

ANÁLISE DE DADOS

REVISÃO I

Prof. Dr. Daniel Caetano

2020 - 1



BASES DA ANÁLISE DE DADOS

Análise de Dados - Conceitos

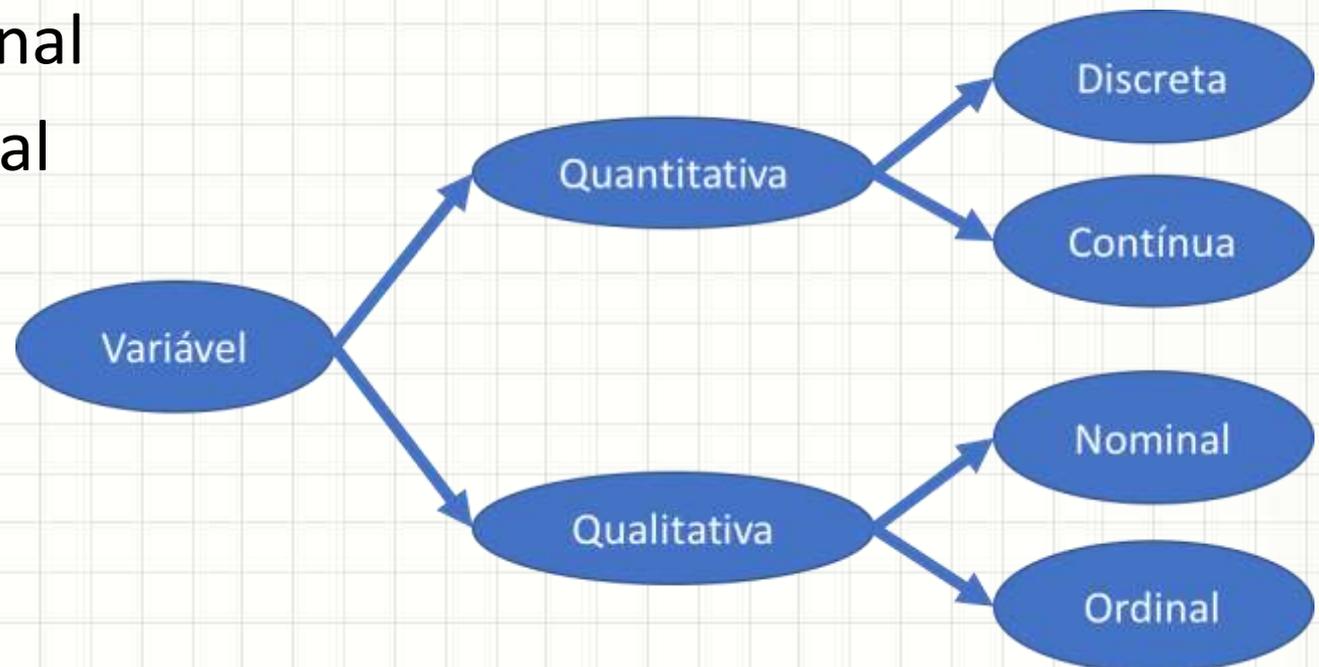
- Há três etapas fundamentais
 - Coleta de dados (Amostragem)
 - Resumo e organização (Estatística Descritiva)
 - Modelo explicativo (Estatística Indutiva)
- Elementos de Estudo
 - Unidade Experimental ou de Observação
 - População ou Universo Estatístico
 - Amostra

Amostragem

- Deve ser representativa – não pode ter viés
- Dois tipos básicos de amostragem
 - Probabilística
 - Amostragem aleatória
 - Amostragem sistemática
 - Amostragem por Conglomerados
 - Amostragem estratificada
 - Não probabilística
 - Amostragem a Esmo
 - Amostragem Intencional

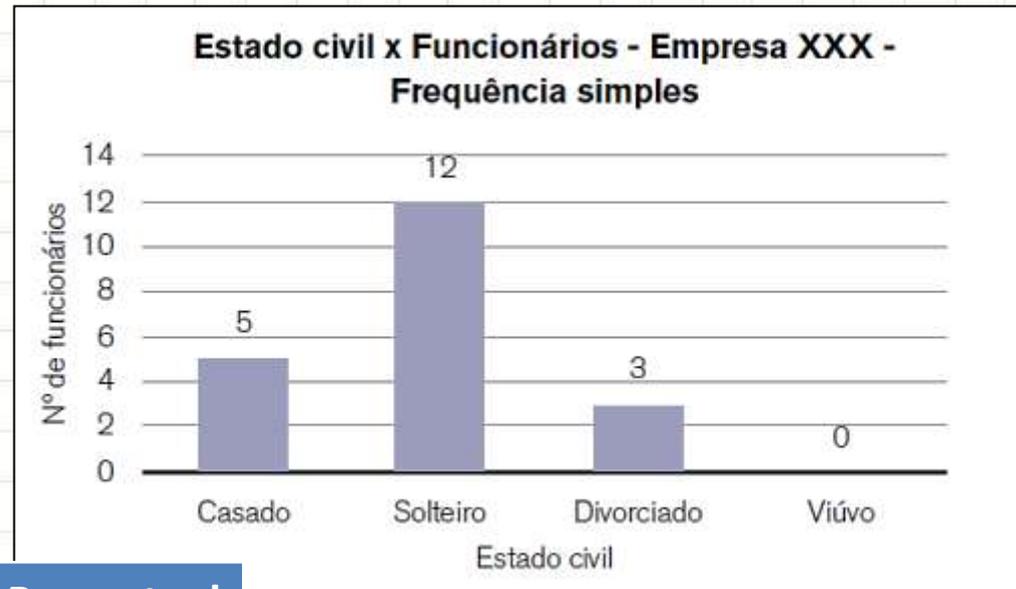
Tipos de Variáveis

- Variáveis Quantitativas
 - Contínua
 - Discreta
- Variáveis Qualitativas
 - Nominal
 - Ordinal



Visualização de Dados

- Existem duas formas clássicas
 - Tabelas
 - Gráficos



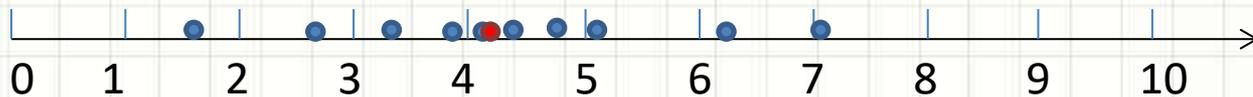
Estado Civil	Nº de Funcionários	Percentual
Casado	5	25%
Solteiro	12	60%
Divorciado	3	15%
Viúvo	0	0%
Total	20	100%



MEDIDAS DE POSIÇÃO E SEPARATRIZES

Medidas de Posição Básicas

- Média
 - Valor “mais próximo de todos”



$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

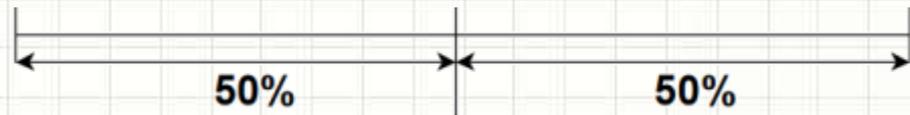
$$\bar{x}_p = \frac{p_1 \cdot x_1 + p_2 \cdot x_2 + \dots + p_n \cdot x_n}{p_1 + p_2 + \dots + p_n}$$

- Moda
 - Valor “mais frequente”
 - Valor que aparece mais vezes
 - Amodal, unimodal, bimodal, trimodal...

Separatrizes

- Dividem a amostra em grupos de elementos

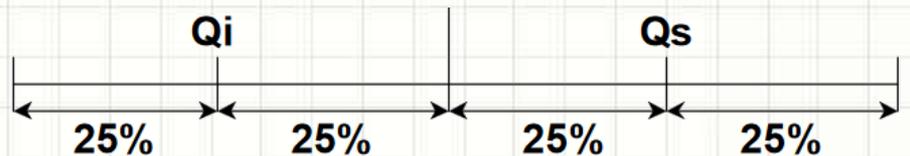
– Mediana



$$P_i = \frac{(n + 1)}{2}$$

2, 3, 4, 4, 5, 5, 6, 6, 6, 7, 8

– Quartis



$$P_{q_{inf}} = \frac{(n + 1)}{4}$$

$$P_i = \frac{2 \cdot (n + 1)}{4}$$

$$P_{q_{sup}} = \frac{3 \cdot (n + 1)}{4}$$

– Decis

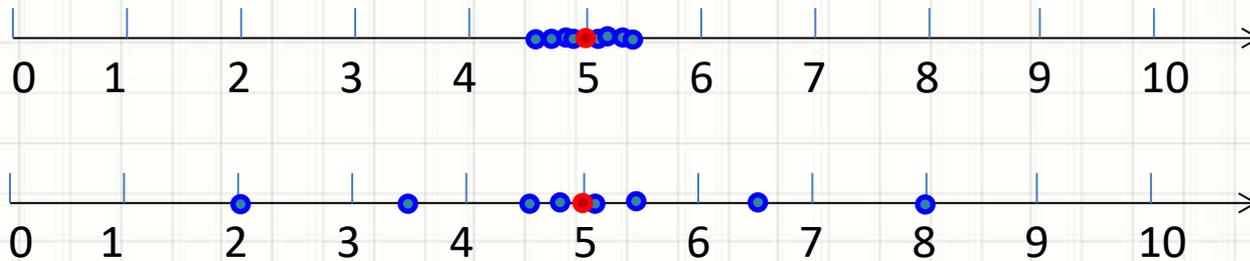
– Centis ou Percentis



MEDIDAS DE DISPERSÃO

Medidas de Dispersão

- Compare as diferentes amostras de média 5:



- Medidas mais comuns
 - Variância: “erro quadrático médio”
 - Desvio Padrão: “erro médio”
 - Coeficiente de Dispersão: “taxa de erro médio”

Cálculo da Variância

- Variância Amostral

$$Var[x] = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

- Variância Populacional

$$Var[x] = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

- Variância Pop. com Frequências Relativas

$$Var[x] = (x_1 - \bar{x})^2 \cdot fr_1 + (x_2 - \bar{x})^2 \cdot fr_2 + \dots + (x_n - \bar{x})^2 \cdot fr_n$$

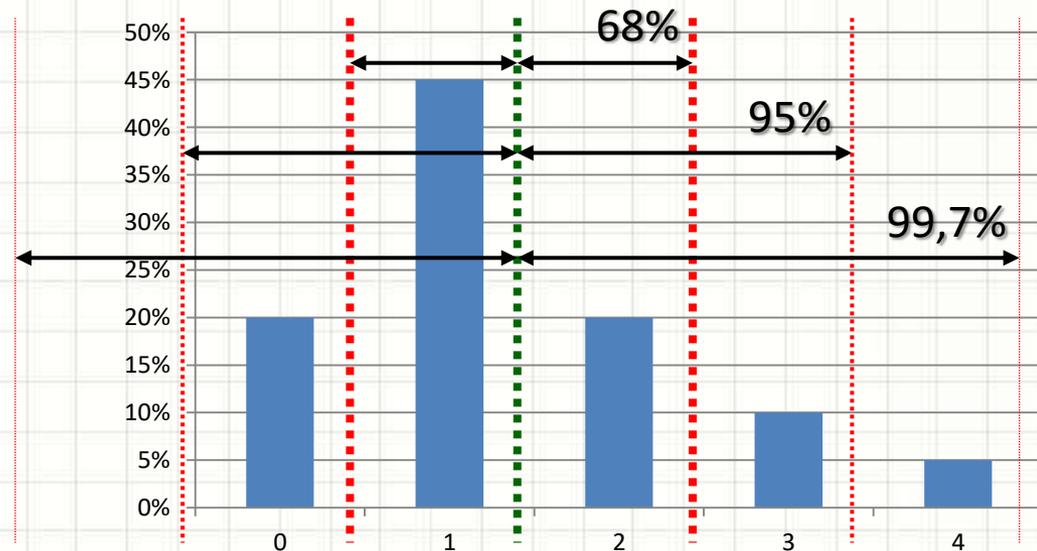
Cálculo do Desvio Padrão

- Cálculo: $\sigma_x = \sqrt{Var[x]}$
 - Para populações, usa-se σ , para amostras, usa-se S
- Interpretação:

Número de Filhos	Frequência Relativa
0	20%
1	45%
2	20%
3	10%
4	5%

$$\bar{x} = 1,35$$

$$\sigma_x = 1,06$$

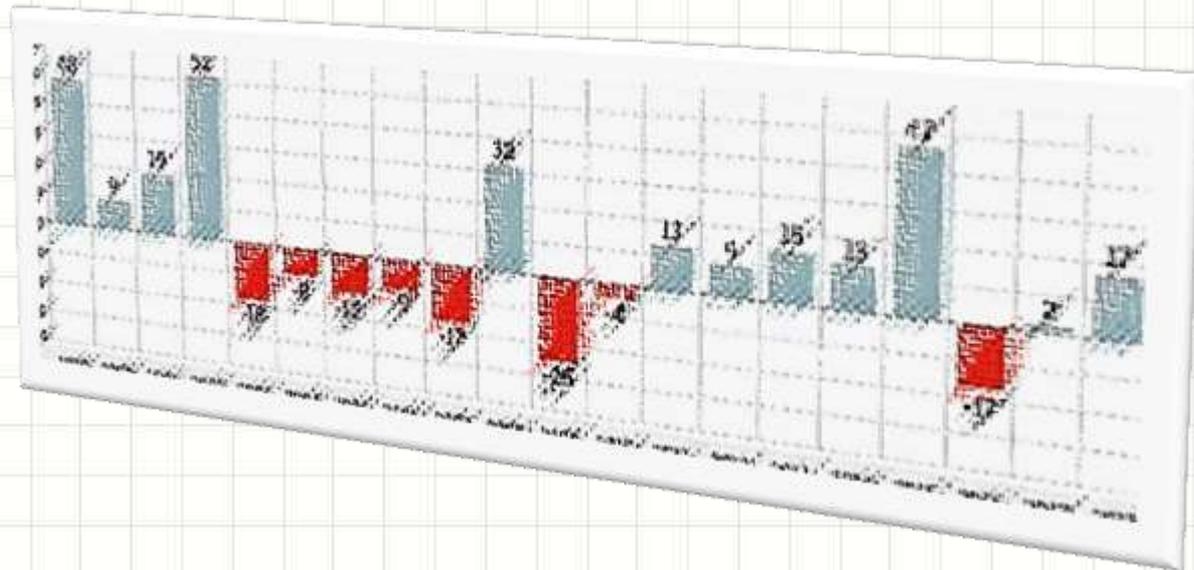


Propriedades

- 4 Propriedades da Variância
 - Variância de variável constante: sempre 0.
 - $\text{Var}[X \cdot C] = \text{Var}[X] \cdot C^2$.
 - $\text{Var}[X + C] = \text{Var}[X]$
 - $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$
- 3 Propriedades do Desvio Padrão
 - Desvio padrão de variável constante: sempre 0.
 - $\sigma[X \cdot C] = \sigma[X] \cdot C$
 - $\sigma[X + C] = \sigma[X]$

Coeficiente de Variação

- Noção relativa: $CV = \frac{\sigma_x}{\bar{x}} \cdot 100\%$
- Classificação da dispersão
 - Baixa dispersão: $0\% \leq CV \leq 15\%$
 - Média dispersão: $15\% \leq CV \leq 30\%$
 - Alta dispersão: $CV > 30\%$



Exercícios de Revisão

- Amostra: 1, 4, 4, 8, 8, 8, 10
- Calcule

– Moda **8**

$$P_i = \frac{(n + 1)}{2} = \frac{(7 + 1)}{2} = 4$$

– Mediana **8**

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} =$$

– Média **6,1**

$$\frac{1 + 4 + 4 + 8 + 8 + 8 + 10}{7} \cong 6,1$$

– Desvio Padrão **3,2** $S_x = \sqrt{Var[x]} \cong 3,2$

$$Var[x] = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1} = \frac{(-5,1)^2 + 2 \cdot (-2,1)^2 + 3 \cdot 1,9^2 + 3,9^2}{6} \cong 10,1$$

– Coef. de Variação **52,5%**

$$CV = \frac{S_x}{\bar{x}} \cdot 100\% \cong 52,5\%$$

Exercícios de Revisão

- População
- Calcule

Número de Autos	Frequência Relativa
0	30%
1	45%
2	15%
3	10%

– Média **1,1**

$$\bar{x}_p = \frac{p_1 \cdot x_1 + \dots + p_n \cdot x_n}{p_1 + \dots + p_n} =$$

$$\frac{0,3 \cdot 0 + 0,45 \cdot 1 + 0,15 \cdot 2 + 0,1 \cdot 3}{0,3 + 0,45 + 0,15 + 0,1} \cong 1,1$$

– Desvio Padrão **1,0**

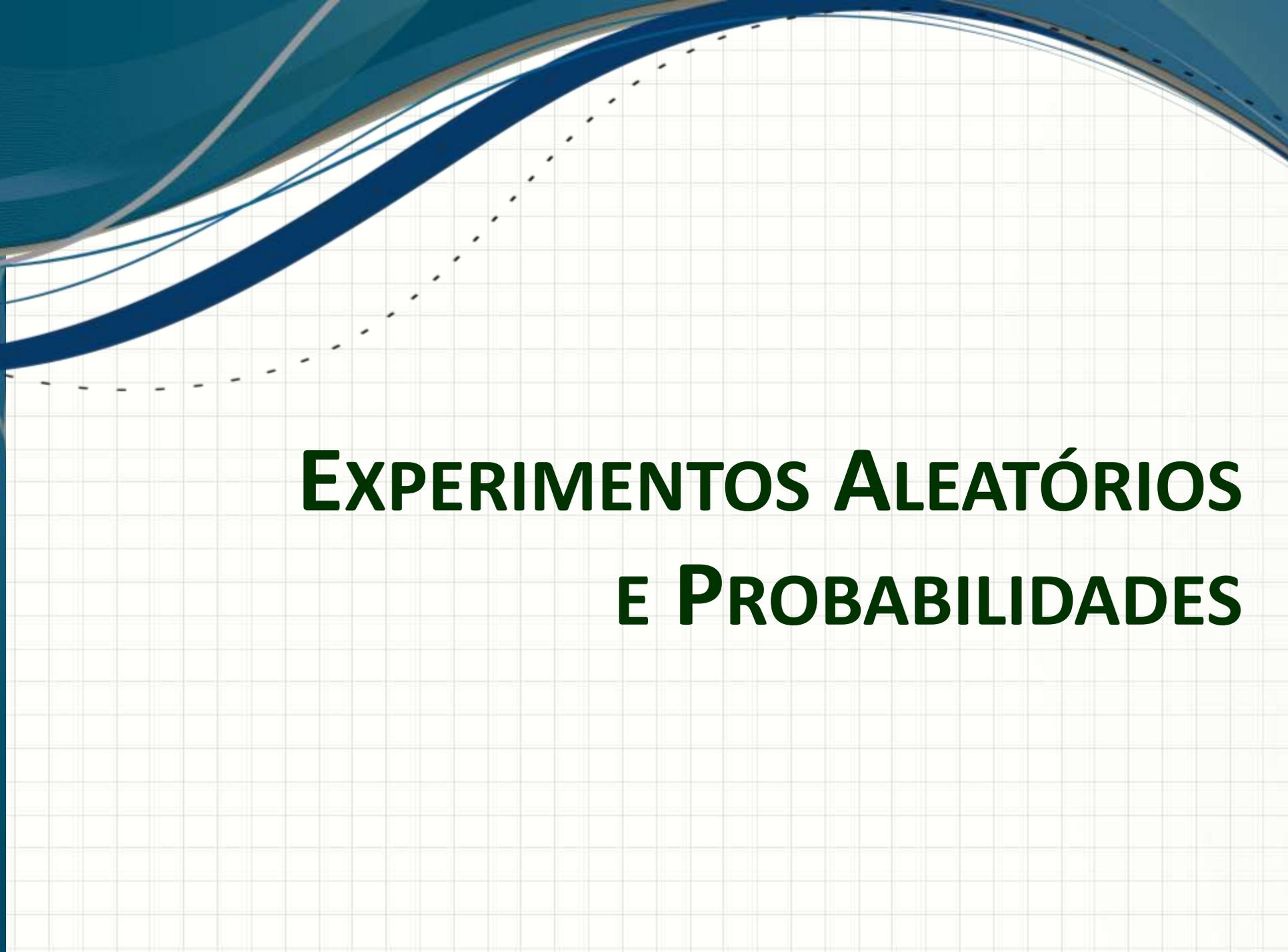
$$\sigma_x = \sqrt{Var[x]} \cong 1,0$$

$$Var[x] = (x_1 - \bar{x})^2 \cdot fr_1 + \dots + (x_n - \bar{x})^2 \cdot fr_n =$$

$$Var[x] = (-1,1)^2 \cdot 0,3 + (-0,1)^2 \cdot 0,45 + 0,9^2 \cdot 0,15 + 2,1^2 \cdot 0,1 = 0,93$$

– Coef. de Variação **90,9%**

$$CV = \frac{\sigma_x}{\bar{x}} \cdot 100\% \cong 90,9\%$$



EXPERIMENTOS ALEATÓRIOS E PROBABILIDADES

Exp. Aleatório e Espaço Amostral

- Conhecida só a probabilidade de um resultado
 - Probabilidade de um evento específico

$$frequência = \frac{Sucessos_Obtidos}{Número\ de\ Experimentos}$$

- Espaço amostral
 - Todas as possibilidades de eventos
 - Espaço Amostral: equiprovável ou não



Probabilidade Teórica

- Medida segundo a expressão:

$$P(A) = \frac{\text{Número de Casos Favoráveis}}{\text{Número Total de Casos}}$$

- Regras e Propriedades da Probabilidade

- $0 \leq P \leq 1$

- $P(S) = 1$

- $P(\emptyset) = 0$

- $P(\bar{A}) = 1 - P(A)$

- Se $A \subset B$, $P(A) \leq P(B)$

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

**Regra da
Soma**

Arranjos e Combinações

- Arranjo com reposição

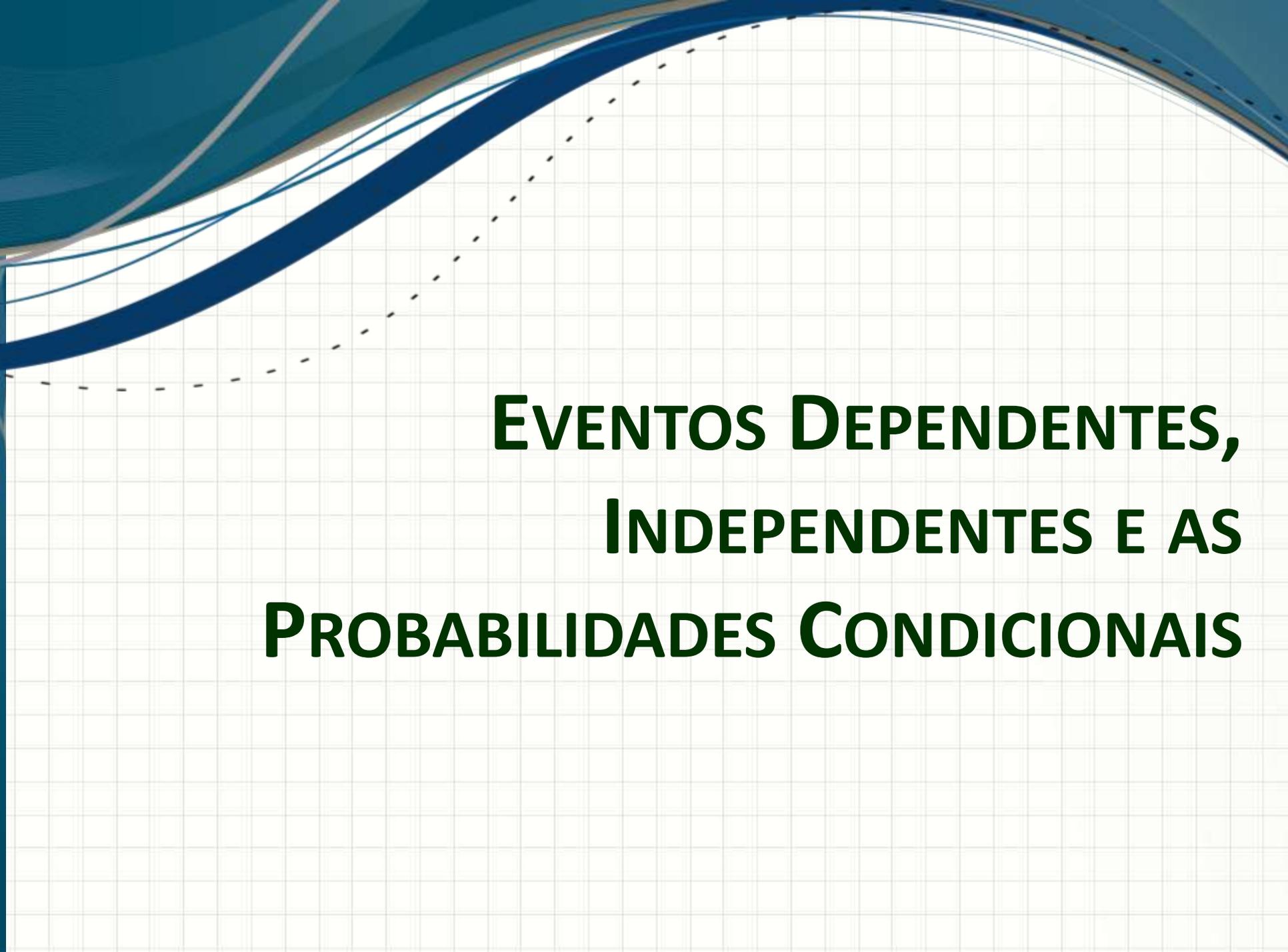
$$A_{r,p} = r^p$$

- Arranjo sem reposição

$$A_{r,p} = \frac{r!}{(r-p)!}$$

- Combinações sem reposição (qualquer ordem!)

$$C_{r,p} = \binom{r}{p} = \frac{r!}{p!(r-p)!}$$



**EVENTOS DEPENDENTES,
INDEPENDENTES E AS
PROBABILIDADES CONDICIONAIS**

Eventos (In)dependentes

- Independentes
 - A ocorrência de um evento não muda a probabilidade do evento seguinte
- Dependentes
 - A ocorrência de um muda a probabilidade do evento seguinte



Prob. Condicional: Eventos Indep.

- Joguei moeda
 - (A) Probabilidade de sair cara? $P(A) = \frac{1}{2}$
- Joguei moeda novamente...
 - (B) Probabilidade de sair cara? $P(B) = \frac{1}{2}$
- Considerando as duas jogadas acima
 - (C) Probabilidade de sair cara na 1ª e na 2ª

$$P(C) = P(A \cap B) = P(A) \cdot P(B) = \frac{1}{4}$$

Regra da Multiplicação

Prob. Condicional: Eventos Depend.

- Tirei uma carta do baralho
 - (A) Prob. da 1ª carta ser de ouros? $P(A) = \frac{13}{52}$

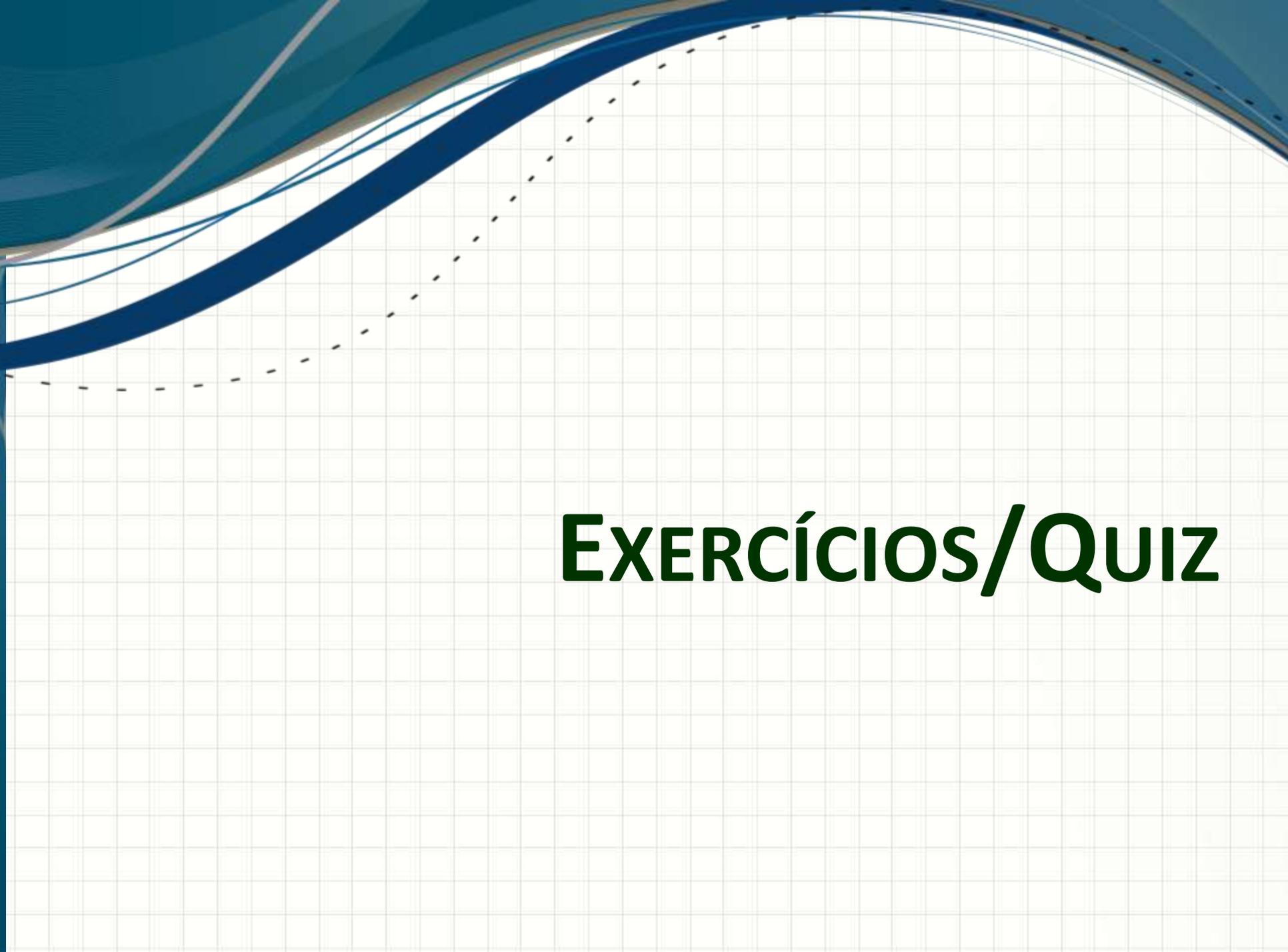
- Sem repor, tirei outra carta...
 - (B) Prob. da 2ª carta ser de copas? $P(B) = \frac{13}{51}$

$$P(B|A) = \frac{13}{51}$$

- Tirando as duas cartas como acima...
 - (C) Prob. Da 1ª de ouros e 2ª de copas?

$$P(A \cap B) = P(A) \cdot P(B|A)$$

Regra da Multiplicação



EXERCÍCIOS/QUIZ

Quiz

<https://kahoot.it/>

Exercícios de Revisão

- No lançamento de um dado não viciado, qual a probabilidade de obter um número maior que 4?
- $S = \{1, 2, 3, 4, 5, 6\}$
- Sucessos = $\{5, 6\}$

$$P(A) = \frac{2}{6} = 0,33$$

Exercícios de Revisão

- Uma moeda foi lançada uma vez, resultando em **cara**; qual a probabilidade de, jogando novamente, sair cara?
- $S = \{K, C\}$
- Sucessos = $\{K\}$

$$P(A) = \frac{1}{2} = 0,5$$

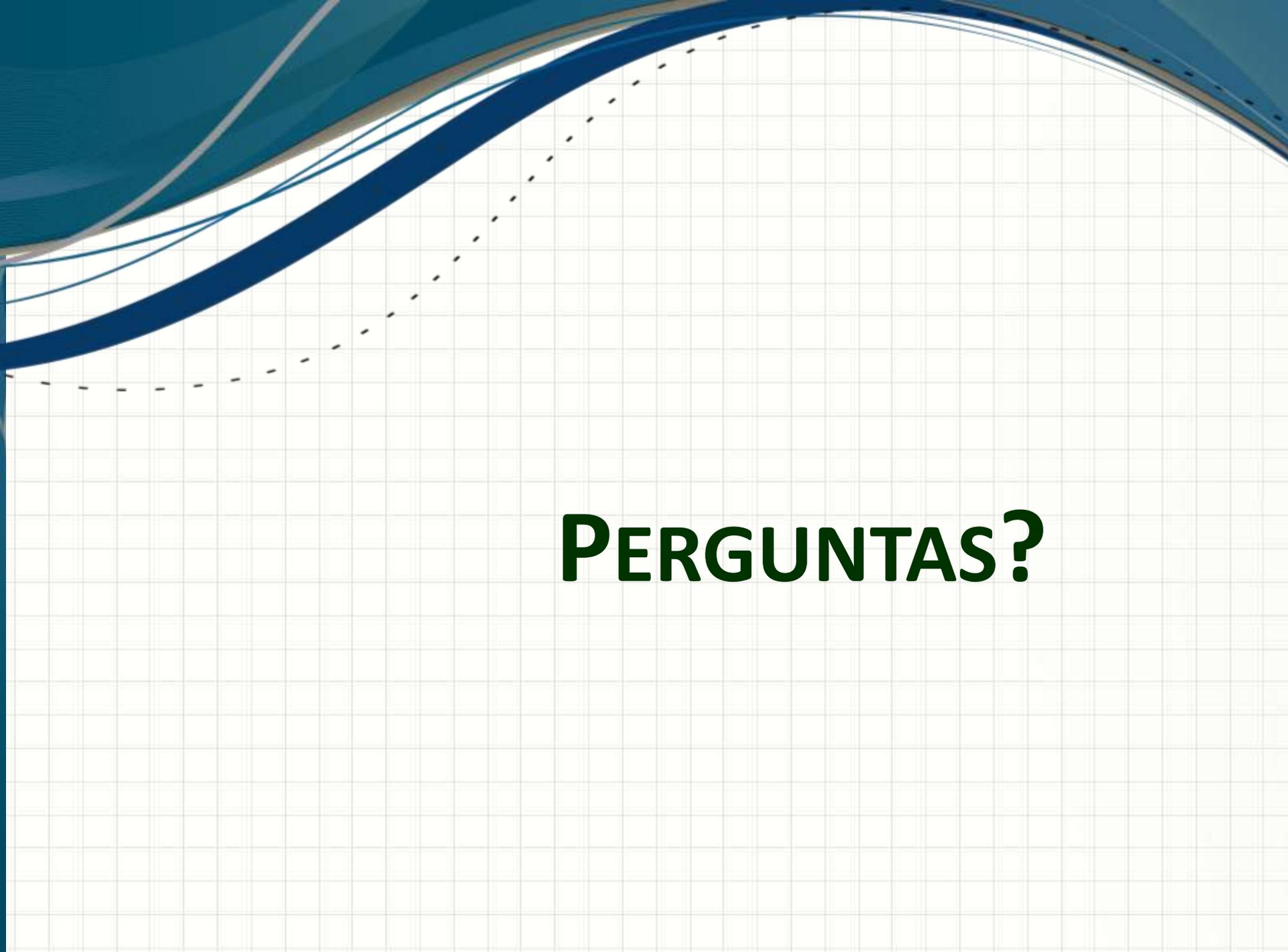
Exercícios de Revisão

- Num estacionamento vazio com vagas de 1 a 15, qual a probabilidade do primeiro motorista escolher uma vaga par ou de número maior que 10?
- $S = \{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15\}$
- Sucessos = $\{2,4,6,8,10,11,12,13,14,15\}$

$$P(A) = \frac{10}{15} = 0,33$$

Exercícios de Revisão

- Há duas bandejas, a) com 2 vinhos e 5 cervejas e b) com 4 vinhos e 3 cervejas. Qual a chance de pegar, aleatoriamente, vinho?
- Probabilidade de pegar bandeja A: $\frac{1}{2}$
 - Probabilidade pegar vinho $P(V|A) = \frac{2}{7}$
 - $P(A \cap V) = \frac{1}{2} * \frac{2}{7} = \frac{1}{7}$
- Probabilidade de pegar bandeja B: $\frac{1}{2}$
 - Probabilidade pegar vinho $P(V|B) = \frac{4}{7}$
 - $P(B \cap V) = \frac{1}{2} * \frac{4}{7} = \frac{2}{7}$
- Probabilidade de pegar vinho: $\frac{3}{7}$



PERGUNTAS?