Probabilidade e Estatística

Matheus Pimenta

Universidade Tecnológica Federal do Paraná
Câmpus Cornélio Procópio

ADNP 2020

Estimação

Definição: As conjunto de técnicas e procedimentos que permitem dar ao pesquisador um grau de confiabilidade, de confiança, nas afirmações que faz para a população baseadas nos resultados das amostras.

Estimação

Definição: As conjunto de técnicas e procedimentos que permitem dar ao pesquisador um grau de confiabilidade, de confiança, nas afirmações que faz para a população baseadas nos resultados das amostras.
O problema fundamental da inferência estatística, portanto, é medir o *grau de incerteza ou risco* dessas generalizações.

Estimação de Parâmetros

É um dos objetivos básicos da experimentação. São dois tipos de estimação: por pontos e por intervalos.

Estimação de Parâmetros

É um dos objetivos básicos da experimentação. São dois tipos de estimação: por pontos e por intervalos.

Estimação por Pontos: a partir das observações, calcula-se uma estimativa, usando o estimador ou “estatística”.

Qualidades de um Bom Estimador

Quanto maior o grau de concentração da distribuição amostral do estimador em torno do verdadeiro valor do parâmetro populacional, tanto melhor será o estimador.

Qualidades de um Bom Estimador

Quanto maior o grau de concentração da distribuição amostral do estimador em torno do verdadeiro valor do parâmetro populacional, tanto melhor será o estimador.

As principais qualidades de um estimador são:

Qualidades de um Bom Estimador

Quanto maior o grau de concentração da distribuição amostral do estimador em torno do verdadeiro valor do parâmetro populacional, tanto melhor será o estimador.

As principais qualidades de um estimador são:

a) consistência;

Qualidades de um Bom Estimador

Quanto maior o grau de concentração da distribuição amostral do estimador em torno do verdadeiro valor do parâmetro populacional, tanto melhor será o estimador.

As principais qualidades de um estimador são:

a) consistência;

b) ausência de vício;

Qualidades de um Bom Estimador

Quanto maior o grau de concentração da distribuição amostral do estimador em torno do verdadeiro valor do parâmetro populacional, tanto melhor será o estimador.

As principais qualidades de um estimador são:

a) consistência;

b) ausência de vício;

c) eficiência;

Qualidades de um Bom Estimador

Quanto maior o grau de concentração da distribuição amostral do estimador em torno do verdadeiro valor do parâmetro populacional, tanto melhor será o estimador.

As principais qualidades de um estimador são:

a) consistência;

b) ausência de vício;

c) eficiência;

d) suficiência.

Qualidades de um Bom Estimador

Quanto maior o grau de concentração da distribuição amostral do estimador em torno do verdadeiro valor do parâmetro populacional, tanto melhor será o estimador.

As principais qualidades de um estimador são:

a) consistência;

b) ausência de vício;

c) eficiência;

d) suficiência.

As definições formais requerem conhecimentos de cálculo.

Estimação por Intervalo

Definição:

A estimação por pontos de um parâmetro não possui uma medida do possível erro cometido na estimação. Para solucionar isto, uma alternativa é estabelecer *limites*, que com certa probabilidade incluam o verdadeiro valor do parâmetro da população.

Estimação por Intervalo

Definição:

A estimação por pontos de um parâmetro não possui uma medida do possível erro cometido na estimação. Para solucionar isto, uma alternativa é estabelecer *limites*, que com certa probabilidade incluam o verdadeiro valor do parâmetro da população.
Estes limites, são definidos como *limites de confiança* e determinam um intervalo de confiança, no qual deverá estar o verdadeiro valor do parâmetro.

Estimação por Intervalo

Definição:

A estimação por pontos de um parâmetro não possui uma medida do possível erro cometido na estimação. Para solucionar isto, uma alternativa é estabelecer *limites*, que com certa probabilidade incluam o verdadeiro valor do parâmetro da população.
Estes limites, são definidos como *limites de confiança* e determinam um intervalo de confiança, no qual deverá estar o verdadeiro valor do parâmetro.
Assim, a estimação por intervalos consiste na fixação de dois valores, tais que $(1-\alpha)$ seja a probabilidade de que o intervalo, por eles determinado, contenha o verdadeiro valor do parâmetro.

Estimação por Intervalo

α : é o nível de incerteza ou grau de desconfiança;

Estimação por Intervalo

α : é o nível de incerteza ou grau de desconfiança;
$1 - \alpha$: é o nível de confiabilidade.

Estimação por Intervalo

α : é o nível de incerteza ou grau de desconfiança;
$1 - \alpha$: é o nível de confiabilidade.
Logo, α nos da o nível de incerteza desta inferência, chamamos de grau de significância.

Intervalos de Confiança (IC) para a média μ de uma população Normal com variância σ^2 conhecida

Considerando uma população normal com média desconhecida que desejamos estimar e com σ^2 conhecida, $X : N(?, \sigma^2)$.

Intervalos de Confiança (IC) para a média μ de uma população Normal com variância σ^2 conhecida

Considerando uma população normal com média desconhecida que desejamos estimar e com σ^2 conhecida, $X : N(?, \sigma^2)$.
O passo a passo para obter intervalos de confiança são:

Intervalos de Confiança (IC) para a média μ de uma população Normal com variância σ^2 conhecida

Considerando uma população normal com média desconhecida que desejamos estimar e com σ^2 conhecida, $X : N(?, \sigma^2)$.
O passo a passo para obter intervalos de confiança são:

1. Retiramos uma amostra casual simples com n elementos;
2. Calculamos a média da amostra $\bar{x}$;
3. Calculamos o desvio padrão da média amostral: $\sigma_{\bar{x}} = \sqrt{\dfrac{\sigma^2}{n}} = \dfrac{\sigma}{\sqrt{n}}$;
4. Fixamos um nível de significância α, e com ele determinamos z_α, tal que $P(|z| > z_\alpha) = \alpha$, ou seja: $P(z > z_\alpha) = \dfrac{\alpha}{2}$ e $P(z < -z_\alpha) = \dfrac{\alpha}{2}$.
   Logo devemos ter: $P(|z| < z_\alpha) = 1 - \alpha$.
   Com isso, desenvolvendo a fórmula anterior chegamos a:
   $$P(\bar{x} - z_\alpha \cdot \sigma_{\bar{x}} < \mu < \bar{x} + z_\alpha \cdot \sigma_{\bar{x}})$$
   Que é a fórmula do *IC* para a média de populações normais com variância conhecidas.

Intervalos de Confiança (IC) para a média μ de uma população Normal com variância σ^2 conhecida

Simplificando a notação temos com os limites anteriores: $\mu_1 = \bar{x} - z_\alpha \cdot \sigma_{\bar{x}}$ e $\mu_2 = \bar{x} + z_\alpha \cdot \sigma_{\bar{x}}$, com isto segue que:

$$IC(\mu, (1-\alpha)\%) = (\mu_1, \mu_2)$$

Intervalos de Confiança (IC) para a média μ de uma população Normal com variância σ^2 conhecida

Simplificando a notação temos com os limites anteriores: $\mu_1 = \bar{x} - z_\alpha \cdot \sigma_{\bar{x}}$ e $\mu_2 = \bar{x} + z_\alpha \cdot \sigma_{\bar{x}}$, com isto segue que:

$$IC(\mu, (1-\alpha)\%) = (\mu_1, \mu_2)$$

Em outras palavras, tomando $\alpha = 5\%$, podemos esperar que 95 dos IC contenham o verdadeiro valor de μ e 5 não contenham o valor de μ, em 100 amostras de mesmo tamanho n, onde obteremos 100 estimativas para $\bar{x}$, com as quais construiremos 100 IC para μ.

Intervalos de Confiança (IC) para a média μ de uma população Normal com variância σ^2 conhecida

Simplificando a notação temos com os limites anteriores: $\mu_1 = \bar{x} - z_\alpha \cdot \sigma_{\bar{x}}$ e $\mu_2 = \bar{x} + z_\alpha \cdot \sigma_{\bar{x}}$, com isto segue que:

$$IC(\mu, (1-\alpha)\%) = (\mu_1, \mu_2)$$

Em outras palavras, tomando $\alpha = 5\%$, podemos esperar que 95 dos IC contenham o verdadeiro valor de μ e 5 não contenham o valor de μ, em 100 amostras de mesmo tamanho n, onde obteremos 100 estimativas para $\bar{x}$, com as quais construiremos 100 IC para μ.
Isto é, em uma amostra qualquer, a probabilidade de que o IC determinado contenha o valor da média é de 95%, ou seja, uma confiança de 95% de que o IC determinado contenha o verdadeiro valor de μ. O risco que corremos de que não contenha o verdadeiro valor é de 5%.

Tabela

Tabela para valores de Z_α.

Nível de Confiança	99,73%	99%	98%	96%	95,45%	95%	90%	80%	68,27%	50%
Z_α	3,00	2,58	2,33	2,05	2,00	1,96	1,645	1,28	1,00	0,6745

Exemplo 01

De uma população normal X, com $\sigma^2 = 9$, tiramos uma amostra de 25 observações, obtendo $\sum_{i=1}^{25} x_i = 152$. Determinar um IC de limites de 90% para μ.

Exemplo 01 - Solução

Solução:

$$\alpha = 10\% \text{ e } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{152}{25} = 6,08$$

Exemplo 01 - Solução

Solução:

$\alpha = 10\%$ e $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i = \dfrac{152}{25} = 6,08$

$\sigma_{\bar{x}} = \sqrt{\dfrac{\sigma^2}{25}} = \sqrt{\dfrac{9}{25}} = \dfrac{3}{5} \implies \sigma_{\bar{x}} = 0,6$

Exemplo 01 - Solução

Solução:

$\alpha = 10\%$ e $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i = \dfrac{152}{25} = 6,08$

$\sigma_{\bar{x}} = \sqrt{\dfrac{\sigma^2}{25}} = \sqrt{\dfrac{9}{25}} = \dfrac{3}{5} \implies \sigma_{\bar{x}} = 0,6$

Utilizando a tabela do inicio das notas de aula, segue que:

Exemplo 01 - Solução

Solução:

$\alpha = 10\%$ e $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i = \dfrac{152}{25} = 6,08$

$\sigma_{\bar{x}} = \sqrt{\dfrac{\sigma^2}{25}} = \sqrt{\dfrac{9}{25}} = \dfrac{3}{5} \implies \sigma_{\bar{x}} = 0,6$

Utilizando a tabela do inicio das notas de aula, segue que:

$z_\alpha = z_{45\%} = z_{0,45} = 1,64$

Exemplo 01 - Solução

Solução:

$\alpha = 10\%$ e $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{152}{25} = 6,08$

$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{25}} = \sqrt{\frac{9}{25}} = \frac{3}{5} \implies \sigma_{\bar{x}} = 0,6$

Utilizando a tabela do inicio das notas de aula, segue que:

$z_\alpha = z_{45\%} = z_{0,45} = 1,64$

Com isto, nosso intervalo de confiança é dado por:

$$P(6,08 - 1,64 \cdot 0,6 < \mu < 6,08 + 1,64 \cdot 0,6) = 0,9$$

$$P(5,096 < \mu < 7,064) = 0,90$$

Exemplo 01 - Solução

Solução:

$\alpha = 10\%$ e $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{152}{25} = 6,08$

$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{25}} = \sqrt{\frac{9}{25}} = \frac{3}{5} \implies \sigma_{\bar{x}} = 0,6$

Utilizando a tabela do inicio das notas de aula, segue que:

$z_\alpha = z_{45\%} = z_{0,45} = 1,64$

Com isto, nosso intervalo de confiança é dado por:

$$P(6,08 - 1,64 \cdot 0,6 < \mu < 6,08 + 1,64 \cdot 0,6) = 0,9$$

$$P(5,096 < \mu < 7,064) = 0,90$$

Ou ainda,

$$IC(\mu, 90\%) = (5,096; 7,064)$$

Exemplo 01 - Solução

Portanto, temos 90% de confiança que o verdadeiro valor μ populacional se encontra entre $5,096$ e $7,064$, ou então corremos um risco de 10% de que o verdadeiro valor da média μ populacional seja menor que $5,096$ ou maior que $7,064$.

$\blacksquare$

Intervalos de Confiança para a Média de Populações Normais com Variâncias Desconhecidas

Quando queremos estimar a média de uma população normal com variância desconhecida, consideramos dois procedimentos:

Intervalos de Confiança para a Média de Populações Normais com Variâncias Desconhecidas

Quando queremos estimar a média de uma população normal com variância desconhecida, consideramos dois procedimentos:

- se $n \leq 30$, então usa-se a distribuição t de Student, que veremos a diante;

Intervalos de Confiança para a Média de Populações Normais com Variâncias Desconhecidas

Quando queremos estimar a média de uma população normal com variância desconhecida, consideramos dois procedimentos:

- se $n \leq 30$, então usa-se a distribuição t de Student, que veremos a diante;
- se $n > 30$, então usa-se a distribuição normal com o estimador s^2 de σ^2.

Intervalos de Confiança para a Média de Populações Normais com Variâncias Desconhecidas

Quando queremos estimar a média de uma população normal com variância desconhecida, consideramos dois procedimentos:

- se $n \leq 30$, então usa-se a distribuição t de Student, que veremos a diante;
- se $n > 30$, então usa-se a distribuição normal com o estimador s^2 de σ^2.

Nesta seção nosso interesse é no segundo caso. Vejamos um exemplo.

Exemplo 02

De uma população normal com parâmetros desconhecidos, tiramos uma amostra de tamanho 100, obtendo-se $\bar{x} = 112$ e $s = 11$. Fazer um IC para μ ao nível de 10%.

Exemplo 02 - Solução

Solução:

Como a amostra é superior a 30, utilizamos:

Exemplo 02 - Solução

Solução:

Como a amostra é superior a 30, utilizamos:

$$\sigma_{\bar{x}} \approx \frac{s}{\sqrt{n}} = \frac{11}{10} = 1,1$$

Exemplo 02 - Solução

Solução:

Como a amostra é superior a 30, utilizamos:

$\sigma_{\bar{x}} \approx \dfrac{s}{\sqrt{n}} = \dfrac{11}{10} = 1,1$

$z_{\alpha} = z_{45\%} = z_{0,45} = 1,64$

Exemplo 02 - Solução

Solução:

Como a amostra é superior a 30, utilizamos:

$\sigma_{\bar{x}} \approx \dfrac{s}{\sqrt{n}} = \dfrac{11}{10} = 1,1$

$z_{\alpha} = z_{45\%} = z_{0,45} = 1,64$

Logo,

Exemplo 02 - Solução

Solução:

Como a amostra é superior a 30, utilizamos:

$\sigma_{\bar{x}} \approx \frac{s}{\sqrt{n}} = \frac{11}{10} = 1,1$

$z_{\alpha} = z_{45\%} = z_{0,45} = 1,64$

Logo,

$$P(112 - 1,64 \cdot 1,1 < \mu < 112 + 1,64 \cdot 1,1) = 0,90$$

$$P(110,20 < \mu < 113,80) = 0,90$$

Exemplo 02 - Solução

Solução:

Como a amostra é superior a 30, utilizamos:

$\sigma_{\bar{x}} \approx \dfrac{s}{\sqrt{n}} = \dfrac{11}{10} = 1,1$

$z_{\alpha} = z_{45\%} = z_{0,45} = 1,64$

Logo,

$$P(112 - 1,64 \cdot 1,1 < \mu < 112 + 1,64 \cdot 1,1) = 0,90$$

$$P(110,20 < \mu < 113,80) = 0,90$$

Ou

Exemplo 02 - Solução

Solução:

Como a amostra é superior a 30, utilizamos:

$\sigma_{\bar{x}} \approx \dfrac{s}{\sqrt{n}} = \dfrac{11}{10} = 1,1$

$z_{\alpha} = z_{45\%} = z_{0,45} = 1,64$

Logo,

$$P(112 - 1,64 \cdot 1,1 < \mu < 112 + 1,64 \cdot 1,1) = 0,90$$
$$P(110,20 < \mu < 113,80) = 0,90$$

Ou

$$IC(\mu, 90\%) = (110,20; 113,80)$$

Exemplo 02 - Solução

Solução:
Como a amostra é superior a 30, utilizamos:
$\sigma_{\bar{x}} \approx \dfrac{s}{\sqrt{n}} = \dfrac{11}{10} = 1,1$
$z_\alpha = z_{45\%} = z_{0,45} = 1,64$
Logo,

$$P(112 - 1,64 \cdot 1,1 < \mu < 112 + 1,64 \cdot 1,1) = 0,90$$
$$P(110,20 < \mu < 113,80) = 0,90$$

Ou

$$IC(\mu, 90\%) = (110,20; 113,80)$$

O que concluímos que apesar de usar o desvio padrão da amostra, temos um grau de certeza de 90% de que o verdadeiro valor da média populacional está entre $110,20$ e $113,80$. ■