



Lisbon School
of Economics
& Management
Universidade de Lisboa

Estatística II

Licenciatura em Gestão do Desporto
2.º Ano/2.º Semestre
2023/2024

Aulas Teórico-Práticas N.ºs 13 e 14 (Semana 8)

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 Teórico-Práticas (Semanas 1 a 5)

- **Capítulo 1:** Revisões e Distribuições de Amostragem

Aulas Teórico-Práticas (Semanas 5 a 7)

- **Capítulo 2:** Estimação

Aulas Teórico-Práticas (Semanas 7 a 9)

- **Capítulo 3:** Testes de Hipóteses

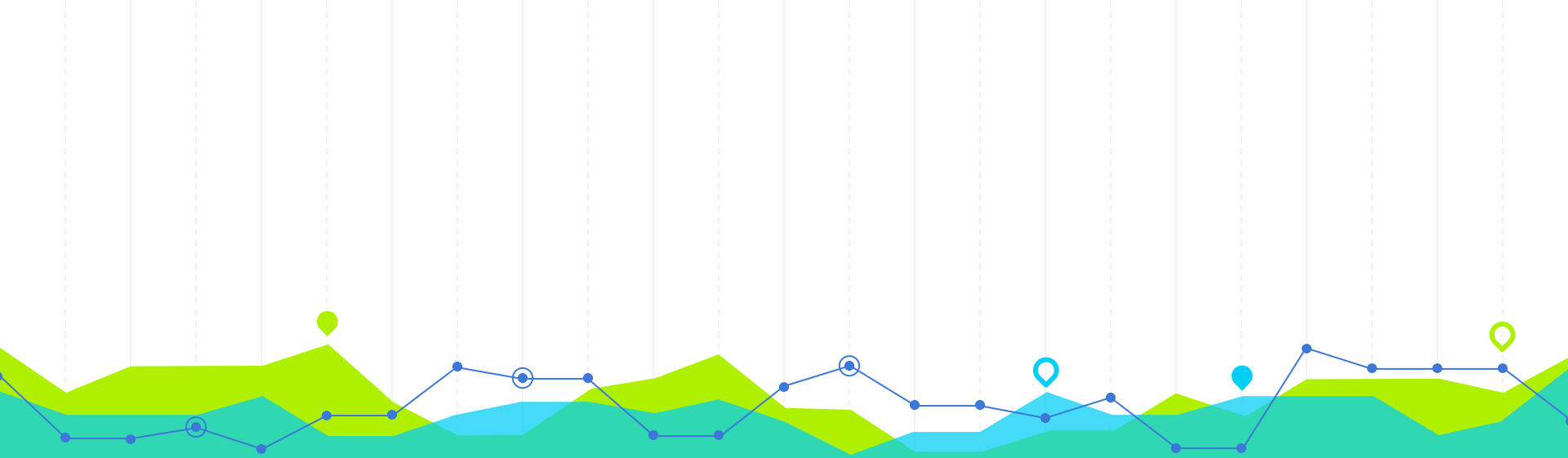
Aulas Teórico-Práticas (Semanas 10 a 13)

- **Capítulo 4:** Modelo de Regressão Linear Múltipla

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.

<https://cas.iseg.ulisboa.pt>



Estimação Intervalar

Intervalos de Confiança (ICs)

1

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**.

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 $\hat{\theta} - \epsilon, \hat{\theta} + \epsilon$.

Intervalo de Confiança (IC)

Definição

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

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

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

Diz-se que (Θ_1^*, Θ_2^*) é um intervalo aleatório, que contém θ com probabilidade $1 - \alpha$.

Intervalo de Confiança

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

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

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

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

Intervalo de Confiança

Definição

A qualquer intervalo (θ_1^*, θ_2^*) , com $\theta_1^* < \theta_2^*$, 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.

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, \dots, X_n) , que vamos usar para obter **Intervalos de Confiança**

[modulol_aula5.pdf](#)

ICs vs Distribuições por Amostragem: Exemplos

AMOSTRAGEM. DISTRIBUIÇÕES POR AMOSTRAGEM

Formulário

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

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

IC vs Distribuições de Amostragem: Exemplos

Relembrando

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

Para obter o I.C. para μ com σ desconhecido

Variável usada	Condições	Distribuição
$\frac{\bar{X} - \mu}{S/\sqrt{n}}$ $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$	$X_i \sim N(\mu, \sigma)$ $i = 1, 2, \dots, n$	$t_{(n-1)}$

Variância corrigida

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

Definição da distribuição t – Student

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

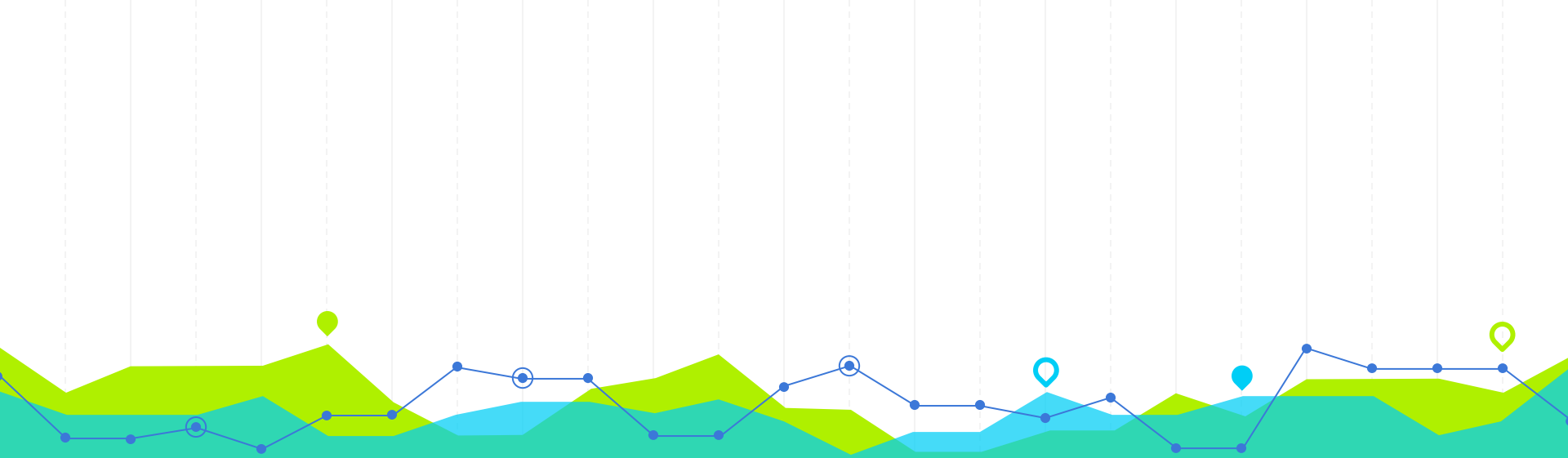
$$T = \frac{Z}{\sqrt{X/n}} \sim t_{(n)}$$

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

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

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

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 $X \sim N(\mu, \sigma)$

- 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

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

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

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

[modulol_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 : $2 \times t_{1-\alpha/2; n-1} (s / n^{1/2})$

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

$$\left(\bar{x} - t_{1-\alpha/2} \frac{s'}{\sqrt{n}}; \bar{x} + t_{1-\alpha/2} \frac{s'}{\sqrt{n}} \right)$$

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 confiança e são usados para testar a hipótese nula (H_0) num teste de hipóteses.

- o 1%, 5%, 10% = alfa => **Níveis de significância.**
- o 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).

Intervalo de Confiança para μ : Exemplo

Exemplo de construção de um I.C. no \mathbb{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[

Intervalo de Confiança para μ e $n \geq 30$: 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 **amostra de dimensão elevada**, i.e., n grande \rightarrow aplicação do Teorema Limite Central

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

Ou, que é o caso mais frequente,

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

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

$$\bar{X} - z_{1-\alpha/2} \frac{s'}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2} \frac{s'}{\sqrt{n}}$$

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

IC para μ : Formulário

Variância corrigida

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

• POPULAÇÕES NORMAIS

Média	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ Variância Conhecida	$\frac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t(n-1)$ Variância Desconhecida
Diferença de médias	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \sim t(v)$
	$T = \frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}}}}{\sqrt{\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	<p>onde v é o maior inteiro contido em r,</p> $r = \frac{\left(\frac{S_1'^2}{m} + \frac{S_2'^2}{n}\right)^2}{\frac{1}{m-1} \left(\frac{S_1'^2}{m}\right)^2 + \frac{1}{n-1} \left(\frac{S_2'^2}{n}\right)^2}$
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	$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X} - \mu}{S' / \sqrt{n}} \stackrel{a}{\sim} N(0,1)$
Diferença de médias	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \stackrel{a}{\sim} N(0,1)$

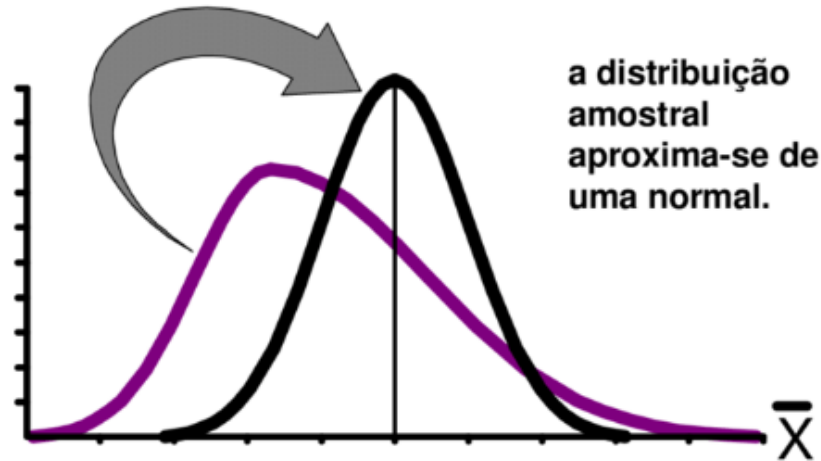
Distribuição Normal: Teorema do Limite Central

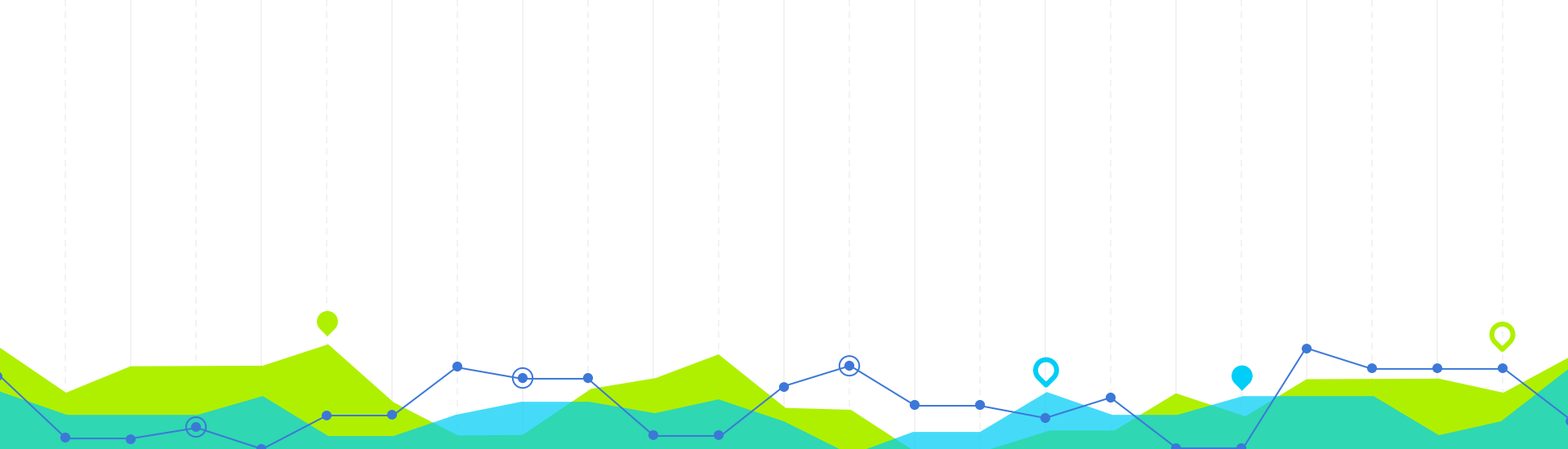


- ✓ The **central limit theorem** (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger.
- ✓ Sample sizes equal to or greater than 30 are considered sufficient for the CLT to hold.
- ✓ A key aspect of CLT is that the average of the sample 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

À medida que o tamanho da amostra se torna suficientemente grande (≥ 30) ...





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 $\bar{x} = 140$ 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:

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

θ (parâmetro desconhecido)

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

Preten-de-se:

Intervalo de confiança a 95% para μ :

I.C. 0.95 (μ)

Exercício 7.1: IC para μ

Passo 0: ("entender" a situação)

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

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

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

Passo 1: (escolha de variável funcional)

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

é conhecida.

Exercício 7.1: IC para μ

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

① variável fidediel que vamos utilizar para construir um intervalo de confiança de uma população normal ($\mu = E[X]$; $\sigma^2 = \text{var}[X]$)

v. fidediel : $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{cn}(0,1)$

Exercício 7.1: IC para μ

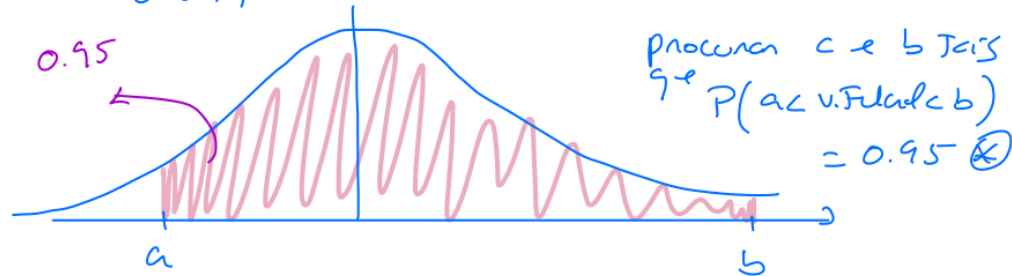
Passo 2 (encontrar os quantis)

diss. de
v. f. $N(\mu, \sigma^2)$

(Neste caso é uma normal
 $N(0,1)$)

confiança

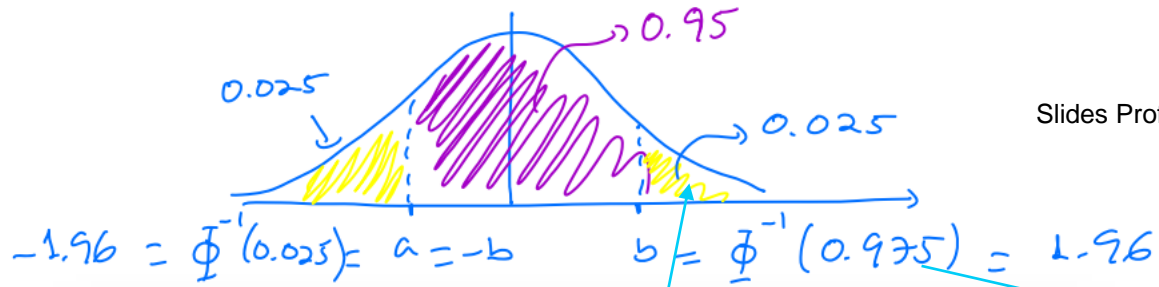
(Neste caso é 95%)



Problema: existe uma infinidade de pontos
 a e b com esta propriedade

Exercício 7.1: IC para μ

Solução: encontrar o par (a, b) tal que $b - a$ seja a menor possível (mas sempre de tal forma que \otimes é válida)



Slides Professora Cláudia Nunes

2 formas de escrever a mesma quantidade:

$$Z_{0,975} = Z^*_{0,025} = 1,96$$

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

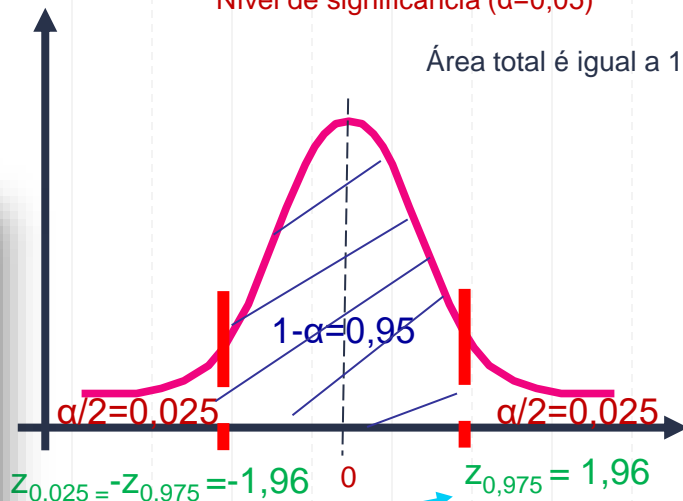
$Z_{1-\alpha/2} = Z_{0,975} = 1,96$ (ver tabela)

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 ($1-\alpha=0,95$)

Nível de significância ($\alpha=0,05$)



Exercício 7.1: IC para μ

Passo 3 (constituição do intervalo)
 \cong encontrar os desvios críticos

v. final \swarrow \searrow 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:

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

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

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

Slides Professora Cláudia Nunes

$$\text{IC}_{95\%}(\mu) = (136.08; 143.92)$$

ICs: Método da Variável Fulcral

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

Passo 0: identificar a distribuição de v.c. X de interesse

- (A) {
- identificar qual o parâmetro θ o qual queremos derivar o i.c.
 - quais os parâmetros conhecidos.

Exemplos

- $X \sim \text{EV}(\mu, \sigma^2)$
 - (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.
- $X \sim \text{Bernoulli}(p) \rightarrow$ I.C. para p . (iv)

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)

$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$	$\frac{\bar{X} - \mu}{S/\sqrt{n}} \stackrel{a}{\sim} N(0,1)$	$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$	$\sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(k-m-1)}$	

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

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 de v. fulcral)

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

(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 $1-\alpha$)

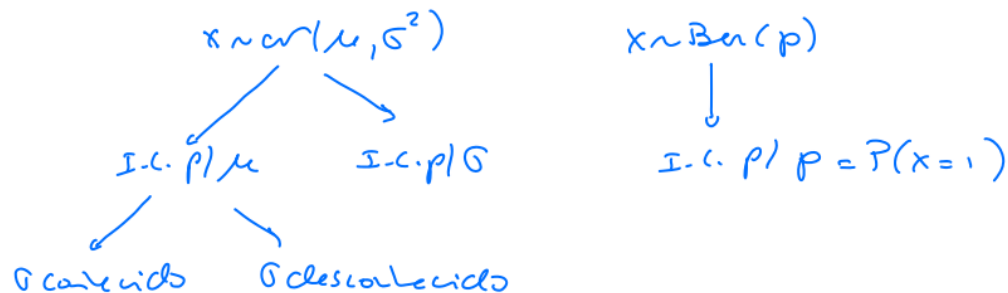
$$I.C._{1-\alpha}(\theta)$$

(Int. de confiança, que resulta de com-
partição do I.C.A. para uma amostra)

ICs: Método da Variável Fulcral - Resumo...

Reconhecer:

Passo 0: identificar a situação



Passo 1: identificar a v. fulcral

Passo 2: encontrar os quantis tendo em conta a distribuição de v. fulcral e de 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.

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

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

~~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 de amostra não significa que se sabe a variância de população!!

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

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

Tabela da t-Student

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

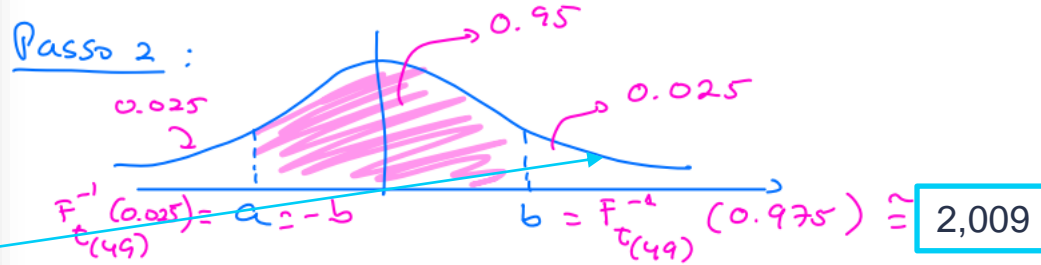
Exercício 7.2: IC para μ

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

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	2,405	2,678	3,261

Passo 0: $X \sim \mathcal{N}(\mu, \sigma^2)$ $\mu, \sigma^2 = ?$
I.C. $_{0.95}(\mu) = 2$

Passo 1: $\frac{\bar{X} - \mu}{S/\sqrt{50}} \sim t_{(49)}$



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

Exercício 7.2: IC para μ

$$\text{Pessoa 3} \quad P\left(-2,009 < \frac{\bar{X} - \mu}{s/\sqrt{50}} < 2,009\right) = 0,95$$

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

$$I.C.A._{0,95}(\mu) = \left] \bar{X} - 2,009 \frac{s}{\sqrt{50}} ; \bar{X} + 2,009 \frac{s}{\sqrt{50}} \right[$$

Concretizando p/c amostra em curso:

$$\bar{x} = 4,544 \quad s^2 = 24,698$$

$$I.C._{0,95}(\mu) = \left] 4,544 - 2,009 \sqrt{\frac{24,698}{50}} ; \right.$$

$$\left. 4,544 + 2,009 \sqrt{\frac{24,698}{50}} \right[,$$

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:

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

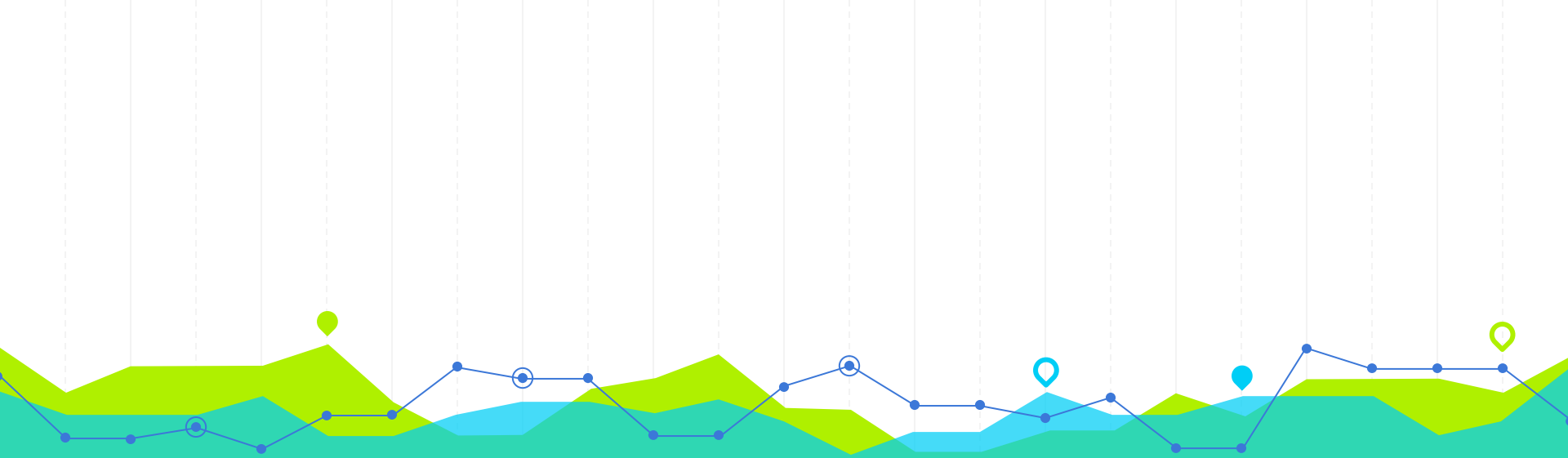
Amplitude do IC para $\mu = 5,956 - 3,132 = 2,824$

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

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 $n = 200$



Intervalo de Confiança para a Variância σ^2

4

Intervalo de Confiança para σ^2

Intervalo a $(1 - \alpha) \times 100\%$ de confiança para σ^2 numa população normal

$$\left[\frac{(n-1) s^2}{\chi^2_{n-1; 1-\frac{\alpha}{2}}}, \frac{(n-1) s^2}{\chi^2_{n-1; \frac{\alpha}{2}}} \right]$$

Variância corrigida

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

$$\frac{(n-1) s^2}{\chi^2_{1-\alpha/2, (n-1)}} < \sigma^2 < \frac{(n-1) s^2}{\chi^2_{\alpha/2, (n-1)}}$$

Intervalo de Confiança para σ^2

Para obter o Intervalo de Confiança para σ^2

Variável usada	Condições	Distribuição
$\frac{(n-1) S'^2}{\sigma^2}$ $S'^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$	$X_i \sim N(\mu, \sigma)$ $i = 1, 2, \dots, n$	$\chi^2_{(n-1)}$

Definição da distribuição χ^2

Se Z_1, Z_2, \dots, Z_n são v.'s a.'s $N(0, 1)$ independentes

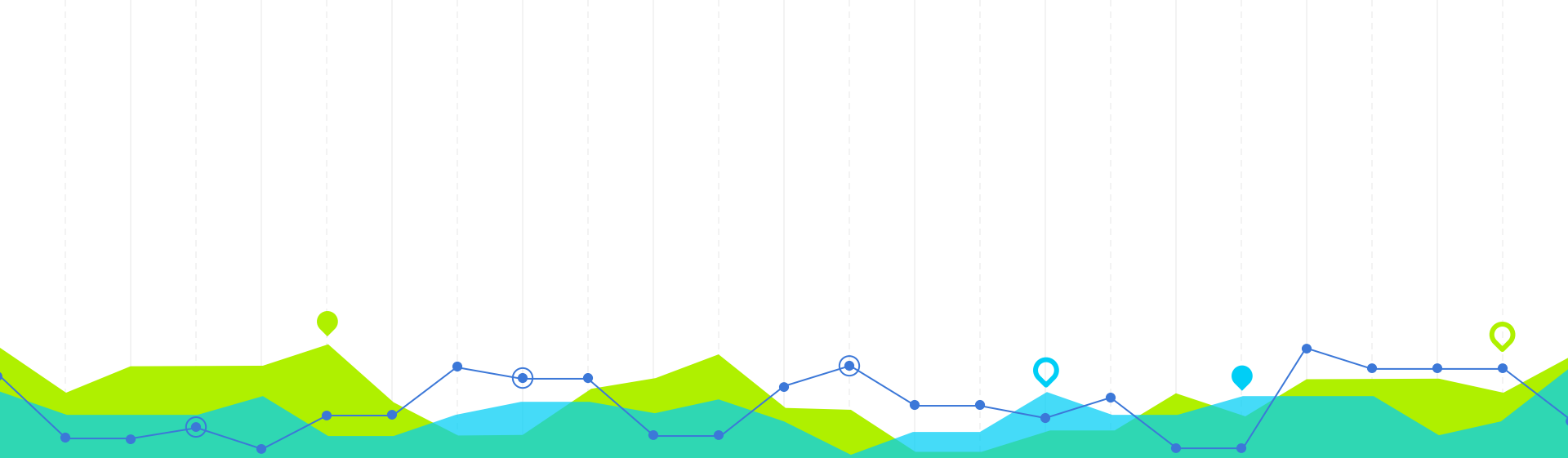
\Downarrow
a v.a. $X = Z_1^2 + \dots + Z_n^2$ é tal que $X \sim \chi^2_{(n)}$

Tem-se $E[X] = n$; $Var[X] = 2n$

IC para σ^2 : Formulário

• POPULAÇÕES NORMAIS

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	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \sim t(v)$
	$T = \frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}}}}{\sqrt{\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	<p>onde v é o maior inteiro contido em r,</p> $r = \frac{\left(\frac{s_1'^2}{m} + \frac{s_2'^2}{n}\right)^2}{\frac{1}{m-1}\left(\frac{s_1'^2}{m}\right)^2 + \frac{1}{n-1}\left(\frac{s_2'^2}{n}\right)^2}$
Variância	$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	Variância corrigida
Relação de variâncias	$\frac{S_1'^2}{S_2'^2} \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$	$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$



Intervalo de Confiança para a Variância

σ^2 : Exercícios

5

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

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

~~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?~~

Exercício 7.2: Admitindo que se pretende construir um intervalo de confiança a 90% p/ σ^2 .



Exercício 7.2: IC para σ^2

Passo 0 : $X \sim \mathcal{N}(\mu, \sigma^2)$ $\mu, \sigma^2 = ?$
I.C. $0,90$ (σ^2)

Passo 1 : v. flud
 $\frac{49 s^2}{\sigma^2} \sim \chi^2(49)$

Cálculo dos Quantis da Distribuição Qui-Quadrado de Probabilidade $1-\alpha/2$ e $\alpha/2$ com $n-1$ g.l.'s

Nível de confiança ($1-\alpha=0,90$)

Nível de significância ($\alpha=0,10$)

Área total é igual a 1

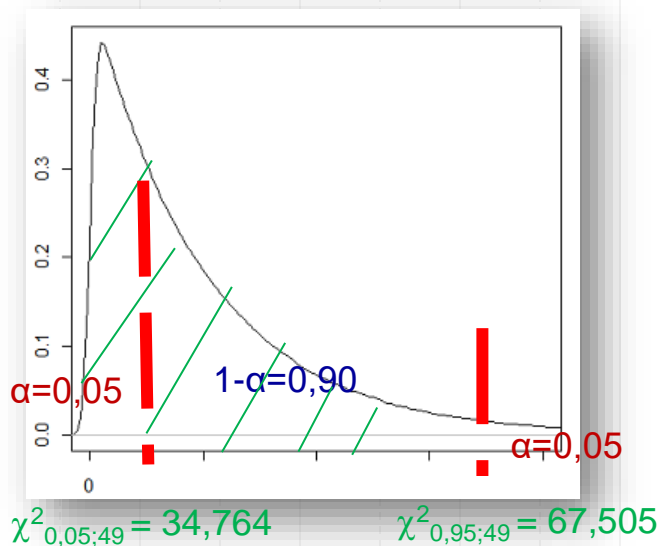
O nível de confiança é $1-\alpha = 0,90 \Leftrightarrow \alpha = 0,10$,
então tem-se $1-\alpha/2 = 0,95$ e $\alpha/2 = 0,05$

Logo, pretende-se calcular o quantil da distribuição
Qui-Quadrado de probabilidade 0,05

$$\chi^2_{0,05;49} = \chi^2_{0,95;49} = 34,764 \text{ (ver tabela a seguir)}$$

Logo, pretende-se calcular o quantil da distribuição
Qui-Quadrado de probabilidade 0,95

$$\chi^2_{0,95;49} = \chi^2_{0,05;49} = 67,505 \text{ (ver tabela a seguir)}$$



Cálculo dos Quantis da Distribuição Qui-Quadrado de Probabilidade $1-\alpha/2$ e $\alpha/2$ com $n-1$ g.l.'s

Tabela do Qui-Quadrado

ϵ	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005	.001
1	.000	.000	.001	.004	.016	.102	.455	1.323	2.706	3.841	5.024	6.635	7.879	10.827
2	.010	.020	.051	.103	.211	.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597	13.815
3	.072	.115	.216	.352	.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838	16.266
4	.207	.297	.484	.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860	18.466
5	.412	.554	.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.832	15.086	16.750	20.515
6	.676	.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548	22.457
7	.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278	24.321
8	1.344	1.647	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955	26.124
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589	27.877
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188	29.588
11	2.603	3.053	3.816	4.575	5.578	7.584	10.341	13.701	17.275	19.675	21.920	24.725	26.757	31.264
12	3.074	3.571	4.404	5.226	6.304	8.438	11.340	14.845	18.549	21.026	23.337	26.217	28.300	32.909
13	3.565	4.107	5.009	5.892	7.041	9.299	12.340	15.984	19.812	22.362	24.736	27.688	29.819	34.527
14	4.075	4.660	5.629	6.571	7.790	10.165	13.339	17.117	21.064	23.685	26.119	29.141	31.319	36.124
15	4.601	5.229	6.262	7.261	8.547	11.037	14.339	18.245	22.307	24.996	27.488	30.578	32.801	37.698
16	5.142	5.812	6.908	7.962	9.312	11.912	15.338	19.369	23.542	26.296	28.845	32.000	34.267	39.252
17	5.697	6.408	7.564	8.672	10.085	12.792	16.338	20.489	24.767	27.587	30.191	33.409	35.718	40.791
18	6.265	7.015	8.231	9.390	10.865	13.675	17.338	21.605	25.989	28.869	31.526	34.805	37.156	42.312
19	6.844	7.633	8.907	10.117	11.651	14.562	18.338	22.718	27.204	30.144	32.852	36.191	38.582	43.819
20	7.434	8.260	9.591	10.851	12.443	15.452	19.337	23.828	28.412	31.410	34.170	37.566	39.997	45.314
21	8.034	8.897	10.283	11.591	13.240	16.344	20.337	24.935	29.615	32.671	35.479	38.932	41.401	46.796
22	8.643	9.542	10.982	12.338	14.041	17.240	21.337	26.039	30.813	33.924	36.781	40.289	42.796	48.268
23	9.260	10.196	11.689	13.091	14.848	18.137	22.337	27.141	32.007	35.172	38.076	41.638	44.181	49.728
24	9.886	10.856	12.401	13.848	15.659	19.037	23.337	28.241	33.196	36.415	39.364	42.980	45.558	51.179
25	10.520	11.524	13.120	14.611	16.473	19.939	24.337	29.339	34.382	37.652	40.646	44.314	46.928	52.619
26	11.160	12.198	13.844	15.379	17.292	20.843	25.336	30.435	35.563	38.885	41.923	45.642	48.290	54.051
27	11.808	12.878	14.573	16.151	18.114	21.749	26.336	31.528	36.741	40.113	43.195	46.963	49.645	55.475
28	12.461	13.565	15.308	16.928	18.939	22.657	27.336	32.620	37.916	41.337	44.461	48.278	50.994	56.892
29	13.121	14.256	16.047	17.708	19.768	23.567	28.336	33.711	39.087	42.557	45.722	49.588	52.335	58.301
30	13.787	14.953	16.791	18.493	20.599	24.478	29.336	34.800	40.256	43.773	46.979	50.892	53.672	59.702
40	20.707	22.164	24.433	26.500	29.051	33.660	39.335	45.616	51.805	55.758	59.342	63.691	66.766	73.403
50	27.991	29.707	32.357	34.764	37.689	42.942	49.335	56.334	63.167	67.505	71.420	76.154	79.490	86.660
60	35.534	37.485	40.482	43.188	46.459	52.294	59.335	66.981	74.397	79.082	83.298	88.379	91.952	99.608
70	43.275	45.442	48.758	51.739	55.329	61.698	69.334	77.577	85.527	90.531	95.023	100.425	104.215	112.317
80	51.172	53.540	57.153	60.391	64.278	71.145	79.334	88.130	96.578	101.879	106.629	112.329	116.321	124.839
90	59.196	61.754	65.647	69.126	73.291	80.625	89.334	98.650	107.565	113.145	118.136	124.116	128.299	137.208
100	67.328	70.065	74.222	77.929	82.358	90.133	99.334	109.141	118.498	124.342	129.561	135.807	140.170	149.449

$$\chi^2_{n,\epsilon} : P(X > \chi^2_{n,\epsilon}) = \epsilon$$

$$\chi^2_{0,05;49} = \chi^2_{0,95;49} = 34,764$$

$$\chi^2_{0,95;49} = \chi^2_{0,05;49} = 67,505$$

NOTA: Saben a variância de amostra não significa se se sabe a variância de população!

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

$$\left[\frac{(n-1) s^2}{\chi^2_{n-1; 1-\frac{\alpha}{2}}}; \frac{(n-1) s^2}{\chi^2_{n-1; \frac{\alpha}{2}}} \right]$$

IC_{90%}(σ²): (7,928; 34,811)

Exercício 7.2: IC para σ^2

NOTA:

v. flud.: $\frac{49 \cdot 52}{\sigma^2} \sim \chi^2_{(49)}$ ✓

①

v. flud. $\frac{49 \times 24.698}{\sigma^2} \sim \chi^2_{(49)}$ ✗

② Interpretação: I.C. $A_{1-\alpha}(\theta) =]L(x_1, \dots, x_n); U(x_1, \dots, x_n)$

significa que se recolhamos muitas amostras e calculamos, p/ cada uma delas, o correspondente intervalo então em cerca de $(1-\alpha) 100\%$ os intervalos assim obtidos contêm o verdadeiro valor do parâmetro desconhecido θ .

③

A amplitude do intervalo é tanto menor, em geral, qto maior for a dimensão da amostra.

Obrigada!

Questões?

