Aula 1
Estatística II

Introdução a estimação paramétrica

Plano para hoje

- Breve apresentação da UC
- Breve revisão de alguns conceitos base: População, amostra, processo de amostragem, distribuição da amostra, estatística e distribuição por amostragem de uma estatística.
- Introdução à estimação paramétrica
- Método dos Momentos - Ideia intuitiva, formalização e exemplos

Plano para hoje

- **Breve apresentação da UC**
- Breve revisão de alguns conceitos base: População, amostra, processo de amostragem, distribuição da amostra, estatística e distribuição por amostragem de uma estatística.
- Introdução à estimação paramétrica
- Método dos Momentos - Ideia intuitiva, formalização e exemplos

Estatística II

- Licenciaturas em Economia e Finanças
- 2º Semestre - 2022/2023
- Equipa dos docentes:
  - **Erida Gjini**, erida.gjini@iseg.ulisboa.pt **(aulas teoricas)**
  - Graça Leão Fernandes gleao@iseg.ulisboa.pt (aulas praticas)
  - Beatrix Malheiros Leal beatrizmleal@iseg.ulisboa.pt (aulas praticas)

- PROGRAMA
  - **Parte 1: Estatística (~ 13/02-20/03)**
    - *Estimação*
    - *Testes de hipóteses*
    - *Testes não-paramétricos*
  - **Parte 2: Econometria (22/03-17/05)**
    - *Introdução*
    - *Modelo de regressão linear multipla*
    - *Inferência estatistica*
    - *Heterocedasticidade*
    - *Previsao*
    - *Tópicos sobre formas funcionais*

Ver página da disciplina no internet

Bibliografia

- Murteira, B., Silva Ribeiro, C., Andrade e Silva, J., Pimenta, C. e Pimenta, F., (2015). Introdução à Estatística, Escolar Editora**. [MSRASP]**
- Wooldridge, J. M. (2016), Introductory Econometrics: a Modern Approach, 6th. ed., Cengage Learning. **[W]**

- PROGRAMA
  - **Parte 1: Estatística (Capítulos 7,8,9)**
    - *Estimação*
    - *Testes de hipóteses*
    - *Testes não-paramétricos*
  - **Parte 2: Econometria (Capítulos 1-8)**
    - *Introdução*
    - *Modelo de regressão linear multipla*
    - *Inferência estatística*
    - *Heterocedasticidade*
    - *Previsão*
    - *Tópicos sobre formas funcionais*

Os slides vão ser disponíveis na página da disciplina

Calendário e avaliação

Período de aulas: 13 Fevereiro – 17 Maio

25 aulas

Feriado Carnaval (19-20 Fevereiro)

Férias de Páscoa (2-7 Abril)

Avaliação

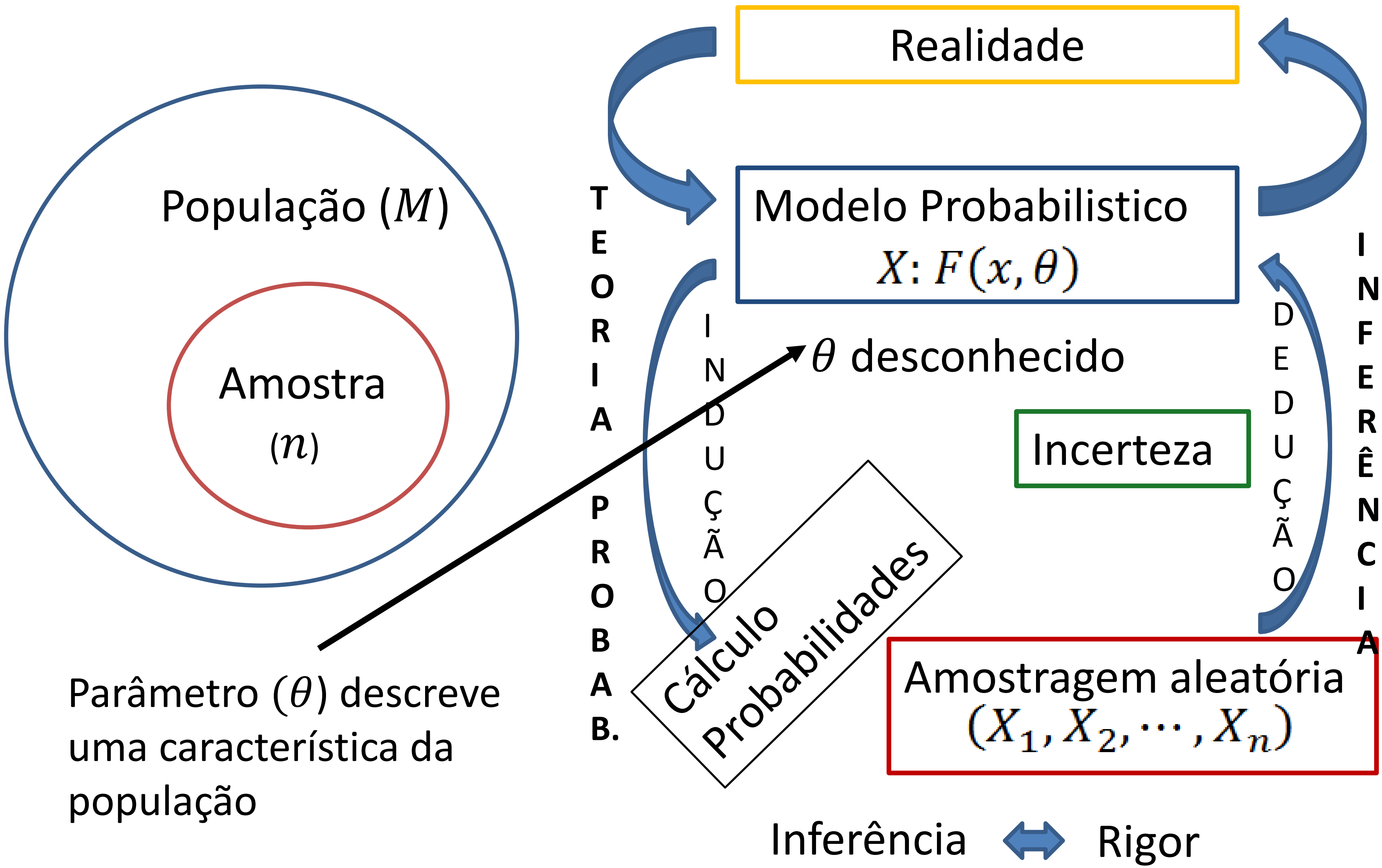
- Exame final (100%)
  - *Exemplos dos anos anteriores vão ser disponíveis*
  - *Época normal: <u>29.05.2023 @ 09:00 horas</u>*
  - *Época recurso: <u>29.06.2023 @12:00 horas</u>*
  - *Outra época em Setembro: <u>08.09.2023 @12:00 horas</u>*

Horário: segundas e quartas das 13h30 às 15h00 no Anfiteatro 4 do edifício das Francesinhas 2.

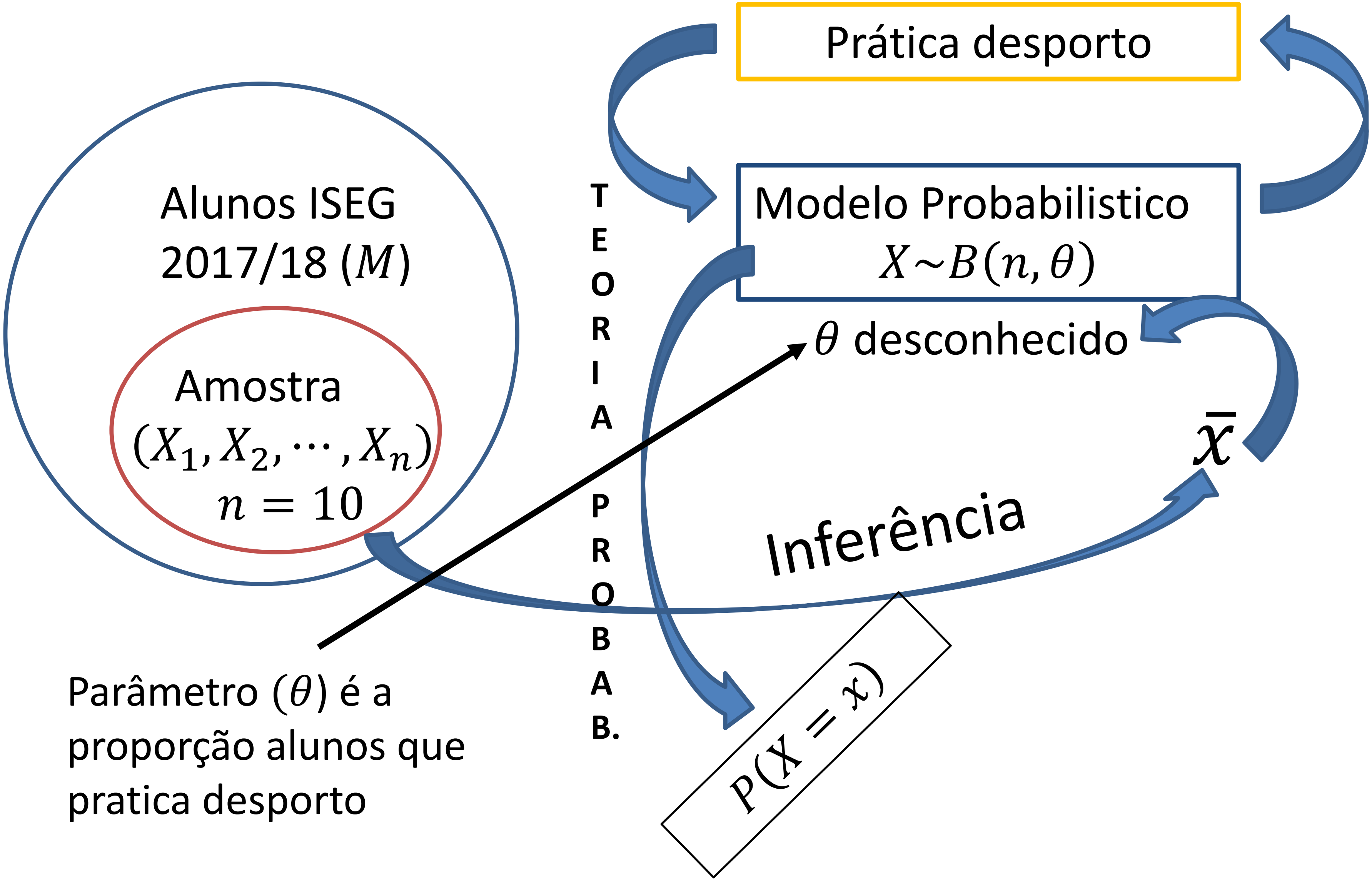
Plano para hoje

- Breve apresentação da UC
- Breve revisão de alguns conceitos base: População, amostra, processo de amostragem, distribuição da amostra, estatística e distribuição por amostragem de uma estatística.
- Introdução à estimação paramétrica
- Método dos Momentos - Ideia intuitiva, formalização e exemplos

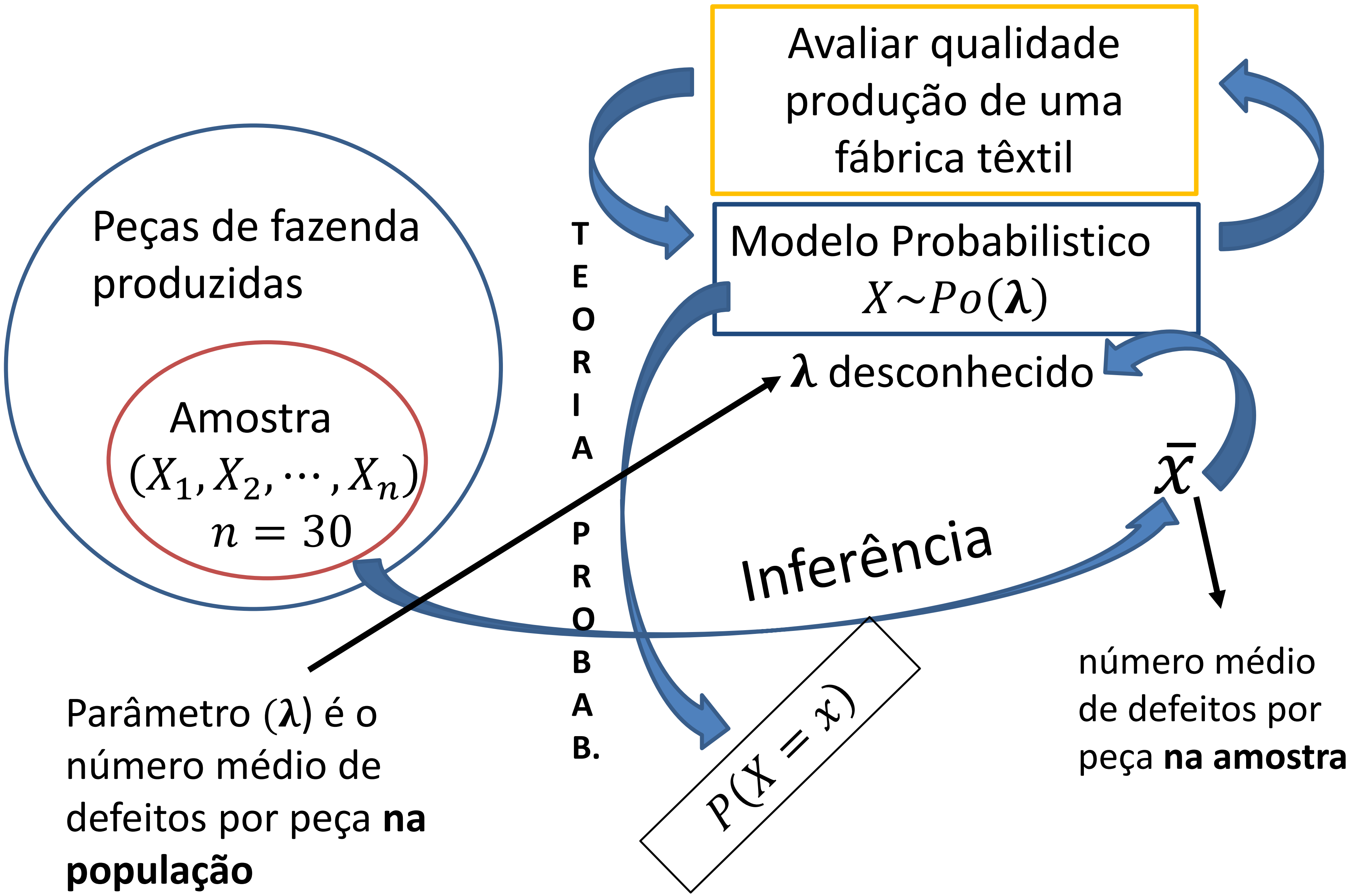
Amostragem: Introdução



Amostragem: Introdução



Amostragem: Introdução



Capítulo 6 - Amostragem

6.2 - Especificação.

- Especificação de um modelo (universo/população)

  Escolha de uma família de modelos probabilísticos para descrever a distribuição da população.

- Distribuição da população

  Descreve o modo como se distribuem os "números" que constituem a população

Amostragem

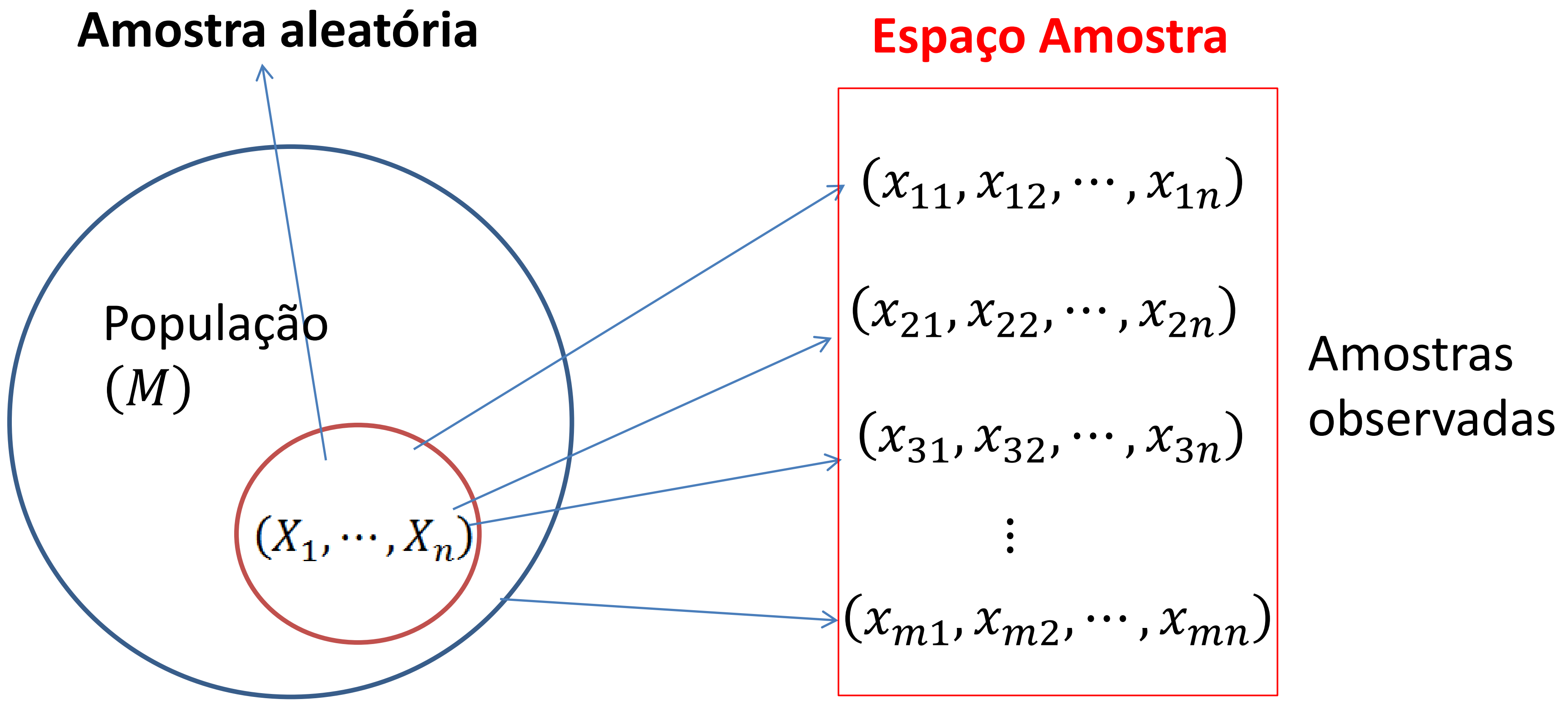
- População
  - Conjunto de “números” dos quais se extrai uma amostra

- Amostra casual
  - Cada elemento da população tem igual probabilidade de ser selecionado

Definição 6.1 – Amostra casual

$(X_1, X_2, \cdots, X_n)$, são **independentes e identicamente distribuídas(**uma **“cópia”** da v.a. X**) – simbolicamente** $\boldsymbol{iid}$

Amostragem



Def: **Espaço Amostra** é o conjunto de todas as amostras de dimensão $\boldsymbol{n}$ que é possível extrair de uma população de dimensão M

Amostragem

- Análise Estatística

  Determinar que generalizações, baseadas na amostra, podem ser feitas acerca da população

- Abordagem Frequencista

  O aspecto central da análise estatística consiste em reconhecer a **variabilidade** dos diferentes **conjuntos amostrais**

  "Statistics is the science of variation." Douglas M. Bates (1949 -)

Exemplo – Assuma-se que X (prática ou não de desporto)$\sim B(1,\theta)$

$Modelo$ será:

$$F_\theta = \{f(x|\theta) = \theta^x (1-\theta)^{1-x} : x \in \{0,1\} \Lambda\ \theta \in \boldsymbol{\Theta} = (0,1)\}$$

Amostra casual $(X_1, X_2, \cdots, X_n)$, sendo:

$X_i = 1$ $(iésimo\ indíviduo\ da\ amostra\ pratica\ desporto)$

$X_i = 0$ $(caso\ contrário)$

$X_i,\ i = 1, 2, \cdots, n$ são $\boldsymbol{iid}$ a X

Suponha-se que a amostra tem dimensão $n = 3$.

Espaço amostra terá 8 elementos:

$(\mathbf{0}, \mathbf{0}, \mathbf{0})$	com probabilidade	$(1-\theta)^3$
$(\mathbf{1}, \mathbf{0}, \mathbf{0})$, $(\mathbf{0}, \mathbf{1}, \mathbf{0})$, $(\mathbf{0}, \mathbf{0}, \mathbf{1})$	“	$\theta(1-\theta)^2$
$(\mathbf{1}, \mathbf{1}, \mathbf{0})$, $(\mathbf{1}, \mathbf{0}, \mathbf{1})$, $(\mathbf{0}, \mathbf{1}, \mathbf{1})$	“	$\theta^2(1-\theta)$
$(\mathbf{1}, \mathbf{1}, \mathbf{1})$	“	θ^3

Amostragem

Uma empresa tem 6 trabalhadores. O número de anos de experiência dos mesmos é $\{2, 4, 6, 6, 7, 8\}$. Seleccionaram-se amostras de 2 trabalhadores sem reposição. O **espaço amostra** contem 15 amostras diferentes representadas na tabela abaixo:

	x_1	x_2
Am1	2	4
Am2	2	6
Am3	2	6
Am4	2	7
Am5	2	8
Am6	4	6
Am7	4	6
Am8	4	7
Am9	4	8
Am10	6	6
Am11	6	7
Am12	6	8
Am13	6	7
Am14	6	8
Am15	7	8

Cada uma das 15 amostras tem a mesma probabilidade de ser selecionada.

Amostragem

6.3 – Estatística

A estatística descreve uma característica da amostra

Permite condensar a informação amostral num único número.

Qualquer função de $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$, por exemplo, $T = h\,(\boldsymbol{X})$ é uma **estatística.**

Definição 6.2 – Estatística
Uma estatística é uma variável aleatória $T(X_1, X_2, \cdots, X_n)$ *função da* amostra aleatória $(X_1, X_2, \cdots, X_n)$, *que não é função de qualquer parâmetro desconhecido.*

Amostragem

Exemplo 6.7 – Se $(X_1, X_2, \cdots, X_n)$ ***é amostra casual de uma população de Bernoulli,*** *as* estatísticas:

$T_1(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} X_i$, *representa o número de* "sucessos" na amostra,

$T_2(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} X_i / n$, *indica a proporção de "sucessos"* na amostra.

Exemplo 6.9 – Se $(X_1, X_2, \cdots, X_n)$ ***é amostra casual de uma população normal*** $\boldsymbol{X \sim N\left(\mu, \sigma^2\right)}$ com parâmetros desconhecidos,

São estatísticas unidimensionais:

$$\sum_{i=1}^{n} X_i \qquad \bar{X} = \sum_{i=1}^{n} X_i / n \qquad \sum_{i=1}^{n} X_i^2 \qquad \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

Não são estatísticas:

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

Amostragem

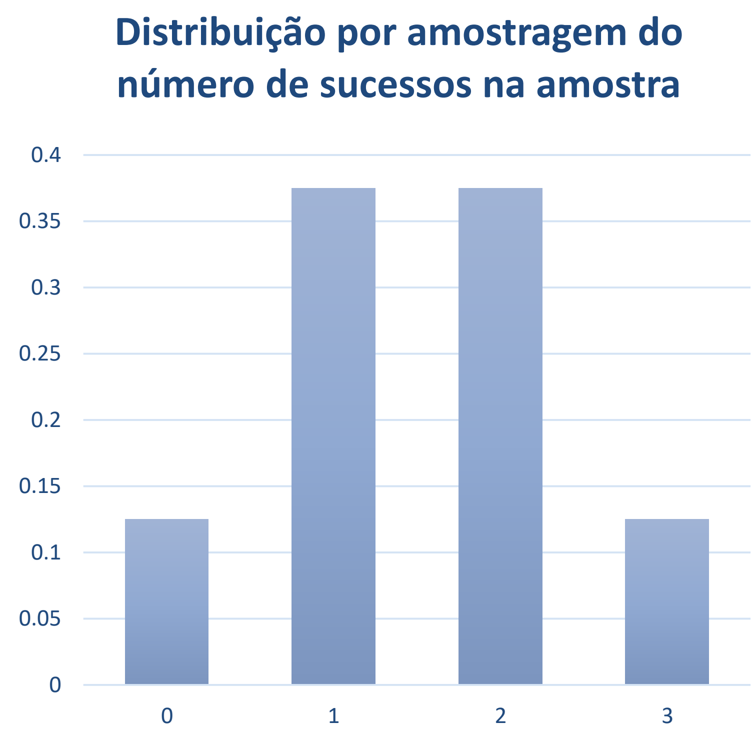
- Distribuições amostrais
  - Distribuição da estatística T.
  - Pode ser difícil de obter
  - Vamos usar a estatística T e a sua distribuição para inferir sobre parâmetros da população usando a amostra
  - Utilidade de qualquer estatística **depende** do **comportamento probabilístico da estatística** e **não** do **seu valor** $t(x_1, x_2, \cdots, x_n)$ para uma amostra particular.

Amostragem

Exemplo – Assuma-se que X (prática ou não de desporto)$\sim B(1,\theta)$

Seja uma amostra de dimensão $n = 3$ e a estatística $T_1 = \sum_{i=1}^{n} X_i$

Espaço amostra	t_1	$P\left(T_1 = \sum_{i=1}^{n} x_i\right)$
$(0,0,0)$	0	$1/8$
$(1,0,0)$, $(0,1,0)$, $(0,0,1)$	1	$3/8$
$(1,1,0)$, $(1,0,1)$, $(0,1,1)$	2	$3/8$
$(1,1,1)$	3	$1/8$



Amostragem

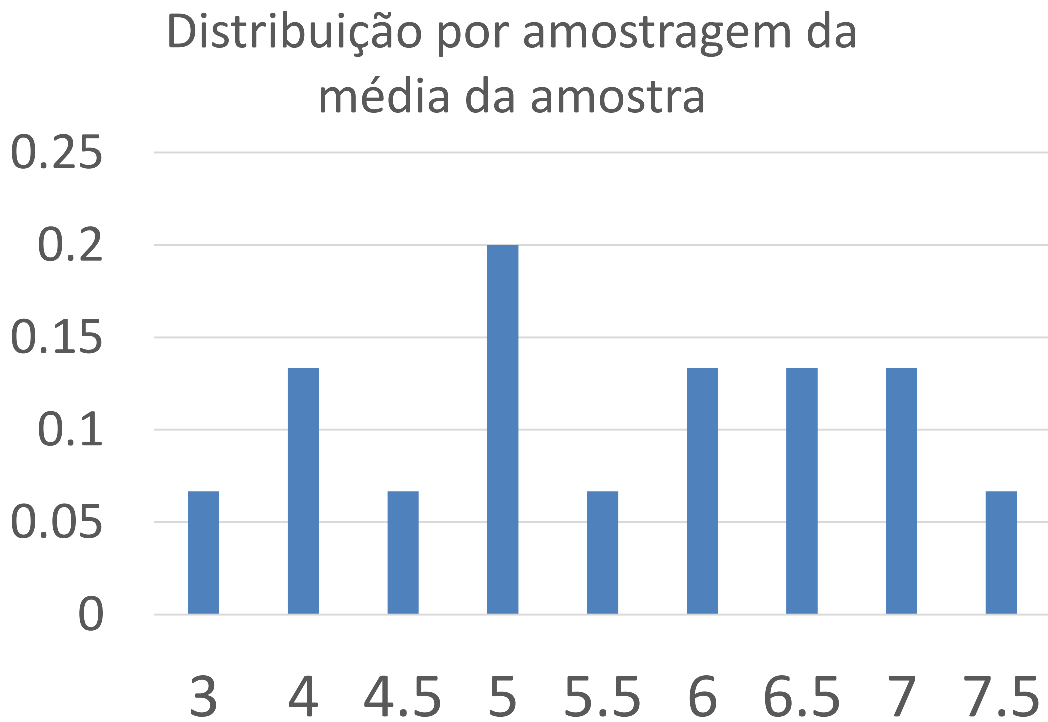
		x_1	x_2	$\bar{x}_j = \sum_{i=1}^{2} x_i/2$	
ESPAÇO AMOSTRA	Am1	2	4	3	ESPAÇO de ESTATÍSTICAS
	Am2	2	6	4	
	Am3	2	6	4	
	Am4	2	7	4.5	
	Am5	2	8	5	
	Am6	4	6	5	
	Am7	4	6	5	
	Am8	4	7	5.5	
	Am9	4	8	6	
	Am10	6	6	6	
	Am11	6	7	6.5	
	Am12	6	8	7	
	Am13	6	7	6.5	
	Am14	6	8	7	
	Am15	7	8	7.5	

Amostragem

$$t(x_{1j}, x_{2j}) = \bar{x}_j = \sum_{i=1}^{2} x_{ij}/2 \ (j = 1, 2, \cdots, 15)$$

$\bar{x}$	$P(\bar{X} = \bar{x})$
3	$1/15$
4	$2/15$
4.5	$1/15$
5	$3/15$
5.5	$1/15$
6	$2/15$
6.5	$2/15$
7	$2/15$
7.5	$1/15$

Estatística $\mathrm{T}(x_{1j}, x_{2j}) = \bar{X}_j$ é uma variável aleatória



Amostragem

Distribuições por amostragem do **mínimo** e do **máximo** da amostra.

Seja a amostra $(X_1, X_2, \cdots, X_n)$ onde $X_i \sim F(x)$, f.d.p ou f.p. $f(x)$.

- **Estatísticas de ordem**: obtêm-se ordenando a amostra:

$$\underbrace{X_{(1)}}_{Min\{X_i\}} \leq X_{(2)} \leq \cdots \leq \underbrace{X_{(n)}}_{Max\{X_i\}}$$

- **Distribuição do mínimo**:

$$G_{(1)}(X) = P\left(X_{(1)} \leq x\right) = 1 - [1 - F(x)]^n$$

- **Distribuição do máximo**:

$$G_{(n)}(X) = P\left(X_{(n)} \leq x\right) = [F(x)]^n$$

Amostragem

Exemplo 6.15 – Seja uma população $X \sim Exp(\lambda)$,e uma amostra casual $(X_1, X_2, \cdots, X_n)$. O mínimo da amostra, $X_{(1)}$, tem pelo T.5.4 uma distribuição exponencial de parâmetro $n\lambda$:

A distribuição do máximo da amostra, $X_{(n)}$ é:

$$G_{(n)}(X) = P\left(X_{(n)} \leq x\right) = \left[1 - e^{-\lambda x}\right]^n$$

Exerc.4 Seja uma amostra casual de dimensão 5 de uma população com função densidade, $f(x) = 3x^2 \quad (0 < x < 1)$. Determine a probabilidade de o valor máximo da amostra não exceder $0{,}9$. E de o valor mínimo da amostra ser inferior a $0{,}1$.

Amostragem

- Média e variância amostrais
  - Média amostral

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

  - Variância amostral

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

Teorema 6.1 – Se $(X_1, X_2, \cdots, X_n)$ é uma amostra casual de uma população para a qual existem média e variância

$$E(\bar{X}) = \mu; Var(\bar{X}) = \frac{\sigma^2}{n}$$

- O teorema apenas exige a existência de μ e de σ^2 (no universo).

Amostragem

$T(X_1, X_2) = \bar{X}$ é uma v.a. Discreta com
$D_{\bar{X}} = \{3, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5\}$

$\bar{x}$	$P(\bar{X} = \bar{x})$
3	$1/15$
4	$2/15$
4.5	$1/15$
5	$3/15$
5.5	$1/15$
6	$2/15$
6.5	$2/15$
7	$2/15$
7.5	$1/15$

Uma empresa tem 6 trabalhadores. O número de anos de experiência dos mesmos é $\{2, 4,6, 6, 7, 8\}$. Então a média de anos de experiência na população é

$$E(X) = \mu = 5.5$$

$$E(\bar{X}) = \sum_{\bar{x} \in D_{\bar{X}}} \bar{x} f_{\bar{X}}(\bar{x}) = 5.5$$

$$E(\bar{X}) = \mu$$

Amostragem

Teorema 6.2 – Se $(X_1, X_2, \cdots, X_n)$ é uma amostra casual de uma população para a qual existem média e variância, tem-se,

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

- Os valores de S^2 têm tendência para saírem inferiores à variância da população; a variância amostral subavalia, **em média**, a variância da população.

- Correcção do problema → **variância corrigida** definida por,

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

Amostragem

• Parâmetros População	• Estatísticas	• Parâmetros da distribuição amostral
Bernoulli: Proporção - θ	Proporção amostral - $\bar{X}$	Média da proporção amostral - $E(\bar{X})$
Poisson Média - λ	Média da amostra - $\bar{X}$	Média da média da amostra - $E(\bar{X})$
Normal Média - μ Variância - σ^2	Variância da amostra S^2	Variância da média da amostra $Var(\bar{X})$
Exponencial Média - μ	Variância corrigida da amostra S'^2	Média da Variância da amostra $E(S^2)$

Amostragem - Populações normais :

- Distribuição da Média amostral com variância conhecida

  - $(X_1, X_2, \cdots, X_n)$ amostra casual de uma população $N(\mu, \sigma^2)$
  - Tem-se, $E(\bar{X}) = \mu$; $Var(\bar{X}) = \sigma^2/n$

Logo, $\bar{X} \sim N(\mu, \sigma^2/n)$ ou $\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

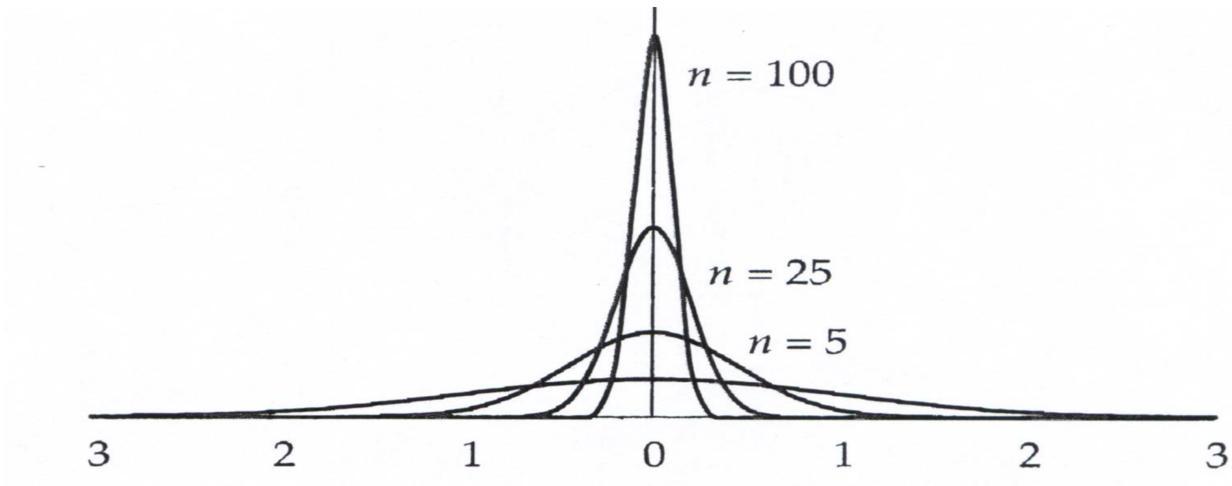


Fig. 6.13 – Distribuição de $\bar{X}$ para *n* = 5, 25 e100.

Amostragem:

Exemplo 6.21 – Suponha-se que a duração das chamadas telefónicas locais em determinada empresa pode ser bem aproximada por uma distribuição normal com média igual a 17 minutos e variância 25. Qual a probabilidade de, numa amostra aleatória de *n* chamadas, a duração média se situar entre (a) 16 e 18 minutos e (b) 14 e 16m?

a) $n = 25$

$$P(16 < \bar{X} < 18) = P\left(\frac{16-17}{5/\sqrt{25}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{18-17}{5/\sqrt{25}}\right)$$

$$= P(-1 < Z < 1)$$

$$= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \approx 0.6826$$

b) $n = 100$

$$P(16 < \bar{X} < 18) = P\left(\frac{16-17}{5/\sqrt{100}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{18-17}{5/\sqrt{100}}\right) = P(-2 < Z < 2)$$

$$= \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \approx 0.9544$$

Amostragem:

b) $n = 25$

$$P(14 < \bar{X} < 16) = P\left(\frac{14-17}{5/\sqrt{25}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{16-17}{5/\sqrt{25}}\right)$$

$$= P(-3 < Z < -1) = \Phi(-1) - \Phi(-3)$$

$$= \Phi(3) - \Phi(1) \approx 0.1573$$

$n = 100$

$$P(14 < \bar{X} < 16) = P\left(\frac{14-17}{5/\sqrt{100}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{16-17}{5/\sqrt{100}}\right) = P(-6 < Z < -2)$$

$$= \Phi(-2) - \Phi(-6) = \Phi(6) - \Phi(2) \approx 0.0228$$

Amostragem:

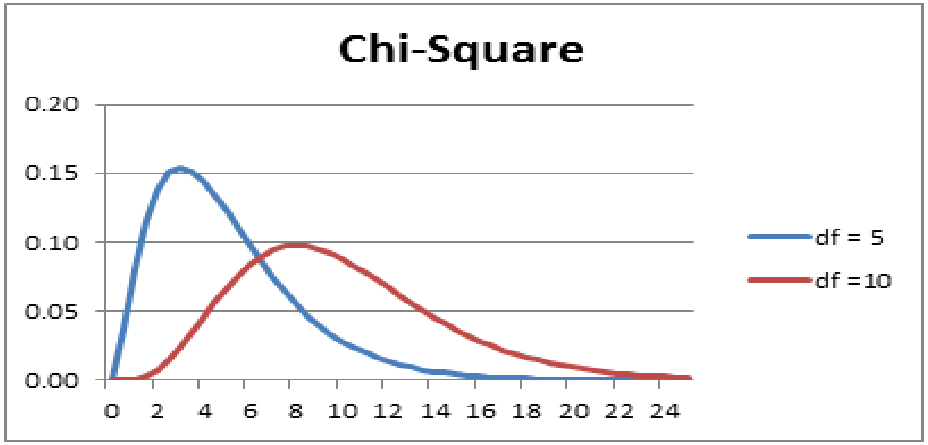
- Populações normais - Distribuição da Variância amostral

  - $(X_1, X_2, \cdots, X_n)$ amostra casual de uma população $N(\mu, \sigma^2)$
  - Relembre-se, $X_i \sim N(0,1) \Rightarrow X_i^2 \sim \boldsymbol{\chi}^2_{(1)} \Rightarrow \sum_{i=1}^{n} X_i^2 \sim \boldsymbol{\chi}^2_{(n)}$

Logo,

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{\boldsymbol{X}})^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

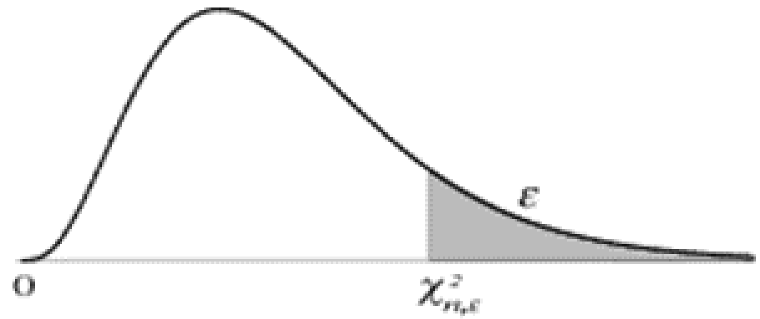
(um grau de liberdade é perdido quando se usa $\overline{X}$ em vez de $\boldsymbol{\mu}$)



Amostragem:

TABELA 6 – DISTRIBUIÇÃO DO QUI-QUADRADO

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



ε / n	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

Amostragem - Populações normais :

- Distribuição da Média amostral com variância desconhecida

  - $(X_1, X_2, \cdots, X_n)$ amostra casual de uma população $N(\mu, \sigma^2)$

    desconhecido

  - Deve substituir-se σ^2 por S^2 ou S'^2

  - Deve usar-se o rácio de Student $\dfrac{N(0,1)}{\sqrt{\left(\chi^2_{(n-1)}\right)/n}} \sim t_{(n-1)}$

Logo, se $\bar{X}$ e S^2 são a média e variância da amostra casual de tamanho n de uma população normal com média μ, então

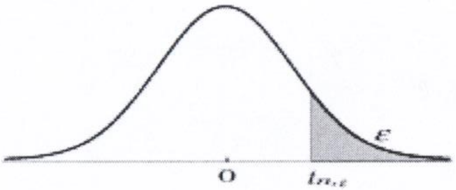
$$T = \frac{\bar{X} - \mu}{S'/\sqrt{n}} = \frac{\bar{X} - \mu}{S/\sqrt{n-1}} \sim t(n-1)$$

- Para grandes amostras a *t*-“*Student*” pode ser aproximada pela normal

Amostragem:

TABELA 7 – DISTRIBUIÇÃO *t*-*"Student"*

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



$n \backslash \varepsilon$	.400	.250	.100	.050	.025	.010	.005	.001
1	0.325	1.000	3.078	6.314	12.706	31.821	63.656	318.289
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	22.328
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	7.173
...								
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Amostragem

Teorema 5.10 Teorema do limite central

Dada a sucessão de variáveis aleatórias $iid, X_1, X_2, \cdots, X_n, \cdots$, com média μ e variância σ^2(finita), então, quando $n \to +\infty$, a função de distribuição da variável aleatória,

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \dot{\sim} N(0,1)$$

Diz-se que Z_n tem distribuição assintótica $N(0,1)$.

Notas: **1. O T.L.C. aplica-se a v.a.(s)** X_i discreta ou contínua.

2. O T.L.C. pode aplicar-se desde que se conheçam a média e variância da v.a. X_i , **<u>mesmo que a sua distribuição seja desconhecida</u>**

Amostragem

Distribuições por amostragem assintóticas

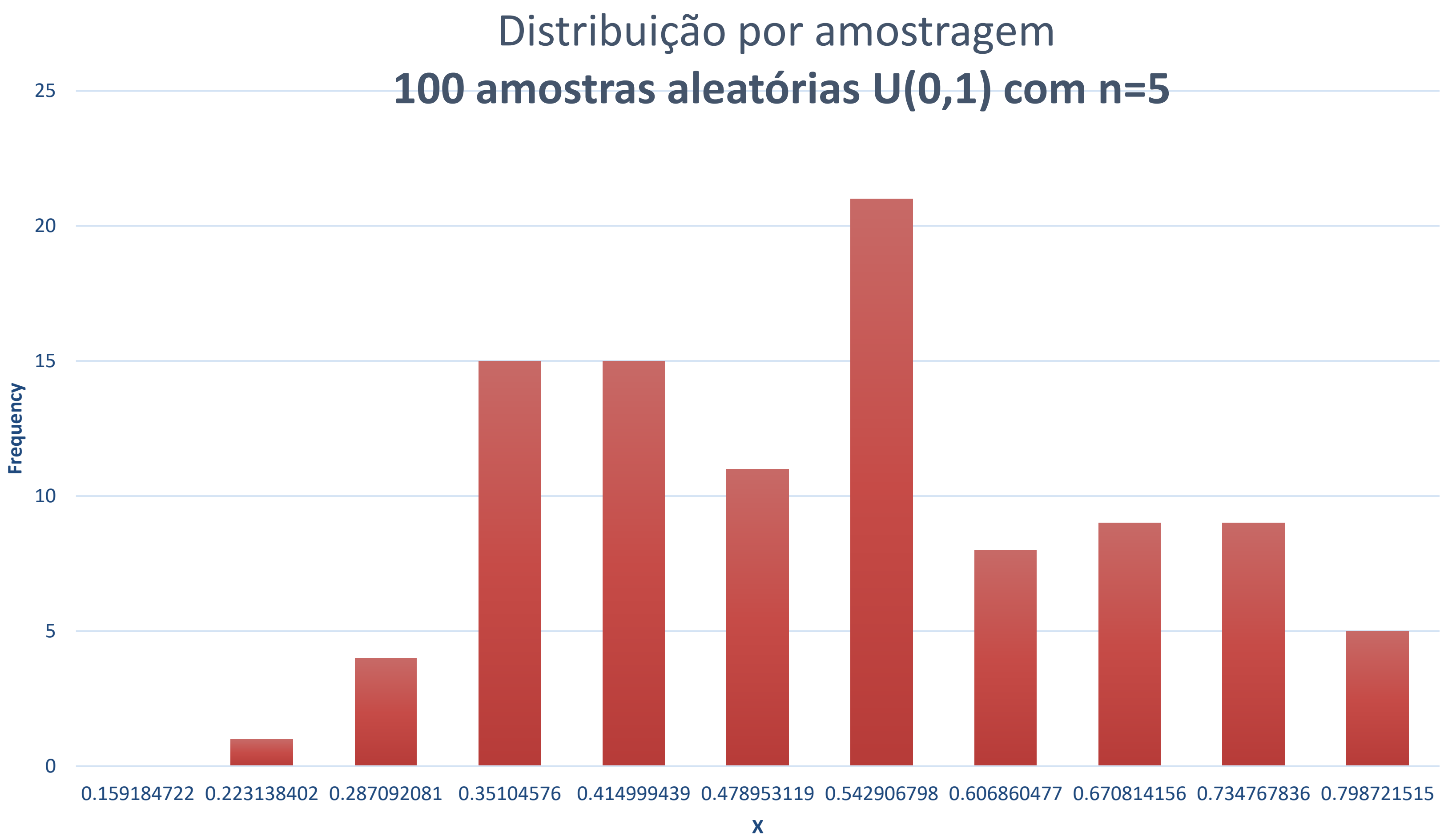
Em muitas situações não é possível obter **distribuições exactas** para as estatísticas $\sum_{i=1}^{n} X_i, \bar{X}, S^2, S'^2$, mas podem obter-se **distribuições aproximadas**, desde que existam os momentos da população até certa ordem.

- Distribuição assintótica da Média amostral
  - $(X_1, X_2, \cdots, X_n)$ amostra casual de uma população
  - $E(X_i) = \mu; Var(X_i) = \sigma^2 \, (i = 1,2,\cdots,n)$

Então, para $n \to \infty$, pelo Teorema do Limite Central

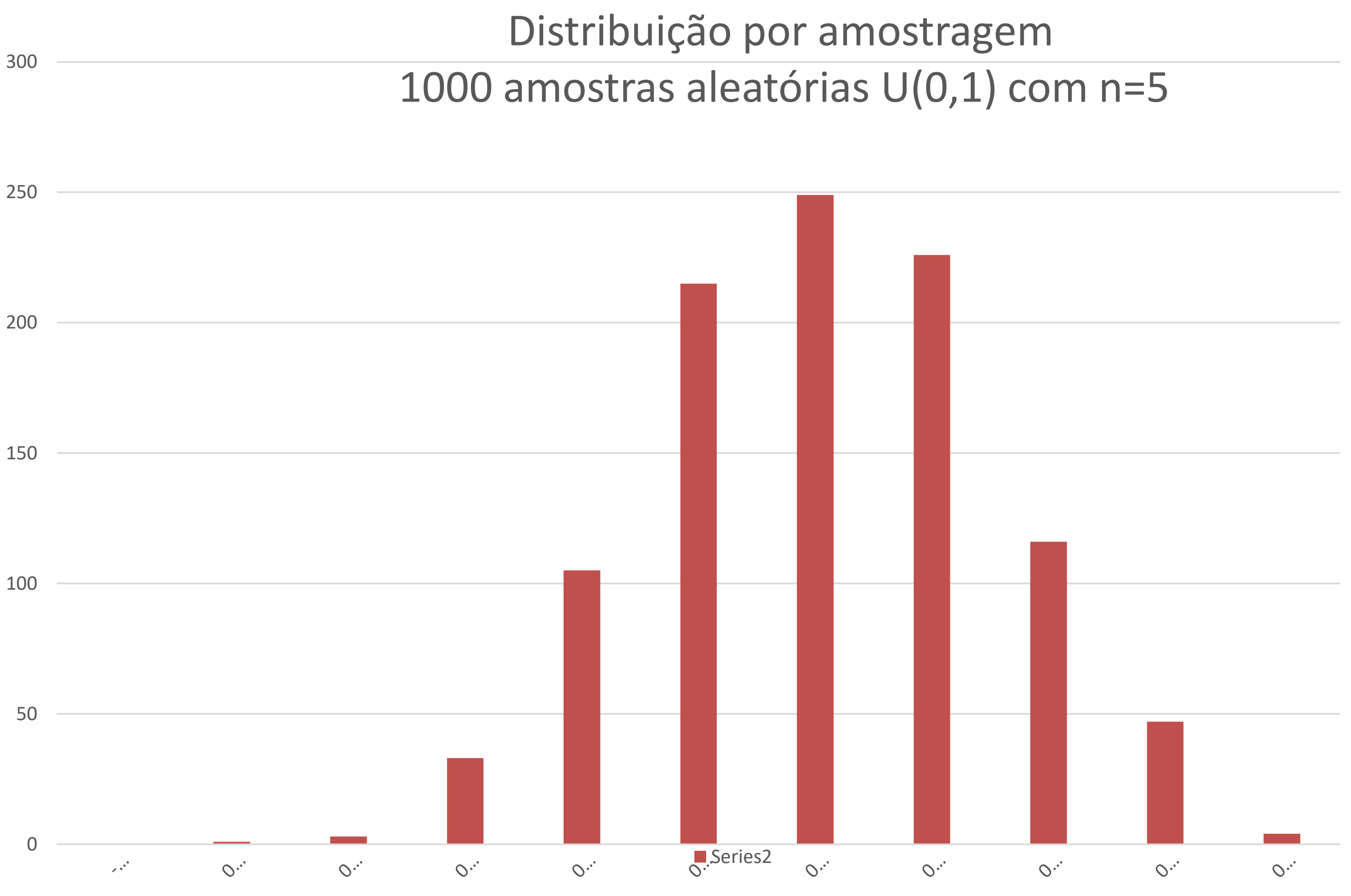
$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \dot{\sim} N(0,1) \qquad (6.21)$$

Amostragem

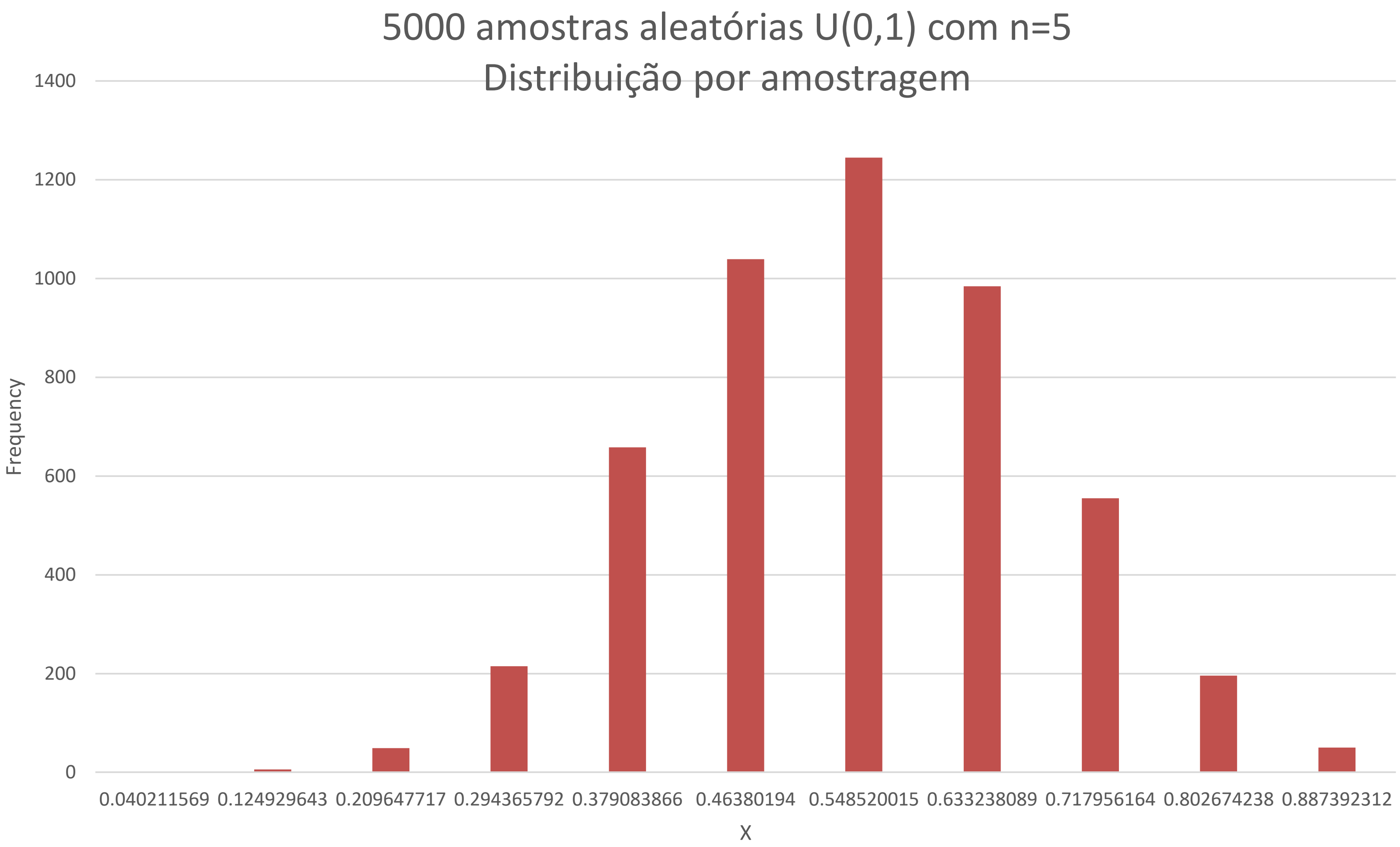


Distribuição por amostragem
100 amostras aleatórias U(0,1) com n=5
25
20
15
10
5
0
Frequency
0.159184722
0.223138402
0.287092081
0.35104576
0.414999439
0.478953119
0.542906798
0.606860477
0.670814156
0.734767836
0.798721515
x

Amostragem


Distribuição por amostragem
1000 amostras aleatórias U(0,1) com n=5
300
250
200
150
100
50
0
Series2

Amostragem



5000 amostras aleatórias U(0,1) com n=5
Distribuição por amostragem
Frequency
1400
1200
1000
800
600
400
200
0
0.040211569
0.124929643
0.209647717
0.294365792
0.379083866
0.46380194
0.548520015
0.633238089
0.717956164
0.802674238
0.887392312
X

Exemplo 6.17 – Sejam as variáveis aleatórias iid, $(X_1, X_2, \cdots, X_{30})$ com distribuição uniforme no intervalo $(0{,}10)$. Pretende determinar-se $P(\bar{X} < \ 5.5)$. Como o valor exacto é de difícil cálculo, recorre-se ao TLC (6.21). Então, tem-se:

$$E(\bar{X}) = \mu = 5, \qquad Var(\bar{X}) = \sigma^2/30 = \frac{10^2/12,}{30} = 0.53.$$

$$P(\bar{X} < 5.5) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{5.5 - 5}{0.53}\right) \approx \Phi(0.94) = 0.8264$$

Amostragem - Populações Bernoulli:

População é composta por elementos de dois tipos: os que possuem e os que não possuem determinado atributo.

- Distribuição da proporção amostral
  - $(X_1, X_2, \cdots, X_n)$ amostra casual de uma população $B(1, \theta)$
  - Tem-se, $E(\bar{X}) = \theta$; $Var(\bar{X}) = \theta(1-\theta)/n$
  - Então, para $n \to \infty$, pelo Teorema do Limite Central

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\theta}{\sqrt{n\theta(1-\theta)}} = \frac{\bar{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \dot{\sim} N(0,1) \qquad (6.21)$$

Amostragem - Populações Bernoulli:

Exemplo 6.19 – Admita-se que uma instituição bancária classifica os seus clientes possuidores de cartões de crédito em “maus” e “bons” riscos, conforme tenham ou não faltado a um pagamento nos últimos 2 anos. Suponha-se que a proporção de “maus” riscos (classificados por $X = 1$*) é de* 0.05 *para as agências da zona de Lisboa. Qual a* probabilidade de se obter pelo menos 10% de maus riscos numa amostra de:

(a) 50 clientes; (b) 400 clientes?

Amostragem - Populações Bernoulli:

A resposta a qualquer das duas alíneas é obtida calculando $P(\bar{X} \geq 0{,}1)$, sabendo-se que $X_i \sim Bi(1; 0{,}05) \quad i = 1,2,\cdots,n$.

(a) $n = 50;\ \theta = 0{,}05 \rightarrow$ aproximar usando o TLC

$$P(\bar{X} \geq 0{,}1) = 1 - \mathrm{P}\left(\frac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} < \frac{0{,}1 - 0{,}05}{\sqrt{\frac{0{,}05(1-0{,}05)}{50}}}\right)$$

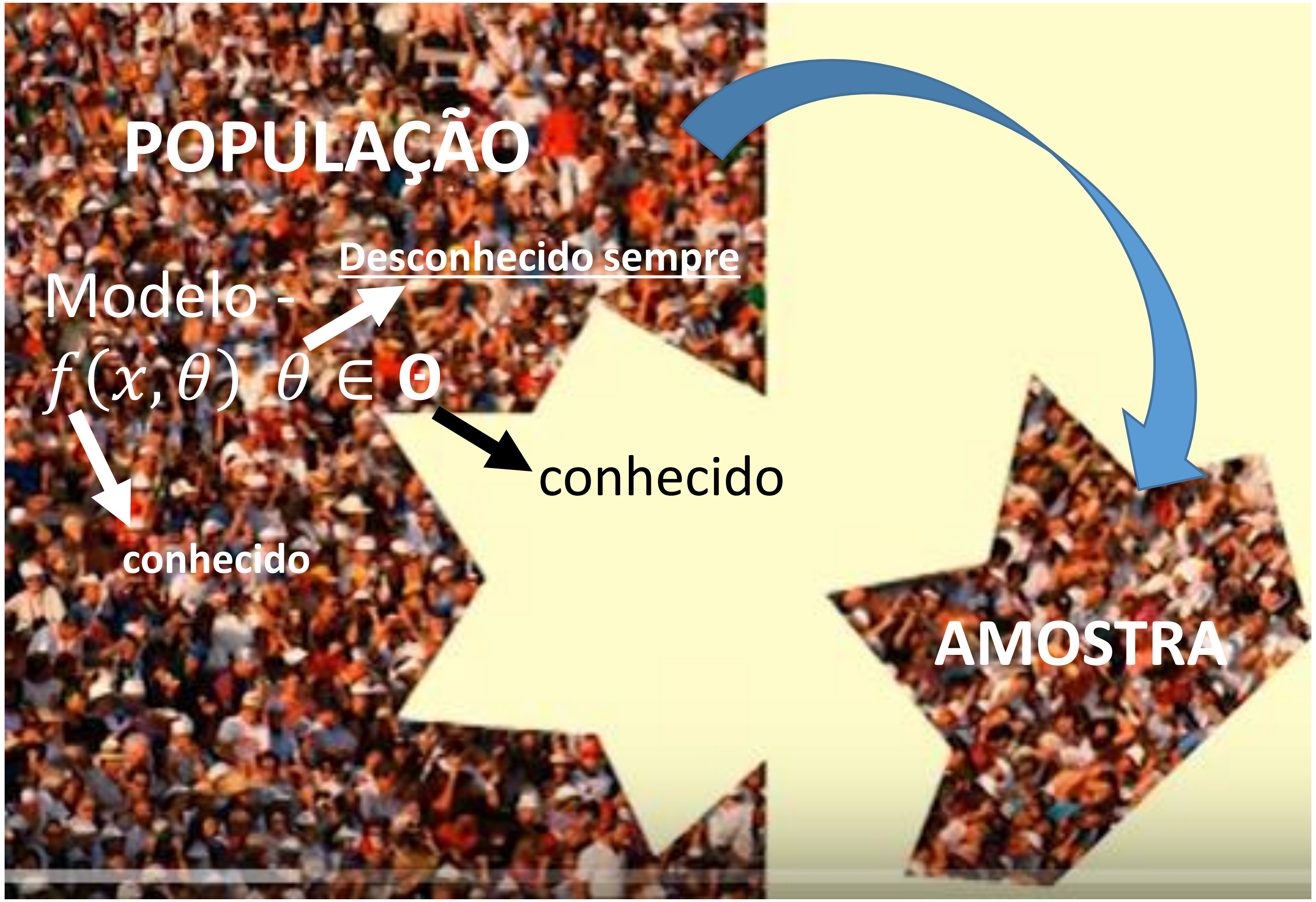
$$\approx 1 - \boldsymbol{\phi}(1.62) = 1 - 0.9474 = 0.0526$$

(b) $n = 400;\ \theta = 0{,}05 \rightarrow$ aproximar usando o TLC

$$P(\bar{X} \geq 0{,}1) = 1 - \mathrm{P}\left(\frac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} < \frac{0{,}1 - 0{,}05}{\sqrt{\frac{0{,}05(1-0{,}05)}{400}}}\right)$$

$$\approx 1 - \boldsymbol{\phi}(4{,}59) = 0$$

ESTIMAÇÃO

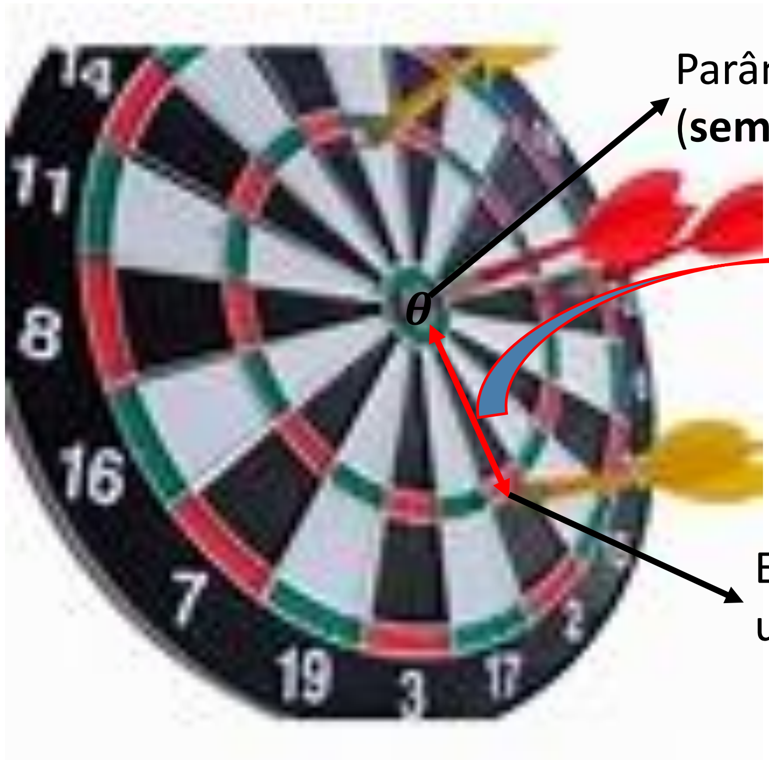


Estimar é utilizar a informação da amostra para "adivinhar" o valor de θ.

Dois aspectos a ter em, conta: a **precisão** e a **confiança.**

Ideia Importante $\rightarrow$ Fixada a dimensão da amostra, quanto mais precisa a resposta, menor a confiança que nela se deposita.

A estimação paramétrica desenvolve-se privilegiando: a precisão (**estimação por pontos**)
a confiança (**estimação por intervalos**).



Parâmetro **desconhecido** (**sempre**)

Precisão tem a ver com a distância entre o centro do alvo - $\boldsymbol{\theta}$ e o ponto onde a seta acertou no alvo – **estimativa**.

Estimativa – valor assumido pelo estimador para uma amostra particular **conhecido** (**sempre**)

Confiança tem a ver com a confiança em acertar num determinado círculo do alvo

ESTIMAÇÃO

Ter presente que o parâmetro de interesse θ pode ser:

- **Multidimensional**

Exemplo $\rightarrow$ Suponha que a valorização de um activo financeiro tem distribuição normal com média μ e variância σ^2.
Observada uma amostra casual pretende-se estimar μ (rendibilidade esperada) e σ^2(risco).

- **Função do(s) parâmetro(s) da distribuição**

. Exemplo $\rightarrow$ Suponha-se que o número de sinistros originados anualmente por uma apólice de seguro automóvel tem distribuição de Poisson de parâmetro $\boldsymbol{\lambda}$ (desconhecido). Em vez de nos interessarmos pelo parâmetro (média do fenómeno) podemos estar interessados numa função de $\boldsymbol{\lambda}$, por exemplo, $P(X=0|\boldsymbol{\lambda})=e^{-\boldsymbol{\lambda}}$, probabilidade de não se verificar nenhum sinistro. Então pretende-se estimar $\tau(\boldsymbol{\lambda})=e^{-\boldsymbol{\lambda}}$ que traduz essa probabilidade

ESTIMAÇÃO POR PONTOS

- **Conceitos Fundamentais:**

Estimador: é uma variável aleatória, função da amostra casual e representa-se por $T(X_1, X_2, \cdots, X_n) \Rightarrow$ é uma estatística.

Variável aleatória

Por exemplo: $T(X_1, X_2, \cdots, X_n) = \bar{X}$ média da amostra

Estimativa: é um número assumido pelo estimador para a particular amostra que se observou. Representa-se por $t(x_1, x_2, \cdots, x_n) \in \mathbb{R}$.

Por exemplo: $t(x_1, x_2, \cdots, x_n) = \bar{x} \longrightarrow$ **Valor constante $\Rightarrow$não varia $\Rightarrow$não tem qualquer distribuição**

- **Dois problemas em aberto**:
  - Como encontrar estimadores para determinado parâmetro?
  - Encontrado um ou mais estimadores, como avaliar a sua qualidade?

ESTIMAÇÃO POR PONTOS

Método dos momentos

IDEIA

Utilizar os momentos da amostra para estimar os correspondentes momentos da população e, a partir daí, estimar os parâmetros de interesse

- Existem outros métodos de estimação por pontos: Método da máxima verosimilhança.

Método dos momentos

Momentos da população tem de existir	• Momentos em relação à origem da População		• Momentos empíricos ou da amostra	Momentos da amostra existem sempre
	$\mu'_1 = E(X)$	$=$	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	
	$\mu'_2 = E(X^2)$	$=$	$\frac{\sum_{i=1}^{n} X_i^2}{n}$	
	$\vdots$		$\vdots$	
	$\mu'_r = E(X^r)$	$=$	$\frac{\sum_{i=1}^{n} X_i^r}{n}$	

ESTIMAÇÃO POR PONTOS

Método dos momentos

- **Formalização**:
  - $(X_1, X_2, \cdots, X_n)$ amostra casual de uma população

$$f(x, \theta_1, \theta_2, \cdots, \theta_k) \quad \Big(\overbrace{\theta_1, \theta_2, \cdots, \theta_k}^{desconhecidos}\Big) \in \boldsymbol{\Theta}$$

  - Constrói-se um sistema igualando os k 1ºs momentos da população aos k 1ºs momentos da amostra. O sistema terá tantas equações quantos os parâmetros a estimar.

$$\underbrace{\mu'_r = \psi_r(\theta_1, \theta_2, \cdots, \theta_k)}_{momento\ em\ relação\ à\ origem} = \underbrace{\frac{\sum_{i=1}^{n} X_i^r}{n} = \psi_r(\theta_1, \theta_2, \cdots, \theta_k) \;\; r = 1, 2, \cdots, k}_{momento\ da\ amostra\ ou\ empírico}$$

  - Resolve-se o sistema em ordem aos k parâmetros desconhecidos. Admite-se que o sistema tem solução única $\tilde{\theta}_j = \boldsymbol{\phi}_j(X_1, X_2, \cdots, X_n)$ $\quad j = 1, 2, \cdots, k$
  - Diz-se que os estimadores $\widetilde{\theta_1}, \widetilde{\theta_2}, \cdots, \widetilde{\theta_k}$ foram obtidos pelo método dos momentos

ESTIMAÇÃO POR PONTOS

Método dos momentos

- Exemplo 1 Considere-se uma população de Bernoulli da qual se extraiu uma amostra casual de dimensão n com o objectivo de estimar θ.

Como se sabe:

- 1º momento da população é $\mu_1' = E(X) = \theta$ → Constante – não conhecida
- 1º momento da amostra é $\bar{X}$ → **Variável aleatória**
- sistema - $\bar{X} = \theta$
- solução: Estimador → $\tilde{\theta}_{(X_1,X_2,\cdots,X_n)} = \bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

  Estimativa → $\tilde{\theta}_{(x_1,x_2,\cdots,x_n)} = \bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ → Constante – conhecida

Cuidado com a notação !

ESTIMAÇÃO POR PONTOS

Método dos momentos

- Exemplo 2 Considere-se uma população normal da qual se extraiu uma amostra casual de dimensão n com o objectivo de estimar μ e σ^2

Constantes desconhecidas

Como se sabe:

- momentos da população: $\mu_1' = E(X) = \mu; \ Var(X) = E(X^2) - \mu^2 \Rightarrow \mu_2' = E(X^2) = \sigma^2 + \mu^2$

- momentos da amostra: $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ e $\frac{\sum_{i=1}^{n} X_i^2}{n}$

Variáveis aleatórias

- sistema - $\begin{cases} \bar{X} = \mu \\ \frac{\sum_{i=1}^{n} X_i^2}{n} = \sigma^2 + \mu^2 \end{cases}$

- solução: Estimadores: $\tilde{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$; $\widetilde{\sigma^2} = \frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}^2 = S^2$

<u>**Cuidado com a notação !**</u>

Estimativas: $\tilde{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$; $\widetilde{\sigma^2} = s^2$

Constantes conhecidas