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 - 2024/2025
- Equipa dos docentes:
  - **Erida Gjini**, erida.gjini@iseg.ulisboa.pt **(aulas teoricas)**
  - **Gabriel Zsurkis**
  - **Wilso Araujo**

- PROGRAMA
  - **Parte 1: Estatística (~ 28/01-27/02)**
    - *Estimação*
    - *Testes de hipóteses*
    - *Testes não-paramétricos*
  - **Parte 2: Econometria (~06/03-29/04)**
    - *Introdução*
    - *Modelo de regressão linear multipla*
    - *Inferência estatistica*
    - *Heterocedasticidade*
    - *Previsao*
    - *Tópicos sobre formas funcionais*

190 alunos inscritos

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: 28 Janeiro– 29 Abril 2025

24 aulas teóricas (2 por semana)

4 aulas praticas por semana– (sexta- e quinta-feira) Ver turmas

Avaliação

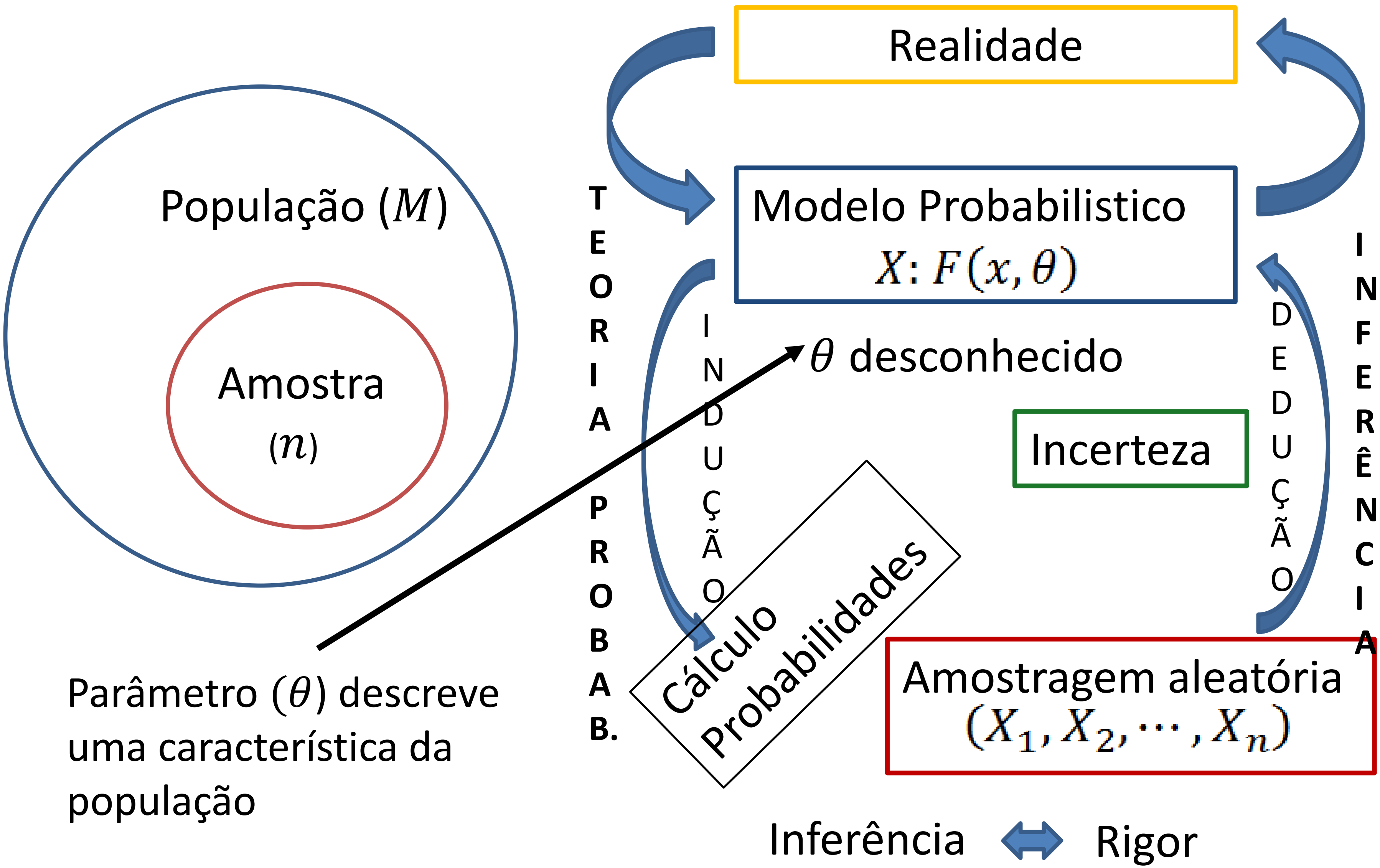
- Exame final (100%)
  - *Exemplos dos anos anteriores vão ser disponíveis*
  - *Época normal: 07.05.2025 @ 12:00-15:00*
  - *Época recurso/melhoria: 03.06.2025 @12:00-15:00*
  - *Exame EE: 04.07.2025 @12:00-15:00 horas*

Horário: terças e quintas das 13h30 às 15h00 no Anfiteatro 3 do edifício das Francesinhas 2.

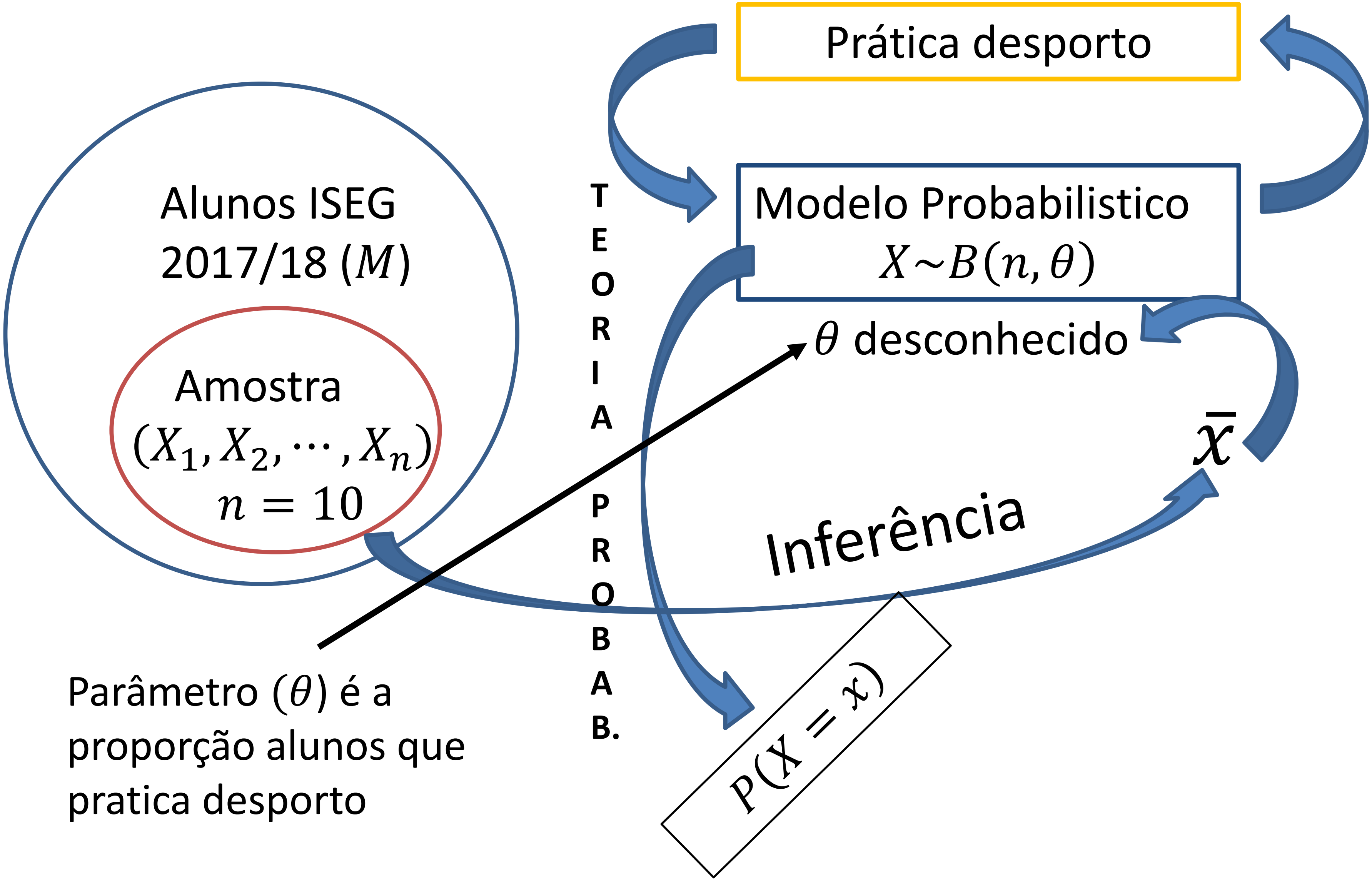
Plano para hoje

- 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

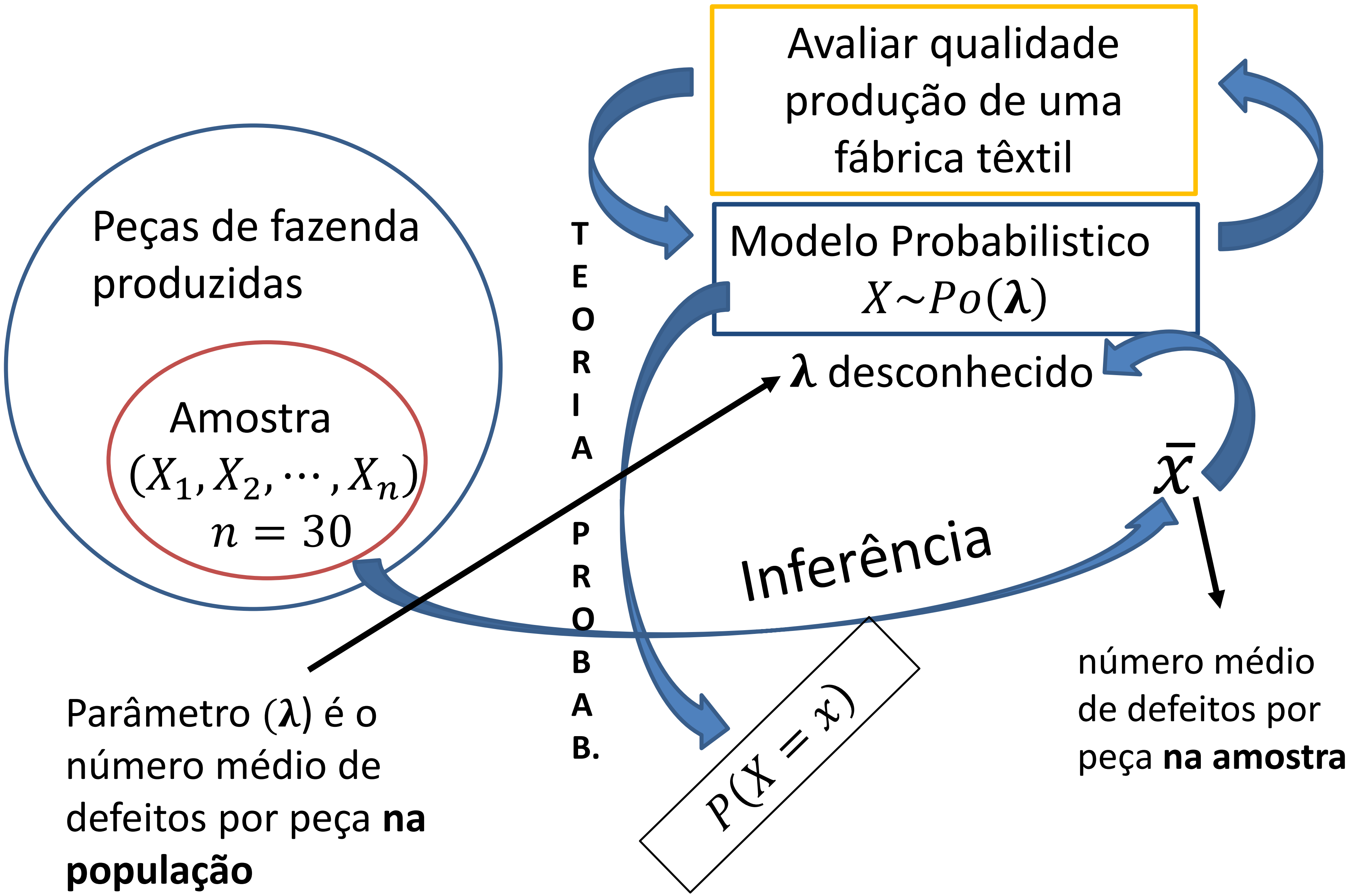
Amostragem: Introdução



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

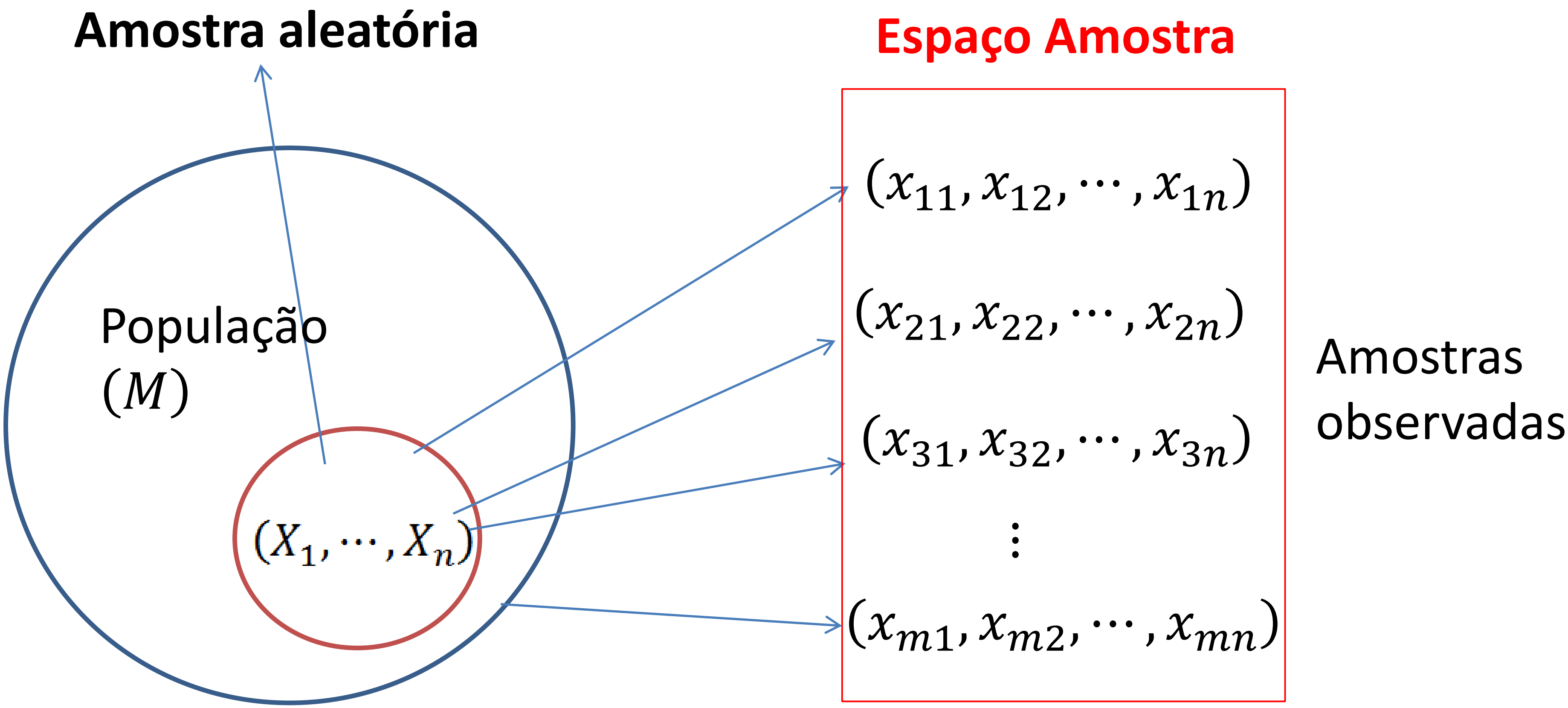
- População
  - Conjunto de “números” dos quais se extrai uma amostra

- 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

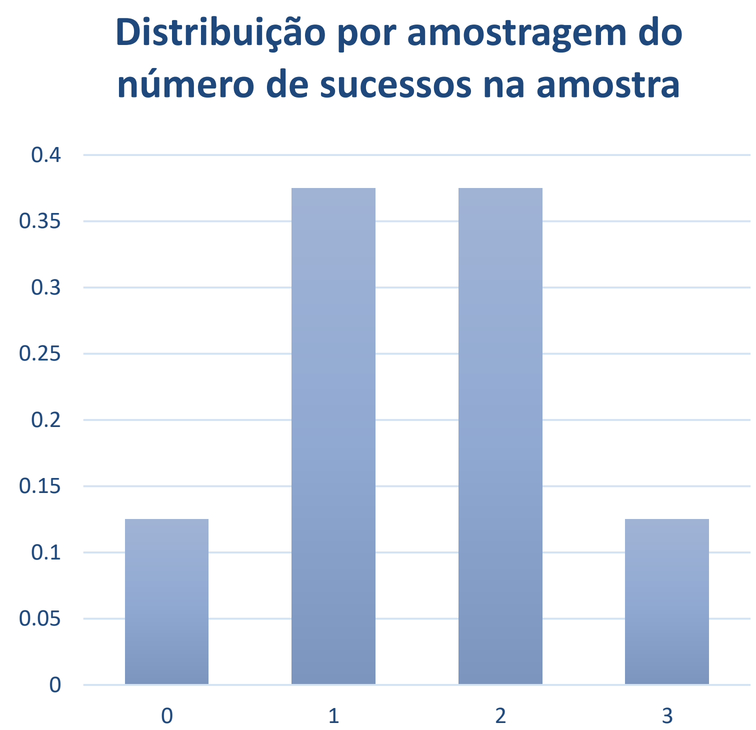
- Distribuições amostrais
  - Distribuição da estatística T.
  - Pode ser difícil de obter
  - Vamos usar a estatística T e a sua distribuição para inferir sobre parâmetros da população usando a amostra
  - Utilidade de qualquer estatística **depende** do **comportamento probabilístico da estatística** 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	$\boldsymbol{t}_1$	$P\left(T_1 = \sum_{i=1}^{n} \boldsymbol{x}_i\right)$
$(\mathbf{0},\mathbf{0},\mathbf{0})$	0	$1/8$
$(\mathbf{1},\mathbf{0},\mathbf{0})$, $(\mathbf{0},\mathbf{1},\mathbf{0})$, $(\mathbf{0},\mathbf{0},\mathbf{1})$	1	$3/8$
$(\mathbf{1},\mathbf{1},\mathbf{0})$, $(\mathbf{1},\mathbf{0},\mathbf{1})$, $(\mathbf{0},\mathbf{1},\mathbf{1})$	2	$3/8$
$(\mathbf{1},\mathbf{1},\mathbf{1})$	3	$1/8$



Amostragem

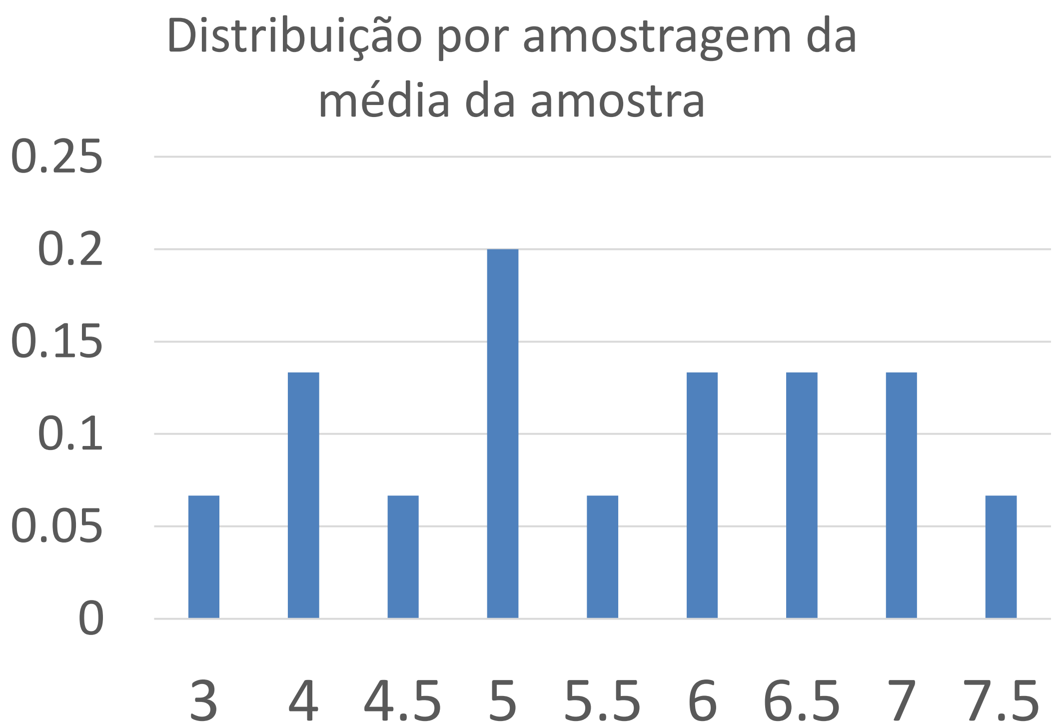
		x_1	x_2	$\bar{x}_j = \sum_{i=1}^{2} x_i/2$	
ESPAÇO AMOSTRA	Am1	2	4	3	ESPAÇO de ESTATÍSTICAS
	Am2	2	6	4	
	Am3	2	6	4	
	Am4	2	7	4.5	
	Am5	2	8	5	
	Am6	4	6	5	
	Am7	4	6	5	
	Am8	4	7	5.5	
	Am9	4	8	6	
	Am10	6	6	6	
	Am11	6	7	6.5	
	Am12	6	8	7	
	Am13	6	7	6.5	
	Am14	6	8	7	
	Am15	7	8	7.5	

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.

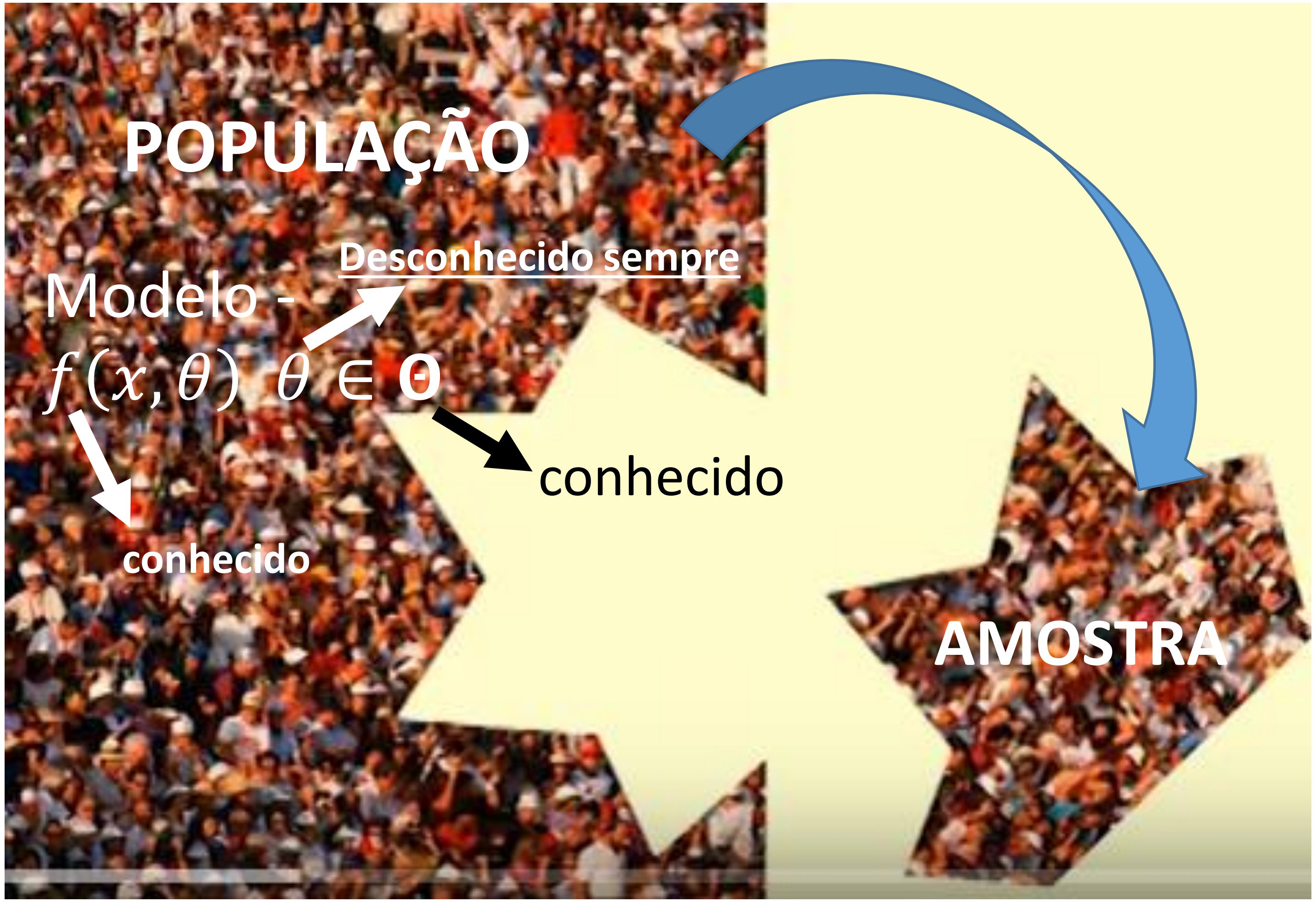
- Correcção do problema → **variância corrigida** definida por,

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

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 $E(S^2)$

ESTIMAÇÃO

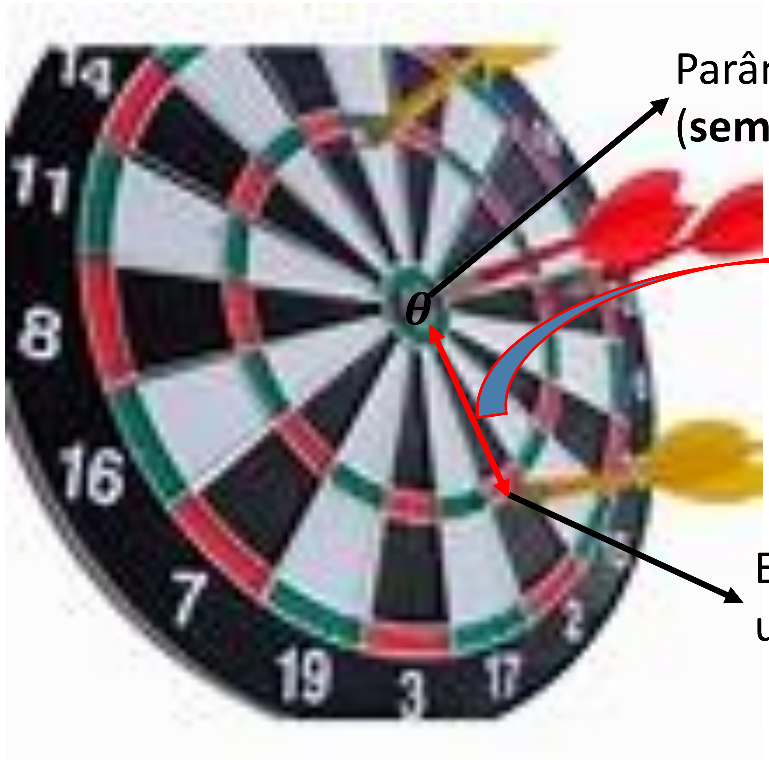


Estimar é utilizar a informação da amostra para “adivinhar” o valor de θ.

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, $P(X = 0|\boldsymbol{\lambda}) = e^{-\lambda}$, probabilidade de não se verificar nenhum sinistro. Então pretende-se estimar $\tau(\boldsymbol{\lambda}) = e^{-\lambda}$ 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: $t(x_1, x_2, \cdots, x_n) = \bar{x}$ → **Valor constante ⇒não varia ⇒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
$\mu_1' = E(X)$	$=$	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
$\mu_2' = E(X^2)$	$=$	$\frac{\sum_{i=1}^{n} X_i^2}{n}$
$\vdots$		$\vdots$
$\mu_r' = E(X^r)$	$=$	$\frac{\sum_{i=1}^{n} X_i^r}{n}$

Momentos da amostra existem sempre

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 $\tilde{\theta}_j = \boldsymbol{\phi}_j(X_1, X_2, \cdots, X_n) \qquad j = 1, 2, \cdots, k$
  - 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:

Constante – não conhecida

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

Constante – conhecida

Estimativa ⟶ $\tilde{\theta}_{(x_1,x_2,\cdots,x_n)} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

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