



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.º 6 e 7 (Semana 4)

**Docente:** Elisabete Fernandes

**E-mail:** efernandes@iseg.ulisboa.pt



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



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

# Conteúdos Programáticos

## Aulas 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

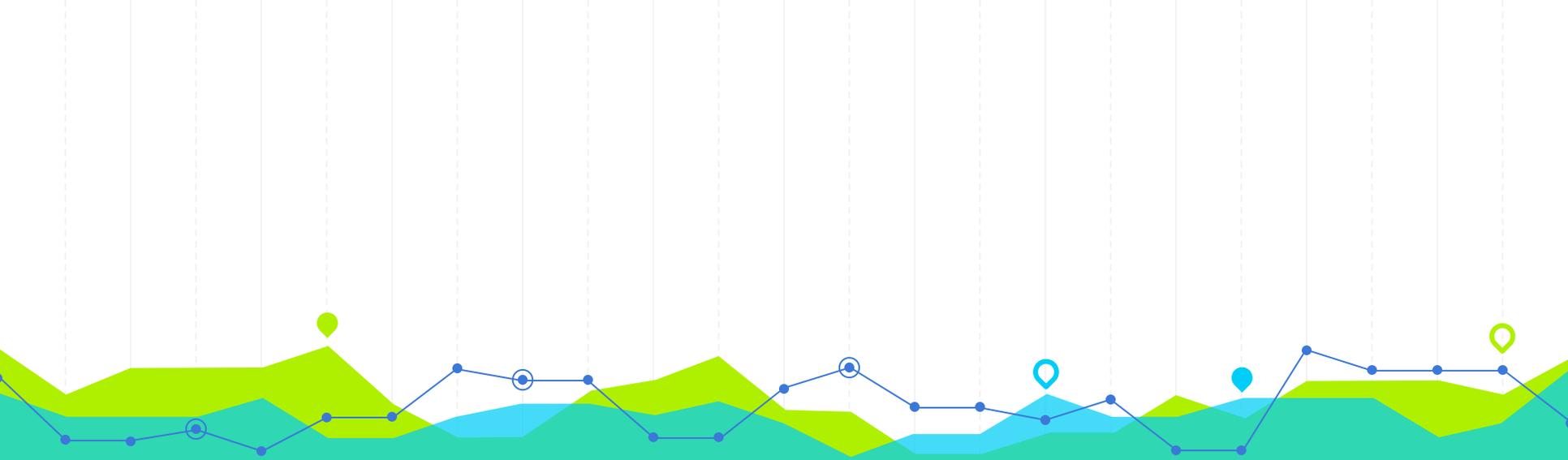
## Aulas Teórico-Práticas (Semanas 10 a 13)

- **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>



# Quociente de Variâncias Amostrais

Distribuições de Amostragem

1

# Quociente de Variâncias Amostrais

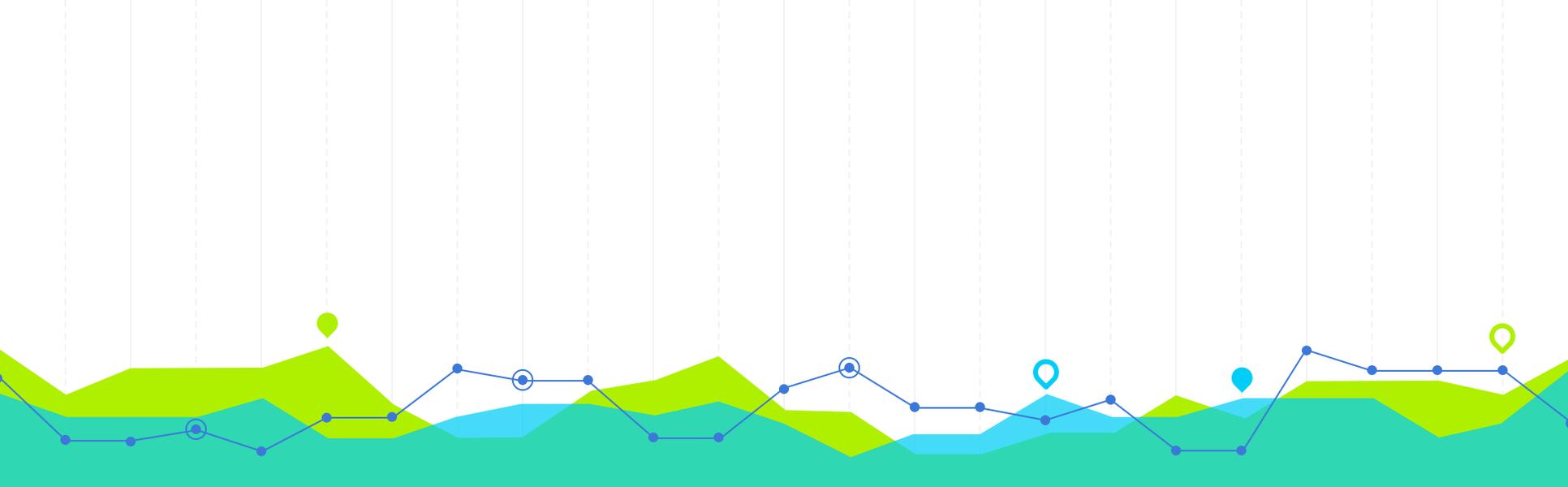
Sejam  $X_{11}, X_{12}, \dots, X_{1n_1}$  e  $X_{21}, X_{22}, \dots, X_{2n_2}$  duas a. a. independentes, de dimensão  $n_1$  e  $n_2$  retiradas de duas populações Normais, sendo  $X_1 \sim N(\mu_1; \sigma_1)$  e  $X_2 \sim N(\mu_2; \sigma_2)$ , respetivamente, e

$$S_1^2 = \sum_{i=1}^{n_1} \frac{(X_{1i} - \bar{X}_1)^2}{n_1 - 1} \quad \text{e} \quad S_2^2 = \sum_{i=1}^{n_2} \frac{(X_{2i} - \bar{X}_2)^2}{n_2 - 1},$$

então, por um teorema da distribuição  $F$ , tem-se que:

$$F = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \sim F_{n_1-1; n_2-1}.$$

[ProbabilidadesEstatistica2019.pdf](#)



# Quociente de Variâncias Amostrais: Exercícios

Distribuições de Amostragem

# 2

Uma determinada empresa farmacêutica lançou no mercado um novo medicamento, para dormir, que tem estado a ser utilizado nos hospitais. Constatou-se que os doentes não sujeitos a este medicamento em média dormiam 7,5 horas sendo o desvio padrão de 1,4 horas, enquanto os doentes aos quais se administrou este medicamento dormiam em média 8 horas sendo o desvio padrão de 2 horas.

Num determinado hospital observaram-se 31 doentes não sujeitos ao referido medicamento e 61 sujeitos ao medicamento. Qual a probabilidade da variância amostral do primeiro grupo ser inferior à do segundo grupo? Admita a normalidade dos dados.



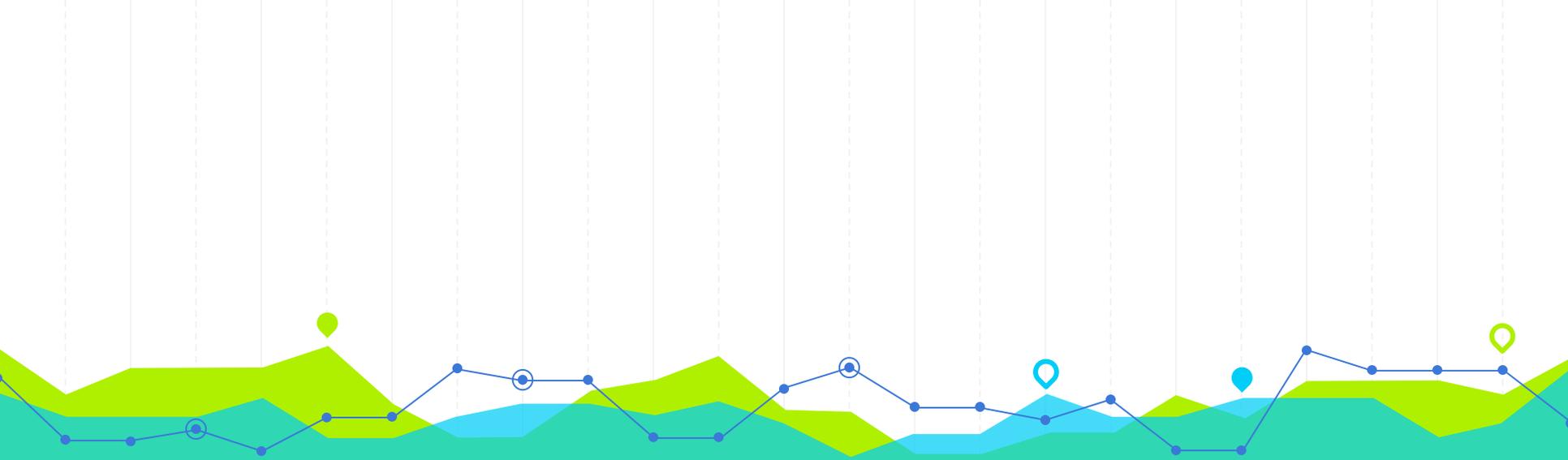
# Exercício: Quociente de Variâncias Amostrais

Sejam:

- $X_1$  a v.a. que representa o número de horas de sono dos doentes não sujeitos ao medicamento,
  - $X_2$  a v.a. que representa o número de horas de sono dos doentes sujeitos ao medicamento,
- com  $X_1 \sim N(\mu_1 = 7,5; \sigma_1 = 1,4)$  e  $X_2 \sim N(\mu_2 = 8; \sigma_2 = 2)$ .

$$\begin{array}{l} X_1 \text{ e } X_2 \text{ dist. Normal} \\ n_1 = 31 \text{ e } n_2 = 61 \end{array} \quad \left| \Rightarrow F = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \sim F_{n_1-1; n_2-1} = 30; 60 \right.$$

$$P(S_1^2 < S_2^2) = P\left(\frac{S_1^2}{S_2^2} < 1\right) = P\left(F < 1 \times \frac{2^2}{1,4^2}\right) = P(F < 2,04) \approx 0,99.$$



# Distribuições do Mínimo e Máximo

Distribuições de Amostragem

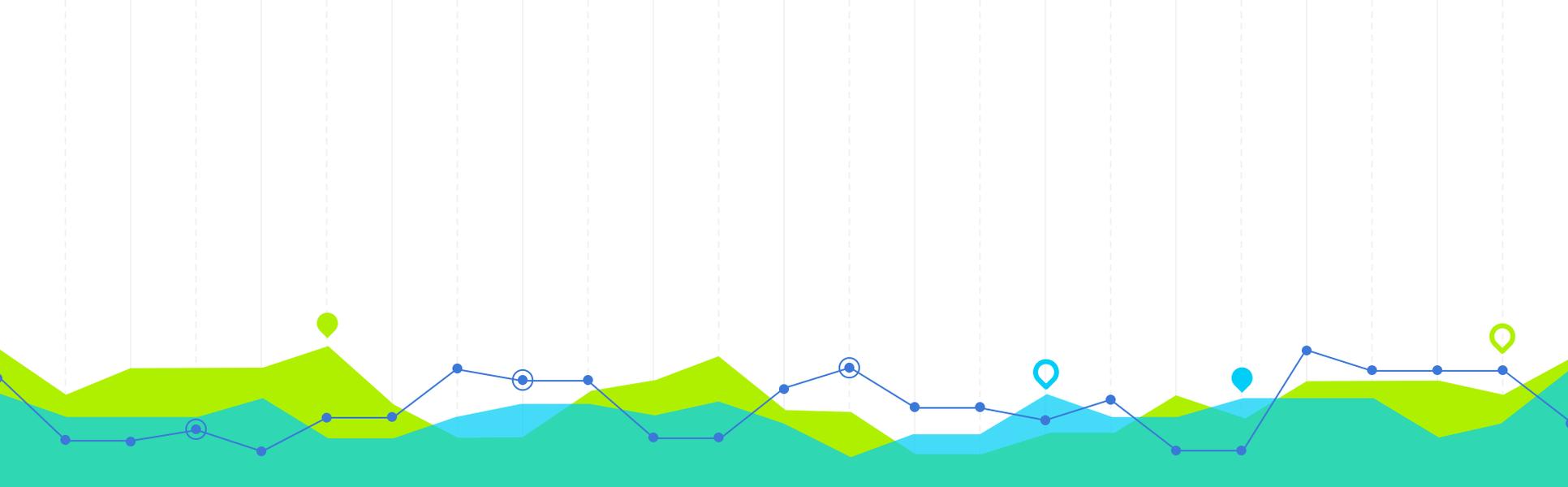
3

# Distribuições de Amostragem

## Formulário

- **DISTRIBUIÇÃO DO MÍNIMO E DO MÁXIMO**

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



# Distribuições de Amostragem: Resumo

# 4

# Distribuições de Amostragem

## 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	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \sim t(\nu)$
	$T = \frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}}}}{\sqrt{\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	<p>onde <math>\nu</math> é o maior inteiro contido em <math>r</math>,</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	$\frac{S_1'^2}{S_2'^2} \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$	

# Distribuições de Amostragem

## Formulário

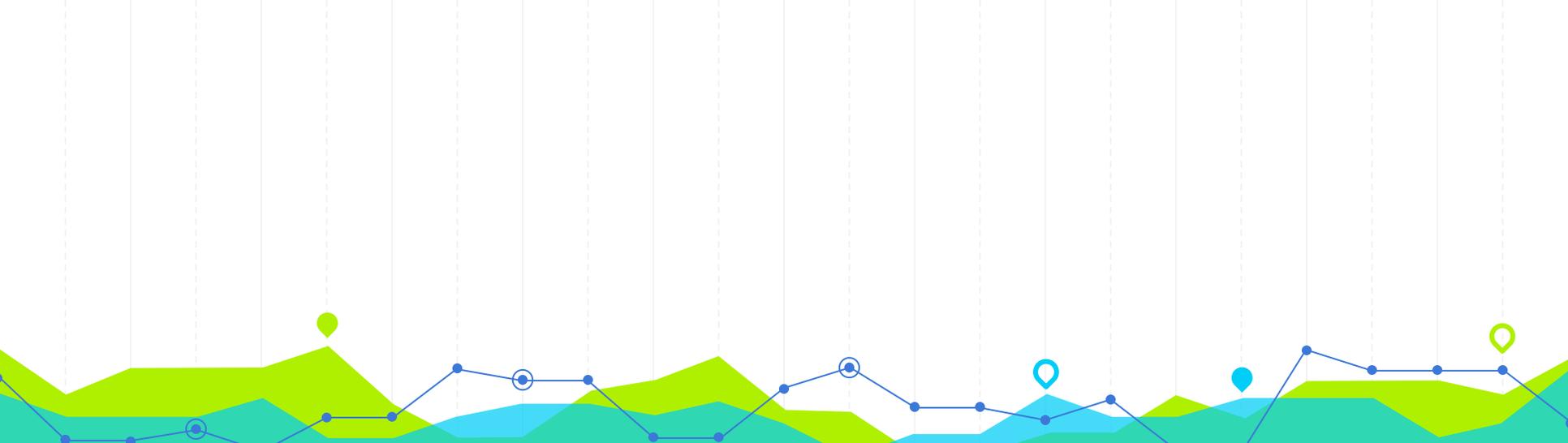
### • GRANDES AMOSTRAS

#### Caso geral

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	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \stackrel{a}{\sim} N(0,1)$

#### População de Bernoulli

Proporção	$\frac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \stackrel{a}{\sim} N(0,1)$
Diferença de proporções	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{m} + \frac{\theta_2(1-\theta_2)}{n}}} \stackrel{a}{\sim} N(0,1)$



# Teorema do Limite Central

# 5

# Distribuição Assintótica da Soma e da Média de V.A.'s: Teorema do Limite Central (TLC)

- Na maioria das situações é **difícil** determinar a distribuição da **soma de variáveis** (mesmo que sejam independentes!). O teorema seguinte **justifica a grande utilidade e importância da distribuição normal** (quer em probabilidades quer em estatística).
- O teorema que vamos ver de seguida diz-nos que, para um **grande número de v.a.'s** com idêntico comportamento (distribucional), seja ele qual for, e mesmo que seja **praticamente desconhecido, a distribuição da soma das v.a.'s aproxima-se de uma distribuição Normal**; e tem uma distribuição que é tão mais próxima da Normal, quanto maior for o número de v.a.'s da soma.

# TLC

## Teorema do Limite Central

Seja  $X_1, \dots, X_n$  uma sequência de v.a. **independentes** e **identicamente distribuídas** com valor esperado  $\mu < \infty$  e variância  $\sigma^2 < \infty$ . Considere-se  $S_n = \sum_{i=1}^n X_i$  e  $\bar{X}_n = \frac{S_n}{n}$ , então quando  $n \rightarrow +\infty$

$$\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \underset{a}{\sim} N(0, 1),$$

ou seja,

$$\lim_{n \rightarrow +\infty} P\left(\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \leq z\right) = \Phi(z),$$

onde  $\Phi(\cdot)$  é a função de distribuição da normal reduzida i.e.,  $N(0, 1)$ .

# TLC

## Teorema do Limite Central

De modo equivalente, se tem

$$\frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} = \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \underset{a}{\approx} N(0, 1),$$

ou seja,

$$\lim_{n \rightarrow +\infty} P\left(\frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} \leq z\right) = \Phi(z)$$

Slides da Professora Conceição Amado

# Valor Médio e Variância de Somas e Médias de V.A.'s Independentes

Observar que:

- $$E(S_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = nE(X_1) = n\mu$$

e como  $X_1, X_2 \dots X_n$  são v.a. independentes tem-se

$$V(S_n) = V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) = nV(X_1) = n\sigma^2$$

- $$E(\bar{X}_n) = E\left(\sum_{i=1}^n X_i/n\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X_1) = \mu$$

e como  $X_1, X_2 \dots X_n$  são v.a. independentes tem-se

$$V(\bar{X}_n) = V\left(\sum_{i=1}^n X_i/n\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n} V(X_1) = \frac{\sigma^2}{n}$$

Formulário

AMOSTRAGEM. DISTRIBUIÇÕES POR AMOSTRAGEM

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

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

# TLC: Aproximações

## Observações:

- O TLC deve ser usado **apenas** quando as v.a.'s  $X_1, \dots, X_n$  **não têm distribuição normal!**
- A demonstração do teorema exige algumas ferramentas matemáticas avançadas.
- As v.a.  $X_1, \dots, X_n$  podem ser **discretas** ou **contínuas**.
- Geralmente considera-se  $n$  grande se  $n \geq 30$
- As distribuições binomial e de Poisson podem ser aproximadas pela distribuição normal (na secção anterior vimos que podem ser escritas como somas de variáveis aleatórias). Assim:
  - $X \sim Bin(n, p)$  pode ser aproximada por  $\tilde{X} \sim N(np, np(1-p))$ , a aproximação tem menor erro quando  $np > 5$  e  $n(1-p) > 5$
  - $X \sim Poisson(\lambda)$  pode ser aproximada por  $\tilde{X} \sim N(\lambda, \lambda)$ , a aproximação tem menor erro quando  $\lambda > 5$

# Teorema de De Moivre-Laplace: Binomial - Normal

**T. Limite Central:** Aplicação a aproximações para distribuições discretas

**Corolário 5.3 (T. de De Moivre-Laplace)** – Dada a sucessão de variáveis aleatórias iid,  $X_1, X_2, \dots, X_n, \dots$ , com distribuição de Bernoulli de média  $E(X_i) = \theta$  e, portanto,  $\text{Var}(X_i) = \theta(1 - \theta)$ , tem-se,

$$\frac{\sum_{i=1}^n X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \stackrel{a}{\sim} N(0,1).$$

Nota:  $X = \sum_{i=1}^n X_i \sim B(n; \theta)$ . Quando  $n$  é grande, utilizar o corolário. Mas...

com **Correção de Continuidade**.

Murteira et al (2015)

# Aproximações Baseadas no TLC: Binomial - Normal

- Probabilidades associadas a uma distribuição Binomial,  $B(n,p)$ , podem ser aproximadas utilizando uma distribuição Normal,  $N(\mu,\sigma)$ , com  $\mu=np$  e  $\sigma = \sqrt{np(1-p)}$ .

Para que a aproximação não seja muito má, devemos ter  $np \geq 5$  e  $n(1-p) \geq 5$ .

## Aproximação da Normal para a Distribuição Binomial

- Quanto mais perto  $p$  estiver de 0,5, melhor será a aproximação da normal para a distribuição binomial
- Quanto maior o tamanho da amostra,  $n$ , melhor será a aproximação da normal para a distribuição binomial
- Regra Geral:
  - A distribuição normal pode ser utilizada para aproximar a distribuição binomial se

$$np \geq 5 \text{ e } n(1-p) \geq 5$$

# Aproximações Baseadas no TLC: Poisson - Normal

- Probabilidades associadas a uma distribuição de Poisson,  $P(\lambda)$ , podem ser aproximadas utilizando uma distribuição Normal,  $N(\mu, \sigma)$ , com  $\mu = \lambda$  e  $\sigma = \sqrt{\lambda}$ .

A aproximação será tanto melhor quanto maior for  $\lambda$ .

# TLC: Correção de Continuidade

A variante do T. L. C. para distribuições discretas introduz a correção de continuidade que permite aproximar uma distribuição discreta por uma distribuição contínua, neste caso a distribuição Normal.

**Correção de continuidade:** Seja  $X$  uma v. a. discreta, com variação  $\Delta$  (i.e.,  $X$  pode tomar os valores  $0, \Delta, 2\Delta, 3\Delta, \dots$ ), com  $E(X) = \mu_X$  e  $Var(X) = \sigma_X^2$ . Então a correção de continuidade é efetuada da seguinte forma:

$$P(X = k) \approx P\left(k - \frac{\Delta}{2} \leq X \leq k + \frac{\Delta}{2}\right).$$

# TLC: Correção de Continuidade

Note-se que nas distribuições usuais, por exemplo, Binomial e Poisson,  $\Delta = 1$ , pois tratam-se de distribuições de contagem logo

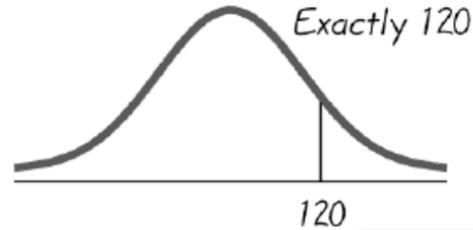
$$P(X = k) \approx P(k - 0,5 \leq X < k + 0,5).$$

Com a aplicação das propriedades da distribuição Normal, podem generalizar-se as seguintes regras:

- $P(X \leq k) \approx P\left(Z \leq \frac{k + 0,5 - \mu_X}{\sigma_X}\right) = \Phi\left(\frac{k + 0,5 - \mu_X}{\sigma_X}\right);$
- $P(X < k) \approx P\left(Z < \frac{k - 0,5 - \mu_X}{\sigma_X}\right) = \Phi\left(\frac{k - 0,5 - \mu_X}{\sigma_X}\right);$
- $P(X \geq k) \approx P\left(Z \geq \frac{k - 0,5 - \mu_X}{\sigma_X}\right) = 1 - \Phi\left(\frac{k - 0,5 - \mu_X}{\sigma_X}\right);$
- $P(X > k) \approx P\left(Z > \frac{k + 0,5 - \mu_X}{\sigma_X}\right) = 1 - \Phi\left(\frac{k + 0,5 - \mu_X}{\sigma_X}\right).$

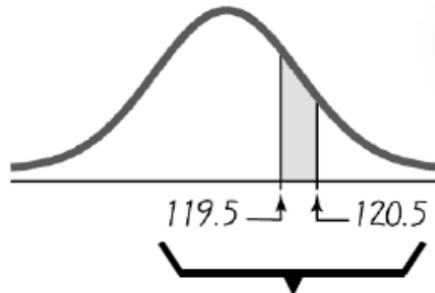
# Aproximações Baseadas no TLC: Correção de Continuidade

$x =$  exatamente 120



$$P(X = 120) \sim P(120-0,5 < \tilde{X} < 120+0,5)$$

$$P(X = a) \approx P(a - \delta < \tilde{X} < a + \delta)$$

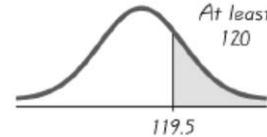


Intervalo que representa o valor discreto 120

Note-se que  $\tilde{X}$  é uma v.a. Normal que “aproxima”  $X$  (sendo esta uma v.a. discreta).

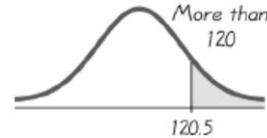
# Aproximações Baseadas no TLC: Correção de Continuidade

$X = \underline{\text{pelo menos}} 120$   
 $= 120, 121, 122, \dots$



$$P(X \geq 120) \sim P(\tilde{X} \geq 120-0,5)$$

$X = \underline{\text{mais do que}} 120$   
 $= 121, 122, 123, \dots$



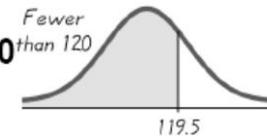
$$P(X > 120) \sim P(\tilde{X} > 120+0,5)$$

$X = \underline{\text{no máximo}} 120$   
 $= 0, 1, \dots, 118, 119, 120$



$$P(X \leq 120) \sim P(\tilde{X} \leq 120+0,5)$$

$X = \underline{\text{menos do que}} 120$   
 $= 0, 1, \dots, 118, 119$



$$P(X < 120) \sim P(\tilde{X} < 120-0,5)$$

Note-se que  $\tilde{X}$  é uma v.a. Normal que “aproxima”  $X$  (sendo esta uma v.a. discreta).

# TLC: Resumo da Correção de Continuidade

## Observação - aplicação do TLC para distribuições discretas

- **Importante:** A utilização do teorema do limite central para **distribuições discretas** apresenta um problema, uma vez que se vai aproximar um fenómeno discreto por uma distribuição contínua. Embora não exista uma solução ótima para todas as situações, na prática é comum adotar a chamada **correção de continuidade**. Considere-se  $0 < \delta < 1$  e  $\tilde{X}$  a v.a. normal que “aproxima”  $X$ , tem-se:

- $P(X = a) \approx P(a - \delta < \tilde{X} < a + \delta)$
- $P(a < X < b) \approx P(a + \delta \leq \tilde{X} \leq b - \delta)$
- $P(X \leq a) \approx P(\tilde{X} \leq a + \delta)$
- $P(X < a) \approx P(\tilde{X} < a - \delta) = P(\tilde{X} \leq a - \delta)$  (recordar que  $\tilde{X}$  é contínua...)

- No caso da aplicação do TLC à binomial e à Poisson o valor típico é  $\delta = 0.5$ , ou seja metade da variação entre dois valores consecutivos.

# TLC

## Formulário

### TEOREMA DO LIMITE CENTRAL E COROLÁRIOS

TLC: Sendo  $X_i$  iid com  $E(X_i) = \mu$  e  $\text{Var}(X_i) = \sigma^2 \Rightarrow \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n} \sigma} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\sim} N(0,1)$

Corolário: Sendo  $X_i \sim B(1; \theta)$ , iid  $\Rightarrow \frac{\sum_{i=1}^n X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \stackrel{a}{\sim} N(0,1)$

Correcção de continuidade:  $P(a \leq X \leq b) \approx \Phi\left(\frac{b + \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right)$ , com  $a$  e  $b$  inteiros

Corolário: Sendo  $X \sim \text{Po}(\lambda)$ , quando  $\lambda \rightarrow +\infty \Rightarrow \frac{X - \lambda}{\sqrt{\lambda}} \stackrel{a}{\sim} N(0,1)$

Correcção de continuidade:  $P(a \leq X \leq b) \approx \Phi\left(\frac{b + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right)$ , com  $a$  e  $b$  inteiros

# Exemplo 1: TLC

## Exemplo 1- aplicação TLC

Uma empresa de chocolates embala caixas com 100 pacotes de bombons. Os pesos por pacote são v.a.'s  $X_i$ , onde  $X_i \perp\!\!\!\perp X_j$ , com valor médio 0.5kg e variância  $0.1\text{kg}^2$ . São colocadas 10 caixas numa palete. Qual é a probabilidade aproximada do peso dos bombons da palete ser superior a 510kg?

Slides da Professora Conceição Amado

# Exemplo 1: TLC

## Resolução:

$X_i$  – 'v.a. peso do  $i$ -ésimo pacote (kg)'

$E(X_i) = 0.5$ ,  $V(X_i) = 0.1$ ,  $\forall i$  e  $X_i \perp\!\!\!\perp X_j$  para  $i \neq j$

$S = \sum_{i=1}^{1000} X_i$  'v.a. peso dos bombons colocados na paleta (kg)'

$E(S) = E\left(\sum_{i=1}^{1000} X_i\right) = \sum_{i=1}^{1000} E(X_i) = 1000E(X_1) = 1000 \times 0.5 = 500$   
(porque são identicamente distribuídas (i.d.) a  $X_1$ )

$V(S) = V\left(\sum_{i=1}^{1000} X_i\right) = \sum_{i=1}^{1000} V(X_i) = 1000V(X_1) = 1000 \times 0.1 = 100$   
porque são independentes e identicamente distribuídas (i.d.) a  $X_1$ .

Slides da Professora Conceição Amado

# Exemplo 1: TLC

**Resolução (cont.):**

Pelo T.L.C. tem-se que:

$$\frac{S - E(S)}{\sqrt{V(S)}} = \frac{S - 500}{\sqrt{100}} \stackrel{a}{\sim} N(0, 1),$$

$$P(S > 510) = 1 - P(S \leq 510) = 1 - P\left(\frac{S - 500}{\sqrt{100}} \leq \frac{510 - 500}{\sqrt{100}}\right) \stackrel{T.L.C.}{\approx} \\ \stackrel{T.L.C.}{\approx} 1 - \Phi(1) \approx 0.1587.$$

Slides da Professora Conceição Amado

## Exemplo 2: TLC

### Exemplo 2 - aplicação TLC e correção de continuidade

O número de chamadas de telemóvel registadas a partir de certa “zona” numa hora tem, em condições estacionárias, distribuição de Poisson de parâmetro 1500. Calcule a probabilidade (aproximada) de ocorrerem mais de 1600 chamadas na próxima hora.

Slides da Professora Conceição Amado

## Exemplo 2: TLC

### Resolução:

$X \sim \text{Poisson}(1500)$ , como o valor de  $\lambda$  é elevado não consta nas tabelas disponibilizadas, e sem a ajuda de um computador não é possível calcular a probabilidade pedida. Mas, podemos usar a distribuição normal para obter uma aproximação.

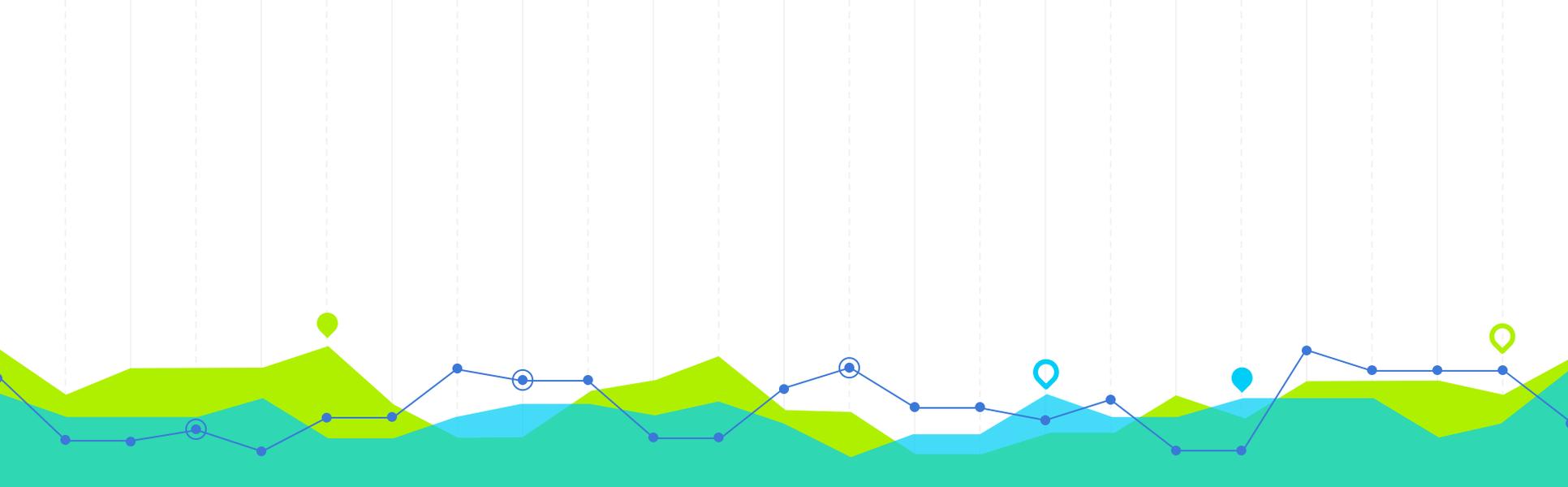
Como  $X \sim \text{Poisson}(1500)$  pode ser aproximada por  $\tilde{X} \sim N(1500, 1500)$

$$\begin{aligned} P(X > 1600) &= 1 - P(X \leq 1600) \simeq 1 - P(\tilde{X} \leq 1600 + 0.5) = \\ &= 1 - P(\tilde{X} \leq 1600.5) = 1 - P\left(\frac{\tilde{X} - 1500}{\sqrt{1500}} \leq \frac{1600.5 - 1500}{\sqrt{1500}}\right) = \\ &= 1 - \Phi(2.59) \simeq 0.0048 \end{aligned}$$

### Nota:

Lambda = 1500 > 5

Logo pode-se aproximar a distribuição Poisson à distribuição Normal



# Teorema do Limite Central: Exercícios

# 6

1. Considere que o tempo de viagem de autocarro entre Lisboa e Évora segue uma distribuição Uniforme entre 100 a 120 minutos. Numa amostra aleatória de 30 viagens, qual a probabilidade de a média dos tempos de viagem ser inferior a 112 minutos?

[ProbabilidadesEstatistica2019.pdf](#)



# Exercício 1: TLC

Seja  $X_i$  a v.a. que representa o tempo da viagem  $i$  de autocarro entre Lisboa e Évora, com  $X_i \sim U(100; 120)$ . Logo,

$$\mu = E(X_i) = \frac{100 + 120}{2} = 110;$$

$$\sigma^2 = \text{Var}(X_i) = \frac{(120 - 100)^2}{12} = \frac{100}{3}.$$

Seja  $\bar{X} = \frac{1}{30}(X_1 + \dots + X_{30})$  a v.a. que representa a média dos 30 tempos de viagem de autocarro entre Lisboa e Évora.

Pelo corolário do T. L. C.,

$$Z = \frac{\bar{X} - 110}{\sqrt{\frac{100}{3 \times 30}}} \underset{\circ}{\sim} N(0; 1).$$

Logo,

$$P(\bar{X} < 112) \approx P\left(Z < \frac{112 - 110}{\sqrt{\frac{100}{3 \times 30}}}\right) = \Phi(1,80) = 0,9641.$$

2. Seja  $X$  a v. a. que representa o tempo decorrido entre a entrada consecutiva de 2 automóveis numa autoestrada (AE), que segue uma distribuição Exponencial com valor médio 15 segundos. Numa amostra aleatória de 100 entradas consecutivas, qual a probabilidade de a soma dos tempos ser superior a 1350 segundos?

[ProbabilidadesEstatistica2019.pdf](#)



## Exercício 2: TLC

Seja  $X_i$  a v.a. que representa o tempo decorrido entre a  $i$ -ésima entrada consecutiva de 2 automóveis numa auto-estrada (AE), com  $X_i \sim \text{Exp}\left(\lambda = \frac{1}{15}\right)$ . Logo,

$$\mu = E(X_i) = \frac{1}{\frac{1}{15}} = 15 \text{ segundos,}$$

$$\sigma^2 = \text{Var}(X_i) = \frac{1}{\left(\frac{1}{15}\right)^2} = 15^2 = 225.$$

Seja  $S_{100} = X_1 + \dots + X_{100}$  a v.a. que representa a soma dos 100 decorrido entre a  $i$ -ésima entrada consecutiva de 2 automóveis numa autoestrada (AE).

Pelo T. L. C.,

$$Z = \frac{S_{100} - 100 \times 15}{\sqrt{100 \times 225}} \underset{\circ}{\sim} N(0; 1).$$

Logo,

$$P(S_{100} > 1350) = 1 - P(S_{100} \leq 1350) \approx 1 - P\left(Z < \frac{1350 - 1500}{\sqrt{100 \times 225}}\right) = 1 - \Phi(-1) = \Phi(1) = 0,8413.$$

3. Com a ingestão de um determinado fármaco para o tratamento da depressão, 20% dos doentes referem sentir efeitos secundários durante a 1ª semana de tratamento. Seja  $X$  a v. a. que representa o número de doentes que sentem os referidos efeitos secundários numa amostra de  $n$  doentes. Numa amostra de 100 doentes, calcule a probabilidade de:

- a) Pelo menos 15 terem sentido os efeitos secundários.
- b) Mais de 16 e menos de 45 terem sentido efeitos secundários.

[ProbabilidadesEstatistica2019.pdf](#)



## Exercício 3: TLC

Sabe-se que  $X \sim B(n = 100; p = 0,2)$ .

Como  $n > 50$  e  $0,1 < p < 0,9$  podemos usar a aproximação à distribuição Normal e

$$X \overset{\circ}{\sim} N(\mu = np = 20; \sigma = \sqrt{npq} = 4).$$

$$\text{a) } P(X \geq 15) = 1 - P(X < 15) \approx 1 - P\left(Z < \frac{15 - 0,5 - 20}{4}\right) = 1 - \Phi(-1,38) = \Phi(1,38) = 0,9162.$$

$$\begin{aligned} \text{b) } P(16 < X < 45) &\approx P\left(\frac{16 + 0,5 - 20}{4} < Z < \frac{45 - 0,5 - 20}{4}\right) = P(-0,88 < Z < 6,13) \\ &= \Phi(6,13) - \Phi(-0,88) = \Phi(6,13) - (1 - \Phi(0,88)) \approx 1 - (1 - 0,8106) = 0,8106. \end{aligned}$$

### Nota:

$$n \times p = 20 > 5$$

$$n \times p \times (1-p) = 16 > 5$$

Logo pode-se aproximar a distribuição Binomial à distribuição Normal

4. Seja  $X$  a v.a. que representa o número de automóveis que entram numa autoestrada (AE) num período de 30 segundos, com  $X \sim P(\lambda = 9)$ . Num período de 5 minutos, qual a probabilidade de:

a) Entrarem mais de 59 automóveis na AE?

b) Entrarem no mínimo 55 e no máximo 80 automóveis na AE?

[ProbabilidadesEstatistica2019.pdf](#)



## Exercício 4: TLC

**Nota:**

5 minutos = 300 segundos

Para 30 segundos, tem-se  $\lambda = 9$ , então para 300 segundos tem-se  $\lambda = 90$

Seja  $X'$  a v. a. que representa o número de automóveis que entram numa AE num período de 5 minutos (= 300 segundos =  $30 \times 10$  segundos), com  $X' \sim P(\lambda' = 10\lambda = 90)$ .

Como  $\lambda > 20$ , podemos usar a aproximação à distribuição Normal e

$$X' \overset{\circ}{\sim} N(\mu = \lambda' = 90; \sigma = \sqrt{\lambda'} = \sqrt{90}).$$

$$\begin{aligned} \text{a) } P(X' > 59) &= 1 - P(X' \leq 59) \approx 1 - P\left(Z \leq \frac{59 + 0,5 - 90}{\sqrt{90}}\right) = 1 - P(Z \leq -3,21) = 1 - \Phi(-3,21) \\ &= \Phi(3,21) = 0,9993. \end{aligned}$$

$$\begin{aligned} \text{b) } P(55 \leq X' \leq 80) &\approx P\left(\frac{55 - 0,5 - 90}{\sqrt{90}} \leq Z \leq \frac{80 + 0,5 - 90}{\sqrt{90}}\right) = P(-3,74 \leq Z \leq -1,00) \\ &= \Phi(-1,00) - \Phi(-3,74) = (1 - \Phi(1,00)) - (1 - \Phi(3,74)) = \Phi(3,74) - \Phi(1,00) \\ &\approx 0,9999 - 0,8413 = 0,1586. \end{aligned}$$

**Nota:**

$\lambda = 90 > 5$

Logo pode-se aproximar a distribuição Poisson à distribuição Normal

**5.20** O tempo (em horas) que João Pestana dorme por noite é uma variável aleatória com distribuição uniforme no intervalo  $(7,12)$ .

- (a) Calcule a probabilidade de João Pestana dormir mais de 11 horas numa noite.
- (b) Calcule a probabilidade de, em 20 noites, João Pestana dormir mais de 11 horas em pelo menos 3 dessas noites.
- (c) Qual a probabilidade de João Pestana dormir mais de 1100 horas em 100 noites?



# Exercício 5.20 (a)

- **V.a. de interesse**

$X$  = tempo (em horas) que JP dorme por noite

- **Distribuição de  $X$**

$X \sim \text{Uniforme}(a, b)$

com

$$a = 7$$

$$b = 12$$

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

$$f_X(x) \stackrel{\text{form}}{=} \begin{cases} \frac{1}{12-7} = 0.2, & x \in (7, 12) \\ 0, & \text{caso contrário} \end{cases}$$

- **Prob. pedida**

$$\begin{aligned} P(X > 11) &= \int_{11}^{+\infty} f_X(x) dx \\ &= \int_{11}^{12} 0.2 dx + \int_{12}^{+\infty} 0 dx \\ &= (0.2x) \Big|_{11}^{12} \\ &= 0.2. \end{aligned}$$

## Exercício 5.20 (a)

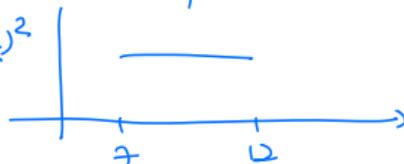
$X =$  Tempo (em horas)  $\bar{y}$  de uma nave noite  
uniforme  $(7, 12)$

$$f_X(x) = \frac{1}{b-a}, a \leq x \leq b$$

$$E(X) = \frac{b+a}{2} \quad \text{Var}(X) = \frac{(b-a)^2}{12}$$

$$f_X(x) = \begin{cases} \frac{1}{5} & x \in (7, 12) \\ 0 & \text{c. e.} \end{cases}$$

$$E(x) = \frac{7+12}{2} = 9.5 \quad \text{var}(x) = \frac{(12-7)^2}{12}$$
$$= \frac{25}{12}$$



$$a) P(X > 11) = \int_{11}^{12} \frac{1}{5} dx = \frac{1}{5}$$

# Exercício 5.20 (b)

- **V.a. de interesse**

$Y$  = no. de noites em que JP dorme mais de 11 horas, num total de 20 noites

- **Distribuição de  $Y$**

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

com

$$n = 20$$

$$p = P(X > 11)$$

$$\stackrel{(a)}{=} 0.2$$

- **Ep. de  $Y$**

$$P(Y = y) = \binom{20}{y} 0.2^y (1 - 0.2)^{20-y}, \quad y = 0, 1, 2, \dots, 20$$

- **Prob. pedida**

$$\begin{aligned} P(Y \geq 3) &= 1 - P(Y < 3) \\ &= 1 - P(Y \leq 2) \\ &= 1 - F_{\text{binomial}(20, 0.2)}(2) \\ &\stackrel{\text{tabela/calc}}{=} 1 - 0.2061 \\ &= 0.7939 \end{aligned}$$

# Exercício 5.20 (c)

- **V.a.**

$X_i$  = tempo (em horas) que JP dorme na noite  $i$ ,  $i = 1, \dots, n$   
 $n = 100$

- **Distribuição, valor esperado e variância comuns**

$X_i \stackrel{i.i.d.}{\sim} X$ ,  $i = 1, \dots, n$

$$E(X_i) = E(X) = \mu \stackrel{form}{=} \frac{a+b}{2} = \frac{7+12}{2} = 9.5, \quad i = 1, \dots, n$$

$$V(X_i) = V(X) = \sigma^2 \stackrel{form}{=} \frac{(b-a)^2}{12} = \frac{(12-7)^2}{12} = \frac{25}{12}, \quad i = 1, \dots, n$$

- **V.a. de interesse**

$S_n = \sum_{i=1}^n X_i$  = tempo (em horas) que JP dorme em  $n$  noites

- **Valor esperado e variância de  $S_n$**

$$E(S_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) \stackrel{X_i \sim X}{=} n E(X) = n \mu$$

$$V(S_n) = V\left(\sum_{i=1}^n X_i\right) \stackrel{X_i \text{ indep.}}{=} \sum_{i=1}^n V(X_i) \stackrel{X_i \sim X}{=} n V(X) = n \sigma^2$$

- **Distribuição aproximada de  $S_n$**

Pelo teorema do limite central (TLC) pode escrever-se

$$\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \approx \text{Normal}(0, 1).$$

- **Valor aproximado da probabilidade pedida**

Atendendo a que  $n\mu = 100 \times 9.5 = 950$  e  $n\sigma^2 = 100 \times \frac{25}{12} = \frac{625}{3}$ , temos

$$\begin{aligned} P(S_n > 1100) &= 1 - P(S_n \leq 1100) \\ &= 1 - P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq \frac{1100 - n\mu}{\sqrt{n\sigma^2}}\right) \\ &\stackrel{TLC}{\approx} 1 - \Phi\left(\frac{1100 - 950}{\sqrt{\frac{625}{3}}}\right) \\ &\approx 1 - \Phi(10.39) \\ &< 1 - \Phi(4.09) \\ &\stackrel{tabela/calc.}{\approx} 1 - 0.999978 \\ &= 0.000022. \end{aligned}$$

- **[Obs.**

Tirando partido do facto da f.d.  $\Phi(z)$  ser estritamente crescente e do “último” valor da tabela da f.d. da normal padrão, podemos adiantar que:

- $10.39 > 4.09 \Rightarrow \Phi(10.39) > \Phi(4.09) \stackrel{tabela/calc.}{\approx} 0.999978;$
- $-10.39 < -4.09 \Rightarrow \Phi(-10.39) < \Phi(-4.09) = 1 - \Phi(4.09) \stackrel{tabela/calc.}{\approx} 1 - 0.999978 = 0.000022.]$

1. A rede informática interna de um banco funciona permanentemente. Considere que o tempo (em minuto) de utilização desta rede por um utilizador autenticado é uma variável aleatória  $X$  com função de distribuição dada por:

$$F_X(x) = \begin{cases} 1 - e^{-\sqrt{\frac{x}{30}}}, & x > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(s) ds$$

- (a) Obtenha a probabilidade de o tempo de utilização  $X$  ser superior a 30 minutos e inferior a 3 horas. (

- (b) Tendo em conta que  $E(X) = 150$  e  $V(X) = 112500$ , obtenha um valor aproximado para a (3.0) probabilidade de o tempo médio de utilização da rede por 36 utilizadores autenticados ser superior a 2 horas.



## Exercício 1 (a)

$$\begin{aligned} P(30 < X < 3 \times 60 = 180) &= F_X(180) - F_X(30) \\ &= 1 - e^{-\sqrt{\frac{180}{75}}} - 1 + e^{-\sqrt{\frac{30}{75}}} = e^{-\sqrt{\frac{30}{75}}} - e^{-\sqrt{\frac{180}{75}}} \end{aligned}$$

## Exercício 1 (b)

$X_i$  = tempo de utilização de rede do  $i$ -ésimo utilizador

$$P\left(\frac{\sum_{i=1}^{36} X_i}{36} > 120\right) = P\left(\sum_{i=1}^{36} X_i > 120 \times 36\right)$$

Porém  $X_i$  não está em nenhuma das situações que acabámos de descrever.

Como fazer?

Resposta: Teorema do Limite Central

# Teorema do Limite Central

Teorema: Seja  $\{X_i, i=1, \dots, n\}$  um conjunto de v.c. independentes e com a mesma distribuição, que não pode ser a distribuição é normal. Então

$$P\left(\sum_{i=1}^n X_i \leq x\right) \xrightarrow{n \rightarrow \infty} \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

onde  $\mu = E(X_i)$  e  $\sigma^2 = \text{Var}(X_i)$ ,  $\forall i$ .

$$\Leftrightarrow \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \overset{a}{\sim} \mathcal{N}(0,1)$$

# Exercício 1 (b)

Voltando ao exercício:

- $X_i$  = tempo de utilização de rede pelo  $i$ -ésimo utilizador
- $X_i \sim \text{CV}(\ )$ ;  $E(X_i) = 150$ ;  $\text{var}(X_i) = 112500$   
 $= \mu$   $= \sigma^2$
- $P\left(\sum_{i=1}^{36} X_i > 120 \times 36\right) =$   
 $= 1 - P\left(\sum_{i=1}^{36} X_i \leq 120 \times 36\right) = \textcircled{*}$
- $n = 36$  [ Para efeitos práticos, podemos utilizar o Teorema do Limite Central para  $n \geq 30$  ]  
(TLC)

$$\textcircled{*} \approx 1 - \Phi\left(\frac{120 \times 36 - \overbrace{36 \times 150}^{n\mu}}{\underbrace{\sqrt{112500}}_{\sigma\sqrt{n}} \sqrt{36}}\right)$$

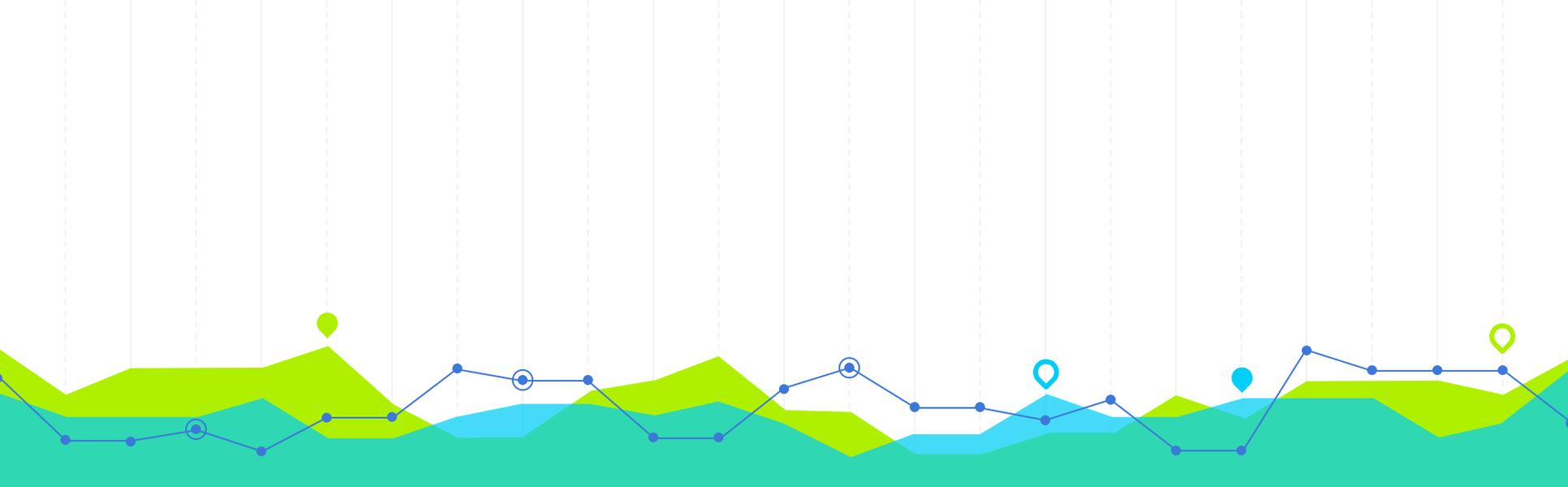
## Exercício 1 (b)

NOTA : se  $X_i \sim \text{CN}(\mu, \sigma^2)$  então não se pode utilizar o Teorema do Limite Central porje em vez de termos:

$$P\left(\sum_{i=1}^n X_i \leq x\right) \xrightarrow{n \rightarrow \infty} \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

Vamos ter

$$P\left(\sum_{i=1}^n X_i \leq x\right) = \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$



# Distribuições de Amostragem e TLC: Exercícios

Murteira et al (2015)

# 7

1. Em determinada universidade verifica-se que 30% dos alunos têm carro. Selecciona-se uma amostra casual de 20 alunos.
  - a) Qual a probabilidade de o primeiro aluno escolhido ter carro, e os outros, não?
  - b) Qual a probabilidade de, no máximo, 10 alunos terem carro?
  - c) Calcule o valor esperado e a variância da proporção de alunos, na amostra, com carro.



## Exercício 1 a)

$$X = \begin{cases} 1, & \text{se aluno tem carro} \\ 0, & \text{caso contrário} \end{cases} \rightarrow X \sim B(1, \theta), \text{ onde } \theta = 0.3$$

Amostra casual:  $(X_1, X_2, \dots, X_{20})$ ,  $n = 20$ , onde  $X_i \sim B(1, 0.3)$  ( $i = 1, 2, \dots, 20$ )

---

$P(X_1=1, X_2=0, X_3=0, \dots, X_{20}=0)$  (a)

$$P = P(X_1=1) P(X_2=0) P(X_3=0) \dots P(X_{20}=0) = P(X=1) [P(X=0)]^{19} = 0.3 \times 0.7^{19} \approx 0.00034$$

## Exercício 1 b)

$\sum_{i=1}^{20} X_i$  - v.a. n.º alunos com carro na amostra de 20 alunos

É uma estatística  $\rightarrow$  Distribuição por amostragem:  $\sum_{i=1}^{20} X_i \sim B(20, 0.3)$

$$P\left(\sum_{i=1}^{20} X_i \leq 10\right) \approx 0.9829$$

## Exercício 1 c)

Proporção de alunos, na amostra, com carro  $\rightarrow \frac{\sum_{i=1}^{20} X_i}{20} = \bar{X}$

$$\bullet E(\bar{X}) = E\left(\frac{\sum_{i=1}^{20} X_i}{20}\right) = \frac{1}{20} E\left(\sum_{i=1}^{20} X_i\right) = \frac{1}{20} \sum_{i=1}^{20} \overbrace{E(X_i)}^{\theta = 0.3} = \frac{1}{20} \sum_{i=1}^{20} 0.3 = \frac{1}{20} \times 20 \times 0.3 = 0.3$$

$$\bullet \text{Var}(\bar{X}) = \text{Var}\left(\frac{\sum_{i=1}^{20} X_i}{20}\right) = \left(\frac{1}{20}\right)^2 \text{Var}\left(\sum_{i=1}^{20} X_i\right) \overset{X_i \text{ iid.}}{=} \frac{1}{20^2} \sum_{i=1}^{20} \overbrace{\text{Var}(X_i)}^{\theta(1-\theta)} = \frac{1}{20^2} \sum_{i=1}^{20} 0.3 \times 0.7 =$$

$$= \frac{1}{20^2} \times 20 \times 0.3 \times 0.7 = \frac{0.21}{20} = 0.0105$$

$$\text{Conclusão: } \begin{cases} E(\bar{X}) = E(X) = \theta \\ \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n} = \frac{\theta(1-\theta)}{n} \end{cases}$$

# Obrigada!

Questões?

