# Especialização em *Data Science* e Estatística Aplicada

## Módulo I - Análise Estatística para *Data Science*

Profa. Dra. Renata M. R. Vasconcelos

Goiânia, 2024

IME
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

FEN
FACULDADE DE ENFERMAGEM

UFG
UNIVERSIDADE FEDERAL DE GOIÁS

# Conteúdo Programático

- **Aula 1:** Introdução à probabilidade: experimento aleatório, espaço amostral, eventos e operações, definição de probabilidade, cálculo de probabilidade, probabilidade condicional, Teorema de Bayes.
- **Aula 2:** Variável Aleatória e principais distribuições de probabilidade.
- **Aulas 3 e 4:** Inferência Estatística para uma população: principais conceitos, distribuição amostral, estimação pontual e intervalar, teste de hipóteses.

# Aula 1 - Introdução à Probabilidade

1. Introdução: Conceitos Importantes
   - Fenômenos ou Experimentos Aleatórios

2. Probabilidade: conceitos fundamentais
   - Espaço amostral e eventos

3. Cálculo de Probabilidades
   - Medida de Probabilidade

# Introdução: Conceitos Importantes

**Fenômenos ou Experimentos Aleatórios**
Situações em que os resultados possíveis são conhecidos, mas não se pode saber *a priori* qual deles ocorrerá (resultados incertos).

Considerações Importantes:

- $\hookrightarrow$ O conjunto de todos os possíveis resultados do experimento aleatório é conhecido;
- $\hookrightarrow$ Ao repetir a experiência sob as mesmas condições, ainda assim, não podemos prever o resultado com exatidão;
- $\hookrightarrow$ Os resultados, em cada experiência, são incertos.

$\Longrightarrow$ A **variabilidade** dos resultados analisados está associada à incerteza;

# Introdução: Conceitos Importantes

**Questões relacionadas à algumas situações de incerteza**

- Qual o valor da face superior após o lançamento de um dado honesto de seis faces?
- Qual o número de filhotes de uma fêmea da raça de cão Rottweiller?
- Qual o peso de um roedor adulto?
- Qual é a audiência estimada de um comercial de TV com 30 segundos de duração para cada exibição?

# Introdução: Conceitos Importantes

**Conhecimento em Probabilidade**
**Objetivo:** Usar a linguagem de probabilidade para mostrar como se mede o nível de certeza da ocorrência de resultados possíveis de um experimento/fenômeno.

$\hookrightarrow$ **Os conceitos de probabilidade podem ser utilizados para oferecer possíveis padrões a eventos imprevisíveis: MODELOS PROBABILÍSTICOS.**

**Observação Importante:**
Experimento Aleatório $\neq$ Experimento Determinístico.

$\hookrightarrow$ O **Experimento Determinístico** não fornece um **modelo** adequado para a explicação do fenômeno.

# Probabilidade: conceitos fundamentais

**Probabilidade**

- Ideia intuitiva;
- Teoria teve início em meados do séc. XVI com Gerolamo Cardano: *Jogos de azar e sua relação com os resultados*;
- Medida associada à chance de algo ocorrer.

$\hookrightarrow$ É consensual interpretar a probabilidade como a possibilidade de um determinado fato ou evento vir a ocorrer, avaliada numericamente e em termos percentuais.

# Probabilidade: conceitos fundamentais

### Ideia subjetiva × Previsão

- Ideia subjetiva: Avaliação pela observação e conhecimento completo dos fatores que influenciam o fato ou evento.

  **Ex.**: Em um dia nublado, imaginamos a chance de que possa chover.

  $\hookrightarrow$ *Ao sairmos de casa, devemos nos precaver para o uso do guarda-chuva*.

- Previsão: Avaliação pela observação do comportamento passado do evento e das circunstâncias nas quais ocorreu.

  Baseada na história anual do local em que se está.

  **Ex.**: no ano anterior, 35% dos dias nublados resultaram em chuvas.

  $\hookrightarrow$ *decisão mais respaldada sobre a chuva*.

# Probabilidade: conceitos fundamentais

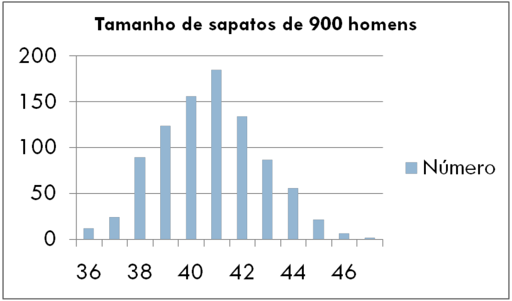
### Variabilidade nos dados

- **A variação é inerente à natureza.**

  $\Longrightarrow$ Existe uma multiplicidade de características que podem ser observáveis nos mais diversos seres vivos e cada uma delas apresenta determinado grau de variabilidade.

# Probabilidade: conceitos fundamentais

| Tamanho do sapato | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Número | 12 | 25 | 90 | 124 | 156 | 185 | 134 | 87 | 56 | 22 | 7 | 2 |




Quanto ao padrão de variabilidade, **embora os dados apontam que alguns tamanhos são mais prováveis que outros,** podemos notar que **a variabilidade do tamanho 36 ao 47** **<u>geraria incerteza</u>** em relação aos tamanhos que um fabricante de sapatos deveria produzir e um estabelecimento de varejo deveria estocar!

Figura 1: Tamanho dos sapatos de 900 homens adultos.

# Probabilidade: conceitos fundamentais

**Figura 1**

Pode ser desejável conhecer a probabilidade de um futuro cliente "procurar" pelo tamanho 47.

- Embora esse número esteja dentro da faixa dos dados coletados, olhando para o gráfico, parece improvável que ele ocorra (uma vez que somente um número relativamente pequeno de clientes usa esse tamanho);
- É necessário então investigar a probabilidade de diferentes valores de dados (tamanhos de sapato) ocorrerem antes de serem tomadas decisões sobre os níveis de estoque.

# Probabilidade: conceitos fundamentais

Além disso...

↬ Quando trabalhamos com a Estatística Inferencial (com os chamados Testes Estatísticos), temos como objetivo *conhecer uma ou mais características de determinada população por meio das características da amostra em estudo.*

Em outras palavras, desejamos, a partir da **amostra**, realizar nossas **conclusões para a população**.

# Probabilidade: conceitos fundamentais

↬ Para que tais **conclusões** sejam **efetivas**, necessitamos utilizar modelos probabilísticos.

**Modelos Probabilísticos**

**Descrição matemática de um fenômeno aleatório** consistindo de:

- Uma lista de possíveis resultados (*espaço amostral*);
- Uma probabilidade medida para cada resultado.

# Probabilidade: conceitos fundamentais

**Espaço amostral**

Conjunto de todos os resultados possíveis de um experimento aleatório.

**Notação:** $\Omega$ ou $S$.

**Evento**

Subconjunto do espaço amostral.

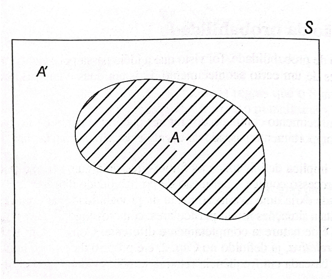



Figura 2: Diagrama de *Venn*: evento A e seu evento complementar A' do espaço amostral S.

# Probabilidade: conceitos fundamentais

**Exemplo 1**

$\mathrm{E}$: Lançar um dado e observar sua face superior.

$\Omega = \{1, 2, 3, 4, 5, 6\}$

*Eventos*:

***Eventos Unitários:*** $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$, $\{6\}$;

***Outros:*** $A = \{1, 3\} = \{$*"face 1 ou face 3"*$\}$;

$A^C = \bar{A} = \{2, 4, 5, 6\} = \{$*"não ocorrer face 1 ou face 3"*$\}$ _ *evento complementar de A*;

$B = \{2, 4, 6\} = \{$*"faces com números pares"*$\}$;

$C = \{1, 3, 5\} = \{$*"faces com números ímpares"*$\}$;

ev unitario

# Probabilidade: conceitos fundamentais

**Exemplo 2**

$\mathrm{E}$: Escolher, aleatoriamente, um aluno do curso de especialização e observar a sua estatura (em *cm*).

$\Omega = \{$*"estaturas possíveis, em cm"*$\}$

$\Longrightarrow$ *O espaço amostral será um intervalo de valores a partir de zero.*

*Eventos*:

$A = \{$*"173 cm a 175 cm"*$\} = \{173 \leq$ *estatura* $\leq 175\}$;

$B = \{$*"maior que 174 cm"*$\} = \{$*estatura* $\geq 174\}$;

$C = \{$*"pelo menos 160 cm"*$\} = \{$*estatura* $\geq 160\}$;

$D = \{$*"no máximo 169 cm"*$\} = \{$*estatura* $\leq 169\}$.

# Probabilidade: conceitos fundamentais

**Exemplo 3**

E: "Lançamento de dois dados".

$\#\Omega = 6 \times 6 = 36$ *elementos;*

<u>*Eventos*</u>:

*Eventos simples:* $F = \{(4,6),(5,5),(6,4)\} = \{$*"a soma dos dois valores é dez"*$\}$;

$G = \{(1,1),(2,2),\ldots,(6,6)\} = \{$*"os dois valores são iguais"*$\}$;

$H = \{(1,4),(1,5),(1,6),\ldots,(6,5),(6,6)\}$

$= \{$*"Uma das faces é maior do que 3"*$\}$;

*Outros:* $F \cup G = \{$*"a soma dos dois valores é dez"* <u>*ou*</u> *"os dois valores são iguais"*$\}$;

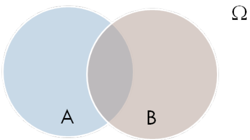
$G \cap H = \{$*"os dois valores são iguais"* <u>*e*</u> *"Uma das faces é maior do que 3"*$\}$.

# Probabilidade: conceitos fundamentais

↬ **Algumas operações importantes envolvendo os eventos de $\Omega$.**

*Sejam A e B dois eventos de um espaço amostral $\Omega$:*

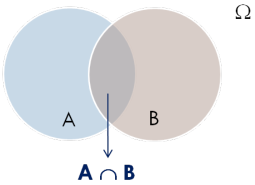
**a) União:** O evento **A ∪ B** ocorre quando ocorre somente o evento A ou ocorre somente o evento B ou ocorrem ambos os eventos A e B.




$$A \cup B = \{x \in \Omega \mid x \in A \quad ou \quad x \in B\}$$

# Probabilidade: conceitos fundamentais

**b) Interseção:** O evento **A ∩ B** ocorre quando ocorrem elementos comuns em ambos os eventos A e B.




$$\mathbf{A \cap B = \{x \in \Omega \mid x \in A \;\; \textit{e} \;\; x \in B\}}$$

# Probabilidade: conceitos fundamentais

**Exemplo 4**

**Teste de Diagnóstico.**

**Sox et al. (1989):** *Um teste de diagnóstico tem como finalidade fornecer informação confiável sobre a condição do paciente e influenciar a decisão do médico de tratar ou não o paciente.*

- ↪ Como não existe teste perfeito, os resultados de um novo teste devem ser comparados com aqueles do teste que fornece o diagnóstico definitivo (*"padrão-ouro" _ teste que teoricamente produz resultados corretos*);
- ↪ No geral, estudos devem ser feitos para a comprovação da **qualidade de testes** a serem utilizados com o intuito de minimizar as chances de diagnósticos incorretos.

# Probabilidade: conceitos fundamentais

## Exemplo 4: Resultado de um teste clínico

- Pesquisa: determinação do grau de confiabilidade de um teste diagnóstico.
- Objetivo: medir o nível de certeza da ocorrência de um evento, por exemplo, a existência de uma doença após a observação de um teste positivo.
  - *Consideraremos o teste positivo quando indicar a presença da doença e teste negativo quando indicar a ausência.*
- Metodologia: na etapa de pesquisa, o pesquisador utiliza o teste primeiramente em dois grupos específicos de pessoas:

  Grupo 1: pacientes realmente doentes.

  Grupo 2: pessoas sem a doença em questão.

$\hookrightarrow$ *O diagnóstico, nessa etapa, deve ser feito por um meio diferente do teste em estudo ("padrão-ouro").*

# Probabilidade: conceitos fundamentais

Tabela 1: Resultados de uma pesquisa para investigar a qualidade de um teste clínico

| Doença (*padrão-ouro*) | Teste | | Total |
|---|---|---|---|
| | Positivo (T+) | Negativo (T-) | |
| Presente (D+) | 90 | 10 | 100 |
| Ausente (D-) | 5 | 95 | 100 |
| Total | 95 | 105 | 200 |

**Pela Tabela 1:**

- Para os realmente doentes, o resultado do teste foi correto em $90\%$ e para os não doentes, $95\%$.

# Probabilidade: conceitos fundamentais

*Evento*:

↬ Para definir os índices que descrevem o grau de confiabilidade de um teste, é preciso trabalhar com os seguintes eventos:

| Evento | Descrição | Evento | Descrição |
|---|---|---|---|
| **T+** | Teste é positivo | **D+** | O paciente é doente |
| **T-** | Teste é negativo | **D-** | O paciente não é doente |

# Probabilidade: conceitos fundamentais

Como calcular a probabilidade de um dado evento, associado a uma certa característica???

# Definição de Probabilidade

A probabilidade pode ser expressa de muitas maneiras como, por exemplo, razão, diferença, porcentagem, fração. ***O <u>padrão matemático</u> é que a probabilidade seja expressa como uma fração ou um número decimal entre 0 e 1***. ***Variação da Probabilidade***

| Um gato dando à luz um cachorro | Uma mulher dando à luz um menino | Choverá um dia em Goiânia no próximo ano |
|---|---|---|
| 0 (definitivamente não ocorrerá) | 0,5 ou 50% | 1 ou 100% (definitivamente ocorrerá) |

# Definição de Probabilidade

## Definição Subjetiva

Valor entre **0** e **1** que representa um ponto de vista pessoal sobre a possibilidade de ocorrer determinado evento.

- Medida racional, embora não se baseie em técnicas computacionais, e tem sentido quando fornecida por alguém que conhece o assunto;
- Predomina nas *decisões administrativas*, nas *aplicações financeiras* e nos *jogos de azar*.
  - Desvantagens:
    - Medida muito pessoal;
    - Pode não ter relação alguma com os resultados realmente obtidos.

## Exemplo 5

Quando alguém diz "a probabilidade de o Brasil ganhar a próxima Copa do Mundo de Futebol é $80\%$", está usando a definição subjetiva de probabilidade

# Cálculo de Probabilidade

**Definição Frequentista (atribuída à Richard von Mises, 1883-1953)**

- Apelo intuitivo: *medição experimental/probabilidade avaliada*;
- A **probabilidade de um evento A** é a sua **frequência relativa** de ocorrência, ou seja, a proporção de vezes que o evento ocorre em um número elevado de repetição do experimento, sob as mesmas condições.

$$\boxed{f_r = \frac{n_A}{n}}$$

Estimação da probabilidade que ocorra o evento A.

- $n$: número de vezes (realizações independentes) que um experimento é realizado sob as mesmas condições;
- $n_A$: número de ocorrências de um acontecimento *A* nas "*n*" repetições do experimento ("*n*" suficientemente grande)

# Cálculo de Probabilidade

### Probabilidade Frequentista

↬ Nas condições apresentadas, registra-se simplesmente que à medida que *n* aumenta, a razão $n_A/n$ converge para um determinado limite, que é a **probabilidade do evento *A* ocorrer** _ Notação: P(A).

$$f_r = \frac{n_A}{n} \xrightarrow{\text{“n” grande}} P(A)$$

# Cálculo de Probabilidade

## Exemplo 5

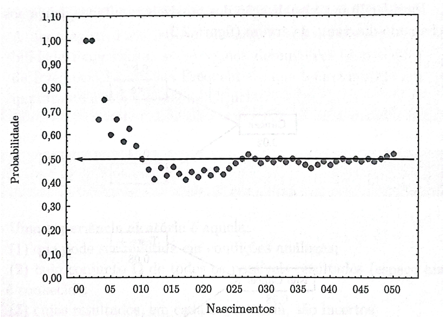



Figura 3: Probabilidade frequentista quando observamos o evento $A = \{$“nascer menina”$\}$

# Cálculo de Probabilidade

**Exemplo 6**

**Noção de probabilidade frequentista quando observamos o evento A "nascer menina";**

- 1º nascimento $\Longrightarrow P^*(A) = 1$, pois nasceu menina e $f_r = \frac{1}{1} = 1$;
- 2º nascimento $\Longrightarrow P^*(A) = 1$, pois nasceu menina e $f_r = \frac{2}{2} = 1$;
- 3º nascimento $\Longrightarrow P^*(A) = 0,66$, pois nasceu menino e $f_r = \frac{2}{3} = 0,66$;
- nº maior de nascimentos $\Longrightarrow P^*(A)$ se altera e, após um nº grande de observações, se estabiliza em torno de 0,50! (depois de 50 observações, $P^*(A) = 0,52$).

exfrequentista

# Cálculo de Probabilidade

**Exemplo 7**

Considere o lançamento de uma moeda.

*Calcular a probabilidade de* $A = \{\text{'resultado obtido é cara'}\}$

| | $fr_1$ | $fr_2$ | $fr_3$ | $fr_4$ | ... | $fr_A$ |
|---|---|---|---|---|---|---|
| **Cara** | 2/5 | 6/10 | 22/50 | 47/100 | ... | 0,5 |
| **Coroa** | 3/5 | 4/10 | 28/50 | 53/100 | ... | 0,5 |
| **n** | **5** | **10** | **50** | **100** | ... | $\infty$ |

# Cálculo de Probabilidade

**Exemplo 8**
Um médico verificou que, de 2.964 nascidos vivos, 73 tinham alguma deficiência ou doença séria. Com base nessa amostra, a estimativa da probabilidade de um recém-nascido ter deficiência ou doença séria é:

**Solução:**

$$\frac{73}{2964} = 0,0246$$

# Cálculo de Probabilidade

**Exemplo 9**

**Risco**

*Em uma amostra de 30.195 registros hospitalares selecionados ao acaso, foram identificados 1.133 pacientes com lesões sérias causadas por imprudência, negligência ou imperícia do médico. O risco estimado de lesão séria por erro médico nesse hospital é:*

**Solução:**

$$\frac{1133}{30195} = 0,0375$$

# Cálculo de Probabilidade

**Eventos Equiprováveis**

Resultados de um experimento aleatório que tem a mesma chance de ocorrência.

**Probabilidade de eventos equiprováveis**

- No **Exemplo 1**, os eventos unitários são equiprováveis.
  - A probabilidade de ocorrência de cada face é $\frac{1}{6}$ e, consequentemente:

$$P(A) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} \cong 0,33; \quad P(B) = \frac{3}{6} = 0,5; \quad P(C) = \frac{3}{6} = 0,5.$$

ev unitario

# Cálculo de Probabilidade

**Definição Clássica (atribuída à Laplace, 1749-1827)**

Seja um **espaço amostral finito** $\Omega$, formado por eventos equiprováveis.
Sendo *A* um evento de $\Omega$, então a probabilidade de ocorrência de *A* é dada por:

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{\#A}{\#\Omega} = \frac{\#A}{n}$$

onde:

- $\#A$: nº de possibilidade favoráveis a ocorrência do evento *A*.
- n: nº total de possibilidades/resultados do espaço amostral $\Omega$.

# Cálculo de Probabilidade

**Exemplo 10**
Qual a probabilidade de sair ao menos uma cara em dois lançamentos consecutivos de uma moeda não viciada?

**Solução:**
**Cara = k** e **Coroa = c**
$\Omega = \{(k,k),(k,c),(c,k),(c,c)\}$
$A = \{\text{“sair ao menos uma cara''}\} = \{(k,k),(k,c),(c,k)\}$

$$P(A) = \frac{\#A}{\#B} = \frac{3}{4} = 75\%.$$

**Conclusão:** A probabilidade/chance/risco de sair ao menos uma cara em dois lançamentos consecutivos de uma moeda não viciada é de 75%.

# Cálculo de Probabilidade

**Exemplo 11**

Pela ilustração referente ao **Tamanho dos sapatos de 900 homens adultos**, utilizando as informações fornecidas, meça a probabilidade de um cliente homem calçar:

(a) tamanho 38?

(b) tamanho 47?

**Solução:**

**(a) Dados:** 90 homens calçam 38.

$$P(\text{‘cliente calçar 38'}) = \frac{90}{900} = 0,1 = 10\%.$$

**(b) Dados:** 2 homens calçam 47.

$$P(\text{‘cliente calçar 47'}) = \frac{2}{900} = 0,0022 = 0,22\%.$$

# Cálculo de Probabilidade

**Considerações Importantes**

**A:** evento de $\Omega$.

Note que **A** estará sempre contido em $\Omega$.

Logo, para qualquer evento **A**,

$$P(\emptyset) = \mathbf{0} \leq \mathbf{P(A)} \leq \mathbf{1} = P(\Omega)$$

onde

- $\emptyset$ = { }: conjunto vazio que corresponde ao **evento impossível**;
- $\Omega$: conjunto dos possíveis resultados de um experimento aleatório que corresponde ao **evento certo**.

# Cálculo de Probabilidade

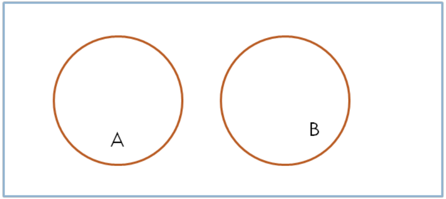
## Eventos Disjuntos ou Mutuamente Exclusivos ou Excludentes

**A** e **B**: eventos de $\Omega$.

Resultado de um experimento aleatório tal que os eventos **A** e **B** não podem ocorrer simultaneamente.

$\Longrightarrow$ a interseção entre **A** e **B** é o conjunto vazio, ou seja,

$$P(A \cap B) = P(\emptyset) = 0.$$



# Cálculo de Probabilidade

**Probabilidade: eventos complementares**

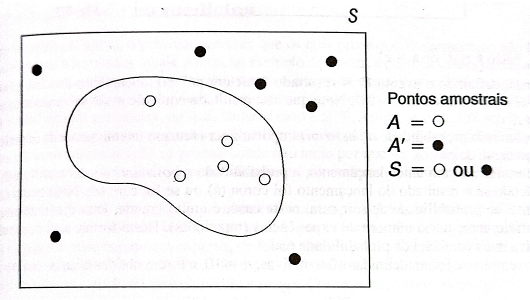
**A:** evento de $\Omega$.

$\bar{\text{A}}$: evento complementar de **A** e, logo, também evento de $\Omega$.

Note que:

- $A \cup \bar{A} = \Omega$
- $A \cap \bar{A} = \emptyset$

$\Longrightarrow$ **A** e $\bar{\text{A}}$ são eventos disjuntos.




Nota: Sendo $A \cap \bar{A} = \emptyset \Longrightarrow P(A \cap \bar{A}) = P(\emptyset) = 0$.

# Cálculo de Probabilidade

Por definição:

$$P(\text{“ocorrer A”}) = P(A) = \frac{\#A}{\#\Omega} \quad ; \quad P(\text{“ñ ocorrer A”}) = P(\overline{A}) = \frac{\#\overline{A}}{\#\Omega}$$

$$\Longrightarrow \mathbf{P(A \cup \overline{A})} =^{(\star)} P(A) + P(\overline{A}) = \frac{(\#A) + (\#\overline{A})}{\#\Omega} = \frac{\#\Omega}{\#\Omega} = \mathbf{1}$$

De onde vem que

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

$(\star)$ : Uma forma mais geral de alocar probabilidade a eventos é **somando as probabilidades** dos resultados que compõem o evento.

# Cálculo de Probabilidade

**Soma de Probabilidades**

**Regra do "ou".**

**Soma de eventos disjuntos (Ilustração)**

Foi feito um estudo de caso-controle com pacientes hospitalizados (7.804 casos e 15.207 controles) para determinar os fatores de risco de câncer do pulmão. Os dados da **Tabela 2** foram obtidos para saber se o risco de câncer do pulmão aumenta com o número de cigarros fumados por dia.

Qual a probabilidade de uma pessoa, tomada ao acaso dessa amostra, fumar um maço de cigarros (20 cigarros) ou mais por dia?

# Cálculo de Probabilidade

Tabela 2: Resultados de uma pesquisa para investigar a qualidade de um teste clínico

| Número de cigarros/dia | Casos | Controle | Total | Risco |
|---|---|---|---|---|
| Nenhum | 164 | 2.616 | 2.780 | 0,059 |
| 1 a 9 | 664 | 2.194 | 2.858 | 0,232 |
| 10 a 19 | 1.704 | 3.385 | 5.089 | 0,335 |
| 20 a 29 | 2.127 | 3.108 | 5.235 | 0,406 |
| 30 ou mais | 1.369 | 1.746 | 3.115 | 0,439 |
| Total | 6.028 | 13.049 | 19.077 | |

# Cálculo de Probabilidade

**Soma de eventos disjuntos (*Continuação*)**

**Solução:**

$$P(\text{“de 20 a 29”}) = \frac{5.235}{19.077} = 0,274$$

$$P(\text{“30 ou mais”}) = \frac{3.115}{19.077} = 0,163$$

**Conclusão:** A probabilidade de uma pessoa fumar um maço ou mais cigarros por dia, nessa amostra, é:

$$P = 0,274 + 0,163 = 0,437$$

# Cálculo de Probabilidade

**Exemplo 12**

Suponha que, para se dar prioridade no atendimento de uma fila bancária, são considerados casos:

***A***: *{'grávidas'};*
***B***: *{'pessoas com mais de 60 anos'}.*

**Solução:**

*Note que* $P(A \cap B) = 0$*, de forma que* ***A*** *e* ***B*** *são disjuntos.*

*Além disso, considerando que* $P(A) = 6\%$ *e que* $P(B) = 12\%$*,*
***a probabilidade de que haja necessidade de dar prioridade na fila*** *seria de*

$$P(A \cup B) = P(A) + P(B) = 0,06 + 0,12 = 0,18 \;\; \textit{ou} \;\; 18\%$$

# Cálculo de Probabilidade

**Soma de eventos não disjuntos (Ilustração)**

Foi feito um estudo de caso-controle com pacientes hospitalizados (299 casos e 292 controles) para determinar os fatores de risco para infarto do miocárdio. Os dados da **Tabela 3** foram obtidos para saber se os indivíduos diabéticos apresentam maior risco de infarto do miocárdio.

Qual a probabilidade de uma pessoa, tomada ao acaso dessa amostra, ser ou não ser diabética ou infartada?

# Cálculo de Probabilidade

Tabela 3: Resultados de uma pesquisa para investigar a qualidade de um teste clínico

| **Diabéticos** | **Infartados** | | **Total** |
|---|---|---|---|
| | **Casos** | **Controles** | |
| **Sim** | 90 | 10 | 100 |
| **Não** | 5 | 95 | 100 |
| **Total** | *95* | *105* | *200* |

*Fonte:* SILVA, M. A. D.; SOUSA, A., G., M., R.; SCHARGODSKY,H. Fatores de risco para infarto do miocárdio no Brasil. **Arq Bras Cardiol.** v. 71, n. 5, p. 667-675, 1998.

# Cálculo de Probabilidade

**Soma de eventos não disjuntos (*Continuação*)**

**Solução:** Probabilidade de ter tido infarto:

$$P(\text{“infartado”}) = \frac{299}{591} = 0,506$$

Probabilidade de ter sido diabético:

$$P(\text{“diabético”}) = \frac{88}{591} = 0,149$$

# Cálculo de Probabilidade

**Solução:** *Continuação*

↬ Veja que as pessoas que **tiveram infarto e são diabéticas** estão no conjunto interseção e, portanto, foram consideradas nos dois cálculos. Então

$$P(\text{“infartado”} \cap \text{“diabético”}) = \frac{59}{591} = 0,0998$$

**Conclusão:** A probabilidade de **ter tido infarto ou ser diabético** é dada pela soma das probabilidades de *ter tido infarto* mais a probabilidade de *ser diabético* menos a probabilidade de que ambos ocorram simultaneamente.
Em outras palavras:

$$P(\text{“infartado”} \cup \text{“diabético”}) = \frac{299}{591} + \frac{88}{591} - \frac{59}{591} = \frac{328}{591} = 0,555$$

# Cálculo de Probabilidade

**Exemplo 13**

**A** $=$ {'obesidade'} ; **B** $=$ {'sedentarismo'}.

*Obtenha a probabilidade de:*

a) *selecionar um indivíduo obeso ou sedentário;*
b) *selecionar um indivíduo que não seja obeso e nem sedentário.*

Tabela 4: Informações de 15 indivíduos classificados quanto às variáveis obesidade e sedentarismo

| Indivíduo | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Obesidade | N | N | S | N | S | S | N | N | N | S | N | N | S | N | N |
| Sedentarismo | S | N | S | S | N | S | N | S | S | S | N | N | S | N | S |

# Cálculo de Probabilidade

Note que 10 indivíduos são obesos ou sedentários. No caso, são os que se apresentam como (S,S), (S,N), (N,S).

**Solução:** (a) selecionar um indivíduo obeso (A) ou sedentário (B).

$$P(\text{“indivíduo ser obeso”}) = P(A) = \frac{5}{15} \cong 0,33 \ \text{ ou } \ 33\%$$

$$P(\text{“indivíduo ser sedentário”}) = P(B) = \frac{9}{15} = 0,6 \ \text{ ou } \ 60\%$$

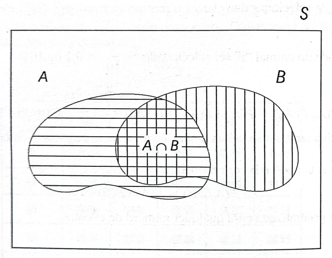
Existem 4 indivíduos são **obesos e sedentários**, de forma que:

$$P(A \cap B) = \frac{4}{15} \cong 0,2667 \ \text{ ou } \ 26,67\%$$

# Cálculo de Probabilidade

**Conclusão:** A probabilidade de se selecionar um indivíduo **obeso ou sedentário** pode ser calculada por:

$$PA \cup B) = P(A) + P(B) - P(A \cup B) = \frac{5}{15} + \frac{9}{15} - \frac{4}{15} = \frac{10}{15} \cong 0,3333 = 33,33\%$$



# Cálculo de Probabilidade

Note que 5 indivíduos não são obesos ($\bar{A}$) e não são sedentários ($\bar{B}$). Esses se apresentam como (N, N).

**Solução:** (b) selecionar um indivíduo que não apresente nenhuma das duas características.

$$P(\bar{A} \cap \bar{B}) = \frac{5}{15} \cong 0,33 \text{ ou } 33\% = P(\overline{A \cup B})$$

# Cálculo de Probabilidade

**Exemplo 14**

**Hipertensão**

*Um critério bastante aceito para definir hipertensão arterial na população é considerar um indivíduo como hipertenso se ele apresentar: pressão arterial sistólica maior que 140 (mm Hg), pressão arterial diastólica maior que 90 (mm Hg), ou ambas.*

*Definindo* $A = PAS > 140$ *(mm Hg);* $B = PAS > 90$ *(mm Hg) e considerando* $P(A) = 0,15$, $P(B) = 0,1$ e $P(A \cap B) = 0,08$.

## Conclusão:

Nessas condições, a probabilidade de um indivíduo hipertenso seria:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,15 + 0,1 - 0,08 = 0,17 = 17\%$$

# Cálculo de Probabilidade

**Lei multiplicativa**

**Regra do “e”.**

**Eventos Independentes**

**A** e **B**: eventos de $\Omega$.

Se **A** e **B** são eventos independentes, a probabilidade de ocorrer **A** e **B** é dada pela probabilidade de ocorrer **A**, multiplicada pela probabilidade de ocorrer **B**. Escreve-se:

$$P(A \cap B) = P(A) \times P(B)$$

# Cálculo de Probabilidade

**Regra 1 da multiplicação: eventos independentes (Ilustração)**

A probabilidade do nascimento de uma criança do sexo feminino é igual a 0,5. Um casal deseja saber qual a probabilidade de que três crianças sejam do sexo feminino tem dois nascimentos.

**Solução:**

Nesse caso, cada nascimento não depende do outro e, consequentemente, ser do sexo feminino também não. Daí,

$$P = 0,5 \times 0,5 = 0,25$$

# Cálculo de Probabilidade

## Regra 1 da multiplicação: eventos independentes (Condição)

### Exemplo 15

Para determinar se existe associação entre implantes mamários e doenças do tecido conjuntivo e doenças correlatas, durante vários anos foram observadas 749 mulheres que haviam recebido implante e 1.498 que não haviam recebido implante. Verificou-se que cinco das que haviam recebido implantes e 10 das que não haviam recebido tiveram doenças do tecido conjuntivo. Você acredita que ter doenças do tecido conjuntivo depende ou não de a mulher ter implantes mamários?

# Cálculo de Probabilidade

Tabela 5: Distribuição de mulheres com implante mamário e o fato de terem ou não doenças do tecido conjuntivo e outras correlatas

| Implante mamário | Doenças do tecido conjuntivo e outras | | Total |
|---|---|---|---|
| | Sim | Não | |
| **Sim** | 5 | 744 | 749 |
| **Não** | 10 | 1.488 | 1.498 |
| **Total** | *15* | *2.232* | *2.247* |

**Solução:**

Nesse caso, precisamos verificar a condição de independência:

$$P(\text{“}implante\text{”} \cap \text{“}doença\text{”}) = P(\text{“}implante\text{”}) \times P(\text{“}doença\text{”})$$

# Cálculo de Probabilidade

**Solução:** Continuação

$$P(\text{“}implante\text{”}) = \frac{749}{2.247}$$

$$P(\text{“}doença\text{”}) = \frac{15}{2.247}$$

Daí,

$$P(\text{“}implante\text{”}) \times P(\text{“}doença\text{”}) = \frac{749}{2.247} \times \frac{15}{2.247} = \frac{1}{3} \times \frac{15}{2.247} = \frac{5}{2.247}$$

Por outro lado,

$$P(\text{“}implante\text{”} \cap \text{“}doença\text{”}) = \frac{\#(\text{“}implante\text{”} \cap \text{“}doença\text{”})}{\text{‘}Total'} = \frac{5}{2.247} = P(\text{“}implante\text{”}) \times P(\text{“}doença\text{”})$$

# Cálculo de Probabilidade

## Considerações

- Na área da saúde (assim como em outras áreas), algumas condições podem **favorecer ou dificultar** a ocorrência de determinado evento.

  Exemplo: Sabe-se que:

  - *a probabilidade de ter câncer do pulmão depende de ter ou não o hábito de fumar;*
  - *a probabilidade de ter algumas doenças contagiosas depende de ter ou não sido imunizado.*

- Outras vezes, a probabilidade de ocorrer determinado evento **não depende** da ocorrência de outro.

  Exemplo:

  - *a probabilidade de ter cárie dentária não depende de a pessoa ser ou não míope;*
  - *a probabilidade de ter cálculos renais não depende da profissão.*

# Cálculo de Probabilidade

### Considerações

- Avaliar se há ou não dependência entre determinados eventos em muitas pesquisas que são realizadas significa buscar os **fatores que alteram as probabilidades.**, como visto no Exemplo 15.

# Exercícios no software R

**Atividade prática no R**

Conjunto de dados: dados sobre vacinação contra a Covid-19 à população indígena brasileira...

Fonte:

# Cálculo de Probabilidade

## Regra 2 da multiplicação: eventos dependentes (Ilustração)

Suponha que um levantamento estatístico efetuado em certa população verificou que **23% de indivíduos do sexo masculino** e **18% do sexo feminino** são hipertensos.

Se nessa mesma população, o **percentual de casais hipertensos é de 7,2%**, então existe dependência (ou associação) entre o fato de o homem e a mulher do casal apresentarem hipertensão, pois

**H** = {'homem hipertenso'};
**M** = {'mulher hipertensa'}.

Daí,

$$P(H \cap M) = P(H) \cdot P(M)$$

$$P(H \cap M) = 0,23 \cdot 0,18 = 4,14\% \neq 7,2\% = \frac{\#(H \cap M)}{\#\Omega}$$

# Cálculo de Probabilidade

**Regra 2 da multiplicação: eventos dependentes**

$\hookrightarrow$ Situações em que alguns eventos **têm efeito** sobre a ocorrência de outros (consequentemente, sobre a chance de ocorrência dos mesmo).

$\hookrightarrow$ Nesse sentido, entende-se que alguns eventos estão **condicionados** a outros.

- Algumas situações que sugerem uma relação de dependência:
  - Você só entra no cinema se comprar a entrada. Então, comprar entrada é *condição* para entrar no cinema;
  - *estar alcoolizado aumenta a probabilidade de provocar acidente de trânsito*;
  - *vida sedentária* aumenta a probabilidade de *sobrepeso*.

# Cálculo de Probabilidade

**Probabilidade condicional**

**A**, **B**: eventos.

A probabilidade condicional de **B** dado **A** corresponde a probabilidade de ocorrer o evento **B** sob a condição de **A** ter ocorrido.

Notação: $P(B|A)$: **Probabilidade de B dado A**.

↬ Se **A** e **B** são eventos dependentes, a probabilidade de ocorrer **B** dado que **A** ocorreu, é dada por:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \Longrightarrow P(A \cap B) = P(A) \times P(B|A)$$

sendo $P(A) > 0$.

# Cálculo de Probabilidade

**Exemplo 16**

Um grupo de pessoas foi classificado quanto a peso e pressão arterial, de acordo com as proporções mostradas na Tabela 6.

Tabela 6: Distribuição de um conjunto de pacientes segundo peso e pressão arterial.

| Pressão arterial | Peso | | | Total |
|---|---|---|---|---|
| | Excesso | Normal | Deficiente | |
| Elevada | 0,10 | 0,08 | 0,02 | 0,20 |
| Normal | 0,15 | 0,45 | 0,20 | 0,80 |
| Total | 0,25 | 0,53 | 0,22 | 1,00 |

# Cálculo de Probabilidade

**Exemplo 16 - Continuação**

Considerando que a pessoa escolhida tem excesso de peso, qual a probabilidade de ter também pressão elevada?

**Solução:**

**A:** ter pressão elevada $\Longrightarrow P(A) = 0,20.$
**B:** ter excesso de peso $\Longrightarrow P(B) = 0,25.$
$\mathbf{A} \cap \mathbf{B}$ : ter pressão elevada e ter excesso de peso $\Longrightarrow P(A \cap B) = 0,10.$

Assim,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,10}{0,25} = 0,4.$$

# Cálculo de Probabilidade

**Exemplo 17**

**Medidas de acurácia de testes de diagnósticos**

*O bom uso de um teste diagnóstico requer:*

- *conhecimento de medidas que caracterizam a sua qualidade intríseca;*
- *a sensibilidade (s);*
- *a especificidade (e);*
- *índices que refletem a sua capacidade de produzir decisões clínicas corretas:*
  - *valor da predição positiva (VPP);*
  - *valor da predição negativa (VPN).*

# Cálculo de Probabilidade

**Exemplo 17 - Continuação**

**Teste ergométrico para diagnóstico de doença coronariana**

Wiener *et. al.* (1979) compararam os resultados do teste ergométrico de tolerância a exercícios entre indivíduos com e sem doença coronariana.

O teste foi considerado positivo quando se observou mais de 1mm de depressão ou elevação do segmento ST por no mínimo 0,08s, em comparação com os resultados obtidos com o paciente em repouso.

O diagnóstico definitivo foi feito por angiografia.

A Tabela 7 mostra os resultados encontrados.

# Cálculo de Probabilidade

Tabela 7: Resultados da aplicação do teste ergométrico de tolerância a exercícios em 1.465 pessoas

| **Doença coronoriana** | **Teste ergométrico** | | **Total** |
|---|---|---|---|
| | **Positivo** $(T_+)$ | **Negativo** $(T_-)$ | |
| **Presente** $(D_+)$ | 815 | 208 | 1.023 |
| **Ausente** $(D_-)$ | 115 | 327 | 442 |
| **Total** | *930* | *535* | *1.465* |

*Fonte:* Wiener *et. al.* (1979).

# Cálculo de Probabilidade

**Exemplo 17 - Continuação**

A sensibilidade e a especificidade são estimadas por:

$$s = P(T_+|D_+) = \frac{327}{1.023} = 0,80;$$

$$e = P(T_-|D_-) = \frac{815}{442} = 0,74.$$

# Cálculo de Probabilidade

**Exemplo 18**

**Metástase de carcinoma hepático**

*Lind & Singer (1986) estudaram a qualidade da tomografia computadorizada para o diagnóstico de metástase de carcinoma de fígado, obtendo os resultados sintetizados na Tabela 8.*

*O total de 150 pacientes foram submetidos a dois exames: a tomografia computadorizada e a laparotomia.*

*Este último é tomado como o padrão-ouro, isto é, classifica o paciente sem erro.*

*A Tabela 8 mostra os resultados encontrados.*

# Cálculo de Probabilidade

Tabela 8: Resultados da tomografia computadorizada em 67 pacientes com metástase e 83 sem metástase do carcinoma hepático

| Metástase de carcinoma hepático | Tomografia computadorizada | | Total |
|---|---|---|---|
| | Positiva $(T_+)$ | Negativa $(T_-)$ | |
| **Presente** $(D_+)$ | 52 | 15 | 67 |
| **Ausente** $(D_-)$ | 9 | 74 | 83 |
| **Total** | *61* | *89* | *150* |

*Fonte:* Lind $\&$ Singer (1986).

# Cálculo de Probabilidade

**Algumas considerações importantes:**

**A**, **B**: eventos.

- Embora $P(A \cap B) = P(B \cap A)$ Note que os universos de referência são diferentes. Nesse sentido, pela definição dada:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \neq \frac{P(B \cap A)}{P(A)} = P(B|A)$$

- Quando **A** e **B** são *independentes*, resulta que:

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

# Cálculo de Probabilidade

**Teorema de Bayes - Ilustração 1**

Carlos, quando almoça fora, tem 40% de chance de escolher uma churrascaria. Quando escolhe ir a uma churrascaria, tem 85% de chance de sair da dieta. Entretanto, mesmo quando Carlos não vai à churrascaria, tem 45% de chance de sair da dieta. Carlos quebrou a dieta. Qual a probabilidade de ele ter ido à churrascaria?

Solução:

**A:** ir à churrascaria; (o que queremos saber)

**B:** sair da dieta; (o que sabemos)

# Cálculo de Probabilidade

| | | | |
|---|---|---|---|
| Ir à churrascaria (0,40) | Sair da dieta (0,85) | $\Rightarrow 0,4 \times 0,85 =$ | 0,34 |
| | Não sair da dieta (0,15) | $\Rightarrow 0,4 \times 0,15 =$ | 0,06 |
| Não ir à churrascaria (0,60) | Sair da dieta (0,45) | $\Rightarrow 0,6 \times 0,45 =$ | 0,27 |
| | Não sair da dieta (0,15) | $\Rightarrow 0,6 \times 0,55 =$ | 0,33 |
| | | | 1,00 |

Figura 4: Árvore decisória que contém os possíveis resultados e suas probabilidades de ocorrência.

# Cálculo de Probabilidade

| | | | |
|---|---|---|---|
| Ir à churrascaria (0,40) | Sair da dieta (0,85) | $\Rightarrow 0,4 \times 0,85 =$ | $0,34$ |
| | Não sair da dieta (0,15) | $\Rightarrow 0,4 \times 0,15 =$ | $0,06$ |
| Não ir à churrascaria (0,60) | Sair da dieta (0,45) | $\Rightarrow 0,6 \times 0,45 =$ | $0,27$ |
| | Não sair da dieta (0,15) | $\Rightarrow 0,6 \times 0,55 =$ | $0,33$ |
| | | | 1,00 |

Figura 5: Árvore decisória que contém os possíveis resultados e suas probabilidades de ocorrência.

# Cálculo de Probabilidade

**Teorema de Bayes - Ilustração 1 (Continuação)**

Note que:

$P(A \cap B) = 0,34;$
$P(B) = 0,34 + 0,27 = 0,61;$ pois Carlos pode sair da dieta quando vai OU quando não vai à churrascaria.

Portanto, a probabilidade de Carlos ter ido à churrascaria é:

$$P(A|B) = \frac{0,34}{0,34 + 0,27} = 0,5574 = 55,74\%$$

# Cálculo de Probabilidade

**Teorema de Bayes - Ilustração 2**

Uma das formas de avaliar a eficiência de um teste para detectar uma doença é quantificar a probabilidade de erro. Em geral, testes sofisticados envolvem vários procedimentos laboratoriais e diversos equipamentos. Denominamos falso-positivo ao erro em que o teste indica positivo para um paciente que não tem a doença. Por outro lado, teremos um erro falso-negativo se o teste não acusar a doença num paciente doente. Os erros originam doentes sadios e sadios doentes. As probabilidades dos erros são calculadas condicionalmente à situação do paciente. Seus complementares fornecem as probabilidades de acerto do teste.

# Cálculo de Probabilidade

**Teorema de Bayes - Ilustração 2 (Continuação)**

Para fixar as ideias, considere que um determinado teste resulta positivo para não doentes, com probabilidade 0,1 . Também com probabilidade 0,1 , o teste será negativo para um paciente doente. As informações fornecidas se referem aos erros que podem ser cometidos ao realizar o teste. Se a incidência da doença na população é de 1 para cada 10 mil habitantes, qual é a probabilidade de uma pessoa estar realmente doente se o teste deu positivo? Definindo os eventos:

**D:** a pessoa está doente

**A:** o teste é positivo

# Cálculo de Probabilidade

## Teorema de Bayes - Ilustração 2 (Continuação)

Assim as informações disponíveis são as seguintes:

$P(D) = 0,0001$

$P(A|\overline{D}) = 0,1$

$P(\overline{A}|D) = 0,1$

Ainda podemos escrever os complementares:

$P(\overline{D}) = 0,9999$

$P(A|D) = 0,9$

# Cálculo de Probabilidade

## Teorema de Bayes - Ilustração 2 (Continuação)

A probabilidade que desejamos calcular é $P(D|A)$. Logo:

$$P(D|A) = \frac{P(A|D) \cdot P(D)}{P(A|D) \cdot P(D) + P(A|\overline{D}) \cdot P(\overline{D})}$$

$$P(D|A) = \frac{0,9 \cdot 0,0001}{0,09 \cdot 0,0001 + 0,1 \cdot 0,9999}.$$

$$P(D|A) = 0,0009$$

Portanto a probabilidade de estar doente dado que o teste deu positivo é 0,0009, ou ainda, é aproximadamente 1 em 1000.

# Exercícios no software R

**Atividade prática no R**

Vamos usar dados extraídos do site do Ministério da Saúde - Vacinômetro COVID-19:
`https://infoms.saude.gov.br/extensions/SEIDIGI_DEMAS_Vacina_C19/SEIDIGI_DEMAS_Vacina_C19.html`

# Referências Bibliográficas


- BUSSAB, W. O; MORETTIN, P. A. Estatística básica. 8. ed. São Paulo: Saraiva, 2013. 548 p.
- FREI, F. Introdução à Inferência Estatística – aplicações em saúde e biologia. 1. ed. São Paulo: Interciência, 2018 546 p.
- VIEIRA, Sônia. Introdução à bioestatística. 6ª Edição. Elsevier Brasil, 1997.
- ARANGO, Héctor Gustavo. Bioestatística teórica e computacional. In: Bioestatística teórica e computacional. 2001. p.
- `https://infoms.saude.gov.br/extensions/SEIDIGI_DEMAS_Vacina_C19/SEIDIGI_DEMAS_Vacina_C19.html`

# Especialização em *Data Science* e Estatística Aplicada

## Módulo I - Análise Estatística para *Data Science*

**Profa. Dra. Renata M. R. Vasconcelos**

**renatamrv@ufg.br**

IME
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

FEN
FACULDADE DE ENFERMAGEM

UFG
UNIVERSIDADE FEDERAL DE GOIÁS