Lisbon School of Economics & Management
Universidade de Lisboa

Estatística I

Licenciatura em Gestão
2.º Ano/1.º Semestre
2023/2024

Aulas Teóricas N.ºs 24 e 25 (Semana 13)

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 3)	Aulas Teóricas (Semanas 3 a 5)	Aulas Teóricas (Semanas 5 a 7)	Aulas Teóricas (Semanas 8 a 13)
• **Capítulo 2**: Probabilidades	• **Capítulo 3**: Variáveis Aleatórias Unidimensionais	• **Capítulo 4**: Variáveis Aleatórias Multidimensionais	• **Capítulo 5**: Distribuições Teóricas • **Capítulo 6**: Amostragem. Distribuições por Amostragem.

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

Aula 20	Início do capítulo 5: probabilidade versus inferência estatística. Universo e amostra. Amostra aleatória. Estatísticas. Distribuição por amostragem.
Aula 21	Distribuição por amostragem do máximo e do mínimos amostrais. Momentos da média e da variância amostrais. Distribuição assintótica da média amostral. Uma população Bernoulli: distribuição do total da amostra e da média amostral.
Aula 22	Distribuição por amostragem no contexto de duas populações Bernoulli. Distribuição por amostragem da média amostra e da variância amostral no contexto de uma população normal. Rácio t-student. Distribuição da diferença entre as médias amostrais no contexto de duas populações normais: início.
Aula 23	Inferência sobre a diferença das médias de duas populações normais. Inferência sobre o rácio das variâncias de duas populações normais.

1

Inferência Estatística

Conceitos: Universo, Amostra Aleatória, Distribuição de uma Amostra Aleatória e Estatísticas

Probabilidade vs. Inferência Estatística

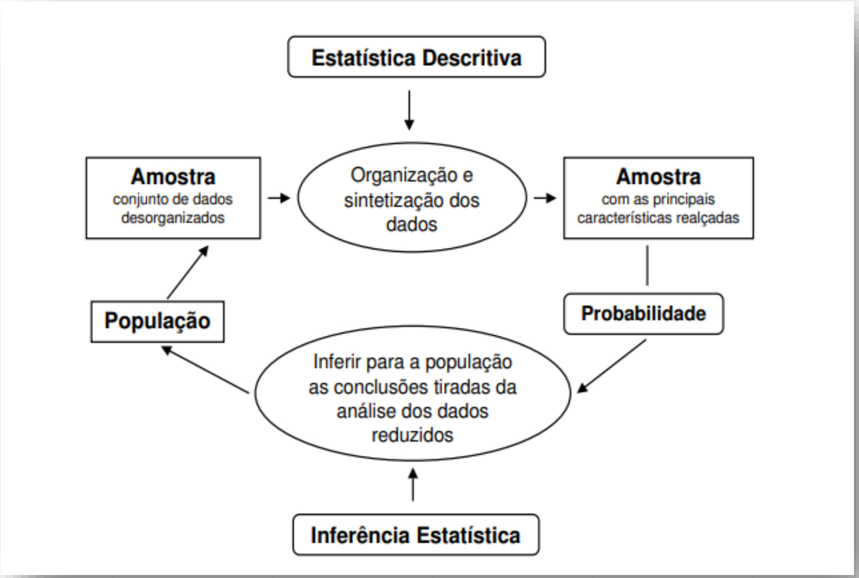
Pode dizer-se que a Probabilidade e a Inferência têm objectivos diferentes: enquanto na Probabilidade se parte de um dado esquema ou modelo para calcular a probabilidade de certos resultados ou acontecimentos se realizarem; na Inferência parte-se de dados ou observações e procura saber-se ou inferir-se algo sobre o modelo.

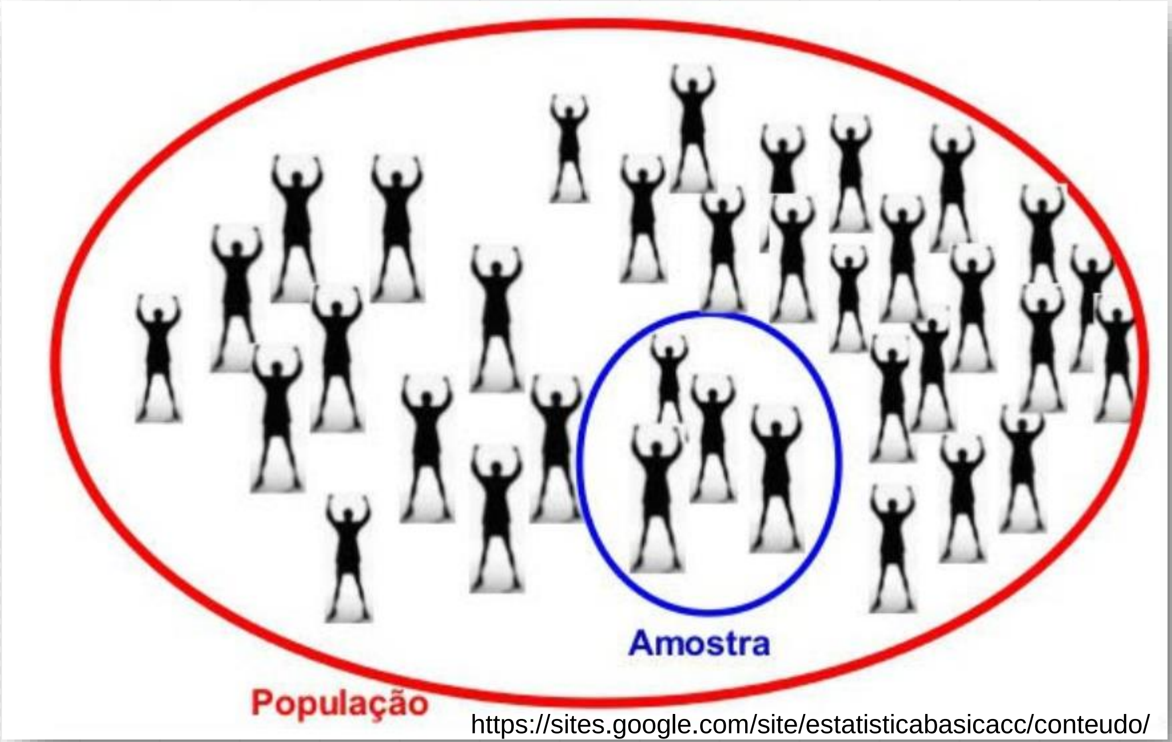
A Inferência é a "passagem do particular ao geral."

A Inferência Estatística tem como objectivo definir procedimentos que aplicados a uma amostra extraída da população, nos permitam estimar parâmetros desconhecidos dessa população ou algo sobre o modelo da população.

De facto uma amostra particular é apenas uma das muitas amostras (em nº infinito se a população for infinita) que se podem obter por um processo de amostragem.

Estatística Descritiva vs. Inferência Estatística





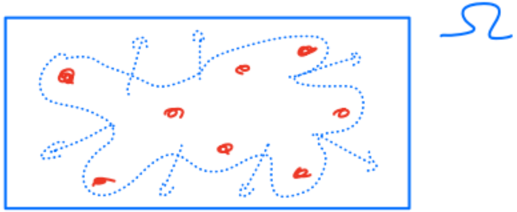
https://sites.google.com/site/estatisticabasicacc/conteudo/

- **POPULAÇÃO:**
  - é uma coleção completa de todos os elementos a serem estudados
- **AMOSTRA:**
  - é um subconjunto da população
- **CENSO:**
  - é uma coleção da dados relativos a todos os elementos de uma população:

https://slideplayer.com.br/slide/2627398

Amostra Aleatória

Perante "falta de informação" sobre a v.a. (população), vamos recolher uma amostra, estudar intensamente esta amostra e tentar extrapolar p/a população.



• - elementos da população que vão ser estudados

$\Downarrow$

amostra

$(x_1, x_2, .., x_n)$

Slides Professora Claúdia Nunes

População vs Amostra Aleatória

população → recolher um subconjunto representativo da população →

(X)
↓
v.a. interesse

① $(X_1, X_2, \ldots, X_n)$

(amostra aleatória)

② estudar a amostra, através da estatística descritiva

→ ③ inferir ("extrapolar") os resultados da amostra p/ a população.

Slides Professora Claúdia Nunes

Amostragem

① Amostragem (Regras p/ assegurar a "bondade" dos resultados)

Amostra é representativa da população se tiver as seguintes características:

$(X_1, X_2, \ldots, X_n)$: conjunto de variáveis aleatórias independentes entre si e têm todas a mesma distribuição, que é a distribuição da v.a. (população) X

Slides Professora Claúdia Nunes

a.a = amostra aleatória
i.i.d. = independentes e identicamente distribuídas

Distribuição de uma a.a.: V.a.´s Discretas

Qual é a distribuição da a.a. ($X1$, …, Xn), sendo as v.a.´s Xi´s (i = 1, …, n) iid a uma v.a. discreta X?

Distribuição de uma a.a.: V.a.´s Discretas

1ª Consequência: Qual a distribuição de uma a.a.?

i) se X for uma v.a. discreta

$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$ [função de prob. conjunta da a.a.]

$\underset{\text{independentes}}{=} P(X_1 = x_1) P(X_2 = x_2) \ldots P(X_n = x_n)$

$\underset{\text{ident. distribuídas}}{=} P(X = x_1) P(X = x_2) \ldots P(X = x_n)$

$= \prod_{i=1}^{n} P(X = x_i)$

Slides Professora Claúdia Nunes

Distribuição de uma a.a.: V.a.´s Poisson

Determine a seguinte função de probabilidade conjunta da a.a. $(X1, X2, X3)$: $P(X1=4, X2=3, X3=5)$, sendo Xi´s ($i = 1, 2, 3$) iid a $X \cap P(\lambda)$.

Distribuição de uma a.a.: V.a.´s Poisson

exemplo: $X \sim Poisson(\lambda)$ $\lambda = ?$

$$P(X_1 = 4, X_2 = 3, X_3 = 5) = \prod_{i=1}^{3} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$= \frac{e^{-\lambda}\lambda^4}{4!}\frac{e^{-\lambda}\lambda^3}{3!}\frac{e^{-\lambda}\lambda^5}{5!} = \frac{e^{-3\lambda}\lambda^{12}}{4!\,3!\,5!}$$

Slides Professora Claúdia Nunes

Distribuição de uma a.a.: V.a.´s Contínuas

Qual é a distribuição da a.a. ($X1$, …, Xn), sendo Xi´s (i = 1, …, n) iid a uma v.a. Contínua X?

Distribuição de uma a.a.: V.a.´s Contínuas

ii) se X for uma v.a. contínua

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)$$ [função de densidade de probabilidade conjunta]

$$\underset{\text{independentes}}{=} f_{X_1}(x_1)\, f_{X_2}(x_2) \ldots f_{X_n}(x_n)$$

$$= f_X(x_1)\, f_X(x_2) \ldots f_X(x_n)$$

identicamente distribuídas

$$= \prod_{i=1}^{n} f_X(x_i)$$

Slides Professora Claúdia Nunes

Distribuição de uma a.a.: V.a.´s Exponenciais

Determine a seguinte função densidade de probabilidade conjunta da a.a. $(X1, X2, X3, X4)$: $f(1.3, 0.5, 2.5, 0.78)$, sendo $Xi´s$ $(i = 1, 2, 3, 4)$ iid a $X \cap Exp(\lambda)$.

Distribuição de uma a.a.: V.a.´s Exponenciais

exemplo: $X \sim \text{Exp}(\mu)$

$x_1 = 1.3$; $x_2 = 0.5$; $x_3 = 2.5$; $x_4 = 0.78$

$$f_{X_1, X_2, X_3, X_4}(1.3, 0.5, 2.5, 0.78) = \mu e^{-\mu \times 1.3} \times \mu e^{-\mu \times 0.5}$$

$$\times \mu e^{-\mu \times 2.5} \times \mu e^{-\mu \times 0.78}$$

$$= \mu^4 e^{-\mu(1.3 + 0.5 + 2.5 + 0.78)} //$$

Média e Variância Amostrais

② Duas características da amostra frequentemente calculadas são:

<u>média amostral</u>: $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$

<u>variância amostral</u>: $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2 \quad (*)$$

(*) porque: $\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 =$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i^2 + \bar{x}^2 - 2 x_i \bar{x}) = \frac{1}{n} \Big[\sum_{i=1}^{n} x_i^2 + n\bar{x}^2 - 2\bar{x} \underbrace{\sum_{i=1}^{n} x_i}_{n\bar{x}} \Big]$$

$$= \frac{1}{n} \Big[\sum_{i=1}^{n} x_i^2 + n\bar{x}^2 - 2n\bar{x}^2 \Big]$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

variância amostral corrigida: $s^2_{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Inferência Estatística

(3) Inferência estatística

- X (v.a. de interesse)
- com distribuição conhecida
- todos ou alguns parâmetros da distribuição são desconhecidos

parâmetros desconhecidos (θ)

objetivo: "adivinhar" o valor de θ, com base na amostra

i.e

estimar θ, com base na amostra

$\Downarrow$

Estimação $\begin{cases} \text{Pontual (cap. 6)} \\ \text{Intervalar (cap. 7)} \end{cases}$

Slides Professora Claúdia Nunes

Estatística

i) ESTATÍSTICA: uma estatística é uma função (qualquer) que depende da amostra e não depende de nenhum parâmetro desconhecido.

Exemplo: $X \sim Poisson(\lambda)$ $\quad \lambda = ?$

a.a. $(X_1, X_2, \ldots, X_{10})$

$T_1 = \overline{X} = \frac{1}{10} \sum_{i=1}^{n} X_i$: É UMA ESTATÍSTICA

$T_2 = 3$: É UMA ESTATÍSTICA

$T_3 = \frac{1}{10} \sum_{i=1}^{10} (X_i - \lambda)^2$ NÃO É ESTATÍSTICA!!

Slides Professora Claúdia Nunes

Estatística

DEF. 4: ESTATÍSTICA:

Uma estatística é uma função das variáveis observáveis, e é por si própria uma variável observável que não contém qualquer parâmetro desconhecido.

EXEMPLO 5: Seja a a.a. $[X_1, X_2, \ldots, X_n]$. A média amostral $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ é uma estatística.

EXEMPLO 6: Seja a v.a $X \sim N(\mu, \sigma^2)$ com μ e σ^2 desconhecidos, então $x - \mu$ não é estatística porque não é observável, é função do parâmetro desconhecido μ.

EXEMPLO 6: Seja uma v.a. com distribuição $N(\mu;\sigma^2)$. Quais são Estatísticas?

a) $X^2-\mu$

b) $\dfrac{X}{\sigma^2}$

c) X^2-3

d) $X-4$

e) $X - \log X^3$

Quais destas funções são estatísticas?

apostila.PDF (ufpr.br)

Distribuições do Mínimo e Máximo

Distribuições de Amostragem

2

Estatísticas de Ordem: Distribuição do Mínimo e Distribuição do Máximo

$P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) = 1 - [P(X_1 > x) \times \ldots \times P(X_n > x)] = 1 - [1 - P(X < x)^n] = 1 - [1-F(x)]^n$

$P(X_{(n)} \leq x) = P(X_1 \leq x) \times \ldots \times P(X_n \leq x) = P(X < x)^n = F(x)^n$

Distribuições de Amostragem

Formulário

- **DISTRIBUIÇÃO DO MÍNIMO E DO MÁXIMO**

$$G_1(x) = 1 - [1 - F(x)]^n \quad ; \quad G_n(x) = [F(x)]^n$$

ProbabilidadesEstatistica2019.pdf

Exercícios Suplementares

3

Uma dada universidade possui 40% de estudantes de áreas TIC e os restantes estudantes são de outras áreas. De entre os estudantes de áreas TIC, 5% têm mais de 25 anos de idade. De entre os estudantes de outras áreas, 2% têm mais de 25 anos de idade.

Tendo sido escolhido ao acaso um estudante da universidade e verificando-se que o mesmo tem mais de 25 anos de idade, calcule a probabilidade de ter sido escolhido um estudante de áreas TIC.

Resolução do Exercício

- **Acontecimentos e probabilidades para um estudante da universidade escolhido ao acaso**

Acontecimento	Probabilidade
T = "escolha de um estudante de áreas TIC"	$P(T) = 0.4$
M = "escolha de um estudante com mais de 25 anos de idade"	$P(M) = ?$
	$P(M \mid T) = 0.05$
	$P(M \mid \bar{T}) = 0.02$

- **Cálculo da probabilidade pedida**

Ao recorrer-se ao teorema de Bayes, temos

$$
\begin{aligned}
\boxed{P(T \mid M)} &= \frac{P(M \mid T) \times P(T)}{P(M)} \\
&= \frac{P(M \mid T) \times P(T)}{P(M \mid T) \times P(T) + P(M \mid \bar{T}) \times P(\bar{T})} \\
&= \frac{0.05 \times 0.4}{0.05 \times 0.4 + 0.02 \times (1 - 0.4)} \\
&= \frac{0.02}{0.02 + 0.012} \\
&= 0.625.
\end{aligned}
$$

Segundo dados recentes, a percentagem de indívíduos da população portuguesa com sangue do tipo O^- é de 6%.

Qual é a probabilidade de, em 20 dadores de sangue escolhidos ao acaso de tal população, pelo menos um mas não mais de quatro dadores ter este tipo sangue?

Resolução do Exercício

- **V.a.**

  X = número de dadores com sangue do tipo O^-, em 20 escolhidos ao acaso

- **Distribuição**

  $X \sim \text{binomial}(n, p)$

  $n = 20$

  $p = P(\text{dador com sangue do tipo } O^-) = 0.06$

- **F.p. de X**

  $P(X = x) = \binom{20}{x} 0.06^x (1 - 0.06)^{20-x}, \quad x = 0, 1, \ldots, 20$

- **Probabilidade pedida**

$$
\begin{aligned}
P(1 \le X \le 4) &= \sum_{x=1}^{4} P(X = x) \\
&= \sum_{x=1}^{4} \binom{20}{x} 0.06^x (1 - 0.06)^{20-x} \\
&\simeq 0.7043.
\end{aligned}
$$

Alternativamente, $P(1 \le X \le 4) = F_x(4) - F_X(1^-) = F_X(4) - F_X(0) \overset{tabelas,calc.}{=} 0.9944 - 0.2901 = 0.7043.$

Considere que o número semanal de acidentes rodoviários, na localidade A, é uma variável aleatória de Poisson com variância igual a 3, e, na localidade B, é uma variável aleatória de Poisson com valor esperado igual a 5. Suponha que estas duas variáveis aleatórias são independentes.

Determine a probabilidade de o número total de acidentes rodoviários nessas duas localidades numa semana exceder três.

Resolução do Exercício

- **V.a.**

$X =$ número semanal de acidentes rodoviários na localidade A

$Y =$ número semanal de acidentes rodoviários na localidade B

$X \sim \text{Poisson}(\lambda_X = 3)$, pois $V(X) = 3 = \lambda_X$
$\perp\!\!\!\perp$
$Y \sim \text{Poisson}(\lambda_Y = 5)$, pois $E(Y) = 5 = \lambda_Y$

- **Importante**

A soma de duas v.a. independentes com distribuição de Poisson possui também distribuição de Poisson: $Z = X + Y \sim \text{Poisson}(\lambda_Z = \lambda_X + \lambda_Y = 3 + 5 = 8)$.

- **Probabilidade pedida**

$$P(Z > 3) = 1 - P(Z \le 3) = 1 - F_Z(3) \overset{tabelas,calc.}{=} 1 - 0.0424 = 0.9576.$$

Seja (X, Y) um par aleatório contínuo com função de densidade de probabilidade conjunta

$$f_{X,Y}(x,y) = \begin{cases} 2, & 0 \le x \le 1, 0 \le y \le x \\ 0, & \text{caso contrário.} \end{cases}$$

Deduza a função de densidade de probabilidade de X.

Resolução do Exercício

- **Par aleatório**

  (X, Y) com a f.d.p. conjunta do enunciado

- **F.d.p. marginal de** X

$$\boxed{f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, dy}$$

$$\boxed{= \int_0^x 2\, dy = 2y\big|_0^x\, dx = 2x, \quad 0 \le x \le 1.}$$

Obrigada!

Questões?