Universidade Federal do Piauí-UFPI
Campus Senador Helvídio Nunes de Barros-CSHNB
Plano Nacional de Formação de Professores da Educação Básica-PARFOR

Estimação por Intervalo

Professor: Josimar Vasconcelos
Contato: josimar@ufpi.edu.br
http://prof-josimar.blogspot.com.br/

18 de agosto de 2012

Roteiro

Intervalo de confiança

Introdução

Podemos visualizar a inferência estatística como um conjunto de ferramentas que objetiva estudar a população por meio de evidências fornecidas pela amostra. Na inferência estatística podemos trabalhar a estrimação pontual, estimação intervalar, testes de hipóteses, dentre outros.

Exemplo: Duvida-se da “honestidade” de um dado e decide-se lançá-lo n vezes antes de utilizá-lo em um jogo.

Conceitos básicos

- Parâmetros: μ, σ, σ^2, P.
- Estimadores: $\overline{X}$, S, S^2, $\widehat{P}$.
- Estimativas: são valores numéricos obtidos pelo estimador.

Esquema tático

Ferramenta	Estimador	Fórmula do estimador
Média	$\overline{X}$	$\overline{X} = \dfrac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$
Variância	S^2	$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$
DP	S	$S = \sqrt{S^2}$
Proporção	$\widehat{P}$	$\widehat{P} = \dfrac{\text{frequência amostral com a característica}}{n}$

Exemplos

1. Para detectar o apoio popular a um projeto governamental de reforma agrária, foram entrevistadas 400 pessoas espalhadas em várias capitais. A amostra contém as 400 respostas que consistem de **sim** (para aqueles que concordam com o projeto) e **não** para o caso contrário.
2. Para estudar o nível de colesterol em uma população de esportístas, coletamos uma amostra de 10 jovens atletas, obtendo os seguintes valores: 180, 196, 185, 165, 190, 195, 180, 176, 165, e 195.

Erro amostral

- É a diferença que existe entre os resultados obtido na amostra e os resultados que poderiam ter sido obtidos na população de interesse. Isto é,

$$e = \overline{X} - \mu.$$

Para o intervalo de confiança temos:

$$e = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = Z_{\alpha/2}\text{EP}\left(\overline{X}\right).$$

Estimação por intervalo

Até o momento discutimos estimadores pontuais, no qual fornecem estimativas com um único valor numérico para o parâmetro de interesse. Como são variáveis aleatórias, temos que os estimadores possuem uma distribuição de probabilidade, logo podemos apresentar uma estimativa mais informativa para o parâmetro populacional. Assim, chamamos esse método de estimação intervalar ou intervalo de confiança.

Estimação por intervalo

Até o momento discutimos estimadores pontuais,

Estimação por intervalo

Até o momento discutimos estimadores pontuais, no qual fornecem estimativas com um único valor numérico para o parâmetro de interesse.

Estimação por intervalo

Até o momento discutimos estimadores pontuais, no qual fornecem estimativas com um único valor numérico para o parâmetro de interesse. Como são **variáveis aleatórias**,

Estimação por intervalo

Até o momento discutimos estimadores pontuais, no qual fornecem estimativas com um único valor numérico para o parâmetro de interesse. Como são **variáveis aleatórias**, temos que os estimadores possuem uma **distribuição de probabilidade**,

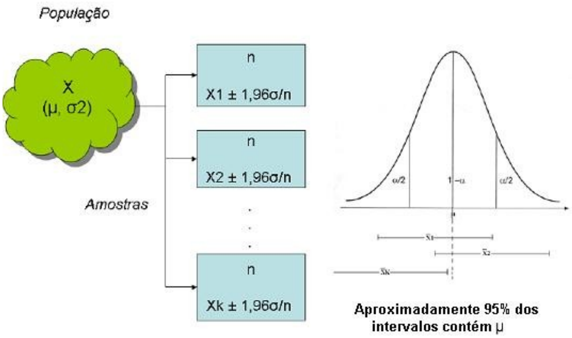
Estimação por intervalo

Até o momento discutimos estimadores pontuais, no qual fornecem estimativas com um único valor numérico para o parâmetro de interesse. Como são **variáveis aleatórias**, temos que os estimadores possuem uma **distribuição de probabilidade**, logo podemos apresentar uma estimativa mais informativa para o parâmetro populacional. Assim, chamamos esse método de estimação intervalar ou **intervalo de confiança**.

Dinâmica para entender o $\mathbb{IC}$

Para interpretar o $\mathbb{IC}$ da média, assumimos que os valores foram amostrados de forma independente e aleatória de um população normal com média μ e variância σ^2. Dado que estas suposições são válidas, temos 95% de "chance" do intervalo conter o verdadeiro valor da média populacional. Isto é, se produzirmos diversos intervalos de confiança provenientes de diferentes amostras independentes de mesmo tamanho, podemos esperar que aproximadamente 95% destes intervalos devem conter o verdadeiro valor da média populacional.

Dinâmica para entender o $\mathbb{IC}$



$\mathbb{IC}$ para a média μ de uma certa população Normal, com variância σ^2 conhecida

Imaginando uma amostra de tamanho n dada por $(X_1, X_2, \ldots, X_n)$, temos que a média amostral tem distribuição Normal com a mesma média μ e variância σ^2/n. Assim,

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

Continuação

Fixando um valor α tal que $0 < \alpha < 1$, podemos encontrar um valor $z_{\alpha/2}$ em que

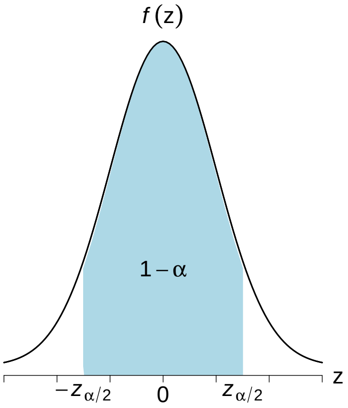
$$P\left(|Z| < z_{\alpha/2}\right) = \alpha.$$

Assim, o intervalo de confiança para μ, com coeficiente de confiança α, é dado por

$$\mathbb{IC}(\mu, \alpha) = \left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

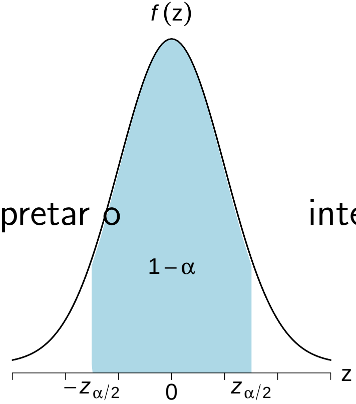
Observação

O $z_{\alpha/2}$ representa o valor de α dividido por 2 no qual a "massa" de probabilidade α deve ser distribuída igualmente em torno de 0.



Observação

O $z_{\alpha/2}$ representa o valor de α dividido por 2 no qual a "massa" de probabilidade α deve ser distribuída igualmente em torno de 0.



Como interpretar o intervalo de confiança?

Exemplos

Suponha que os comprimentos dos jacarés adultos de uma certa raça siga o modelo Normal com média μ desconhecida e variância igual a 0,01m^2. Uma amostra de dez animais foi sorteada e forneceu média 1,69m. Desejamos uma estimativa para o parâmetro desconhecido μ.

Exemplos

Suponha que os comprimentos dos jacarés adultos de uma certa raça siga o modelo Normal com média μ desconhecida e variância igual a 0,01m^2. Uma amostra de dez animais foi sorteada e forneceu média 1,69m. Desejamos uma estimativa para o parâmetro desconhecido μ.

$$\mathbb{IC}(\mu, 95\%) = [1,63; 1,75].$$

Exemplos

Suponha que os comprimentos dos jacarés adultos de uma certa raça siga o modelo Normal com média μ desconhecida e variância igual a 0,01m^2. Uma amostra de dez animais foi sorteada e forneceu média 1,69m. Desejamos uma estimativa para o parâmetro desconhecido μ.

$$\mathbb{IC}(\mu, 95\%) = [1,63; 1,75].$$

Portanto, com 95% de confiança podemos afirmar que a média dos comprimentos dos jacarés está entre 1,63 e 1,75.

Amplitude do intervalo

A aplitude do intervalode confiança é dada pela diferença entre o extremo superior e inferior, ou seja,

$$\overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) =$$

Amplitude do intervalo

A aplitude do intervalode confiança é dada pela diferença entre o extremo superior e inferior, ou seja,

$$\overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 2 \cdot z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Exemplos

1. Podemos tomar o exemplo dos comprimentos dos jacarés.

Exemplos

1. Podemos tomar o exemplo dos comprimentos dos jacarés.
2. A vida média de baterias automotivas de uma certa marca está sendo estudada. Baseado em estudos similares, com outras marcas, é possível admitir que a vida dessas baterias segue a distribuição Normal com desvio padrão de 4,5 meses. De qual tamanho deverá ser a amostra, para que a amplitude do intervalo de 90% de confiança para a vida média seja de 3 meses?

$\mathbb{IC}$ para a média μ de uma certa população Normal, com variância S^2 desconhecida

Dada uma AAS $(X_1, X_2, \ldots, X_n)$, obtida de uma população com distribuição Normal, com média μ e variância σ^2 desconhecidas, temos que

$$\mathrm{T} = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim \mathrm{t}_{(n-1)}.$$

Isto é, a variável T tem distribuição t de *Student* com $(n-1)$ graus de liberdade.

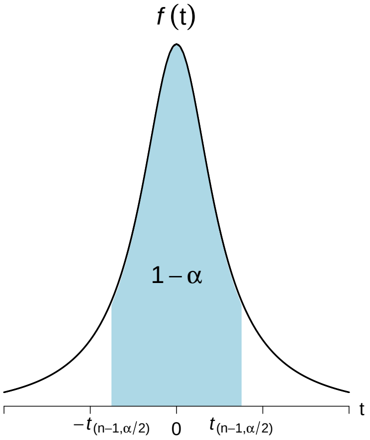
Continuação

Então, ao fixarmos o nível de significância α $(0 < \alpha < 1)$, obtemos da Tabela da distribuição t de Student com $(n-1)$ graus de liberdade, o valor $t_{((n-1),\alpha/2)}$, que satisfaz

$$P[-t_{((n-1),\alpha/2)} \leq \mathrm{T} \leq t_{((n-1),\alpha/2)}] = 1 - \alpha.$$

ou graficamente temos,

Continuação



Continuação

Analogamente ao caso anterior, obtemos que

$$\mathbb{IC}(\mu, \alpha) = \left(\overline{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} ; \overline{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right).$$

Exemplos

Desejamos investigar se uma certa moléstia que ataca o rim altera o consumo de oxigênio desse órgão. Para indivíduos sadios, admitimos que esse consumo tem distribuição Normal com média $12cm^3/min$. Os valores medidos em cinco pacientes com a moléstia foram: $14,4; 12,9; 15,0; 13,7$ e $13,5$.

Exemplos

Desejamos investigar se uma certa moléstia que ataca o rim altera o consumo de oxigênio desse órgão. Para indivíduos sadios, admitimos que esse consumo tem distribuição Normal com média $12cm^3/min$. Os valores medidos em cinco pacientes com a moléstia foram: $14,4; 12,9; 15,0; 13,7$ e $13,5$.

$$\mathbb{IC}(\mu, 90\%) = [13,09; 14,71].$$

$\mathbb{IC}$ para proporções

Dada uma variável aleatória X na presença (ou não) de determinada característica de uma população. Assim temos que X tem distribuição de Bernoulli com parâmetro p, onde p representa a probabilidade de um determinado elemento da amostra ter a característica de interesse.

$\mathbb{IC}$ para proporções

Dada uma variável aleatória X na presença (ou não) de determinada característica de uma população. Assim temos que X tem distribuição de Bernoulli com parâmetro p, onde p representa a probabilidade de um determinado elemento da amostra ter a característica de interesse.

Retiramos uma AAS $(X_1, \ldots, X_n)$ desta população. Cada X_i, tem distribuição de Bernoulli com parâmetro p, ou seja,

$X_1, \ldots, X_n \sim \mathit{Bernoulli}(p)$, com média $(\mu = p)$ e variância $\sigma^2 = p(1-p$

Continuação

Neste caso, o estimador de ($\hat{p}$) para o parâmetro populacional (p) é dado por

$$\hat{p} = \frac{\text{N}^{\text{o}} \text{ de elementos da amostra com a característica}}{\text{Total de elementos da amostra}} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}.$$

Logo, utilizaremos a aproximação da distribuição Normal com média p e variância $p(1-p)/n$. Desse modo,

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

Continuação

Observemos que a variância de $\hat{p}$ depende do parâmetro desconhecido p. No entanto, pelo fato de n ser grande, podemos substituir p por $\hat{p}$. Com isso temos que

$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1).$$

Considerando o mesmo procedimento de montagem do intervalo para a média, construímos o intervalo de confiança para a proporção p:

$$\mathbb{IC}(p,\alpha) = \left[\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right].$$

Exemplos

Pretende-se estimar a proporção p de cura, através do uso de um certo medicamento em doentes contaminados com cercária, que é uma das formas do verme da esquitossomose. Um experimento consistiu em aplicar o medicamento em 200 pacientes, escolhidos ao acaso, e observar que 160 deles foram curados.

Exemplos

Pretende-se estimar a proporção p de cura, através do uso de um certo medicamento em doentes contaminados com cercária, que é uma das formas do verme da esquitossomose. Um experimento consistiu em aplicar o medicamento em 200 pacientes, escolhidos ao acaso, e observar que 160 deles foram curados. O que podemos dizer da proporção p na população em geral?

Exemplos

Pretende-se estimar a proporção p de cura, através do uso de um certo medicamento em doentes contaminados com cercária, que é uma das formas do verme da esquitossomose. Um experimento consistiu em aplicar o medicamento em 200 pacientes, escolhidos ao acaso, e observar que 160 deles foram curados. O que podemos dizer da proporção p na população em geral?

$$\mathbb{IC}(\mu, 95\%) = [0,745; 0,855].$$

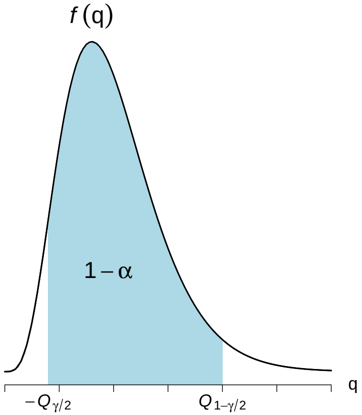
$\mathbb{IC}$ para variância

Consideremos uma amostra aleatória $(X_1, \ldots, X_n)$ de tamanho n de uma população com distribuição normal com média μ e desvio padrão σ. Um estimador para σ é a variância amostral s^2. Assim, sabemos que a quantidade pivotal

$$Q = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Continuação

Seja α a prob. da variável Q, com $(n-1)$ gl, tomar valores entre $Q_{\alpha/2}$ e $Q_{1-\alpha/2}$, valores obtidos na tabela da dist. Qui-quadrado tais que $P[Q < Q_{\alpha/2}] = P[Q > Q_{1-\alpha/2}] = \alpha/2$.



Continuação

Logo, pelo mesmo processo das seções anteriores, obtemos o intervalo de confiança para σ^2 por meio da seguinte expressão:

$$\mathbb{IC}(\sigma^2, \alpha) = \left(\frac{(n-1)s^2}{Q_{1-\alpha/2}}, \frac{(n-1)s^2}{Q_{\alpha/2}} \right).$$

Exemplos

O peso de componentes mecânicos produzidos por uma determinada empresa é uma va que se supõe ter dist. Normal. Pretende-se estudar a variabilidade do peso dos componentes. Para isso, uma amostra de tamanho 11 foi obtida,cujos valores em grama são: $\{98, 97, 102, 100, 98, 101, 102, 105, 95, 102, 100\}$. Construa um intervalo de confiança para a variância do peso, com um grau de confiança igual a 95%.

Exemplos

O peso de componentes mecânicos produzidos por uma determinada empresa é uma va que se supõe ter dist. Normal. Pretende-se estudar a variabilidade do peso dos componentes. Para isso, uma amostra de tamanho 11 foi obtida,cujos valores em grama são: {98, 97, 102, 100, 98, 101, 102, 105, 95, 102, 100}. Construa um intervalo de confiança para a variância do peso, com um grau de confiança igual a 95%.

$$\mathbb{I}\mathbb{C}(\sigma^2, 95\%) = [3,90; 24,61].$$