



**Programa de Pós Graduação
em Ciências**

Coordenadoria de Controle de Doenças-CCD
Secretaria de Estado da Saúde – São Paulo/SP

Infectologia em Saúde Pública – Pesquisas Laboratoriais em Saúde Pública

MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS



– Revisão – Conceitos de Estatística aplicados à Epidemiologia

Carlos R. V. Kiffer

Médico Infectologista

Professor Doutor / Pesquisador Visitante – LEMC / UNIFESP

Sumário

- O que é Estatística?
- Conceitos
 - População e Amostra
 - Variáveis
 - Probabilidade
- Distribuições de Freqüência
 - Freqüências absolutas e relativas
 - Curvas de Freqüência
- Medidas de Tendência Central
 - Média
 - Mediana
 - Moda
- Medidas de Dispersão
 - Amplitude Total
 - Desvio Médio
 - Variância
 - Desvio Padrão
- O que é epidemiologia?
 - Tipos de estudos epidemiológicos
 - Dados de Saúde: A natureza dos dados de saúde: DI

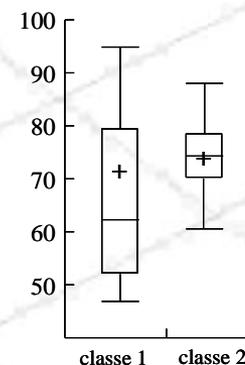
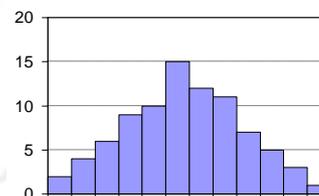
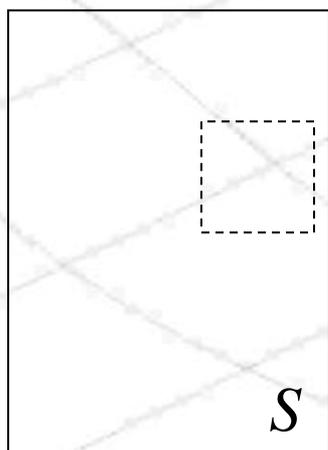
O que é estatística?

“ciência que investiga os processos de obtenção, organização e análise de dados sobre uma população, e os métodos de tirar conclusões ou fazer previsões com base nesses dados.”



Estatística

Estatística Descritiva X Estatística Inferencial



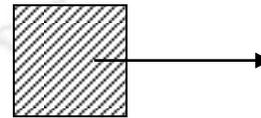
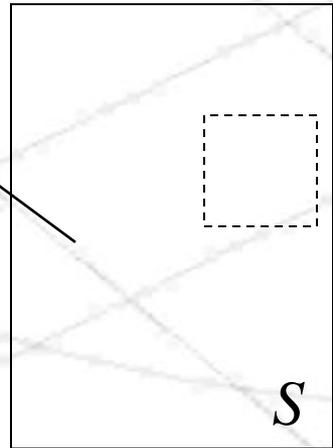
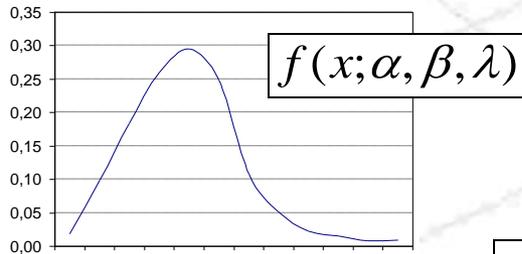
Classes	F.A.	F.R.
10 - 20	2	0,02
20 - 30	4	0,04
30 - 40	6	0,06
40 - 50	12	0,11
50 - 60	10	0,09
60 - 70	25	0,24
70 - 80	12	0,11
80 - 90	19	0,18
90 - 100	7	0,07
100 - 110	5	0,05
110 - 120	3	0,03
120 - 130	1	0,01

106

- média
- moda
- mediana
- desvio médio
- desvio padrão
- assimetria
- curtose

Estatística

Estatística Descritiva X **Estatística Inferencial**



intervalos de confiança
testes de hipóteses

$\hat{\alpha}$
 $\hat{\beta}$
 $\hat{\lambda}$

$f(x; \alpha, \beta, \lambda)$

População e amostra

Tipos de variáveis

População ou Universo é a totalidade dos objetos concebíveis de uma certa classe.

Amostra são os objetos selecionados da população

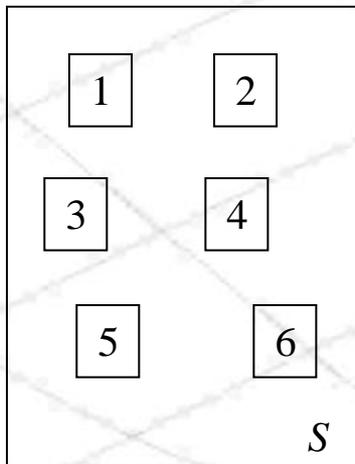
Se esses objetos são selecionados de tal maneira que cada objeto tem a mesma chance de ser selecionado do que o outro, temos uma amostra aleatória

Tipos de Variáveis

- ◆ Variável (característica de interesse)
- ◆ Qualitativa:
 - Resultante de uma classificação por tipos ou atributos
 - Óbitos em um hospital, nos últimos cinco anos
 - Variável: causa mortis (moléstias cardiovasculares, cânceres, etc)
- ◆ Quantitativa :
 - Valores expressos em números.
 - Discreta: valores pertencem a um conjunto enumerável
 - População: habitações de uma cidade.
 - Variável: número de banheiros.
 - Contínua: qualquer valor em um certo intervalo de variação
 - População: estação meteorológica de uma cidade.
 - Variável: precipitação pluviométrica durante um mês.

Probabilidade

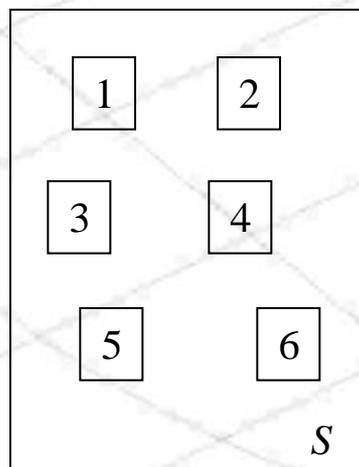
Eventos – Teoria de conjuntos



$$P = \frac{\# \text{eventos favoráveis}}{\# \text{eventos possíveis}}$$

Probabilidade

Eventos – Teoria de conjuntos



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

$$A = \{1, 2, 3\} \text{ (números menores que 4)} \Leftrightarrow P(A) = 0,5$$

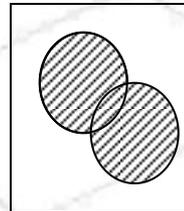
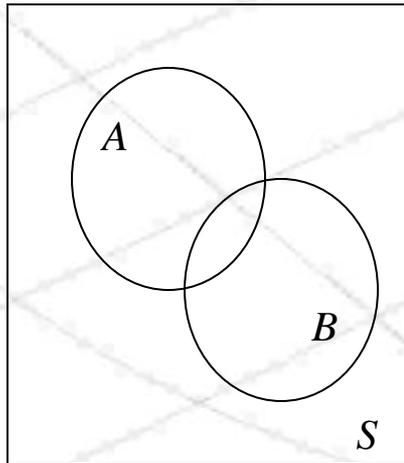
$$B = \{1, 3, 5\} \text{ (números ímpares)} \Leftrightarrow P(B) = 0,5$$

$$C = \emptyset \text{ (números múltiplos de 7)} \Leftrightarrow P(C) = 0$$

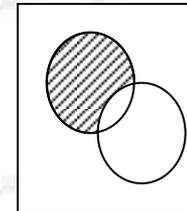
$$D = S \text{ (números maiores que 0)} \Leftrightarrow P(D) = 1$$

Probabilidade

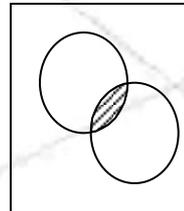
Diagrama de Venn



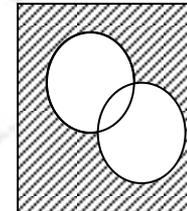
$$A \cup B$$



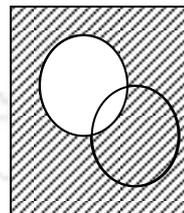
$$A \cap \bar{B}$$



$$A \cap B$$



$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$



$$\bar{A}$$

Probabilidade

Exercícios

- 1) A probabilidade de que um homem esteja vivo daqui a 30 anos é $2/5$;
a probabilidade de que a mulher esteja viva daqui a 30 anos é $2/3$.
Determinar a probabilidade de que, daqui a 30 anos,
- a) ambos estejam vivos;
 - b) somente o homem esteja vivo;
 - c) somente a mulher esteja viva;

H : homem vivo \bar{H} : homem morto
 M : mulher viva \bar{M} : mulher morta

Probabilidade

Exercícios

- 1) A probabilidade de que um homem esteja vivo daqui a 30 anos é $\frac{2}{5}$; a probabilidade de que a mulher esteja viva daqui a 30 anos é $\frac{2}{3}$. Determinar a probabilidade de que daqui a 30 anos:

- a) ambos estejam vivos;

$$P(H \cap M) = P(H).P(M) = \frac{2}{5} \frac{2}{3} = \frac{4}{15}$$

- b) somente o homem esteja vivo;

$$P(H \cap \bar{M}) = P(H).P(\bar{M}) = \frac{2}{5} \frac{1}{3} = \frac{2}{15}$$

- c) somente a mulher esteja viva;

$$P(\bar{H} \cap M) = P(\bar{H}).P(M) = \frac{3}{5} \frac{2}{3} = \frac{6}{15}$$

Distribuições de freqüência

Número de filhos dos 50 funcionários da empresa Fictícia S.A.

Num.de Filhos	Freqüência	Freq. Relativa
0	15	0,30
1	10	0,20
2	13	0,26
3	6	0,12
4	3	0,06
5	3	0,06
Total	50	1,00

As freqüências são:

$f_0 = 15$ (corresponde ao valor 0)

$f_1 = 10$ (corresponde ao valor 1)

$f_2 = 13$ (corresponde ao valor 2)

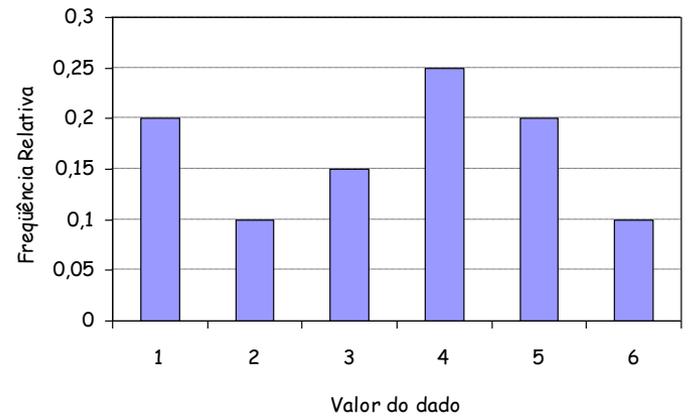
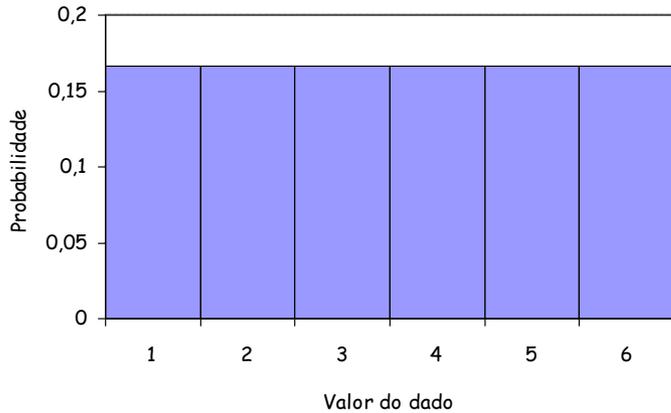
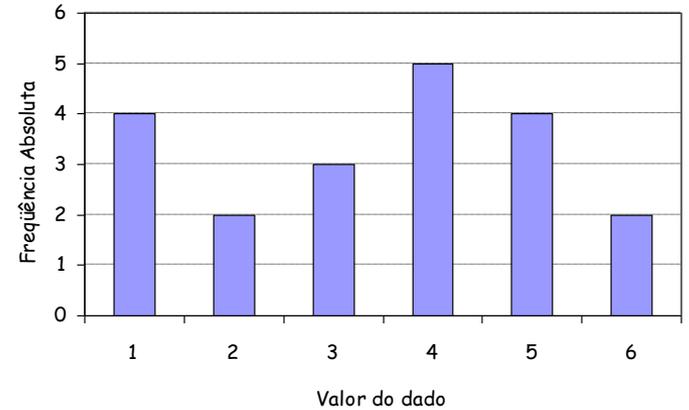
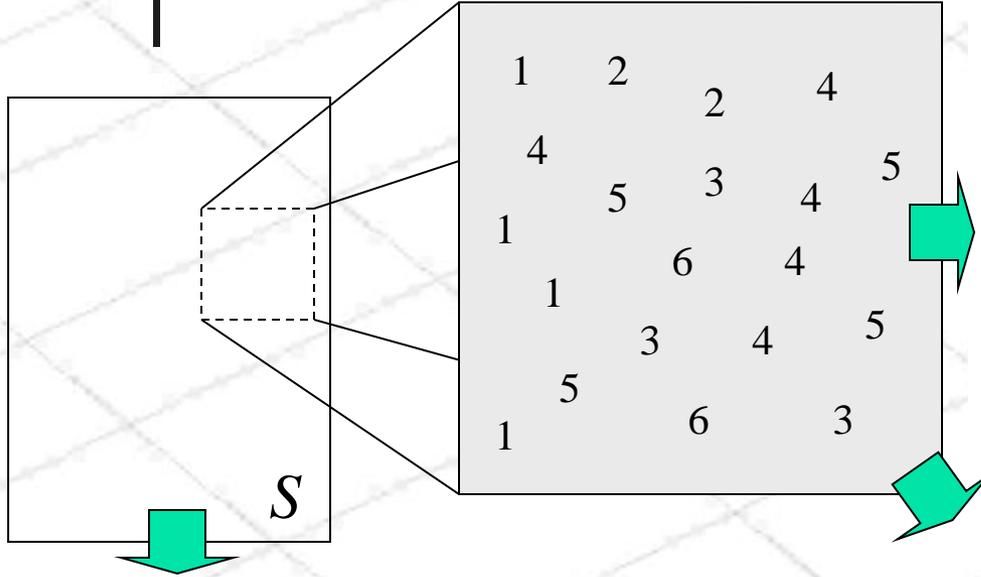
$f_3 = 6$ (corresponde ao valor 3)

$f_4 = 3$ (corresponde ao valor 4)

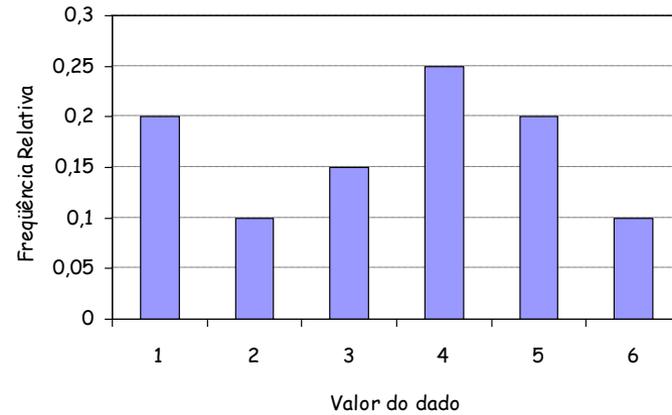
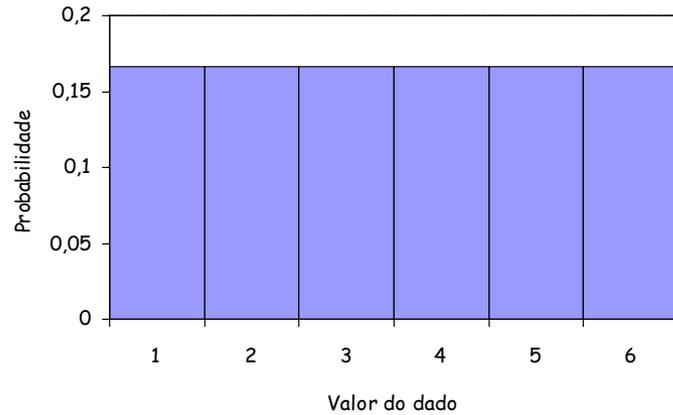
$f_5 = 3$ (corresponde ao valor 5)



Distribuição de Freqüência X Distribuição de Probabilidade



Medidas de Tendência Central



- Média

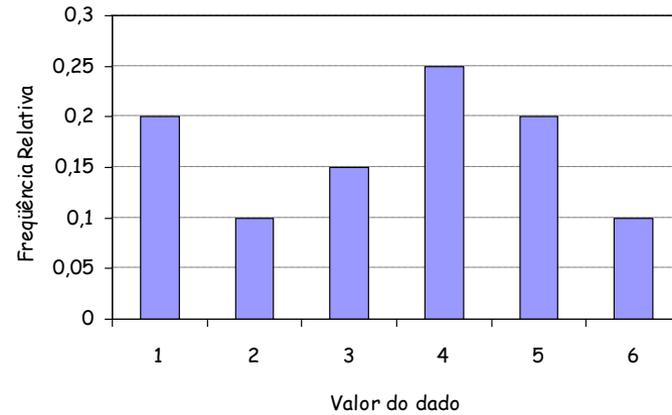
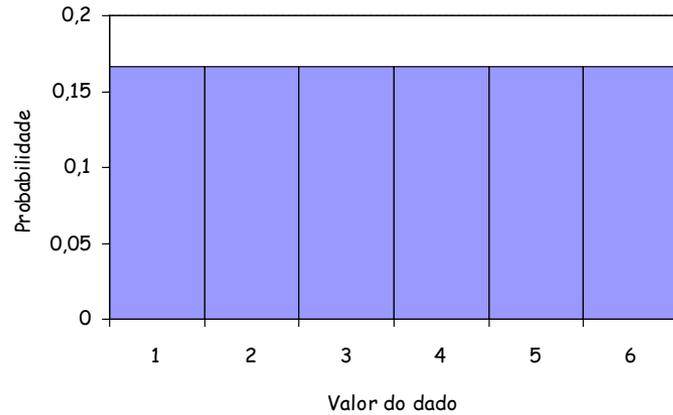
$$\mu = \sum_{i=1}^n x_i P(X = x_i)$$

$$\mu = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5$$

$$\bar{X} = \sum_{i=1}^n x_i fr(x_i)$$

$$\bar{X} = 1 \cdot 0,2 + 2 \cdot 0,1 + 3 \cdot 0,15 + 4 \cdot 0,25 + 5 \cdot 0,2 + 6 \cdot 0,1 = 3,45$$

Medidas de Tendência Central



- Mediana

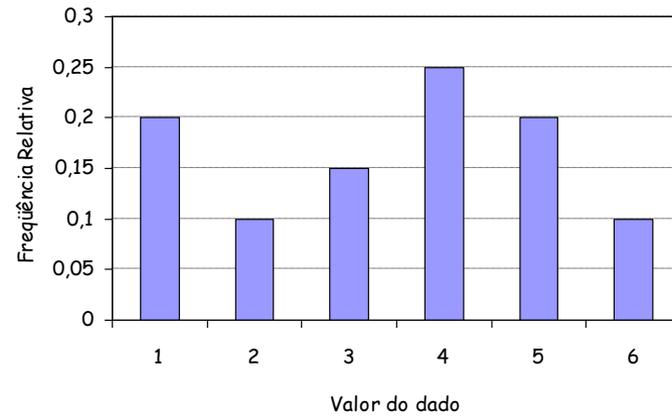
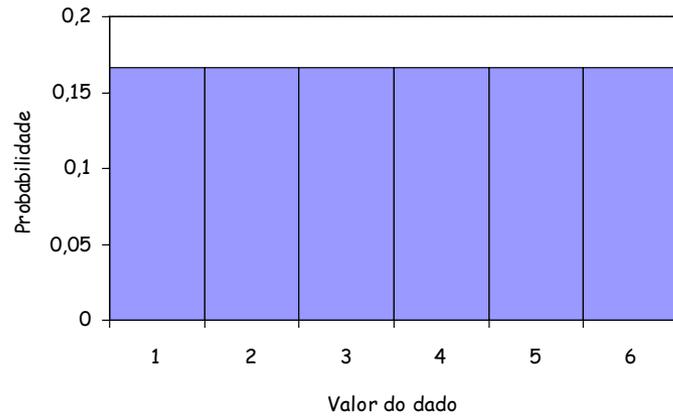
A mediana de um conjunto de números, ordenados em ordem de grandeza, é o valor médio (N ímpar) ou a média aritmética dos dois valores centrais (N par) (é o ponto que divide a distribuição em duas partes equiprováveis)

$$\text{med} = 3,5$$

$$\text{med} = 4$$

OBS: mediana: divide em 2 partes
 quartis: divide em 4 partes (mediana = 2º quartil)
 decis: divide em 10 partes

Medidas de Tendência Central



- Moda

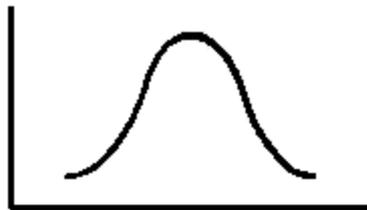
É o valor mais freqüente. A moda é o valor que ocorre com mais freqüência. A moda pode não existir e, mesmo que exista, pode não ser única.

moda = *indeterminada*

moda = 4

OBS: 2 modas (bimodal)
3 modas (trimodal)
muitas modas (multimodal)

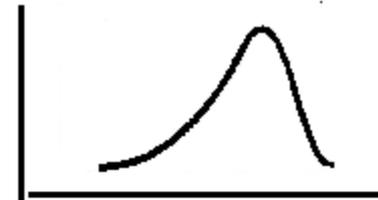
Medidas de Tendência central



Simétrica



Assimétrica positiva



Assimétrica negativa



Forma de J



J invertido



Forma de U



Bimodal



Multimodal

Medidas de Dispersão

Amplitude Total

1	2	2	4	
4	5	3	4	5
1	6	4		
1	3	4	5	
5	3	4	5	
1	6	3		

$$\bar{X} = 3,45$$

$$\text{amplitude total} = \text{máximo} - \text{mínimo} = 6 - 1 = 5$$

1	1	1	1	
1	1	1	1	1
1	1	1	1	
1	1	1	1	1
1	1	1	1	1
1	1	1	6	

Medidas de Dispersão

Desvio Médio

$$\bar{X} = 3,45$$

analisar os desvios em relação à média

1	2	2	4	
4		3	4	5
1	5		4	
	1	6	4	
		3	4	5
	5			
1		6	3	

valor	valor - média
1	2,45
1	2,45
1	2,45
1	2,45
2	1,45
2	1,45
3	0,45
3	0,45
3	0,45
4	0,55
4	0,55
4	0,55
4	0,55
4	0,55
4	0,55
4	0,55
5	1,55
5	1,55
5	1,55
5	1,55
6	2,55
6	2,55

$$\Sigma = 28,1$$

$$\text{Desvio Médio} = \frac{\sum_{i=0}^n |X_i - \bar{X}|}{n}$$

$$\text{Desvio Médio} = \frac{28,1}{20} = 1,405$$

Medidas de Dispersão

Variância
Desvio Padrão

$$\bar{X} = 3,45$$

analisar os desvios em relação à média

1	2	2	4	
4	5	3	4	5
1	5	6	4	
1	3	4	5	
1	5	6	3	

valor	(valor - média) ²
1	6,0025
1	6,0025
1	6,0025
1	6,0025
2	2,1025
2	2,1025
3	0,2025
3	0,2025
3	0,2025
4	0,3025
4	0,3025
4	0,3025
4	0,3025
4	0,3025
4	0,3025
5	2,4025
5	2,4025
5	2,4025
5	2,4025
6	6,5025
6	6,5025

$$\Sigma = 52,95$$

$$Variância = \frac{\sum_{i=0}^n (X_i - \bar{X})^2}{n}$$

$$Variância = \frac{52,95}{20} = 2,6475$$

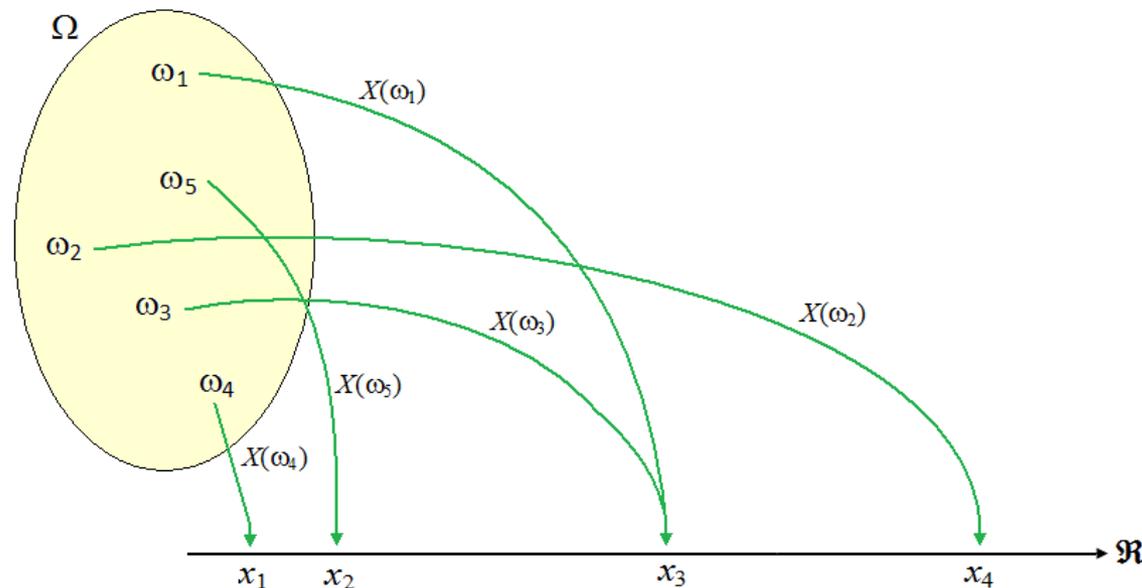
$$Desvio Padrão = \sqrt{Variância}$$

$$Desvio Padrão = \sqrt{2,6475} = 1,6271$$

Variável Aleatória

Uma **variável aleatória** pode ser considerada como uma função que associa elementos do espaço amostral a valores numéricos.

Intuitivamente, uma variável aleatória é uma medição de algum parâmetro que pode gerar um valor diferente a cada medida.



VA discreta: é aquela para a qual o conjunto A é um conjunto finito ou infinito enumerável

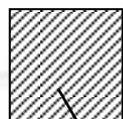
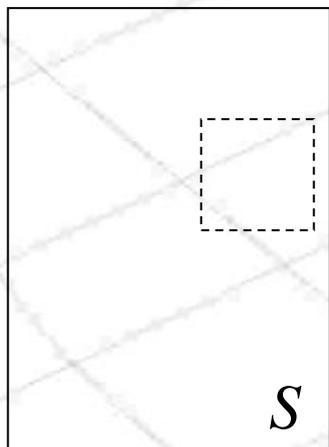
Exs.: $A = \{1, 2, 3, 4, 5, 6\}$, $A = \{0, 1, 2, 3, 4, \dots, \infty\}$, etc.

VA contínua: é aquela para a qual o conjunto A é um conjunto infinito não enumerável, ou seja, é uma v.a. que assume valores em intervalos de números reais

Exs.: $A = \mathfrak{R} = (-\infty, \infty)$, $A = [0, 1] \subset \mathfrak{R}$, etc.

Variável Aleatória

O resultado de jogar uma moeda pode dar Cara (K) ou Coroa (C).



amostra
aleatória

característica

⇒ **variável aleatória**

variável

qualitativa

quantitativa

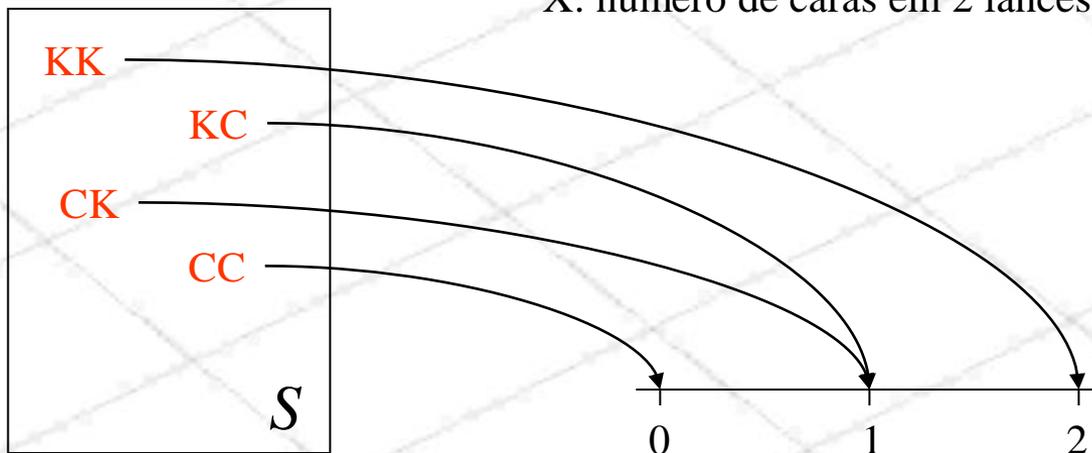
discreta

contínua

- A variável aleatória X pode tomar os valores C ou K
- O conjunto $\Omega = \{ C, K \}$ é o **espaço amostral**
- Os subconjuntos de Ω chamam-se **eventos** ou **acontecimentos** $\{C\}$, $\{K\}$.
- A probabilidade de cada evento elementar = $1/2$.

Variável Aleatória

X: número de caras em 2 lances de moeda



A probabilidade em 2 lançamentos = ao menos 1 {K}

$$X = 0 \Leftrightarrow CC \quad \Leftrightarrow P(X = 0) = P(CC)$$

$$X = 1 \Leftrightarrow KC \cup CK \quad \Leftrightarrow P(X = 1) = P(KC \cup CK)$$

$$X = 2 \Leftrightarrow KK \quad \Leftrightarrow P(X = 2) = P(KK)$$

Função de Probabilidade de uma variável aleatória

Função de probabilidade de uma v.a. discreta

A função que associa probabilidades aos possíveis valores de uma v.a. discreta X , é chamada de função de probabilidade discreta e é representada por:

$$p(x) = P(X = x), x \in A.$$

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n - x}, \quad x = 0, 1, 2, \dots, n.$$

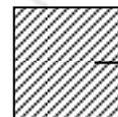
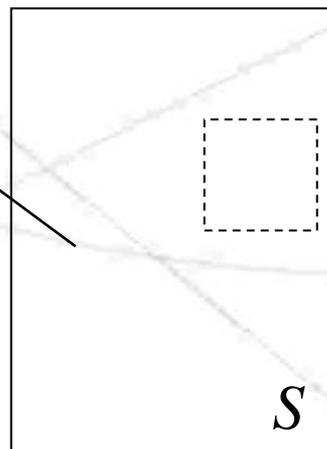
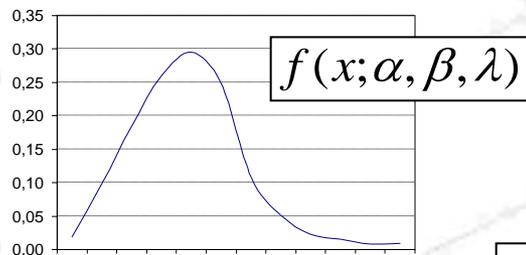
Esse modelo é conhecido como modelo binomial. O modelo binomial está associado a ensaios com apenas dois resultados possíveis: sim/não; ocorre/não ocorre; 0/1.

Esses ensaios quando são independentes recebem o nome de ensaios de Bernoulli.

Nos ensaios de Bernoulli sempre estamos interessados em apenas um dos resultados ao qual chamaremos de sucesso.

O que é epidemiologia?

A Epidemiologia é a ciência que estuda os padrões da ocorrência de doenças em populações humanas e os fatores determinantes destes padrões (Lilienfeld, 1980). Enquanto a clínica aborda a doença no indivíduo, a epidemiologia aborda o processo saúde-doença em grupos, que podem ser pequenos ou populações inteiras.

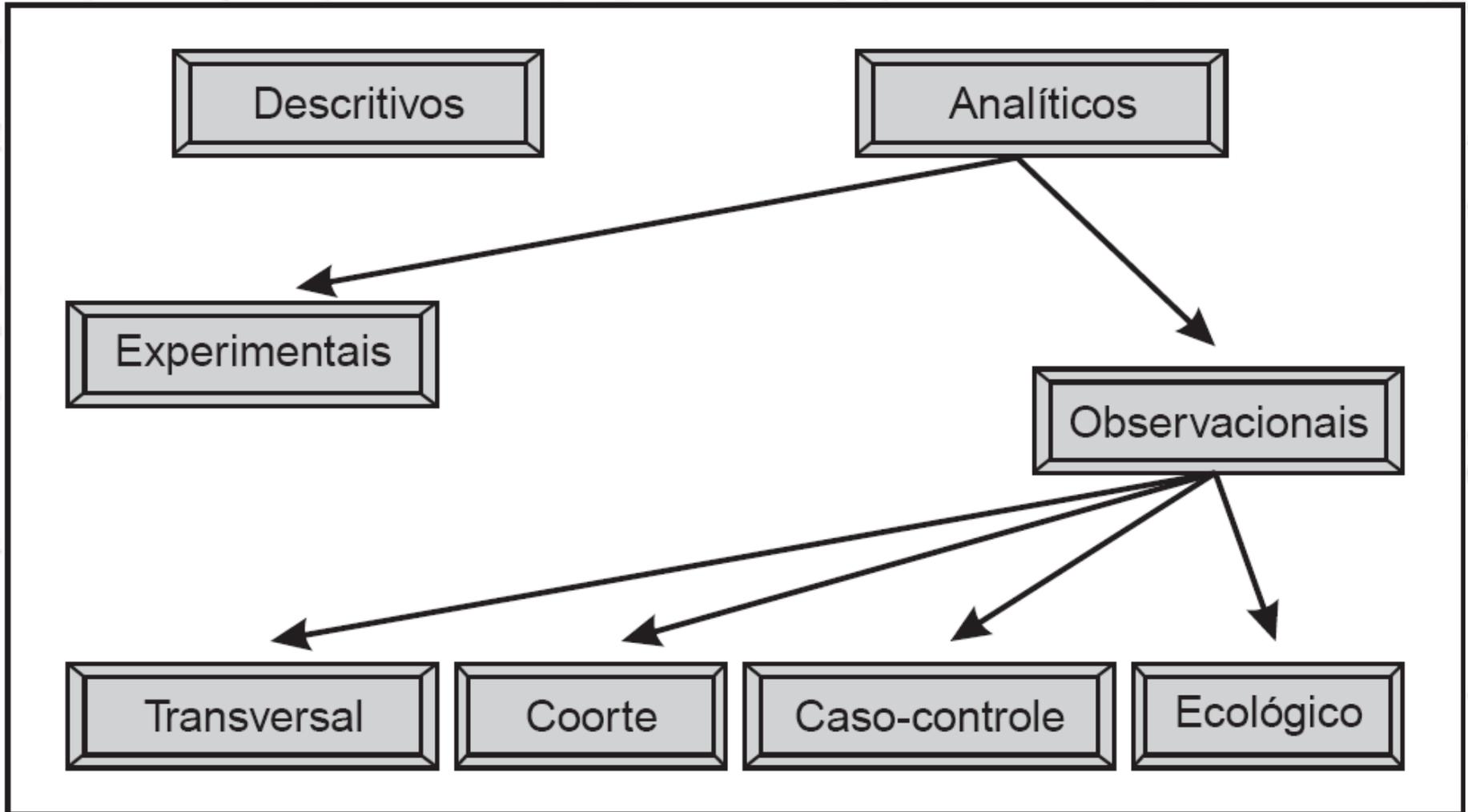


$\hat{\alpha}$

$\hat{\beta}$

$\hat{\lambda}$

Tipos de estudos



A Natureza dos Dados da “Saúde”: O Caso das Doenças Infecciosas

- ◆ Universo conceitual

- Dados de Saúde:

- Observações

- Variáveis de desfecho (aleatórias)

- Medida, Informação qualitativa, etc.
- Taxa / Número / Dado qualitativo

- Intrinsecamente espaciais

- Condição do objeto de estudo da “saúde”
- Ex.: indivíduo, comunidade, ambiente, fatores/riscos
 - Localização
 - Temporal

A Natureza dos Dados da "Saúde": O Caso das Doenças Infecciosas

- ◆ Universo representacional
 - Dados Espaciais
 - Existência definida por:
 - Lugar (localização)
 - Tempo (temporal)
 - Possível representá-lo:
 - Sistema de coordenadas

Portanto, todo Dado de Saúde com possibilidade de representação em um espaço definido por um sistema de coordenadas é um Dado Espacial no contexto da saúde

A Natureza dos Dados da "Saúde": O Caso das Doenças Infecciosas

- ◆ Universo representacional
 - Dados Geográficos
 - Existência definida por:
 - Lugar (localização)
 - Tempo (temporal)
 - Atributos
 - Necessariamente representado:
 - Sistema de projeção cartográfica
 - Representa posições na superfície terrestre

Portanto, todo Dado Espacial com representação em um sistema de projeção cartográfica é um Dado Geográfico no contexto da saúde

Dados de Saúde

	Dado Espacial	Dado Geográfico	Problemas
Casos de infecção hospitalar	X	-	Restrito ao ambiente hospitalar
Casos de infecção hospitalar	X	X	Inserção comunitária
Estudos soropidemiológicos populacionais	X	X	Representação domiciliar
Taxas de morbimortalidade	X	X	Necessariamente tem unidade geográfica de representação

Representações Computacionais para Dados Espaciais em Saúde

A "questão" das Escalas para os Estudos em Saúde

- ◆ Universo conceitual

- Dados de Saúde
 - Intrinsecamente espaciais



- ◆ Universo representacional

- Dados Espaciais
 - Geográficos

Mas afinal, que importância tem os dados de saúde serem espaciais e, mais ainda, terem representações geográficas?

Aí entram em jogo ...

Os Atributos do objeto estudado (Dados de Saúde) e suas relações com o ambiente

- ◆ Relações entre objeto e ambiente
 - Risco
 - “Perigo” potencial
 - Probabilidade de ocorrência de um evento
 - Chance de ocorrência em relação à exposição

Envolve uma relação entre efeitos e causas

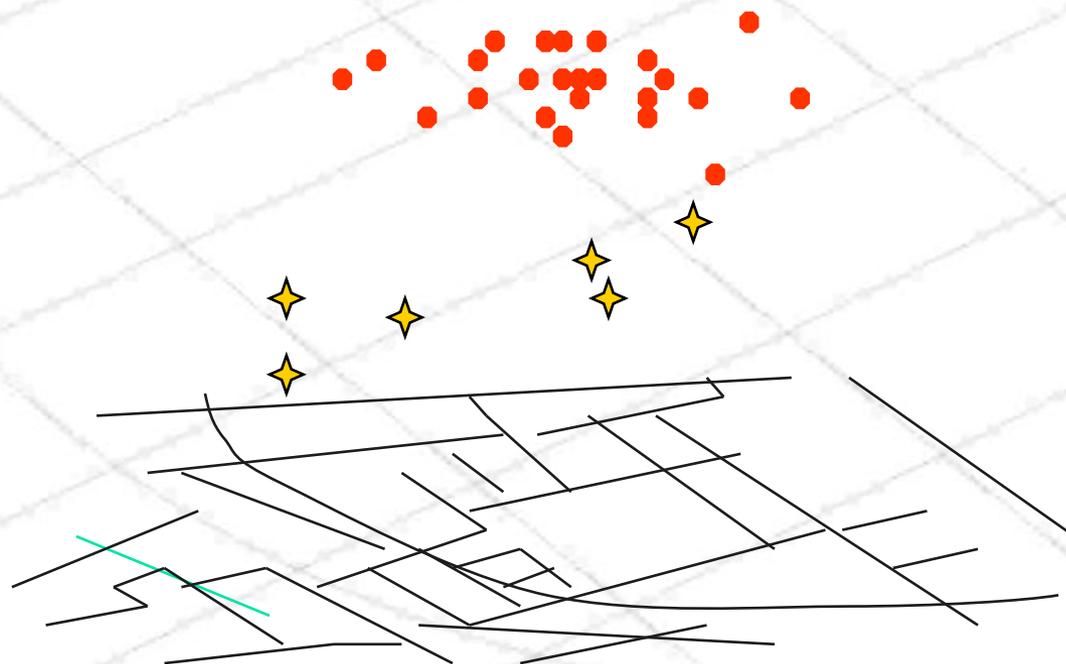
Ambiente

Exposição

Efeito sobre a saúde



Camadas de informações (layers)



Eventos de saúde
Mortes

Condições ambientais
Poços

Base cartográfica
Ruas

Voltando às 'escalas'...

Interrelacionamento de dados epidemiológicos e ambientais

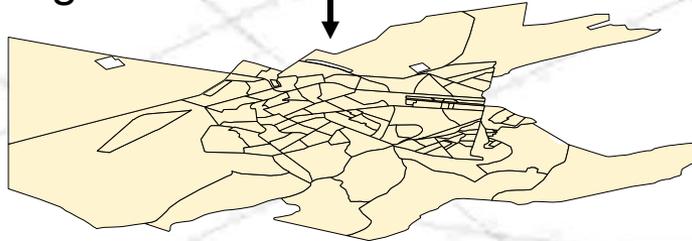
O caso



Atributos
individuais

Efeitos sobre
a saúde

O lugar do caso



Determinantes
sócio-ambientais

Exposição

Portanto, as escalas dizem respeito ao “Lugar do Caso”