# *Introdução à Probabilidade*

Silvia Shimakura

*silvia.shimakura@ufpr.br*

# *Probabilidade*

- **O que é probabilidade?**

  **Medida que quantifica a incerteza** de um acontecimento futuro.

- **Como quantificar incerteza?**

  **Definição clássica x Definição frequentista**

# *Problema 1*

- Experimento 1: Lançamento de um dado balanceado

- Espaço amostral: conjunto de todos os resultados possíveis

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

- Evento B: face par

$$B=\{2, 4, 6\}$$

# Cálculo de probabilidades

- Se os elementos de E forem equiprováveis a probabilidade de um evento A ocorrer:

$$P(A)=\frac{\text{número de elementos em } A}{\text{número de elementos em } E}$$

- Experimento 1: E={1,2,3,4,5,6} → 6 elementos (equiprováveis)
  - Evento B: B={2, 4, 6} → P(B)=3/6

# Problema 2

- Experimento 2: Lançamento de 2 dados balanceados

- Espaço amostral:

$E=\{(1,1), (1,2), \ldots, (1,6), (2,1), (2,2), \ldots, (2,6), \ldots, (6,6)\}$

$\rightarrow$ 36 elementos $\rightarrow$ equiprováveis

- Evento F: a soma dos dois valores é 10

$$F=\{(4,6), (5,5), (6,4)\} \rightarrow P(F)=3/36$$

# Problema 3

- Experimento 3: Lançamento de uma moeda
- Espaço amostral: E={Cara, Coroa}
- Evento C: Cara

  C={Cara}

- Elementos de E são equiprováveis?
- P(C)=?

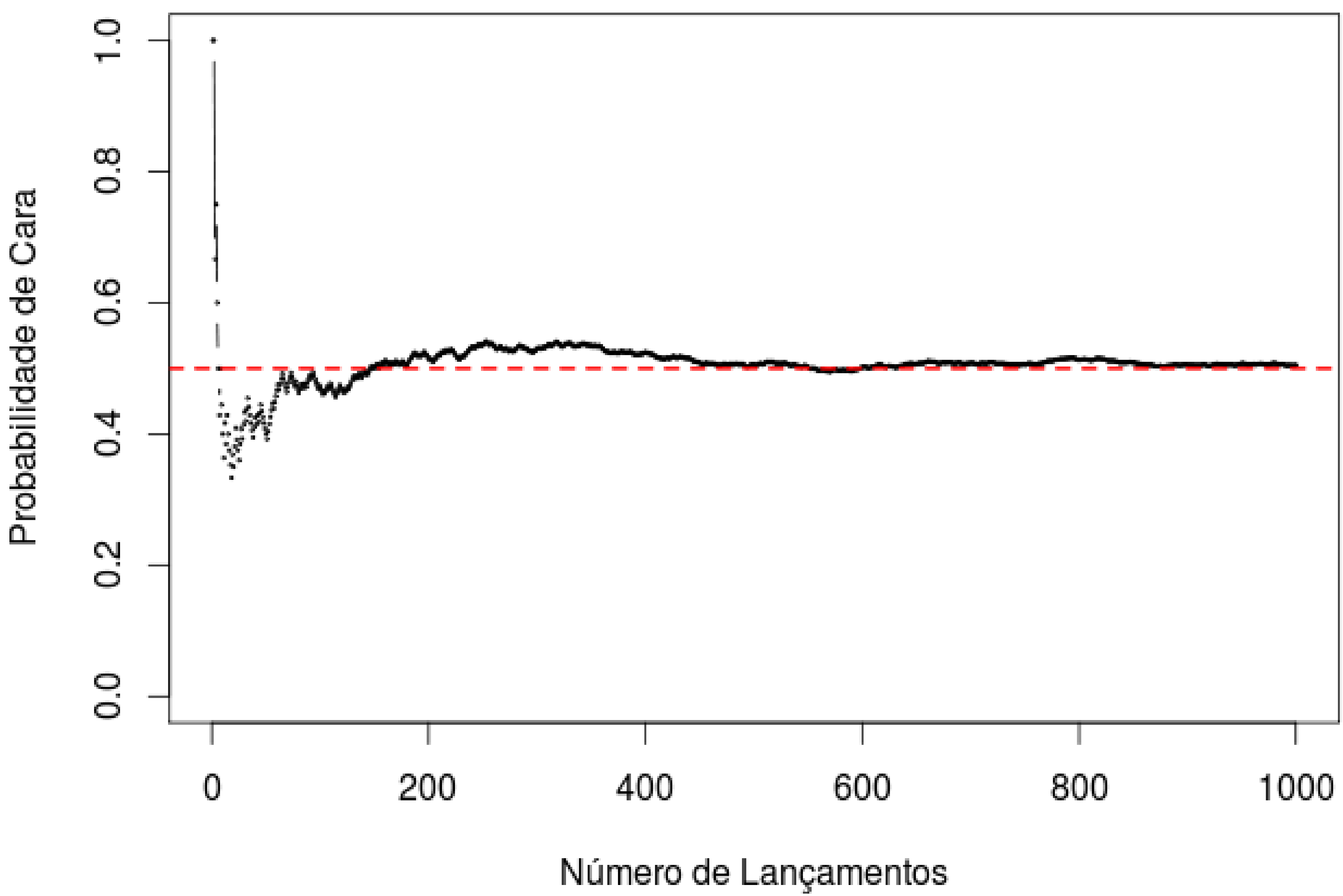
# Visão frequentista de probabilidade

- Probabilidade: frequência relativa de ocorrência do evento para um grande número de sorteios

# Frequência relativa

- C: Cara    O: Coroa

| Resultado | C | C | C | O | C | O | O | O | O | O | O | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequência acumulada de Caras | 1 | 2 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 |
| Número de lançamentos | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Freq. relativa de caras | 1/1 | 2/2 | 3/3 | 3/4 | 4/5 | 4/6 | 4/7 | 4/8 | 4/9 | 4/10 | 4/11 | 5/12 |
| % | 100 | 100 | 100 | 75 | 80 | 67 | 57 | 50 | 44 | 40 | 36 | 42 |



Probabilidade de Cara
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000
Número de Lançamentos

# Tipos especiais de eventos

- Evento complementar de B: elementos de E que não estão em B

E={1, 2, 3, 4, 5, 6}   B={2, 4, 6}

$$\bar{B} = \{1, 3, 5\}$$

- Evento interseção: elementos estão em A e em B

A={1, 3}   B={2, 4, 6}   $A \cap B = \emptyset$

- Evento união: elementos que estão em A ou B

$A \cup B = \{1, 2, 3, 4, 6\}$

# *Tipos especiais de eventos*

- Eventos mutuamente exclusivos: não existem elementos comuns em A e B

$$A=\{1, 3\} \quad B=\{2, 4, 6\} \quad A \cap B=\emptyset$$

# Propriedades de probabilidade

- $0 \leq P(A) \leq 1$, para qualquer evento A
- $P(E)=1$, em que E é o espaço amostral
- $\boxed{P(\bar{A})=1-P(A)}$
- Para dois eventos A e B quaisquer, a probabilidade de que A ou B ocorra:

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

- Se A e B são mutuamente exclusivos, a probabilidade de que A ou B ocorra é a soma das probabilidades.

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

# *Probabilidade condicional*

- $P(A|B)$=P(A ocorrer dado que B ocorreu)
- Para A e B quaisquer:

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

- Para A e B independentes: $P(A|B)=P(A)$

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

# Exemplo: Lançamento de dois dados não viciados

- Espaço amostral: 36 elementos com prob 1/36 cada
- Evento A: soma dos dados é 8
- A=?
- P(A)=?

# Exemplo: Lançamento de dois dados não viciados

- Espaço amostral: 36 elementos com prob 1/36 cada
- Evento A: soma dos dados é 8
- A={(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)}
- P(A)=?

# Exemplo: Lançamento de dois dados não viciados

- Espaço amostral: 36 elementos com prob 1/36 cada
- Evento A: soma dos dados é 8
- A={(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)}
- P(A)=5/36

# Exemplo: Lançamento de dois dados não viciados

- Espaço amostral: 36 elementos com prob 1/36 cada
- Evento A: soma dos dados é 8
- A={(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)}
- P(A)=5/36

Se soubermos que o resultado no primeiro dado é 3, qual será a probabilidade da soma dos dois dados ser 8?

# *Exemplo (cont.)*

- Evento A:

  **A={(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)} → P(A)=5/36**

- Evento B: primeiro dado é 3

  **B={(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)} → P(B)=6/36**

- Evento interseção: **A∩B={(3,5)} → P(A∩B)=1/36**

- P(A|B)=?
- A e B são independentes?
- A e B são mutuamente exclusivos?

# *Exemplo (cont.)*

- Evento A:

  **A={(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)} → P(A)=5/36**

- Evento B: primeiro dado é 3

  **B={(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)} → P(B)=6/36**

- Evento interseção: **A∩B={(3,5)} → P(A∩B)=1/36**

- P(A|B)= **P(A∩B)/P(B)=(1/36)/(6/36)=1/6**
- A e B são independentes?
- A e B são mutuamente exclusivos?

# Exemplo: Distribuição de peso e pressão arterial

| Pressão arterial | Peso | | | |
|---|---|---|---|---|
| | Excesso | Normal | Deficiente | Total |
| Elevada | 0,10 | 0,08 | 0,02 | 0,2 |
| Normal | 0,15 | 0,45 | 0,20 | 0,8 |
| Total | 0,25 | 0,53 | 0,22 | 1 |

- Probabilidade de uma pessoa escolhida ao acaso ter pressão elevada?

- Probabilidade de uma pessoa escolhida ao acaso ter pressão elevada e excesso de peso?

- Sabendo que a pessoa tem excesso de peso, qual a probabilidade de uma pessoa escolhida ao acaso ter pressão elevada?

# Exemplo: Distribuição de peso e pressão arterial (cont.)

- Peso em excesso e pressão arterial normal são eventos mutuamente exclusivos?

- Pressão arterial e peso são independentes?

| Pressão arterial | Peso | | | |
|---|---|---|---|---|
| | Excesso | Normal | Deficiente | Total |
| Elevada | 0,10 | 0,08 | 0,02 | 0,2 |
| Normal | 0,15 | 0,45 | 0,20 | 0,8 |
| Total | 0,25 | 0,53 | 0,22 | 1 |

# *Exemplo: Qualidade de teste diagnóstico*

Num estudo sobre a qualidade do teste ergométrico, Wrinner et al. (1979) compararam os resultados obtidos entre indivíduos com (D+) e sem (D-) doença coronariana.

- T+ :teste positivo        T- :teste negativo

- **T+**: mais de 1mm de depressão ou elevação do segmento ST, por pelo menos 0,08s, em comparação com paciente em repouso.
- **D+** e **D-**: angiografia (teste padrão ouro).

# Exemplo: Qualidade de teste diagnóstico

| Doença coronariana | Teste Ergométrico | | | | | |
|---|---|---|---|---|---|---|
| | T+ | | T- | | Total | |
| D+ | 815 | a | 208 | b | 1023 | a+b |
| D- | 115 | c | 327 | d | 442 | c+d |
| Total | 930 | a+c | 535 | b+d | 1465 | n |

Temos interesse em conhecer as probabilidades de acerto do teste:

- Probabilidade de teste positivo num paciente doente
- Probabilidade de teste negativo num paciente não doente

# Exemplo: Qualidade de teste diagnóstico

| Doença coronariana | Teste Ergométrico | | | | | |
|---|---|---|---|---|---|---|
| | T+ | | T- | | Total | |
| D+ | 815 | a | 208 | b | 1023 | a+b |
| D- | 115 | c | 327 | d | 442 | c+d |
| Total | 930 | a+c | 535 | b+d | 1465 | n |

Temos interesse em conhecer as probabilidades de acerto do teste:

- Probabilidade de teste positivo num paciente doente
  $P(T+|D+)$ = sensibilidade = s
- Probabilidade de teste negativo num paciente não doente
  $P(T-|D-)$ = especificidade = e

# Exemplo: Qualidade de teste diagnóstico

| Doença coronariana | Teste Ergométrico | | | | | |
|---|---|---|---|---|---|---|
| | T+ | | T- | | Total | |
| D+ | 815 | a | 208 | b | 1023 | a+b |
| D- | 115 | c | 327 | d | 442 | c+d |
| Total | 930 | a+c | 535 | b+d | 1465 | n |

Temos interesse em conhecer as probabilidades de acerto do teste:

- Probabilidade de teste positivo num paciente doente
  $s = 815/1023 = 0{,}80$
- Probabilidade de teste negativo num paciente não doente
  $e = 327/442 = 0{,}74$

# Exemplo: Qualidade de teste diagnóstico (cont.)

Do ponto de vista do médico que precisa fechar o diagnóstico além da sensibilidade e especificidade do teste é importante conhecer as seguintes probabilidades:

- Probabilidade de que uma pessoa com resultado positivo no teste realmente tenha a doença
- Probabilidade de que uma pessoa com resultado negativo no teste realmente não tenha a doença

# Exemplo: Qualidade de teste diagnóstico (cont.)

Do ponto de vista do médico que precisa fechar o diagnóstico além da sensibilidade e especificidade do teste é importante conhecer as seguintes probabilidades:

- Probabilidade de que uma pessoa com resultado positivo no teste realmente tenha a doença

$P(D+|T+)=VPP$

- Probabilidade de que uma pessoa com resultado negativo no teste realmente não tenha a doença

$P(D-|T-)=VPN$

# Exemplo: Qualidade de teste diagnóstico

| Doença coronariana | Teste Ergométrico | | | | | |
|---|---|---|---|---|---|---|
| | T+ | | T- | | Total | |
| D+ | 815 | a | 208 | b | 1023 | a+b |
| D- | 115 | c | 327 | d | 442 | c+d |
| Total | 930 | a+c | 535 | b+d | 1465 | n |

Podemos obter VPP e VPN direto da tabela acima?

- VPP=P(D+|T+)=
- VPN=P(D-|T-)=

# Teorema de Bayes

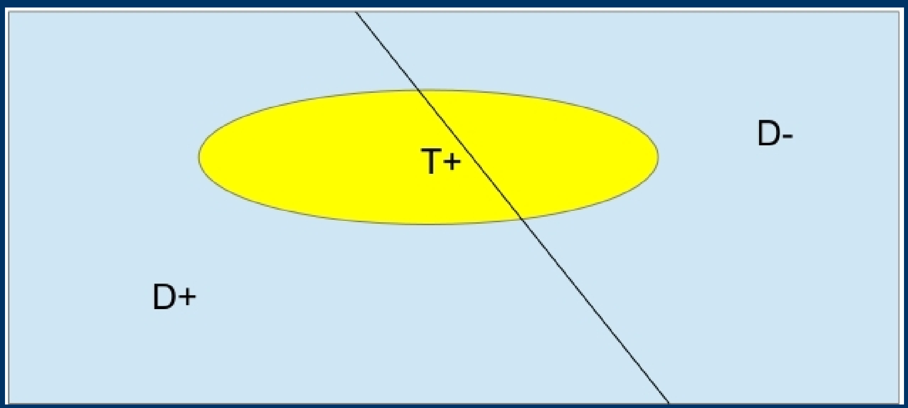
- Se A1, A2 são eventos mutuamente exclusivos e exaustivos:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) = 1$$

- Pelo Teorema de Bayes:

$$P(A_i|B) = \frac{P(A_i)\,P(B|A_i)}{P(A_1)\,P(B|A_1) + P(A_2)\,P(B|A_2)}$$

# Exemplo: Qualidade de teste diagnóstico

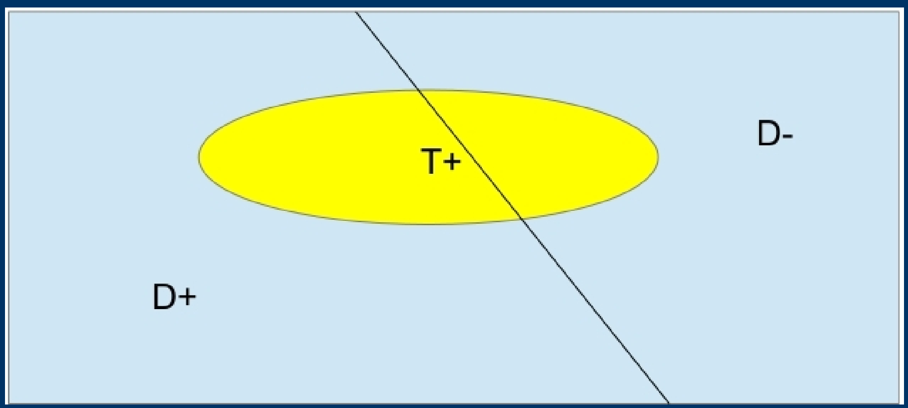



- D+ ∩ D- = ∅
- D+ ∪ D- = E
- Conhecidos:

P(T+|D+)=s    P(T-|D-)=e

$$VPP = P(D+|T+) = \frac{P(D+ \cap T+)}{P(T+)}$$

# Exemplo: Qualidade de teste diagnóstico



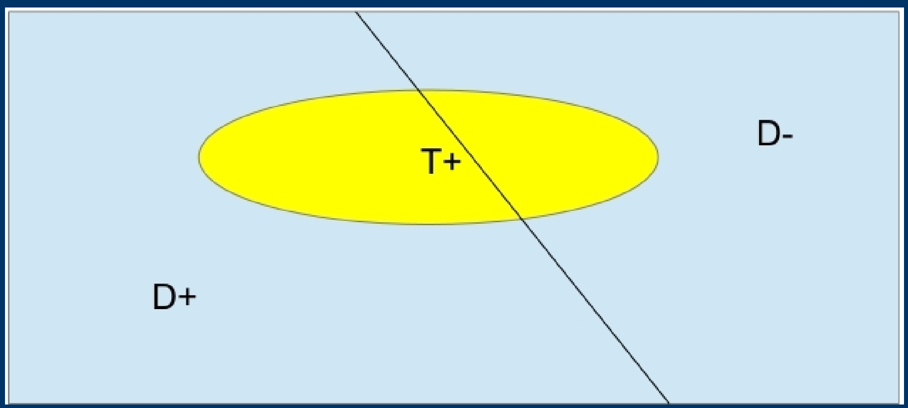

- D+ ∩ D- = ∅
- D+ ∪ D- = E
- Conhecidos:

$P(T+|D+)=s \quad P(T-|D-)=e$

$$P(T+|D+)=\frac{P(T+\cap D+)}{P(D+)} \Rightarrow P(T+\cap D+)=P(T+|D+)P(D+)$$

$$VPP=P(D+|T+)=\frac{P(D+\cap T+)}{P(T+)}$$

# Exemplo: Qualidade de teste diagnóstico



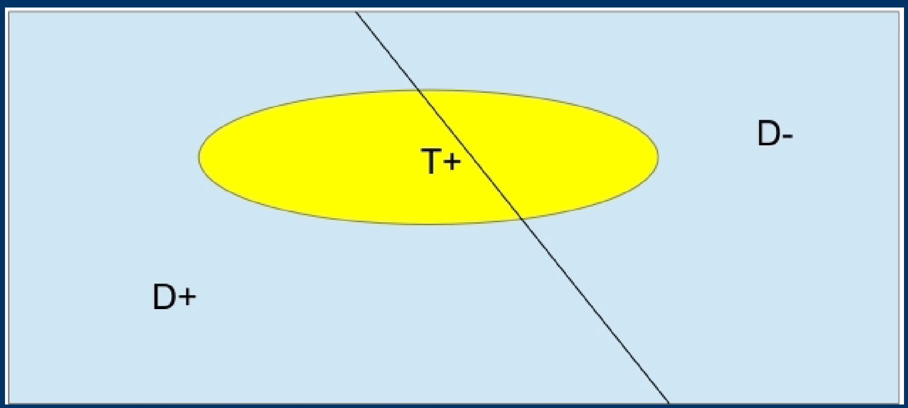

- $D+ \cap D- = \emptyset$
- $D+ \cup D- = E$
- Conhecidos:

$P(T+|D+)=s \quad P(T-|D-)=e$

$$VPP = P(D+|T+) = \frac{P(D+\cap T+)}{P(T+)}$$

$$P(T+|D+) = \frac{P(T+\cap D+)}{P(D+)} \Rightarrow P(T+\cap D+) = P(T+|D+)P(D+)$$
$$= s \times p$$

# Exemplo: Qualidade de teste diagnóstico




- D+ ∩ D- = ∅
- D+ ∪ D- = E
- Conhecidos:

P(T+|D+)=s    P(T-|D-)=e

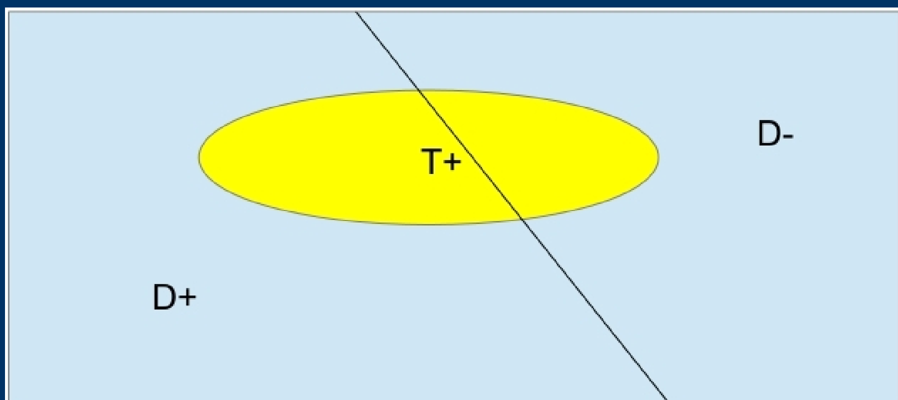
$$VPP = P(D+|T+) = \frac{P(D+\cap T+)}{P(T+)}$$

$$P(T+|D+) = \frac{P(T+\cap D+)}{P(D+)} \Rightarrow P(T+\cap D+) = P(T+|D+)P(D+)$$

$$= s \times p$$

$$P(T+) = P(T+\cap D+) + P(T+\cap D-)$$

# Exemplo: Qualidade de teste diagnóstico




- D+ ∩ D- = ∅
- D+ ∪ D- = E
- Conhecidos:

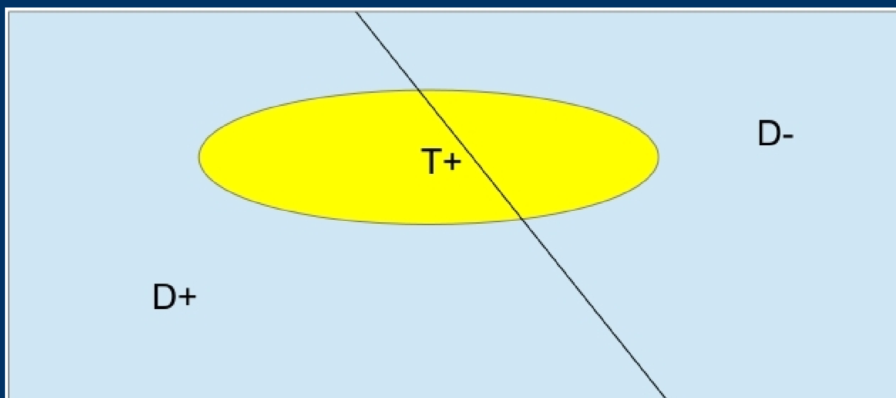
P(T+|D+)=s P(T-|D-)=e

$$VPP = P(D+|T+) = \frac{P(D+\cap T+)}{P(T+)}$$

$$P(T+|D+) = \frac{P(T+\cap D+)}{P(D+)} \Rightarrow P(T+\cap D+) = P(T+|D+)P(D+)$$
$$= s \times p$$

$$P(T+) = P(T+\cap D+) + P(T+\cap D-)$$
$$= P(T+|D+)P(D+) + P(T+|D-)P(D-)$$

# Exemplo: Qualidade de teste diagnóstico




- $D+ \cap D- = \emptyset$
- $D+ \cup D- = E$
- Conhecidos:

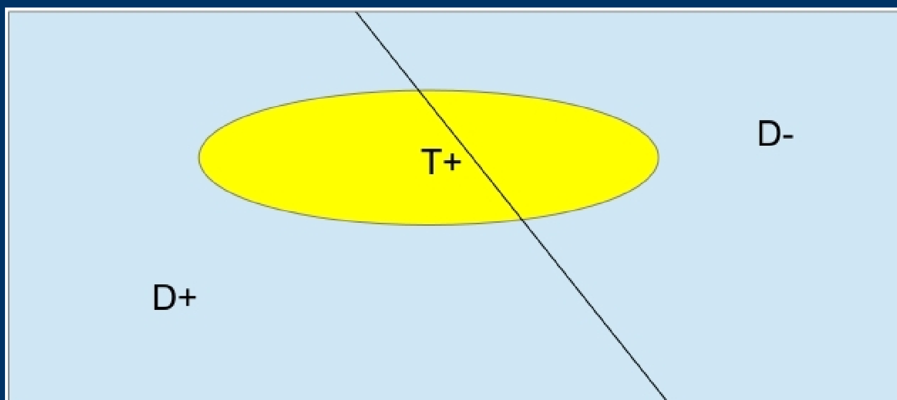
$P(T+|D+)=s \quad P(T-|D-)=e$

$$VPP = P(D+|T+) = \frac{P(D+\cap T+)}{P(T+)}$$

$$P(T+|D+) = \frac{P(T+\cap D+)}{P(D+)} \Rightarrow P(T+\cap D+) = P(T+|D+)P(D+)$$
$$= s \times p$$

$$P(T+) = P(T+\cap D+) + P(T+\cap D-)$$
$$= P(T+|D+)P(D+) + P(T+|D-)P(D-)$$
$$= s\times p + (1-e)\times(1-p)$$

# Exemplo: Qualidade de teste diagnóstico




- D+ ∩ D- = ∅
- D+ ∪ D- = E
- Conhecidos:

P(T+|D+)=s    P(T-|D-)=e

$$VPP = P(D+|T+) = \frac{P(D+ \cap T+)}{P(T+)} = \frac{s\ p}{s\ p + (1-e)(1-p)}$$

# Exemplo: Qualidade de teste diagnóstico

| Doença coronariana | Teste Ergométrico | | | | | |
|---|---|---|---|---|---|---|
| | T+ | | T- | | Total | |
| D+ | 815 | a | 208 | b | 1023 | a+b |
| D- | 115 | c | 327 | d | 442 | c+d |
| Total | 930 | a+c | 535 | b+d | 1465 | n |

Podemos obter VPP e VPN direto da tabela acima?

- VPP=P(D+|T+)=
- VPN=P(D-|T-)=

$$p = P(D+) = \frac{1023}{1465} = 0{,}70\,!\,!\,!$$

# Exemplo: Qualidade de teste diagnóstico – tabela fictícia

| Doença coronariana | Teste Ergométrico | | | |
|---|---|---|---|---|
| | T+ | T- | Total | |
| D+ | 469 | 117 | 586 | s=0,8 |
| D- | 229 | 650 | 879 | e=0,74 |
| Total | 698 | 767 | 1465 | p=0,4 |

Podemos obter VPP e VPN direto da tabela acima?

- VPP=P(D+|T+)= 469/698= 0,67
- VPN=P(D-|T-)= 650/767= 0,85