}_2 = 0{,}2X_1 + 0{,}8X_2$

Solução

Como são ambos não-tendenciosos temos:

$\text{Var}(\overline{X}_1) = \text{Var}(0{,}3X_1 + 0{,}7X_2) = 0{,}3^2\text{Var}(X_1) + 0{,}7^2\text{Var}(X_2) = (0{,}09 + 0{,}49)\sigma^2 = 0{,}58\sigma^2$.

$\text{Var}(\overline{X}_2) = \text{Var}(0{,}2X_1 + 0{,}8X_2) = 0{,}2^2\text{Var}(X_1) + 0{,}8^2\text{Var}(X_2) = (0{,}04 + 0{,}64)\sigma^2 = 0{,}68\sigma^2$.

Portanto

$\overline{X}_1$ é mais eficiente que $\overline{X}_2$

Em igualdade de circunstâncias, é obvio, que um estimador não-tendencioso é preferível a um estimador tendencioso. Mas se tivermos que escolher entre um estimador tendencioso, cuja distribuição é concentrada na vizinhança do verdadeiro valor do parâmetro e um não-tendencioso com grande variância, o estimador tendencioso pode ser preferível, principalmente se é possível determinar a grandeza e a direção da tendenciosidade.

Para julgar situações como esta é definida a propriedade da validade (ou acurácia).

1.2.3. VALIDADE (OU ACURÁCIA)

Proximidade das observações (estimativas) ao parâmetro desejado. Dados dois estimadores $\hat{\theta}_1$ e $\hat{\theta}_2$ de um mesmo parâmetro θ, diremos que $\hat{\theta}_1$ é mais acurado que $\hat{\theta}_2$ se $\text{EQM}(\hat{\theta}_1) < \text{EQM}(\hat{\theta}_2)$. A eficiência relativa de $\hat{\theta}_1$ em relação a $\hat{\theta}_2$ é definida como sendo $\text{EQM}(\hat{\theta}_1)/\text{EQM}(\hat{\theta}_2)$.

1.2.4. COERÊNCIA OU CONSISTÊNCIA

Um estimador é dito coerente (consistente) para qualquer quantidade muita pequena $\delta > 0$ se a probabilidade de que o desvio absoluto entre $\hat{\theta}$ e θ seja menor que δ tende para 1 quando o número de observações "n" tende ao infinito, isto é:

$P(|\hat{\theta} - \theta| < \delta) \to 1$, quando $n \to \infty$

A propriedade acima é equivalente a $\lim\limits_{n\to\infty} \text{EQM}(\hat{\theta}) = 0$ ou então, as duas seguintes, consideradas em conjunto: $\begin{cases} \lim\limits_{n\to\infty} E(\hat{\theta}) = 0, & \text{a tendenciosidade tende a zero e} \\ \lim\limits_{n\to\infty} \text{Var}(\hat{\theta}) = 0, & \text{a variância tende a zero.} \end{cases}$

Exemplo:

Verificar que S^2 é um estimador coerente de σ^2.

Tem-se que $E(S^2) = \sigma$, para qualquer "n" e

$Var(S^2) = \frac{2\sigma^4}{n-1}$. Então, como $\lim_{n\to\infty} \frac{2\sigma^4}{n-1} = 0$, S^2 é consistente.

1.3. MÉTODOS DE ESTIMAÇÃO

Os estimadores considerados até agora foram sugeridos por intuição. Seria desejável, entretanto, um método ou princípio que levasse sempre a bons estimadores. Infelizmente não existe um método geral aplicável a todas as situações. Os principais métodos de estimação são o dos momentos, o da máxima verossimilhança e o dos mínimos quadrados.

1.3.1. MÉTODOS DOS MOMENTOS

É o mais antigo dos métodos para determinar estimadores (determinado por K. Pearson em 1894). Se existem "k" parâmetros para serem estimados, o método consiste em expressar os primeiros "k" momentos da população em termos dos "k" parâmetros através de "k" equações. A solução do sistema formado por estas equações leva a determinação dos estimadores.

As estimativas obtidas desta forma são claramente funções dos momentos da amostra. Uma vez que os momentos da amostra são estimativas consistentes dos momentos da população, os parâmetros estimados serão geralmente coerentes.

1.3.2. MÉTODOS DA MÁXIMA VEROSSIMILHANÇA (MAXIMUM LIKELIHOOD)

Função de verossimilhança

Seja $f(x: \theta)$ a função de probabilidade X (discreta ou contínua) calculada no ponto X = x. O valor de θ está incluído na notação para lembrar que a distribuição da variável X depende do parâmetro θ. Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória da variável aleatória X e sejam $x_1, x_2, ..., x_n$ seus valores amostrais. Define-se a função de verossimilhança L, como sendo a seguinte função da amostra e do parâmetro θ:

$$L(X_1, X_2, ..., X_n; \theta) = f(X_1; \theta).f(X_2; \theta). ... . f(X_n; \theta) \qquad \text{Equação 1.1}$$

Se a amostra $(X_1, X_2, ..., X_n)$ tiver sido obtida, os valores amostrais $(x_1, x_2, ..., x_n)$ serão conhecidos. Já que θ é desconhecido pode-se propor a seguinte questão? Para qual valor de θ $L(x_1, x_2,$

...., x_n; θ) terá o valor máximo? Colocando de outra forma. Suponha que existam dois valores diferentes de θ, digamos θ_1 e θ_2, e que $L(x_1, x_2, ..., x_n; \theta_1) < L(x_1, x_2, ..., x_n; \theta_2)$. Neste caso θ_2 seria preferível, pois para a amostra obtida o evento dado é mais provável de ocorrer com θ_2 do que com θ_1. Este raciocínio pode ser resumido por:

Definição: A estimativa de *máxima verossimilhança* de θ, baseada numa amostra aleatória obtida da população X é aquele valor de θ torna máxima $L(X_1, X_2, ..., X_n; \theta)$, para uma dada amostra $X_1, X_2, ..., X_n$ onde L é definida pela equação 1.1.

Observações:

(i) θ é uma variável aleatória, pois seu valor depende da amostra $X_1, X_2, ..., X_n$.

(ii) Na maioria das situações θ representa um valor isolado, mas se a distribuição depender de mais de um parâmetro (dois como no caso da normal), θ representará um vetor, por exemplo, $\theta = (\alpha, \sigma)$.

(iii) Para determinar a estimativa de MV, deve-se determinar o valor máximo de uma função. No entanto em muitas situações é mais fácil determinar o máximo da função

$\ln L(X_1, X_2, ..., X_n; \theta)$

que possui o ponto de máximo idêntico ao da função L.

As estimativas de MV apresentam algumas propriedades importantes:

(i) A estimativa de MV pode ser tendenciosa. Normalmente esta tendenciosidade pode ser eliminada pela multiplicação de uma constante apropriada.

(ii) Sob condições gerais as estimativas de MV são coerentes. Isto é, se o tamanho da amostra aumentar a estimativa se aproximará do valor a ser estimado.

Exemplo:

Suponha que uma variável aleatória X tenha uma distribuição normal de média μ e desvio padrão σ, isto é a fdp de X seja:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Se $X_1, X_2, ..., X_n$ for uma amostra aleatória de X, sua função de verossimilhança será dada por:

$$L(X_1, X_2, ..., X_n; \mu, \sigma) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left[\frac{X_i-\mu}{\sigma}\right]^2\right\}$$

Portanto:

$\ln L(X_1, X_2, ..., X_n; \mu, \sigma) = \left(-\frac{n}{2}\right)\ln\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$

Deve-se resolver simultaneamente:

$\frac{\partial \ln L}{\partial \mu} = 0$ e $\frac{\partial \ln L}{\partial \sigma} = 0$. Tem-se:

$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{n}\frac{(X_i - \mu)^2}{\sigma^2} = 0$, que fornece $\mu = \overline{X}$, a média amostral e

$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^{n}\frac{(X_i - \mu)^2}{\sigma^3} = 0$, o que fornece $\hat{\sigma}^2 = \frac{1}{n}.\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{n}.\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$.

Observe que a estimativa de máxima verossimilhança de σ^2 é tendenciosa, pois já foi visto que a não-tendenciosa precisa ser dividida por "n-1", isto é, $\hat{\sigma}^2$ ao invés de S^2.

1.3.3. MÉTODOS DOS MÍNIMOS QUADRADOS

Este método é útil para estimar momentos em torno de zero de uma distribuição populacional. O princípio envolvido é um pouco mais complicado do que o do método dos momentos. Considere-se uma variável aleatória X e se r-ésimo momento em torno de zero:

$E(X^r) = \mu_r'$ onde r = 0, 1, 2, ...

A amostra a ser utilizada é $X_1, X_2, ..., X_n$. Para se obter o estimador dos mínimos quadrados de μ_r' é formada a seguinte soma:

$\sum_{i=1}^{n}\left[X_i^r - \mu_r'\right]^2$.

Deve-se escolher o valor μ_r' que torna a soma acima tão pequena quanto possível. Por exemplo, para se achar o estimador dos mínimos quadrados para a média populacional μ $(=\mu_r')$, determina-se para qual valor de μ a soma acima será mínima, isto é, deve-se minimizar a equação:

$\sum_{i=1}^{n}\left(X_i^r - \mu\right)^2$.

Para isto é necessário derivar esta equação em relação μ e igualar esta derivada a zero, obtendo-se:

$-2\sum_{i=1}^{n}(X_i - \mu) = 0$

Que resolvido em função de $\mu = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$.

Desta forma, verifica-se pelo método dos mínimos quadrados que a estimativa da média populacional é dada pela média da amostra, que coincide com o estimador obtido pelo método dos momentos.

As propriedades dos estimadores dos mínimos quadrados devem ser estabelecidas caso a caso.

2. ESTIMAÇÃO POR INTERVALO

O estimador por ponto não permite ter uma idéia do erro cometido ao se fazer a estimativa do parâmetro. Para que se possa associar uma confiança (probabilidade) a uma estimativa é necessário construir um intervalo em torno da estimativa por ponto. Este intervalo é construído baseado na distribuição amostral do estimador.

Para construir intervalos de confiança é necessário inicialmente fixar-se uma probabilidade "$1 - \alpha$" de que o intervalo construído contenha o parâmetro populacional. Esta probabilidade é denominada de confiança do intervalo. Desta forma, "α" será a probabilidade de que o intervalo obtido não contenha o valor do parâmetro, isto é, "α" será a probabilidade de erro.

2.1. DA MÉDIA POPULACIONAL (μ)

A construção de um intervalo de confiança para a média populacional (μ) envolve duas situações típicas. (1) Quando o desvio padrão populacional (σ) for conhecido e (2) Quando o desvio padrão populacional (σ) for desconhecido. A segunda situação é a mais comum, pois é pouco provável que não se conheça a média de uma população, por isto a necessidade de estimá-la, mas, no entanto, se conhece o desvio padrão. Entretanto por razões históricas e didáticas vamos manter as duas situações.

2.1.1. DESVIO PADRÃO POPULACIONAL (σ) CONHECIDO

O intervalo de confiança para a média (μ) de uma população é construído em torno da estimativa pontual $\overline{X}$. Sabe-se que a média da amostra tem distribuição normal de média μ e desvio padrão $\sigma/\sqrt{n}$ se a população de onde for extraída a amostra for normal (ou se a amostra for superior a 30 e retirada de qualquer população) de média μ e de desvio padrão σ, pode-se então utilizar a curva normal para estabelecer os limites para o intervalo de confiança.

Lembrando que o que se quer é um intervalo que contenha o parâmetro populacional μ com probabilidade "$1 - \alpha$" tem-se então:

$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$, onde $z_{\alpha/2}$ é o valor da normal padrão com área à direita é igual a $\alpha/2$.

Mas $Z = (\overline{X} - \mu) / \sigma/\sqrt{n}$ substituindo na expressão acima vem:

$P(-z_{\alpha/2} < (\overline{X} - \mu) / \sigma/\sqrt{n} < z_{\alpha/2}) = 1 - \alpha$. Trabalhando esta desigualdade, segue que:

$P(\overline{X} - z_{\alpha/2}\, \sigma/\sqrt{n} < \mu < \overline{X} + z_{\alpha/2}\, \sigma/\sqrt{n}) = 1 - \alpha$. Que é o intervalo procurado. Assim o intervalo de confiança (probabilidade) de "$1 - \alpha$" para a média de uma população, quando σ for conhecido, é dado por:

$[\overline{X} - z_{\alpha/2}\, \sigma/\sqrt{n}\, ;\, \overline{X} + z_{\alpha/2}\, \sigma/\sqrt{n}]$ onde:

$\overline{X}$ é a estimativa por ponto da média da população.

σ é o desvio padrão da população e

$z_{\alpha/2}$ é o valor da distribuição normal padrão cuja área à direita é igual a $\alpha/2$, isto é, é o valor de Z tal que: $P(Z > z_{\alpha/2}) = \alpha/2$, ou então: $\Phi(-z_{\alpha/2}) = \alpha/2$.

Exemplo:

Uma população tem um desvio padrão igual a 10 e média desconhecida. Uma amostra de tamanho n = 100 é retirada e fornece uma média $\overline{x} = 50$. Qual o intervalo de 95% de confiança para a média desta população?

Solução:

Têm-se $1 - \alpha = 95\%$, então $\alpha = 5\%$ e $\alpha / 2 = 2{,}5\%$. O coeficiente de confiança que deve ser buscado na normal padrão é valor $z_{\alpha/2}$ de Z tal que:

$P(Z > z_{\alpha/2}) = 2{,}5\%$, ou então: $\Phi(-z_{\alpha/2}) = 2{,}5\% \Rightarrow -z_{\alpha/2} = \Phi^{-1}(2{,}5\%) = -1{,}96$.

Então o intervalo de confiança de 95% para a média desta população será:

$[\overline{X} - z_{\alpha/2}\, \sigma/\sqrt{n}\, ;\, \overline{X} + z_{\alpha/2}\, \sigma/\sqrt{n}]$ = [50 - 1,96.10/10; 50 + 1,96.10/10] = [50 - 1,96; 50 + 1,96] = = [48,04; 51,96], ou seja, pode-se afirmar com uma certeza de 95% de que este intervalo conterá a média desta população.

Obs.: O valor $\varepsilon = z_{\alpha/2}\, \sigma/\sqrt{n}$ é denominado de erro padrão da estimação. Não confundir com o valor $\sigma/\sqrt{n}$ que é o erro padrão da amostragem. O erro padrão da estimação é a semi-amplitude do intervalo de confiança. A amplitude do intervalo de confiança (IC) será; 2ε.

2.1.2. DESVIO PADRÃO POPULACIONAL (σ) DESCONHECIDO

Quando o desvio padrão da população (σ) é desconhecido é necessário utilizar sua estimativa "$\hat{\sigma}$". Só que ao substituir-se o desvio padrão populacional pela sua estimativa no quociente:

$(\bar{X} - \mu) / \sigma/\sqrt{n}$ não se terá mais uma normal padrão. De fato, conforme demonstrado pelo estatístico inglês W. S. Gosset[1], conhecido por "Student" o comportamento do quociente:

$(\bar{X} - \mu) / \hat{\sigma}/\sqrt{n}$ segue uma distribuição simétrica em torno de zero, porém com uma variabilidade maior do que a da normal padrão. A distribuição do quociente acima é conhecida como distribuição "T" de Student.

2.1.3. A DISTRIBUIÇÃO T (DE STUDENT)

Na realidade existem infinitas distribuições "T", uma para cada tamanho de amostra. Estas distribuições a exemplo da normal padrão encontram-se tabeladas.

A tabela para a distribuição "T" segue uma metodologia um pouco diferente daquela da normal padrão. De fato, como as distribuições de Student não podem ser padronizadas, isto é, sofrer uma transformação de variável, a exemplo do ocorre na normal, não é possível construir uma única tabela. Assim, neste caso, cada linha de uma tabela representa uma distribuição diferente e cada coluna representa um valor de confiança que poderá ser "α" ou "$\alpha/2$", isto é, a tabela poderá ser unilateral ou bilateral. A linha de cada tabela fornece a distribuição "T" com parâmetro "n - 1" denominado de graus de liberdade, isto é, o grau de liberdade = ν = n - 1 = linha da tabela. (Rever na Apostila de Probabilidade)

2.1.4. O INTERVALO

Neste caso, o intervalo de confiança com probabilidade "$1 - \alpha$" para a média será:

$[\bar{X} - t_{\alpha/2}\,\hat{\sigma}/\sqrt{n}\ ;\ \bar{X} + t_{\alpha/2}\,\hat{\sigma}/\sqrt{n}]$ onde:

$\bar{X}$ é a estimativa por ponto da média da população;

[1] William S. Gosset: Estatístico inglês que trabalhava na cervejaria Guiness.

$\hat{\sigma}$ é o desvio padrão da amostra, calculado com "n-1" no denominador e uma estimativa do desvio padrão da população σ e $t_{\alpha/2}$ é o valor da distribuição T cuja área à direita é igual a $\alpha/2$, isto é, é o valor de T tal que:

$P(T > t_{\alpha/2}) = \alpha/2$, ou então: $P(- t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$.

Na realidade as tabelas fornecem os valores de "T" em função do parâmetro $\nu = n - 1$ = grau de liberdade e da confiança desejada (dada em função de uma ou duas caudas). Assim dado uma confiança unicaudal de $\alpha/2$, o valor tabelada é: $T^{-1}(n - 1, \alpha/2) = -t_{\alpha/2}$.

Exemplo:

Uma amostra de tamanho 25 foi retirada de uma população com o objetivo de estimar a sua média e forneceu os valores $\bar{x} = 50$ e $\hat{\sigma} = 10$. Qual o intervalo de 95% de confiança para a média desta população?

Solução:

Tem-se $1 - \alpha = 95\%$, então $\alpha = 5\%$ e $\alpha / 2 = 2{,}5\%$. O coeficiente de confiança que deve ser buscado na distribuição t com $\nu = n - 1 = 25 - 1 = 24$ (linha da tabela). A coluna deverá ser o valor $\alpha = 5\%$ (tabelas bilaterais) ou então o valor $\alpha / 2 = 2{,}5\%$ (tabelas unilaterais). Em qualquer caso o que se procura é o valor "T" com grau de liberdade igual a 24, isto é, o valor T_{24} tal que:

$P(- t_{\alpha/2} < T_{24} < t_{\alpha/2}) = 95\%$.

Este valor vale 2,064. (note-se que na normal este mesmo valor valia 1,96). Então o intervalo de confiança de 95% para a média desta população será:

$[\bar{X} - t_{\alpha/2}\, \hat{\sigma}/\sqrt{n}\ ;\ \bar{X} + t_{\alpha/2}\, \hat{\sigma}/\sqrt{n}\,] = [50 - 2{,}064.10/5;\ 50 + 2{,}064.10/5] = [50 - 4{,}13;\ 50 + 4{,}13] =$ $[45{,}87;\ 54{,}13]$, ou seja, pode-se afirmar com uma certeza de 95% de que este intervalo conterá a média desta população.

Convém notar que a última linha da tabela da distribuição "T" apresenta valores coincidentes com aqueles que seriam obtidos se fosse utilizada a distribuição normal padrão. Isto ocorre porque a distribuição "T" tende a distribuição normal à medida que o tamanho da amostra aumenta, isto é, a distribuição normal é o limite da distribuição "T" quando o tamanho da amostra tende ao infinito. Esta aproximação já será bastante boa para amostras de tamanho n > 30. No entanto, para evitar confusões sobre quando utilizar um ou outro modelo, não se recomenda a substituição de T por Z, a exemplo, do que fazem outros textos. Quando o valor do grau de liberdade não for encontrado na tabela, toma-se o valor imediatamente inferior.

2.2. DA PROPORÇÃO POPULACIONAL (σ)

Seja P = proporção amostral. Sabe-se que para n > 50 a distribuição amostral de P é aproximadamente normal com média $\mu_P = \pi$ e desvio padrão (erro padrão) $\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}}$. Pode-se então utilizar a curva normal para estabelecer os limites para o intervalo de confiança.

Lembrando que o que se quer é um intervalo que contenha o parâmetro populacional π com probabilidade "1 - α" então tem-se:

$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$, onde $z_{\alpha/2}$ é o valor da normal padrão com área à direita é igual a $\alpha/2$.

Mas $Z = (P - \mu_P) / \sigma_P$ então substituindo na expressão acima vem:

$P(-z_{\alpha/2} < (P - \mu_P) / \sigma_P < z_{\alpha/2}) = 1 - \alpha$. Trabalhando esta desigualdade, segue que:

$P(P - z_{\alpha/2}\sigma_P < \mu_P < P + z_{\alpha/2}\sigma_P) = P(P - z_{\alpha/2}\sigma_P < \pi < P + z_{\alpha/2}\sigma_P) = 1 - \alpha$. Que é o intervalo procurado. Assim o intervalo de confiança (probabilidade) de "1 - α" para a proporção "P" de uma população é dado por:

$$[P - z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} ; P + z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}].$$

Observando-se a expressão acima pode-se perceber que o intervalo de confiança para a proporção populacional π, depende dele mesmo, isto é, é necessário calcular o erro amostral que está expresso em função de π. Como o objetivo é estimar este valor, evidentemente ele não é conhecido. Assim é necessário utilizar, sua estimativa $\hat{\sigma}_P$, isto é, é necessário substituir π por P na expressão $\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}}$. Desta forma o intervalo acima ficará:

$$[P - z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}} ; P + z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}], \text{ onde:}$$

P é a estimativa por ponto da proporção populacional π.

$\hat{\sigma}_P = \sqrt{\frac{P(1-P)}{n}}$ é uma estimativa do erro padrão, isto é, do desvio padrão amostral e

$z_{\alpha/2}$ é o valor da distribuição normal padrão cuja área à direita é igual a $\alpha/2$. É o valor de Z tal que: $P(Z > z_{\alpha/2}) = \alpha/2$, ou então: $\Phi(-z_{\alpha/2}) = \alpha/2$.

Exemplo 1:

Numa pesquisa de mercado, 400 pessoas foram entrevistadas sobre sua preferência por determinado produto. Destas 400 pessoas, 240 disseram preferir o produto. Determinar um intervalo de confiança de 95% de probabilidade para o percentual de preferência dos consumidores em geral para este produto.

Solução:

Têm-se $1 - \alpha = 95\%$, então $\alpha = 5\%$ e $\alpha / 2 = 2,5\%$. O coeficiente de confiança que deve ser buscado na normal padrão é valor $z_{\alpha/2}$ de Z tal que:

$P(Z > z_{\alpha/2}) = 2,5\%$, ou então: $\Phi(-z_{\alpha/2}) = 2,5\%$.

Este valor vale 1,96. A estimativa por ponto para a proporção populacional será: $p = f/n = 240/400 = 0,60 = 60\%$.

Então o intervalo de confiança de 95% para a proporção populacional será:

$$[P - z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}} ; P + z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}] = [0,60 - 1,96\sqrt{\frac{0,60(1-0,60)}{400}} .; 0,60 + 1,96\sqrt{\frac{0,60(1-0,60)}{400}}]$$

= [60% - 4,80% ; 60% + 4,80%] = [55,20%; 64,80%], ou seja, pode-se afirmar com uma certeza de 95% de que este intervalo conterá a proporção populacional, isto é, a verdadeira percentagem dos consumidores que preferem o produto pesquisado.

Exemplo 2:

Numa pesquisa de mercado para estudar a preferência da população de uma cidade em relação ao consumo de um determinado produto, colheu-se uma amostra aleatória de 300 consumidores da cidade e observou-se que 180 consumiam o produto. Determinar um IC de 99% para a proporção populacional de consumidores do produto.

Solução:

Têm-se $1 - \alpha = 99\%$, então $\alpha = 1\%$ e $\alpha / 2 = 0,5\%$. O coeficiente de confiança que deve ser buscado na normal padrão é valor $z_{\alpha/2}$ de Z tal que:

$P(Z > z_{\alpha/2}) = 0{,}5\%$, ou então: $\Phi(-z_{\alpha/2}) = 0{,}5\%$.

Este valor vale 2,575. A estimativa por ponto para a proporção populacional será: p = f/n = 180/300 = 0,60 = 60%.

Então o intervalo de confiança de 99% para a proporção populacional será:

$$[P - z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}\ ;\ P + z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}\ ] = [0{,}60 - 2{,}58\sqrt{\frac{0{,}60(1-0{,}60)}{300}}\ .;\ 0{,}60 + 2{,}58\sqrt{\frac{0{,}60(1-0{,}60)}{300}}\ ]$$

= [60% - 7,28% ; 60% + 7,28%] = [52,72%; 67,28%], ou seja, pode-se afirmar com uma certeza de 99% de que este intervalo conterá a proporção populacional, isto é, a verdadeira percentagem dos consumidores que preferem o produto pesquisado.

2.3. DA VARIÂNCIA POPULACIONAL (σ^2)

Sabe-se que o estimador não-tendencioso de σ^2 é S^2 e que $E(S^2) = \sigma^2$, enquanto $V(S^2) = 2\sigma^2/(n-1)$. No entanto, para se construir um intervalo de confiança para σ^2 é necessário, ainda conhecer qual é o comportamento de S^2, isto é, qual é o modelo teórico (probabilístico) seguido pelo estimador. Assim antes de se construir um intervalo de confiança para a variância populacional é necessário se conhecer um novo modelo probabilístico denominado de qui-quadrado e representado por χ^2 (c grego).

A distribuição ou modelo qui-quadrado pode ser obtida de uma soma de variáveis normais padronizadas, isto é, $\chi^2_n = \sum_{i=1}^{n} Z_i^2$.

A distribuição χ^2 é assimétrica positiva (possuí uma cauda à direita) e de depende do parâmetro ν. Sabe-se também que:

$E(\chi^2) = \nu$ e que $V(\chi^2) = 2\nu$.

A figura 2.1 mostra alguns exemplos de modelos qui-quadrado. A comportamento, distribuição de probabilidade, apresentado pela variância amostral (S^2) está relacionado com a distribuição (modelo) χ^2 através do seguinte resultado:

$\chi^2_{n-1} = \frac{(n-1)S^2}{\sigma^2}$, isto é, a variância segue uma distribuição χ^2 com "n - 1" graus de liberdade a menos de uma constante. Neste caso $\nu = n - 1$.

2.3.1. TABELAS

A distribuição χ^2 está tabelada em função do grau de liberdade n - 1 = ν (linha da tabela) e área à sua direita, isto é, $P(\chi^2 > c) = \alpha$. Na realidade o que está tabelado é a função inversa da χ^2, isto é, entrando com o valor do parâmetro (graus de liberdade) e uma determinada probabilidade (área), a tabela fornece um valor da variável (abscissa) tal que a probabilidade à direita (área) deste valor seja igual a área especificada. (Ver Apostila de Probabilidade para mais detalhes).

2.3.2. O INTERVALO

Suponha que seja fixado um nível de confiança de "1 - α" e que χ_1^2 e χ_2^2 sejam dois valores da distribuição χ^2 tais que $P(\chi_1^2 < \chi^2 < \chi_2^2) = 1 - \alpha$.

$$P(\chi_1^2 < \chi^2 < \chi_2^2) = 1 - \alpha$$

$$P(\chi_1^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_2^2) = 1 - \alpha$$

$$P(\frac{1}{\chi_2^2} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi_1^2}) = 1 - \alpha$$

$$P(\frac{(n-1)S^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_1^2}) = 1 - \alpha$$

Assim o intervalo de confiança (probabilidade) de "1 - α" para a variância da população é dado por:

$$\left[\frac{(n-1)S^2}{\chi_2^2}; \frac{(n-1)S^2}{\chi_1^2}\right]$$

2.4. DO DESVIO PADRÃO POPULACIONAL (σ)

Para determinar um intervalo de confiança de "1 - α" de probabilidade para o desvio padrão populacional basta apenas tomar a raiz quadrada positiva dos termos do intervalo para a variância populacional. Assim o intervalo será:

$$\left[\sqrt{\frac{(n-1)S^2}{\chi_2^2}}; \sqrt{\frac{(n-1)S^2}{\chi_1^2}}\right]$$

O significado deste intervalo é:

$$P\left(\sqrt{\frac{(n-1)S^2}{\chi_2^2}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi_1^2}}\right) = 1-\alpha\,.$$

Exemplo:

Uma amostra extraída de uma população normal forneceu uma variância de s^2 = 8,38. Determinar um intervalo de confiança de 90% para a variância da população e um intervalo de mesma confiabilidade para o desvio padrão da população.

Solução

Neste caso é necessário inicialmente determinar os valores da distribuição χ^2, de modo, que χ_1^2 tenha uma área (probabilidade) à direita igual a 95% e χ_2^2 tenha uma área (probabilidade) à direita igual a 5%. Estes valores são: χ_1^2 = 3,940 e χ_2^2 = 18,307.

O intervalo de confiança, para a variância, será:

$$\left[\frac{(n-1)S^2}{\chi_2^2};\frac{(n-1)S^2}{\chi_1^2}\right]$$

$$\left[\frac{(11-1).8{,}38}{18{,}307};\frac{(11-1).8{,}38}{3{,}940}\right]$$

[4,58; 21,27]

O intervalo de confiança, para o desvio padrão, será:

$$\left[\sqrt{\frac{(n-1)S^2}{\chi_2^2}};\sqrt{\frac{(n-1)S^2}{\chi_1^2}}\right]$$

$$\left[\sqrt{\frac{(11-1).8{,}38}{18{,}307}};\sqrt{\frac{(11-1).8{,}38}{3{,}940}}\right]$$

$\sqrt{4{,}58;}\ \sqrt{21{,}27;}$]

[2,14; 4,61].

3. EXERCÍCIOS

(01) Seja X uma variável aleatória normal com média μ e desvio padrão σ. Sejam:

$\hat{\mu}_1 = \dfrac{\sum_{i=1}^{n} X_i}{n}$, $\hat{\mu}_2 = \dfrac{\sum_{i=1}^{n} X_i}{n+1}$ e $\hat{\mu}_3 = \dfrac{1}{2} + \dfrac{\sum_{i=2}^{n} X_i}{2n}$, três estimadores de μ. Determinar as propriedades dos três estimadores, isto é, suas expectâncias e variabilidades.

(02) Com base na amostra aleatória (X_1, X_2), considerar dois estimadores de μ.

$$\overline{X} = \frac{X_1 + X_2}{2} \text{ e } \mathbf{W} = \frac{1}{3}\mathbf{X}_1 + \frac{2}{3}\mathbf{X}_2$$

(02.1) Verificar se são não tendenciosos.

(02.2) Qual a eficiência de W em relação a $\overline{X}$? Qual o melhor?

(02.3) Provar que $\overline{X}$ é mais eficiente que qualquer outra combinação linear não-tendenciosa. (Sugestão: determinar a variância de $cX_1 + (1 - c)X_2$ em função de "c" e achar o seu mínimo, derivando-a e igualando-a a zero.)

(03) De uma distribuição normal com variância 2,25, obteve-se a seguinte amostra:

27,5; 25,6; 28,2; 26,1 e 25,0

Determinar um intervalo de confiança para a média desta população com confianças de:

(03.1) 95% **(03.2)** 99%

(04) Utilizando-se de uma aas de 145 profissionais de certa região, verificou-se que o salário médio é de 8 salários mínimos (s.m.) com um desvio padrão de 1,8 s.m. A amostra também forneceu a informação de que 70% dos profissionais eram casados.

(04.1) Determine e interprete o intervalo de confiança de 95% para o salário médio de todos os profissionais desta região.

(04.2) Determine e interprete o intervalo de confiança de 99% para a proporção de profissionais casados desta região?

(04.3) Determine e interprete um Intervalo de Confiança de 90% para σ^2.

(05) A amostra apresenta os valores da variável "tamanho da família" coletados através de uma aas em uma vila popular.

X	f
3	10
4	14
5	19
6	15
7	07

(05.1) Determine e interprete o intervalo de confiança de 95% para o parâmetro tamanho familiar médio por domicílio da vila.

(05.2) Determine e interprete o intervalo de confiança de 90% para o parâmetro proporção de domicílios da vila com tamanho igual ou superior a cinco.

(06) A variância de uma população é 150. Deseja-se obter um intervalo de confiança para a média da população com uma confiabilidade de 95% e um erro máximo de 2. Quantos valores desta população devem ser retirados aleatoriamente?

(07) Quer-se estimar a média de uma população de variância desconhecida através de um intervalo de confiança de 95% e com erro de estimação máximo de 5 unidades. Através de uma amostra piloto de 100 valores a variância foi estimada em 400 unidades. Que tamanho deve ter a amostra final?

(08) Uma amostra preliminar de pessoas de uma determinada comunidade apresentou 18% de analfabetos. Com este resultado quer-se estimar a proporção de analfabetos da população com uma confiabilidade de 95% e com um erro de estimação máximo de 2,5%. Qual o tamanho da amostra a ser utilizada?

(09) De uma população normalmente distribuída foi extraída uma aas de n = 10 que apresentou os valores abaixo:

4 8 12 5 7 9 10 11 6 8

(09.1) Determine uma estimativa da variância populacional.

(09.2) Determine uma estimativa da média populacional e do correspondente erro amostral?

(09.3) Determine um intervalo de confiança de 95% para a média desta população.

(10) A tabela apresenta os valores de uma amostra retirada de uma população normal. Determine:

X	f
04 |-- 08	8
08 |-- 12	8
12 |-- 16	6
16 |-- 20	4

(10.1) Um intervalo de confiança de 95% para a média desta população.

(10.2) Um intervalo de confiança de 99% para a média desta população.

4. RESPOSTAS DOS EXERCÍCIOS

(01) Quantos a expectância tem-se:

$$E(\hat{\mu}_1) = E(\overline{X}) = \mu$$

$$E(\hat{\mu}_2) = \frac{n\mu}{n+1}$$

$$E(\hat{\mu}_1) = \frac{\mu}{2} + \frac{(n-1)\mu}{2n} = \frac{(2n-1)\mu}{2n}.$$

Como $\hat{\mu}_2$ e $\hat{\mu}_3$ são tendenciosos só $\hat{\mu}_1 = \overline{X}$ é candidato à eficiência.

Quanto à acurácia, tem-se: EQM($\hat{\mu}$) = Var($\hat{\mu}$) + (Tend$\hat{\mu}$)^2

$$V(\hat{\mu}_1) = V(\overline{X}) = \frac{\sigma^2}{n}$$

$$V(\hat{\mu}_2) = V(\frac{\sum X_i}{n+1}) = \frac{n\sigma^2}{(n+1)^2}$$

$$V(\hat{\mu}_3) = V(\frac{1}{2} + \frac{\sum_{i=2}^{n} X_i}{2n}) = \frac{\mu}{2} + \frac{(n-1)\mu}{2n} = \frac{2(n-1)\mu}{2n}$$

(02) $E(\overline{X}) = E(\frac{X_1 + X_2}{2}) = \mu.$

$$E(W) = E(\frac{1}{3}X_1 + \frac{2}{3}X_2) = \frac{1}{3}E(X_1) + \frac{2}{3}E(X_2) = \frac{\mu}{3} + \frac{2\mu}{3} = \mu.$$

$$V(\overline{X}) = V(\frac{X_1 + X_2}{2}) = \frac{\sigma^2}{2}$$

$$V(W) = V(\frac{1}{3}X_1 + \frac{2}{3}X_2) = \frac{1}{9}V(X_1) + \frac{4}{9}V(X_2) = \frac{\sigma^2}{9} + \frac{4\sigma^2}{9} = \frac{5\sigma^2}{9}$$

A eficiência de $\overline{X}$ em relação a W é dada por: $\frac{V(\overline{X})}{V(W)}$ = 90%. Assim $\overline{X}$ é melhor.

(03) **(03.1)** [25,17; 27,79] **(03,2)** [24,75; 28,21]

(04) **(04.1)** [7,71; 8,29] Tem-se 95% de certeza de que o salário médio de todos os profissionais da área está entre 7,71 s.m. e 8,29 s.m.

(04.2) [60,20%; 79,80%] Tem-se 99% de confiança de que a percentagem de profissionais da área que são casados esteja entre 60,20% e 79,80%.

(04.3) [2,70; 3,98]. Tem-se 90% de confiança de que o valor da variância populacional pertença a este intervalo.

(05) **(05.1)** [4,62; 5,22] Tem-se 95% de confiança de que o valor médio do tamanho familiar da vila esteja entre 4,62 e 5,22 membros.

(05.2) [53,23%; 72,93%] Há 90 de certeza de que o percentual de famílias com 5 ou mais membros esteja entre 53,23% e 72,93%.

(06) n = 145

(07) n = 62, como a amostra piloto utilizada foi de n = 100 é mais confiável ficar com a amostra piloto.

(08) n = 908

(09) **(09.1)** 6,67 **(09.2)** 8 e 0,82 **(09.3)** [6,15; 9,85]

(10) **(10.1)** [9,19; 12,65] **(10.2)** [8,58; 13,26]

5. BIBLIOGRAFIA

KACHIGAN, Sam Kash. Statistical Analysis: An Interdisciplinary Introduction to Univariate & Multivariate Methods. New York: Radius Press, 1986, 589 p.

KMENTA, Jan. *Elementos de Econometria*. Tradução: Carlos Roberto Vieira Araújo. São Paulo: Atlas, 1978, 686 p.

MASON, Robert D., DOUGLAS, Lind A. *Statistical Techniques in Business And Economics*. IRWIN, Boston, 1990.

STEVENS, James. *Applied Multivariate Statistics For The Social Sciences*. Mahwah, New Jersey: LEA – Lawrence Erbaum Associates, Publishers