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 22 e 23 (Semana 12)

**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<br>• **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.<br><br>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.<br><br>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.<br><br>Inferência sobre o rácio das variâncias de duas populações normais. |

# 1 Teorema do Limite Central

# 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 $X_1, \ldots, X_n$ 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 $n \to +\infty$

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

ou seja,

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

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

Slides da Professora Conceição Amado

# TLC

**Teorema do Limite Central**

De modo equivalente, se tem

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

ou seja,

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

Slides da Professora Conceição Amado

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

**Observar que:**

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

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

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

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

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

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

Formulário

**AMOSTRAGEM. DISTRIBUIÇÕES POR AMOSTRAGEM**

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

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

# TLC: Aproximações

**Observações:**

- O TLC deve ser usado apenas quando as v.a.'s $X_1, \ldots, X_n$ não têm distribuição normal!
- A demonstração do teorema exige algumas ferramentas matemáticas avançadas.
- As v.a. $X_1, \ldots, X_n$ 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 $\lambda > 5$

Slides da Professora Conceição Amado

# 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, $X_1, X_2, \ldots, X_n, \ldots$, com distribuição de Bernoulli de média $E(X_i) = \theta$ e, portanto, $\mathrm{Var}(X_i) = \theta(1-\theta)$, tem-se,

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

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(μ,σ), 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

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

SlidesBioestatística6 (ua.pt)

# Aproximações Baseadas no TLC: Poisson - Normal

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

  A aproximação será tanto melhor quanto maior for λ.

SlidesBioestatística6 (ua.pt)

# 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 $0, \Delta, 2\Delta, 3\Delta, \ldots$), com $E(X) = \mu_X$ e $Var(X) = \sigma_X^2$. Então a correção de continuidade é efetuada da seguinte forma:

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

ProbabilidadesEstatistica2019.pdf

# 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

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

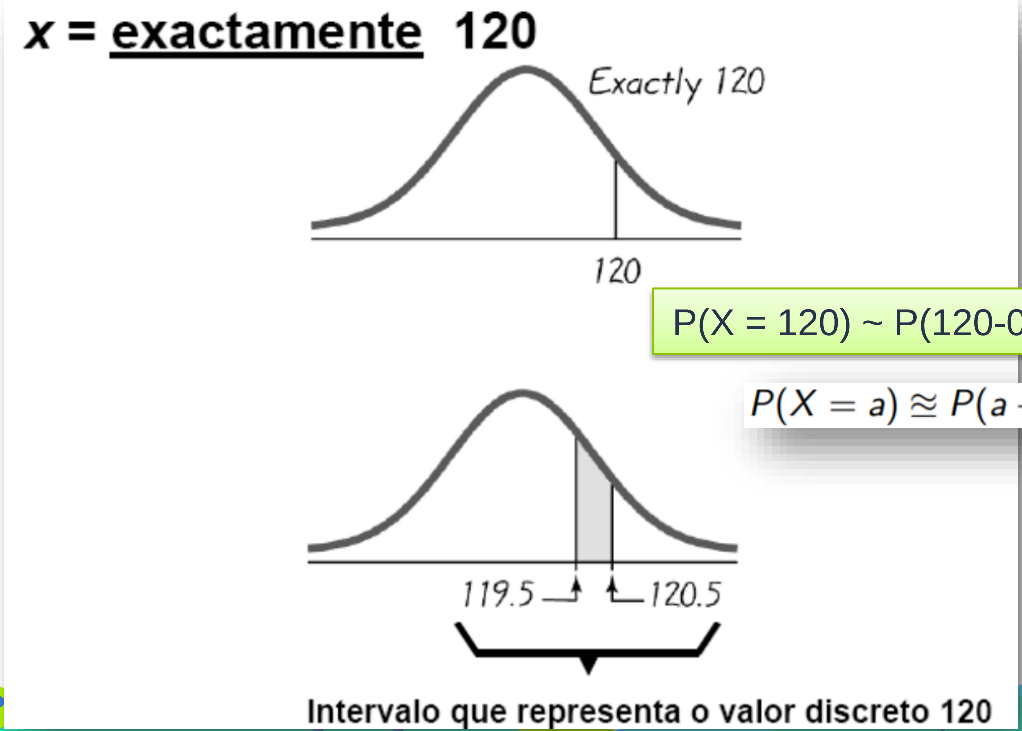
Com a aplicação das propriedades da distribuição Normal, podem generalizar-se as seguintes regras:

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

ProbabilidadesEstatistica2019.pdf

# Aproximações Baseadas no TLC: Correção de Continuidade

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



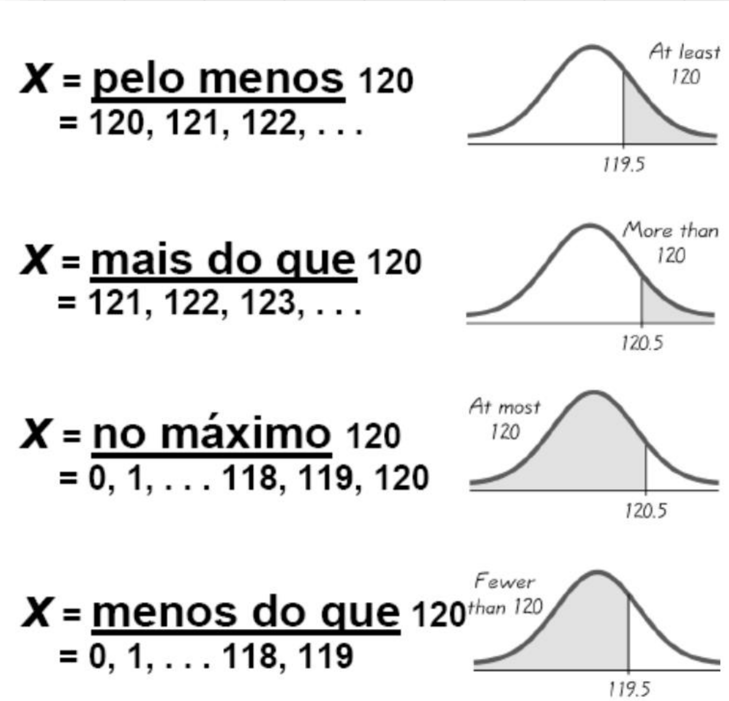

P(X = 120) ~ P(120-0,5 < $\tilde{X}$ < 120+0,5)

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

SlidesBioestatística6 (ua.pt)

# Aproximações Baseadas no TLC: Correção de Continuidade

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




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

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

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

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

# 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:
  - $P(X = a) \approxeq P(a - \delta < \tilde{X} < a + \delta)$
  - $P(a < X < b) \approxeq P(a + \delta \leq \tilde{X} \leq b - \delta)$
  - $P(X \leq a) \approxeq P(\tilde{X} \leq a + \delta)$
  - $P(X < a) \approxeq P(\tilde{X} < a - \delta) = P(\tilde{X} \leq a - \delta)$ (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.

Slides da Professora Conceição Amado

# TLC

Formulário

**TEOREMA DO LIMITE CENTRAL E COROLÁRIOS**

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

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

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

Corolário: Sendo $X \sim \mathrm{Po}(\lambda)$, quando $\lambda \to +\infty \Rightarrow \dfrac{X-\lambda}{\sqrt{\lambda}} \overset{a}{\sim} N(0,1)$

Correcção de continuidade: $P(a \le X \le b) \approx \Phi\left(\dfrac{b+\frac{1}{2}-\lambda}{\sqrt{\lambda}}\right) - \Phi\left(\dfrac{a-\frac{1}{2}-\lambda}{\sqrt{\lambda}}\right)$, 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 palete (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:

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

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

$$\underset{T.L.C.}{\approx} 1 - \Phi(1) \approx 0.1587.$$

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

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

**Nota**:
Lambda = 1500 > 5
Logo pode-se aproximar a distribuição Poisson à distribuição Normal

Slides da Professora Conceição Amado

# Teorema do Limite Central: Exercícios

# 2

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,

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

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

Seja $\overline{X} = \frac{1}{30}(X_1 + \cdots + X_{30})$ a v.a. que representa a média dos 30 tempos de viagem de autocarro entre Lisboa e Évora.
Pelo corolário do T. L. C.,

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

Logo,

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

ProbabilidadesEstatistica2019.pdf

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 $X_i \sim \; Exp\left(\lambda = \frac{1}{15}\right)$. Logo,

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

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

Seja $S_{100} = X_1 + \cdots + X_{100}$ 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.,

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

Logo,

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

ProbabilidadesEstatistica2019.pdf

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

**Nota**:
n×p = 20 > 5
n×p×(1-p) = 16 > 5
Logo pode-se aproximar a distribuição Binomial à distribuição Normal

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

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

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

b) $P(16 < X < 45) \approx P\left(\frac{16 + 0{,}5 - 20}{4} < Z < \frac{45 - 0{,}5 - 20}{4}\right) = P(-0{,}88 < Z < 6{,}13)$

$$= \Phi(6{,}13) - \Phi(-0{,}88) = \Phi(6{,}13) - \big(1 - \Phi(0{,}88)\big) \approx 1 - (1 - 0{,}8106) = 0{,}8106.$$

ProbabilidadesEstatistica2019.pdf

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 × 10 segundos), com $X' \sim P(\lambda' = 10\lambda = 90)$.

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

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

**Nota**:
Lambda = 90 > 5
Logo pode-se aproximar a distribuição Poisson à distribuição Normal

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

$= \Phi(3,21) = 0,9993.$

b) $P(55 \leq X' \leq 80) \approx P\left(\frac{55 - 0,5 - 90}{\sqrt{90}} \leq Z \leq \frac{80 + 0,5 - 90}{\sqrt{90}}\right) = P(-3,74 \leq Z \leq -1,00)$

$= \Phi(-1,00) - \Phi(-3,74) = \left(1 - \Phi(1,00)\right) - \left(1 - \Phi(3,74)\right) = \Phi(3,74) - \Phi(1,00)$

$\approx 0,9999 - 0,8413 = 0,1586.$

ProbabilidadesEstatistica2019.pdf

2º semestre – 2018/2019
04/05/2019 – **9:00**

**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:

$$F_X(x) = \begin{cases} 1 - e^{-\sqrt{\frac{x}{75}}}, & x > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

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

(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 $\overbrace{E(X)}^{\mu} = 150$ e $\overbrace{V(X)}^{\sigma^2} = 112500$, obtenha um valor aproximado para a probabilidade de o tempo médio de $\underbrace{\text{utilização da rede por}}_{X_i \ (i\text{-ésimo})}$ 36 utilizadores autenticados ser superior a 2 horas. (3.0)

# Exercício 1 (a)

$$P(30 < X < 3 \times 60 = 180) = F_X(180) - F_X(30)$$

$$= 1 - e^{-\sqrt{\frac{180}{75}}} - 1 + e^{-\sqrt{\frac{30}{75}}} = e^{-\sqrt{\frac{30}{75}}} - e^{-\sqrt{\frac{180}{75}}}$$

# Exercício 1 (b)

$X_i$ = tempo de utilização da 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.a. independentes e com a mesma distribuição, que não pode ser a distribuição normal. Então

$$\boxed{P\left(\sum_{i=1}^{n} X_i \le x\right) \xrightarrow[n\to\infty]{} \Phi\left(\frac{x-n\mu}{\sigma\sqrt{n}}\right)}$$

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

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

# Exercício 1 (b)

Voltando ao exercício:

- $X_i$ = tempo de utilização da rede pelo i-ésimo utilizador
- $X_i \sim ?(\ )$ ; $\underbrace{E(X_i)}_{=\mu} = 150$; $\underbrace{var(X_i)}_{=\sigma^2} = 112500$
- $P\left(\sum_{i=1}^{36} X_i > 120 \times 36\right) =$

$$= 1 - P\left(\sum_{i=1}^{36} X_i \le 120 \times 36\right) = (*)$$

- $n = 36$ [Para efeitos práticos, podemos utilizar o Teorema do Limite Central (TLC) para $n \geq 30$]

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

Slides Professora Claúdia Nunes

# Exercício 1 (b)

NOTA: se $X_i \sim N(\mu, \sigma^2)$ então não se pode utilizar o Teorema do Limite Central porque em vez de termos:

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

Vamos ter

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

Slides Professora Claúdia Nunes

**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 $X$**

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

- **Prob. pedida**

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

Slides Professora Sofia Naique

# Exercício 5.20 (a)

X = Tempo (em horas) q dorme numa noite

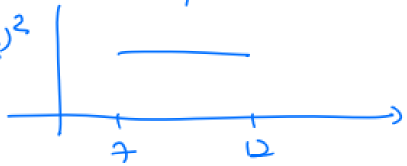
Unif (7, 12)

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

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

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

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




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

Slides Professora Claúdia Nunes

# 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

$$\begin{aligned} n &= 20 \\ p &= P(X > 11) \\ &\overset{(a)}{=} 0.2 \end{aligned}$$

- **F.p. de $Y$**

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

- **Prob. pedida**

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

Slides Professora Sofia Naique

# Exercício 5.20 (c)

- **V.a.**

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

  $n = 100$

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

  $X_i \overset{i.i.d.}{\sim} X, \quad i = 1, \ldots, n$

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

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

- **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 $S_n$**

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

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

- **Distribuição aproximada de $S_n$**

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

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

- **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

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

- **[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:

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

# Obrigada!

## Questões?