Bioestatística F

Conceitos de Probabilidade

Enrico Antonio Colosimo
Depto. Estatística – UFMG
http://www.est.ufmg.br/~enricoc/

Probabilidade

- **Análise Descritiva**: exploração através de gráficos e tabelas dos dados brutos.
- **Análise Inferencial ou confirmatória**: controle da incerteza para tomar decisões. Baseado em teoria da probabilidade.

Reações à Probabilidade

Tipos de Fenômenos

1. **Fenômeno aleatório**: situação ou acontecimentos que não podem ser previstos com certeza.
2. **Fenômeno determinístico**: situação ou acontecimentos que são previstos com certeza.

   Exemplos: leis físicas de Newton

Teoria de Probabilidade

1. Teoria de Probabilidade é um conjunto de métodos para lidar com incerteza existente nos fenômenos aleatórios.
2. É fundamental para entender as técnicas estatísticas.
3. Exemplos: prevalência, testes diagnósticos (sensibilidade, especificidade, PFP, PFN)

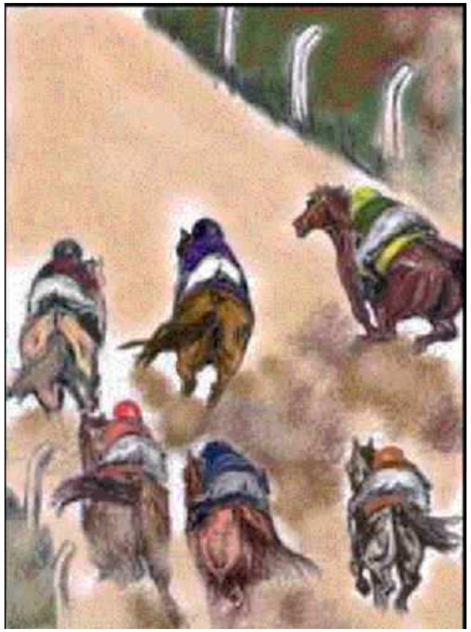
Situação Corrida de Cavalos

- Suponha que você tenha R$ 10.000,00 para apostar em uma corrida de cavalos. Abaixo estão as probabilidades dos cavalos A, B e C de ganharem. Em qual cavalo você apostaria?

Cavalo A: 75%

Cavalo B: 20%

Cavalo C: 5%



Probabilidade

- Conceitos Básicos:

1. **Espaço amostral**: conjunto de todos os resultados possíveis de um experimento (Ω)
2. **Eventos**: subconjuntos de Ω (representados por letras maiúsculas: A, B, ...); conjunto vazio:ϕ.
3. **Operações com Eventos/Conjuntos** : união, intersecção, complementar, etc.

Exemplo (a)

- Um dado equilibrado é lançado:
  - **Experimento**: lançar um dado (ξ_1).
  - **Característica de interesse**: nº. face superior.
  - **Espaço amostral**: $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - **Eventos**:
    - $A = \{1, 3, 5\}$, resultados ímpares
    - $B = \{2, 4, 6\}$, resultados pares
    - $C = \{1, 2, 3\}$, os três menores resultados

Operações sobre conjuntos

- A *união* de dois eventos A e B representa a ocorrência de, pelo menos, um dos eventos A ou B. A *união* é denotada por $A \cup B$
- Exemplo: (experimento do dado)

  - A = "observa-se um número ímpar"
  - C = "observa-se um número ≤ 3"
  - $A \cup C = \{1, 2, 3, 5\}$

Operações sobre conjuntos

- A *interseção* do evento A com B é a ocorrência simultânea de A e B. A *interseção* é denotada por $A \cap B$
- Exemplo: (experimento do dado)

  - A = "observa-se um número ímpar"
  - C = "observa-se um número ≤ 3"
  - $A \cap C = \{1, 3\}$

Operações sobre conjuntos

- Dois eventos A e B são *disjuntos* ou *mutuamente exclusivos* quando não têm elementos em comum, isto é, $A \cap B = \emptyset$
- Exemplo: (experimento do dado)
  - A = "observa-se um número ímpar"
  - B = "observa-se um número par"
  - $A \cap B = \emptyset$

Operações sobre conjuntos

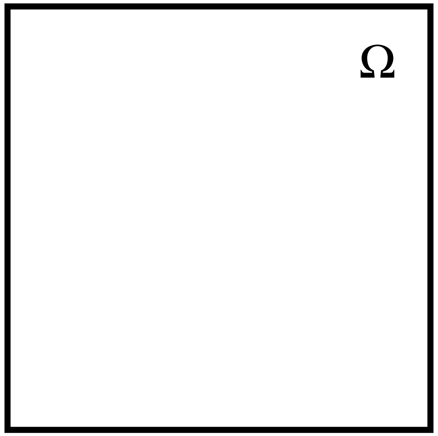
- Dizemos que A e B são *complementares* se sua união é o espaço amostral e sua interseção é vazia. O complementar de A é representado por A^c e temos
  $$A \cup A^c = \Omega \quad \text{e} \quad A \cap A^c = \emptyset$$
- Exemplo: (experimento do dado)
  - $\text{C} = \text{“observa-se um número} \leq 3\text{”}$
  - $C^c = \{4, 5, 6\}$

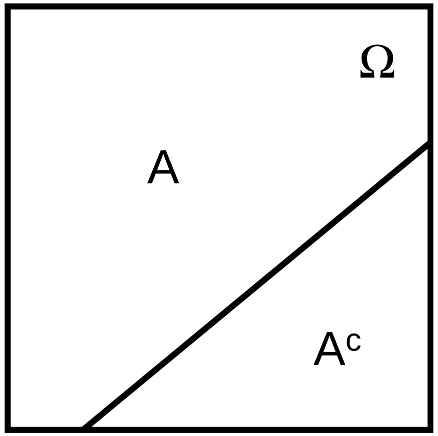
Axiomas de Probabilidade

- É uma função P(.) é denominada probabilidade se satisfaz as seguintes condições:

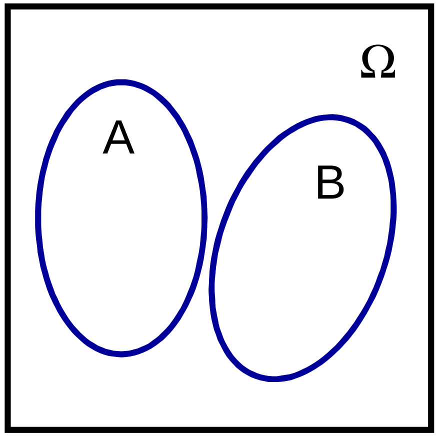
1. $0 \leq P(A) \leq 1$, para qualquer evento A em Ω
2. $P(\Omega) = 1$
3. $P(A \cup B) = P(A) + P(B)$ se A e B são disjuntos

Diagrama de Venn

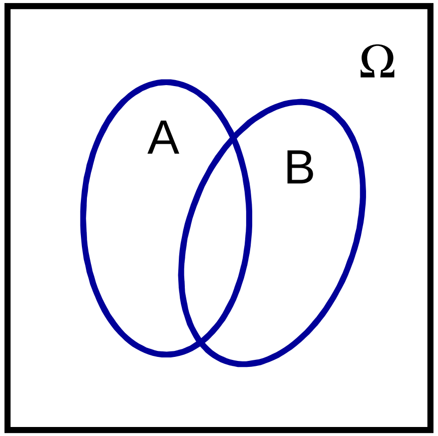




$A \cup A^c = \Omega$, $A \cap A^c = \phi$



$A \cap B = \phi$



$A \cup B$

$A \cap B$

Pelo Diagrama de Venn mostre que:
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Exemplo (Hipertensão)

Seja A o evento que uma pessoa tenha pressão diastólica normotensiva (DBP < 90), e seja B o evento que a pessoa tenha borderline DBP (90 ≤ DBP ≤ 95). Suponha que P(A)=0,7 e P(B)=0,1. Seja Z o evento que a pessoa tem DBP ≤ 95. Assim:

P(Z) = P(A∪B) = P(A)+P(B)=0,8

pois os eventos não podem ocorrer simultaneamente $(A \cap B = \emptyset)$

Exemplo (Hipertensão)

Sejam os eventos A e B definidos como no exemplo anterior:

$A = \{DBP < 90\}$, e $B = \{90 \leq DBP \leq 95\}$

Então, $A \cup B = \{DBP < 95\}$

Sejam C o evento $\{DBP \geq 90\}$ e D o evento

$\{75 \leq DBP \leq 100\}$. Os eventos C e D *não são mutuamente* excludentes, ambos podem ocorrer quando $90 \leq DBP \leq 100$, ou seja $C \cap D = \{90 \leq X \leq 100\}$.

Probabilidade Condicional

Dados dois eventos A e B, a probabilidade condicional de A **dado que ocorreu B**, **P(A|B)** é dada por:

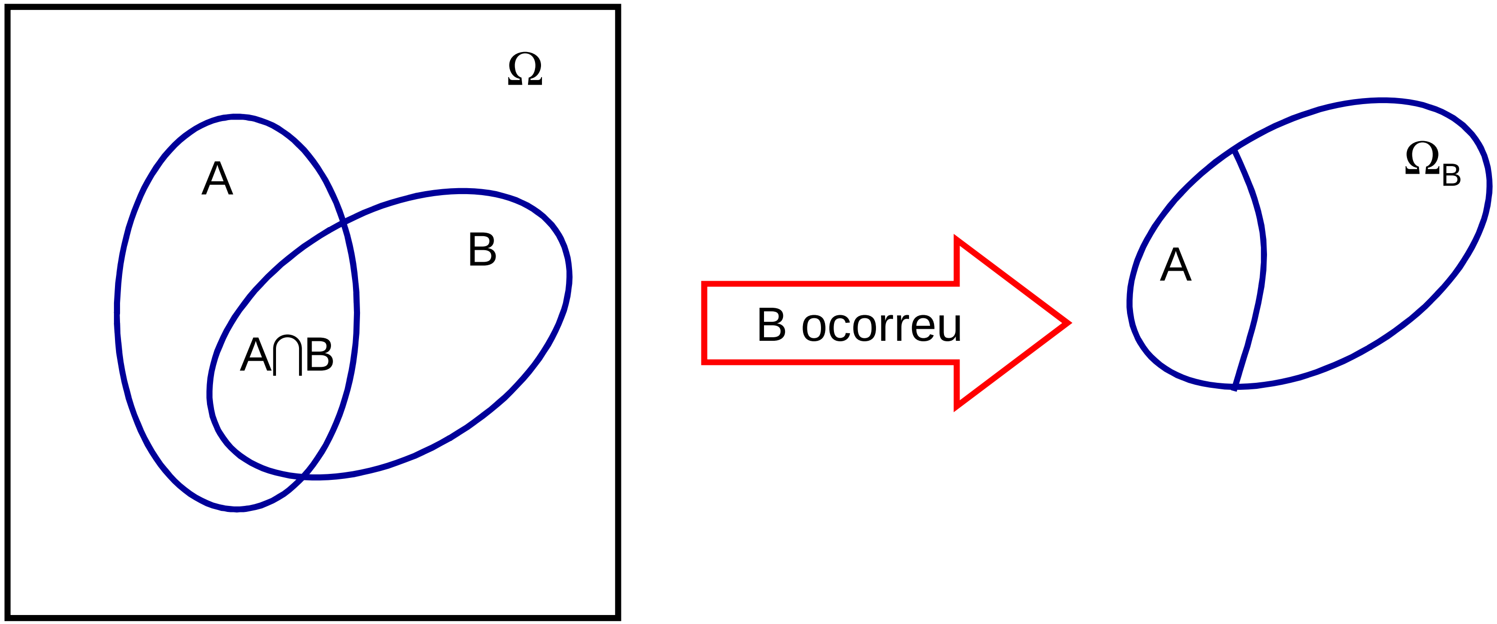
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0 \quad \text{ou seja} \quad P(B) \neq 0$$

Regra do Produto: $P(A \cap B) = P(A \mid B) \cdot P(B)$

P(A|B) é a proporção de A dentro de B onde o "novo" espaço amostral (Ω_B) é o espaço amostral de B tal que:

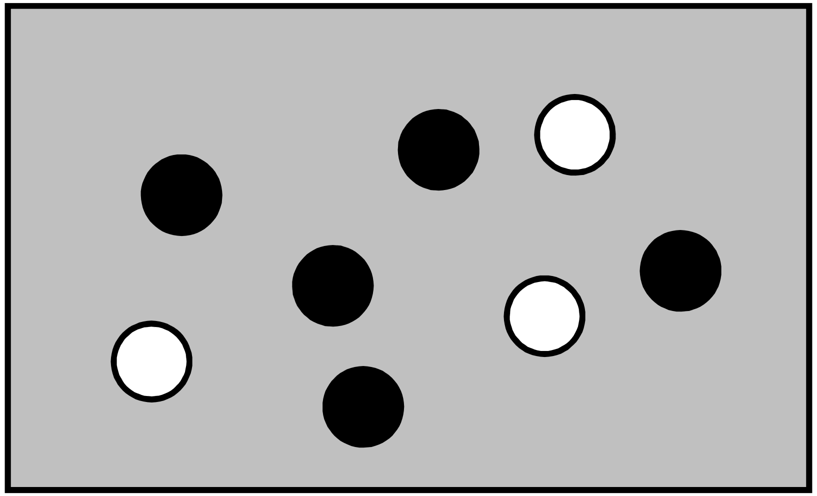
$$0 \leq P(A \mid B) \leq 1$$

Probabilidade Condicional
Interpretação Gráfica



Exemplo Didático

- Uma caixa contém 3 bolas brancas e 5 pretas, todas idênticas em peso e tamanho.
  - Qual a probabilidade de tirar ao acaso uma bola branca?
  - Qual a probabilidade de tirar ao acaso uma bola preta?
  - Qual a probabilidade de tirar uma segunda bola branca se a primeira foi preta (sem reposição)?



Independência de Eventos

Dois eventos **A** e **B** são independentes se a informação da ocorrência (ou não) de B não altera a probabilidade da ocorrência do evento **A**.

$$P\left(A \mid B\right) = P(A)$$

ou na forma:

$$P\left(A \cap B\right) = P(A) \cdot P(B)$$

Exemplo (Hipertensão)

Suponha que estamos realizando uma avaliação de hipertensão em famílas. Seja o espaço amostral constituído de todos os pares da forma (X,Y), onde $X>0$ e $Y>0$, que representam as medições de DBP. Seja o evendo A={DBP da mãe ≥ 95}, B={DBP do pai ≥ 95}, caso de hipertensão. Suponha também que $P(A)=0,1$ e $P(B)=0,2$.

Se A e B são independentes, então a probabilidade que a mãe e pai sejam hipertensos será:

$$P(A \cap B) = P(A)*P(B) = 0,1*0,2 = 0,02 = 2\%$$

Exemplo (Hipertensão)

Uma forma de interpretar isso é considerar que o status hipertensivo da mãe não depende do status hipertensivo do pai. Assim, considerando independência, em 10% das casas onde o pai é hipertensivo a mãe é também hipertensiva e em 10% das casas onde o pai *não* é hipertensivo a mãe será hipertensiva.

Admitimos nesse exemplo que as causas da hipertensão são genéticas; se fossem, em certo ponto, ambientais, poderíamos admitir mais probabildiade de a mãe ser hipertensa quando o pai é hipertenso (B ocorre), do que quando o pai não é hipertenso (B não ocorre).

Exemplo (DST)

Suponha que dois médicos, A e B, testem os pacientes para verificar se possuem sífilis. Sejam os eventos:

A^+={médico A dá diagnóstico positivo}, e

B^+={médico B dá diagnóstico positivo}.

Suponha que o médico A diagnostica 10% dos pacientes como positivo e o médico B 17% e ambos os médicos diagnostiquem 8% dos pacientes como como positivo. Os eventos A^+ e B^+ são independentes?

$P(A^+)*P(B^+)=0,10*0,17=0,017$

$P(A^+ \cap B^+)=0,08$

Ou seja, os eventos não são independentes.

Exemplo (DST)

Qual a probabilidade de que o médico B apresente um diagnóstico positivo de sífilis, dado que o médico A fez um diagnóstico positivo?

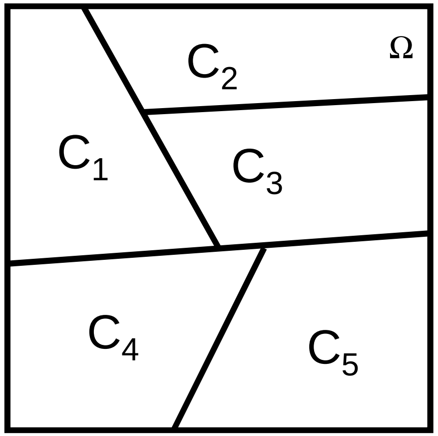
$$P(B^+|A^+) = \frac{P(B^+ \cap A^+)}{P(A^+)} = \frac{0,08}{0,1} = 0,8$$

Assim, o médico B confirmará o diagnóstico positivo dado pelo médico A em 80% das vezes.

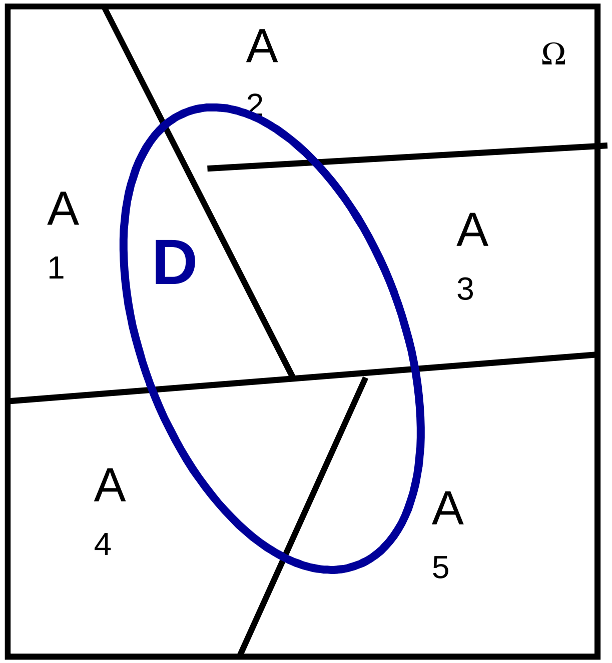
Partição do Espaço Amostral

Os eventos C1, C2, ..., Ck formam uma partição do espaço amostral se eles não tem interseção (são disjuntos) entre si e se sua união é igual ao espaço amostral.

$$C_i \cap C_j = \phi, \quad \forall\, i \neq j \qquad\qquad \bigcup_{i=1}^{k} C_i = \Omega$$



Partição do Espaço Amostral



$$D = (A_1 \cap D) \cup (A_2 \cap D) \cup (A_3 \cap D) \cup (A_4 \cap D) \cup (A_5 \cap D)$$

$$\boxed{\begin{aligned} P(D) &= P(A_1 \cap D) + P(A_2 \cap D) + P(A_3 \cap D) + P(A_4 \cap D) + P(A_5 \cap D) \\ &= P(D|A_1)P(A_1) + P(D|A_2)P(A_2) + P(D|A_3)P(A_3) + P(D|A_4)P(A_4) + P(D|A_5)P(A_5) \end{aligned}}$$

Partição do Espaço Amostral

As probabilidades condicionais P(B/A) e P(B/A^c) e não condiconal P(B) estão relacionadas da seguinte forma:

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Exemplo (Câncer)

Suponha que entre 100.000 mulheres com mamografias negativas 20 terão câncer de mama em dois anos, ou seja

$$P(B/A^c)=20/100.000=0,0002$$

enquanto que 1 em 10 com mamografias positivas terão câncer em dois anos, ou seja

$$P(B/A)=1/10=0,1$$

Suponha, ainda, que 7% da população geral das mulheres terão mamografia positiva.

Exemplo (Câncer)

Qual a probabilidade de uma mulher desenvolver câncer de mama nos próximos dois anos?

$P(B)=P(\text{câncer})$

$=P(\text{câncer}/\text{mam.}^{+})*P(\text{mam.}^{+})+P(\text{câncer}/\text{mam.}^{-})*P(\text{mam.}^{-})$

$=0,1*0,07+0,0002*0,93=0,00719$

Exemplo (Oftalmologia)

Estamos planejando um estudo de 5 anos sobre catarata numa população de cinco mil pessoas com 60 anos ou mais. Sabemos do Censo que 45% dessa população tem entre 60-64 anos, 28% entre 65-69, 20% entre 70-74 e 7% tem 75 anos ou mais.

Também sabemos de um estudo anterior que 2,4%, 4,6%, 8,8% e 15,3% das pessoas nos respectivos grupos de idade desenvolverão catarata nos próximos 5 anos.

Exemplo (Oftalmologia)

Qual a porcentagem dessa população desenvolverá catarata em 5 anos e quantas pessoas isso representa?

Então A_1={idades 60-64}, $P(A_1)=0,45$,

A_2={idades 65-69}, $P(A_2)=0,28$

A_3={idades 70-64}, $P(A_3)=0,20$

A_4={idades ≥75}, $P(A_4)=0,07$

B={desenvolverá catarata em 5 anos}

Exemplo (Oftalmologia)

$P(B/A_1)=0{,}024$, $P(B/A_2)=0{,}046$,
$P(B/A_3)=0{,}088$, $P(B/A_4)=0{,}153$

Desta forma:

$$P(B) = P(B/A_1)*P(A_1)+P(B/A_2)*P(A_2) +P(B/A_3)*P(A_3)+P(B/A_4)*P(A_4)$$

$$= 0{,}024*0{,}45 + 0{,}046*0{,}28 + 0{,}088*0{,}20 + 0{,}153*0{,}07 = 0{,}052$$

Conclusão: 5,2% da população, ou seja, 260 (=5.000*0,052) pessoas desenvolverão catarata nos próximos 5 anos.