Estatística II

Licenciatura em Gestão

2.º Ano/1.º Semestre

2023/2024

Aulas Teóricas N.ºs 6 e 7 (Semana 4)

Docente: Elisabete Fernandes
E-mail: efernandes@iseg.ulisboa.pt

https://doity.com.br/estatistica-aplicada-a-nutricao

https://basiccode.com.br/produto/informatica-basica/

Conteúdos Programáticos

Aulas Teóricas (Semanas 1 a 5)	Aulas Teóricas (Semanas 5 a 7)	Aulas Teóricas (Semanas 7 a 9)	Aulas Teóricas (Semanas 10 a 13)
• **Capítulo 1**: Estimação	• **Capítulo 2**: Testes de Hipóteses	• **Capítulo 3**: Modelo de Regressão Linear	• **Capítulo 4**: Complementos ao Modelo de Regressão Linear

Material didático: Exercícios do Livro Murteira et al (2015), Formulário e Tabelas Estatísticas

Bibliografia: B. Murteira, C. Silva Ribeiro, J. Andrade e Silva, C. Pimenta e F. Pimenta; *Introdução à Estatística*, 2ª ed., Escolar Editora, 2015.

1. Estimação

a. Introdução
b. Métodos dos momentos
c. Método da máxima verosimilhança
d. Propriedade dos estimadores por pontos
e. Estimação por intervalos para populações normais
f. Estimação por intervalos para populações não normais (grandes amostras)

4ª semana (10/10 e 12/10)

T06 - Intervalos de confiança

Introdução. Método da variável fulcral. Aplicação a universos normais: média e variância. Exemplos

T07 - Estimação por intervalos

Método da variável fulcral: Aplicação a universos normais (2 amostras) para a estimação da diferença de média e rácio de variâncias. Exemplos.

1 Estimação Intervalar

Intervalos de Confiança (ICs)

Estimação Pontual vs Estimação Intervalar

Existem dois processos de estimação paramétrica:

✓**Estimação Pontual**: produção de um valor (estimativa), que se pretende que seja o melhor, para um determinado parâmetro da população, com base na informação amostral;

✓**Estimação Intervalar**: construção de um intervalo que, com certo grau de certeza previamente estipulado, se pretende que contenha o verdadeiro valor do parâmetro da população.

Um **estimador** dum parâmetro da população é uma variável aleatória (v. a.) ou função que depende da informação amostral e cujas realizações fornecem aproximações para o parâmetro desconhecido. A um valor específico assumido por este estimador para uma amostra em concreto chama-se **estimativa**.

Afonso & Nunes, 2019

Estimação Pontual vs Estimação Intervalar

Falámos até aqui da estimação pontual e métodos de determinar estimadores de um parâmetro desconhecido θ.

Iremos agora tratar a questão da estimação intervalar.

Os intervalos são preferíveis quando, em vez de se propôr uma estimativa isolada, $\hat{\theta}$, podemos associar-lhe uma medida de erro $\hat{\theta} \pm \epsilon$, para significar que provavelmente o verdadeiro valor do parâmetro estará em $\hat{\theta} - \epsilon\,, \hat{\theta} + \epsilon$.

moduloI_aula5.pdf

Intervalo de Confiança (IC)

Definição

Considere-se uma amostra aleatória $(X_1, X_2, \cdots X_n)$ de uma população com função de distribuição $F(x|\theta)$. Sejam $\Theta_1^\star(X_1, \cdots X_n)$ e $\Theta_2^\star(X_1, \cdots X_n)$ duas estatísticas, tais que

$$P(\Theta_1^\star < \theta < \Theta_2^\star) = 1 - \alpha, \qquad 0 < \alpha < 1,$$

onde α é uma constante, não dependente do parâmetro θ.

Diz-se que $(\Theta_1^\star, \Theta_2^\star)$ é um intervalo aleatório, que contém θ com probabilidade $1 - \alpha$.

moduloI_aula5.pdf

Intervalo de Confiança

Com a utilização de um intervalo de confiança para estimarmos um parâmetro ficamos a ganhar?

Efectivamente, pensemos por exemplo, no estimador $\overline{X}$.
Tem-se $P[\overline{X} = \mu] = 0$, mas já temos uma probabilidade positiva se considerarmos

$$P\{\mu \in]\overline{X} - a, \overline{X} + a[\ \} \quad \text{com a} > 0$$

ou seja, há uma probabilidade positiva de o intervalo aleatório conter o parâmetro desconhecido.

modulol_aula5.pdf

Intervalo de Confiança

Definição

A qualquer intervalo $(\theta_1^\star, \theta_2^\star)$, com $\theta_1^\star < \theta_2^\star$, números reais, que resulta da concretização do intervalo aleatório chama-se **intervalo de confiança** a $(1-\alpha)100\%$ para θ.

Observações: Chama-se **precisão da estimativa** à semi-amplitude do intervalo de confiança e **confiança** ou **grau de confiança** a $(1-\alpha) \times 100\%$ Quanto maior for o intervalo, maior é o grau de confiança, mas menor a precisão da estimativa.

A **margem de erro ou precisão da estimativa** é metade da largura do intervalo de confiança.

moduloI_aula5.pdf

ICs vs Distribuições por Amostragem

A determinação de intervalos de confiança para os parâmetros necessita do conhecimento da distribuição dos estimadores envolvidos distribuições por amostragem, isto é, são distribuições de funções da amostra aleatória $(X_1, X_2, \cdots, X_n)$, que vamos usar para obter **Intervalos de Confiança**

moduloI_aula5.pdf

ICs vs Distribuições por Amostragem: Exemplos

AMOSTRAGEM. DISTRIBUIÇÕES POR AMOSTRAGEM

Formulário

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad ; \quad S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n} = \frac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2 \quad ; \quad (n-1)S'^2 = n\,S^2$$

$$E(\overline{X}) = \mu \quad ; \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} \quad ; \quad E(S^2) = \frac{n-1}{n}\sigma^2 ; \quad E(S'^2) = \sigma^2$$

IC vs Distribuições de Amostragem: Exemplos

Relembrando

- Se $X \frown N(\mu, \sigma)$ e $\boxed{\sigma \text{ conhecido}}$ $\longrightarrow \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \frown N(0,1)$

Para obter o I.C. para μ com $\boxed{\sigma \text{ desconhecido}}$

Variável usada	Condições	Distribuição
$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$ $S'^2 = \dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$	$X_i \frown N(\mu, \sigma)$ $i = 1, 2, \cdots, n$	$t_{(n-1)}$

Variância corrigida

$$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Definição da distribuição t – Student

Se $Z \frown N(0,1)$ e $X \frown \chi^2_{(n)}$ são v.a. independentes

$$\Downarrow$$

$$T = \frac{Z}{\sqrt{X/n}} \frown t_{(n)}$$

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

$$\frac{(n-1)\ S'^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$\frac{\bar{X}-\mu}{S'/\sqrt{n}} \sim \boxed{t_{(n-1)}}$$

modulol_aula5.pdf

Intervalo de Confiança para o Valor Médio μ

2

Intervalo de Confiança para μ

Intervalo de confiança a $(1-\alpha)\times 100\%$ para μ quando $X \frown N(\mu, \sigma)$

- Se σ conhecido $\left]\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\ \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right[.$

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Variância corrigida

$$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- Se σ desconhecido $\left]\bar{X} - t_{n-1,1-\frac{\alpha}{2}}\frac{S'}{\sqrt{n}};\ \bar{X} + t_{n-1,1-\frac{\alpha}{2}}\frac{S'}{\sqrt{n}}\right[.$

$$\bar{x} - t_{1-\alpha/2(n-1)}\frac{S'}{\sqrt{n}} < \mu < \bar{x} + t_{1-\alpha/2(n-1)}\frac{S'}{\sqrt{n}}$$

moduloI_aula5.pdf

Intervalo de Confiança para μ

Amplitude do IC para μ: $2 \times z_{1-\alpha/2;}\,(\sigma / n^{1/2})$

Amplitude do IC para μ: $2 \times t_{1-\alpha/2;\, n-1}\,(s/ n^{1/2})$

$\left(\bar{x} - z_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}};\ \bar{x} + z_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right)$	$\left(\bar{x} - t_{1-\alpha/2}\dfrac{S'}{\sqrt{n}};\ \bar{x} + t_{1-\alpha/2}\dfrac{S'}{\sqrt{n}}\right)$

A **margem de erro** é metade da largura do intervalo de confiança.

O **grau ou nível de confiança** ($1-\alpha$) de um intervalo de confiança é a probabilidade de este vir a incluir o verdadeiro valor do parâmetro populacional.

Os **níveis de significância** (α) são as probabilidades complementares dos níveis de significância e são usados para testar a hipótese nula (H0) num teste de hipóteses.

- 1%, 5%, 10% = alfa => **Níveis de significância**.
- 99%, 95%, 90% = (1-alfa) => Níveis de confiança

O erro padrão da média é uma medida de variação de uma média amostral em relação à média da população. Sendo assim, é uma medida que ajuda a verificar a confiabilidade da média amostral calculada.

Quanto menor o erro padrão mais precisas são as estimativas!

O desvio padrão (s) é uma medida que indica a dispersão dos dados dentro de uma amostra com relação à média.

Este tipo de ICs só são válidos se a variável em estudo tem distribuição normal ou se a amostra em análise é de grande dimensão (i.e., $n \geq 30$) (ver Teorema do Limite Central).

Curiosidade: Intervalos de Confiança no R package

Intervalo de Confiança para μ: Exemplo

Exemplo de construção de um I.C. no R, para o valor médio de uma normal com variância conhecida (exemplo académico!)

Exemplo Dada a amostra referente a 10 alturas, admita-se que os erros de medição são normais de média 0 e desvio padrão 1.5.

```
> x<-c(175,176, 173, 175, 174, 173, 173, 176, 173, 179)
> int.conf.z<-function(x,sigma,conf.level=0.95)
  n <-length(x);xbar<-mean(x)
  alpha <- 1 - conf.level
  zstar <- qnorm(1-alpha/2)
  SE <- sigma/sqrt(n)
  xbar + c(-zstar*SE,zstar*SE)    ## definimos uma função
> int.conf.z(x,1.5)        # basta fazer isto
```

Obteve-se o I.C a 95% para μ $]173.7703; 175.6297[$

modulol_aula5.pdf

Intervalo de Confiança para μ e n ≥ 30: X tem qualquer Distribuição

Intervalo de confiança a $(1-\alpha) \times 100\%$ para μ

Se X tem **dist. qualquer não normal**

É necessário dispor de uma $\boxed{\text{amostra de dimensão elevada}}$, i.e., ***n*** grande $\longrightarrow$ aplicação do Teorema Limite Central

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \qquad \text{se } \underline{\sigma \text{ conhecido}}$$

Ou, que é o caso mais frequente,

$$\frac{\overline{X}-\mu}{s/\sqrt{n}} \sim \mathcal{N}(0,1) \qquad \text{se } \underline{\sigma \text{ desconhecido}}$$

Intervalo a $(1-\alpha) \times 100\%$ de confiança para μ

$$\overline{x} - z_{1-\alpha/2}\frac{S'}{\sqrt{n}} < \mu < \overline{x} + z_{1-\alpha/2}\frac{S'}{\sqrt{n}}$$

$$\left]\overline{X} - z_{1-\frac{\alpha}{2}}\frac{S'}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}}\frac{S'}{\sqrt{n}}\right[.$$

moduloI_aula5.pdf

IC para μ: Formulário

Variância corrigida

$$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- **POPULAÇÕES NORMAIS**

	Variância Conhecida	Variância Desconhecida
Média	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{X}-\mu}{S'/\sqrt{n}} \sim t(n-1)$
Diferença de médias	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \sim N(0,1)$ $T = \frac{\frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{1}{m}+\frac{1}{n}}}}{\sqrt{\frac{(m-1)S_1'^2+(n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	$Z = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1'^2}{m}+\frac{S_2'^2}{n}}} \overset{a}{\sim} t(\nu)$ onde ν é o maior inteiro contido em r, $r = \frac{\left(\frac{s_1'^2}{m}+\frac{s_2'^2}{n}\right)^2}{\frac{1}{m-1}\left(\frac{s_1'^2}{m}\right)^2+\frac{1}{n-1}\left(\frac{s_2'^2}{n}\right)^2}$
Variância	$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	
Relação de variâncias	$\frac{S_1'^2}{S_2'^2}\frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$	

IC para μ: Formulário

Variância corrigida

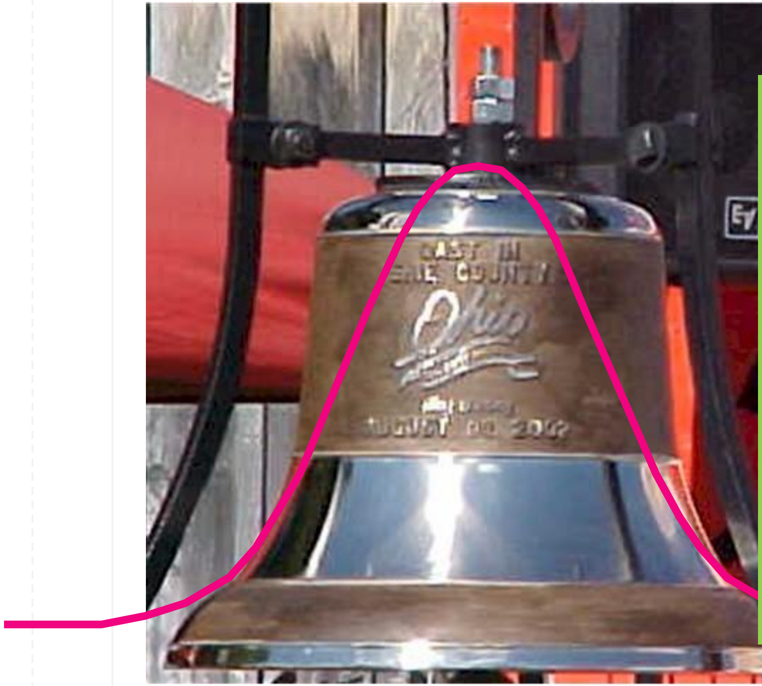
$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

- **GRANDES AMOSTRAS**

Caso geral

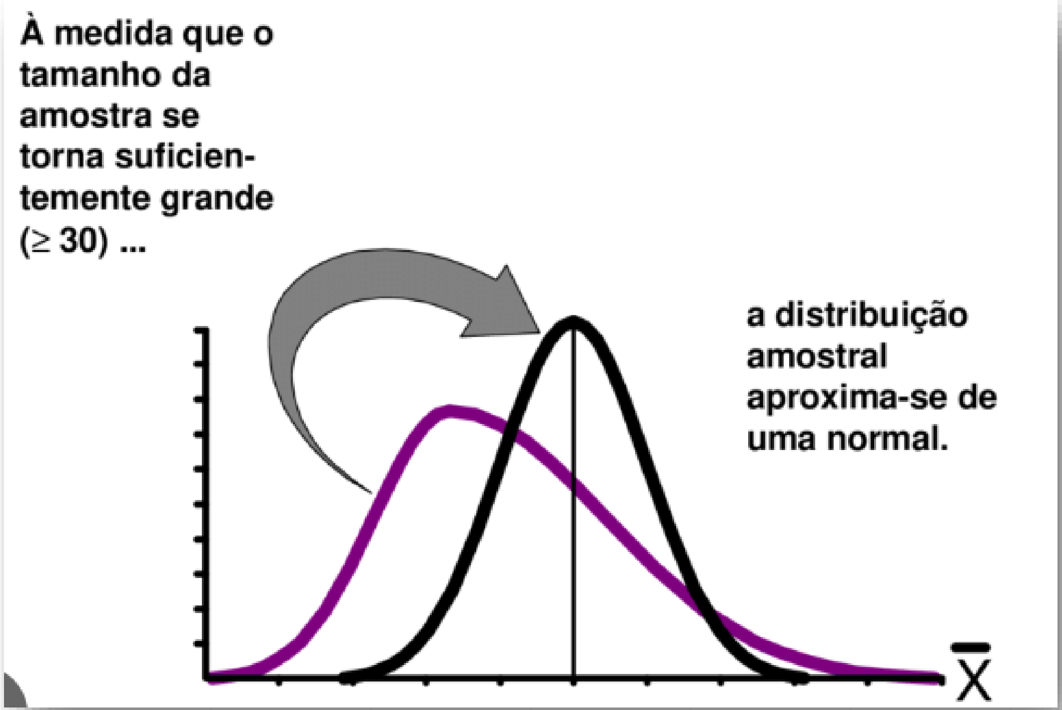
	Variância Conhecida	Variância Desconhecida
Média	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0,1)$	$\frac{\bar{X}-\mu}{S'/\sqrt{n}} \overset{a}{\sim} N(0,1)$
Diferença de médias	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \overset{a}{\sim} N(0,1)$	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1'^2}{m}+\frac{S_2'^2}{n}}} \overset{a}{\sim} N(0,1)$

Distribuição Normal: Teorema do Limite Central



- ✓ The **central limit theorem** (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger.
- ✓ Sample sizes equal to or greater than 30 are considered sufficient for the CLT to hold.
- ✓ A key aspect of CLT is that the average of the sample means and standard deviations will equal the population mean and standard deviation.
- ✓ A sufficiently large sample size can predict the characteristics of a population accurately.

Distribuição Normal: Teorema do Limite Central



https://www.researchgate.net/figure/Figura-46-Teorema-do-Limite-Central_fig7_330703739

Intervalo de Confiança para o Valor Médio μ: Exercícios

3

Exercício Suplementar que não consta do livro Murteira et al (2015)

7.1 Medições do comprimento de 25 peças produzidas por uma máquina conduziram a uma média $\overline{x} = 140$ mm. Admita que cada peça tem comprimento aleatório com distribuição normal de valor esperado μ e desvio padrão $\sigma = 10$ mm, e que o comprimento de cada peça é independente das restantes. Construa um intervalo de confiança a 95% para o valor esperado da população.

Exercício 7.1: IC para μ

Dados:

X = comprimento de uma peça $\sim \mathcal{N}(\mu, 10^2)$

θ (parâmetro desconhecido)

$n = 25 \quad (x_1, x_2, \ldots, x_{25}) \qquad \bar{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 140$

Pretende-se:

Intervalo de confiança a 95% para μ:

$I.C._{0.95}(\mu)$

Slides Professora Claúdia Nunes

Exercício 7.1: IC para μ

Passo 0: ("entender" a situação)

$X \sim N(\mu, 10^2)$

$I.C._{0.95}(\mu)$

[pretende-se um I.C. para o valor esperado de uma população normal de variância conhecida]

Passo 1: (escolha da variável fulcral)

Definição: é uma v.a. que depende do parâmetro desconhecido mas cuja distribuição é conhecida.

Exercício 7.1: IC para μ

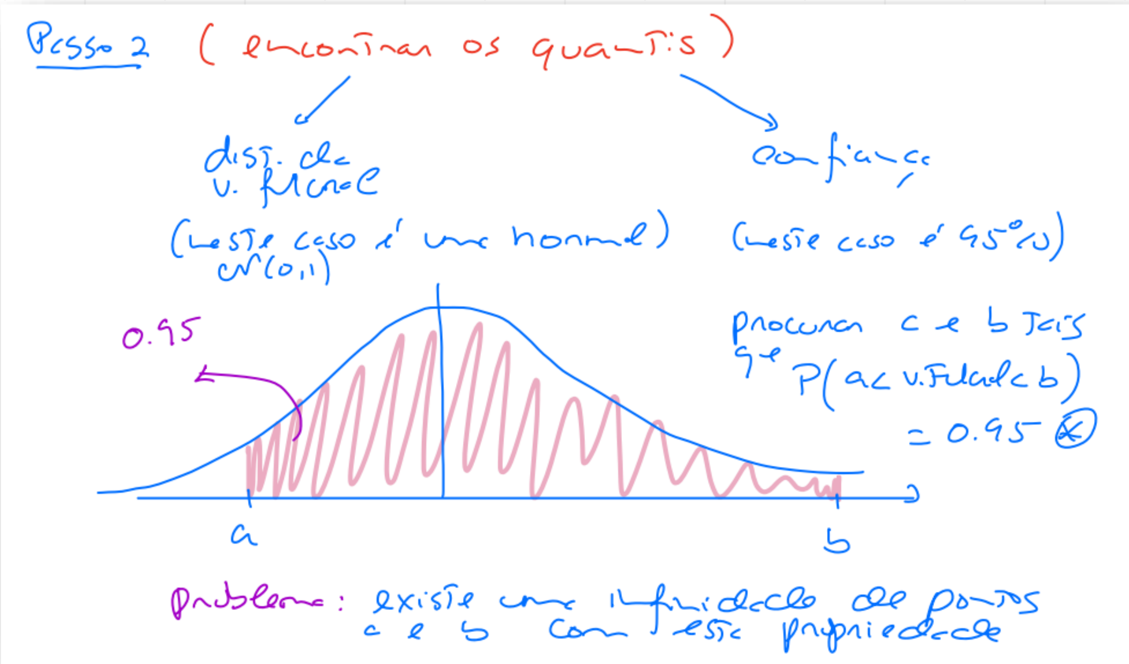
POPULAÇÕES NORMAIS

Média	$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\dfrac{\overline{X}-\mu}{S'/\sqrt{n}} \sim t(n-1)$

① variável fulcral que vamos utilizar para construir um recido, de uma população normal ($\mu = E[X]$; $\sigma^2 = Var[X]$)

v. fulcral : $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

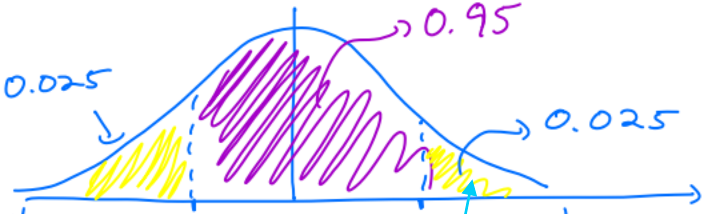
Exercício 7.1: IC para μ



Slides Professora Claúdia Nunes

Exercício 7.1: IC para μ

solução: encontrar o par (a,b) tal que b−a seja o menor possível (mas sempre de tal forma que (*) é válida)



$-1.96 = \Phi^{-1}(0.025) = a = -b \qquad b = \Phi^{-1}(0.975) = 1.96$

Slides Professora Claúdia Nunes

2 formas de escrever a mesma quantidade:

$z_{0,975} = z^*_{0,025} =,96$

Tabela da Normal

ε	.0005	.0010	.0050	.0100	.0200	.0250	.0500	.1000	.2000	.3000	.4000
z_ε	3.290	3.090	2.576	2.326	2.054	1.960	1.645	1.282	.842	.524	.253
$z_{\varepsilon/2}$	3.481	3.290	2.807	2.576	2.326	2.241	1.960	1.645	1.282	1.036	.842

$z_\varepsilon : P(Z > z_\varepsilon) = \varepsilon\,;\quad z_{\varepsilon/2} : P(|Z| > z_{\varepsilon/2}) = \varepsilon\,.$

Cálculo do Quantil da Distribuição Normal Padrão de Probabilidade α

Alternativamente

O nível de confiança é $1-\alpha = 0,95 \Leftrightarrow \alpha = 0,05$, então tem-se $1-\alpha/2 = 0,975$

Logo, pretende-se calcular o quantil da distribuição normal padrão de probabilidade 0,975
$z_{1-\alpha/2} = z_{0,975} = 1,96$ (ver tabela)

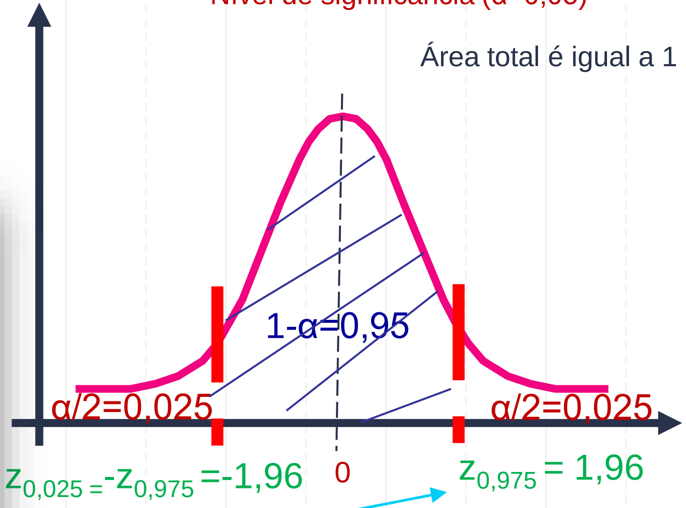
Tabela da Normal

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

Nível de confiança (1-α=0,95)

Nível de significância (α=0,05)

Área total é igual a 1



Exercício 7.1: IC para μ

Passo 3 (construção do intervalo)

≅ resolver umas desigualdades

v. fulcral — quantis

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

$$\Leftrightarrow$$

$$P\left(-1.96\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$\Leftrightarrow$$

$$P\left(-\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Exercício 7.1: IC para μ

(=)

$$P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Conclusão:

$$I.C.A._{0.95}(\mu) = \left]\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} ; \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right[$$

$\bar{x} = 140$; $\sigma = 10$; $n = 25$

$$I.C._{0.95}(\mu) = \left]140 - 1.96 \times \frac{10}{5} ; 140 + 1.96\frac{10}{5}\right[$$

Slides Professora Claúdia Nunes

$$IC_{95\%}(\mu) = (136.08;143.92)$$

ICs: Método da Variável Fulcral

Em geral, os intervalos de confiança são "derivados" através do Método da Variável Fulcral, que a seguir se descreve:

Passo 0: (A)
- identificar a distribuição da v.a. X de interesse
- identificar qual o parâmetro p/ o qual queremos derivar o i.c.
- quais os parâmetros conhecidos.

Exemplos

- $X \sim N(\mu, \sigma^2)$
  - (i) I.C. p/ μ, com σ conhecido
  - (ii) I.C. p/ μ, com σ desconhecido
  - (iii) I.C. p/ σ^2, com μ desconh.
- $X \sim$ Bernoulli (p) $\to$ I.C. para p. (iv)

Slides Professora Claúdia Nunes

ICs: Método da Variável Fulcral

Passo 1 : identificar a v. fulcral
(usando A)

I.C. p/ μ, σ conhecido, $X \sim N(\mu, \sigma^2)$	I.C. p/ μ, σ desconhecido, $X \sim N(\mu, \sigma^2)$	idem, n grande	variância amostral corrigida
(i)	(ii)		
$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim \boxed{t_{(n-1)}}$	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \overset{a}{\sim} N(0,1)$	$\underbrace{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}$
$\frac{(n-1)S^2}{\sigma^2} \sim \boxed{\chi^2_{(n-1)}}$		$\cancel{\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i} \overset{a}{\sim} \chi^2_{(k-m-1)}}$	

$= \left(\frac{1}{n-1}\sum x_i^2\right) \left(-\frac{n}{n-1}\bar{x}^2\right)$

I.C. p/ σ, com μ desconhecido $X \sim N(\mu, \sigma^2)$

(iii)

ICs: Método da Variável Fulcral

Passo 2 : encontrar os quantis
(distribuição da v. fulcral

$N(0,1)$ $\quad t_{(n-1)}$ $\quad \chi^2_{(n-1)}$

(t de student) (qui-quadrado)

? ?

ICs: Método da Variável Fulcral

Passo 3 : inverter as desigualdades
(v. fulcral e dos quantis)

$I.C.A_{1-\alpha}(\theta)$

(int. de confiança aleatório, p/ θ, com confiança $1-\alpha$)

$I.C._{1-\alpha}(\theta)$

(int. de confiança, que resulta da concreti_
zação do I.C.A. para uma amostra)

ICs: Método da Variável Fulcral - Resumo...

Recordar:

Passo 0: identificar a situação

- $X \sim N(\mu, \sigma^2)$
  - I.C. p/ μ
    - σ conhecido
    - σ desconhecido
  - I.C. p/ σ
- $X \sim Ber(p)$
  - I.C. p/ $p = P(X=1)$

Passo 1: identificar a v. fulcral

Passo 2: encontrar os quantis, tendo em conta a distribuição da v. fulcral e da confiança pretendida.

Passo 3: inverter as desigualdades $[a < V.F < b]$ de forma a colocar em evidência o parâmetro p/ o qual vamos construir o I.C.

Exercício Suplementar que não consta do livro Murteira et al (2015)

7.2 Admita que a densidade de construção, X, num projecto de urbanização tem distribuição normal. Uma amostra aleatória de 50 lotes desse projecto conduziu a

n

$$\sum_{i=1}^{50} x_i = 227.2\,;\ \sum_{i=1}^{50} x_i^2 = 2242.6$$

~~Assumindo que o desvio padrão de X é igual a~~ 4, construa um intervalo de confiança a 95% para a densidade média de construção. Que dimensão deveria ter a amostra para que a amplitude desse intervalo fosse reduzida a metade?

Vamos supor que a variância populacional é desconhecida.

Exercício 7.2: IC para μ

Supondo que a variância populacional é desconhecida, tem-se:

NOTA: Saber a variância da amostra não significa que se saiba a variância da população!!

$$s^2 = \frac{1}{50-1}\, 2242.6 - \frac{50}{50-1}\left(\frac{227.2}{50}\right)^2 = 24.698$$

$$\bar{x} = \frac{227.2}{50} = 4.544$$

X tem distribuição normal
Supondo-se que a variância populacional é desconhecida, tem-se:

Exercício 7.2: IC para μ

Tabela da t-Student

$$t_{n,\varepsilon} : P(X > t_{n,\varepsilon}) = \varepsilon$$

n \ ε	.400	.250	.100	.050	.025	.010	.005	.001
1	.325	1,000	3.078	6.314	12.706	31.821	63.656	318.289
2	.289	.816	1.886	2.920	4.303	6.965	9.925	22.328
3	.277	.765	1.638	2.353	3.182	4.541	5.841	10.214
4	.271	.741	1.533	2.132	2.776	3.747	4.604	7.173
5	.267	.727	1.476	2.015	2.571	3.365	4.032	5.894
6	.265	.718	1.440	1.943	2.447	3.143	3.707	5.208
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.785
8	.262	.706	1.397	1.860	2.306	2.896	3.355	4.501
9	.261	.703	1.383	1.833	2.262	2.821	3.250	4.297
10	.260	.700	1.372	1.812	2.228	2.764	3.169	4.144
11	.260	.697	1.363	1.796	2.201	2.718	3.106	4.025
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.930
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.852
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.787
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.733
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.686
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.646
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.610
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.579
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.552
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.527
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.505
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.485
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.467
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.450
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.435
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.421
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.408
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.396
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.385
40	.255	.681	1.303	1.684	2.021	2.423	2.704	3.307
50	.255	.679	1.299	1.676	2.009	2.403	2.678	3.261

$t_{0,975;49} = t^*_{0,025;49} = 2,009$

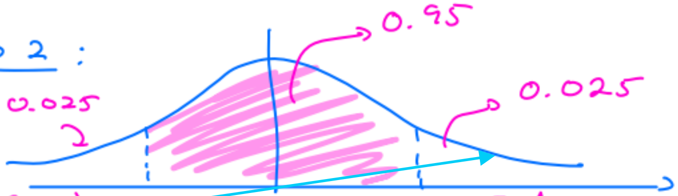
Passo 0: $X \sim N(\mu, \sigma^2)$ $\quad \mu, \sigma^2 = ?$

$I.C._{0.95}(\mu) = 2$

Passo 1:

$$\frac{\bar{X} - \mu}{S/\sqrt{50}} \sim t_{(49)}$$

Passo 2:



$F^{-1}_{t(49)}(0.025) = a = -b$ $\quad b = F^{-1}_{t(49)}(0.975) \simeq \boxed{2,009}$

Slides Professora Claúdia Nunes

Exercício 7.2: IC para μ

Passo 3

$$P\left(-2{,}009 < \frac{\bar{X}-\mu}{S/\sqrt{50}} < 2{,}009\right) = 0.95$$

$$\bar{X} - 2{,}009\frac{S}{\sqrt{50}} < \mu < \bar{X} + 2{,}009\frac{S}{\sqrt{50}}$$

$$I.C.A._{0.95}(\mu) = \left]\bar{X} - 2{,}009\frac{S}{\sqrt{50}} ; \bar{X} + 2{,}009\frac{S}{\sqrt{50}}\right[$$

Concretizando p/ a amostra em causa:

$$\bar{x} = 4.544 \qquad s^2 = 24.698$$

$$I.C._{0.95}(\mu) = \left] 4.544 - 2{,}009\sqrt{\frac{24.698}{50}} ; 4.544 + 2{,}009\sqrt{\frac{24.698}{50}}\right[$$

Supondo que a variância populacional é desconhecida, tem-se:

$IC_{95\%}(\mu)$: (3,132; 5,956)

Exercício 7.2: Amplitude Amostral

Supondo variância populacional desconhecida, tem-se:

$IC_{95\%}(\mu)$: (3,132; 5,956)

Amplitude do IC para μ = 5,956 – 3,132 = 2,824

Amplitude do IC para μ: $2 \times t_{1-\alpha/2;\, n-1}\,(s'/\,n^{1/2})$

n?

$2 \times t_{1-\alpha/2;\, n-1}\,(s'/\,n^{1/2}) = 2 \times 2{,}009 \times 4{,}970\,/\,n^{1/2} \leq 2{,}824/2 \Leftrightarrow n \geq 200$

Pelo menos n = 200

Obrigada!

Questões?