

O papel da Estatística na Investigação Científica

Curso de Fisiologia Clínica – 4º ano

Carina Silva, ESTeSL-IPL e CEAUL

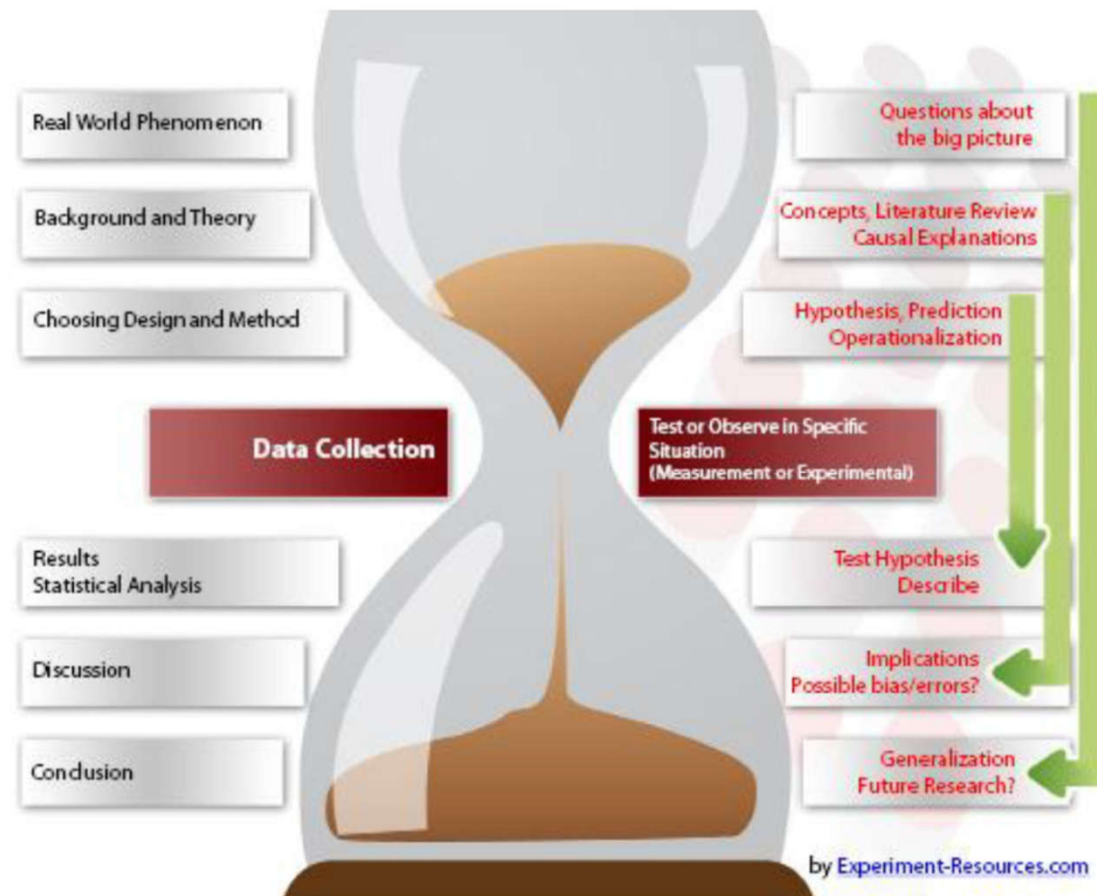
Resumo

- ▶ Introdução
- ▶ Objetivos
- ▶ A estatística no processo de investigação
- ▶ Alguns conceitos gerais
- ▶ Aplicação no SPSS - Exercícios

Introdução

- ▶ O que é a Estatística?
- ▶ Porquê estudar Estatística?
- ▶ Qual a importância da Estatística na investigação científica?
- ▶ Qual o papel da Estatística na Investigação?
- ▶ Alguns exemplos de termos estatísticos usados em investigação?
- ▶ Que tipos de má utilização da Estatística se verificam com maior ocorrência na investigação?

O papel da Estatística na Investigação?



Source: <http://www.experiment-resources.com/>

Qual a importância da Estatística na investigação científica?

- ▶ A Estatística permite aos investigadores caracterizarem, sumariarem, apresentarem e interpretarem os resultados da investigação.

Quais são os termos estatísticos (mais) usados em investigação?

População/amostra	Generalização	Variabilidade
Testes estatísticos	Correlação	Análise de regressão
Valor-p	Nível de significância	significativo
Viés	erro	média
Desvio padrão	precisão	Intervalo de confiança
...

Que tipos de má utilização da Estatística se verificam com maior ocorrência na investigação?

Remoção de dados
que não favorecem a
investigação

Amostras enviesadas

Falsa causalidade

Manipulação dos dados

Gráficos mal construídos

Confusão entre
significância estatística e
significância prática

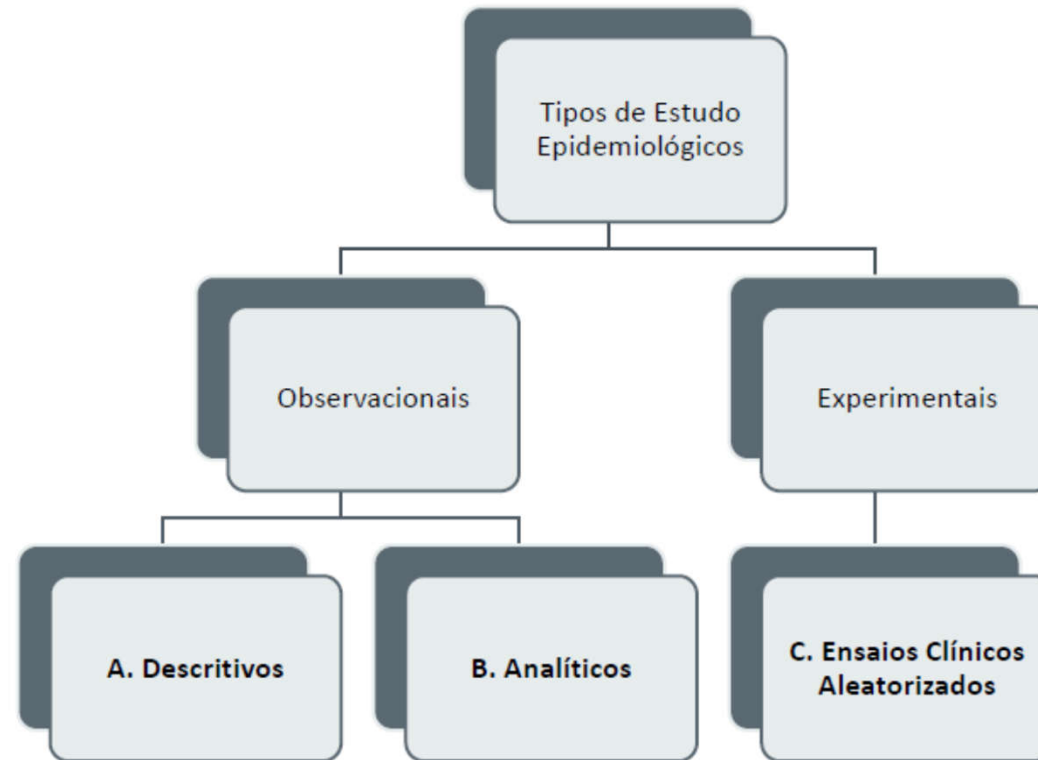
Má interpretação do
erro estimado

Generalização
desadequada

Alguns conceitos gerais

Tipos de Estudos

O desenho do estudo está relacionado com o objetivo específico de cada investigação, mas na investigação clínica, usualmente o objetivo final é o estabelecimento de causalidade (relação causa-efeito)



Tipos de Estudos

A. Estudos Descritivos

- Informam sobre a distribuição de um evento na população, em termos quantitativos:

Incidência ou Prevalência

- Possibilitam a caracterização da doença:
 - **Tempo:** curso da epidemia/doença, o tipo de curva e período de incubação (evolução histórica)
 - **Lugar:** extensão geográfica do problema
 - **Pessoa:** grupo de pessoas, faixa etária, exposição aos fatores de risco
- Não investigam uma relação de causa e efeito
- Descrição de um problema/realidade
- Uso de medidas de tendência central, medidas de dispersão, frequências, proporções e indicadores como prevalência, incidência, taxas ...

Tipos de Estudos

B. Estudos Analíticos

- Estudos comparativos que lidam com “hipóteses” - estudos de causa e efeito, exposição e doença;
- Testam hipóteses.

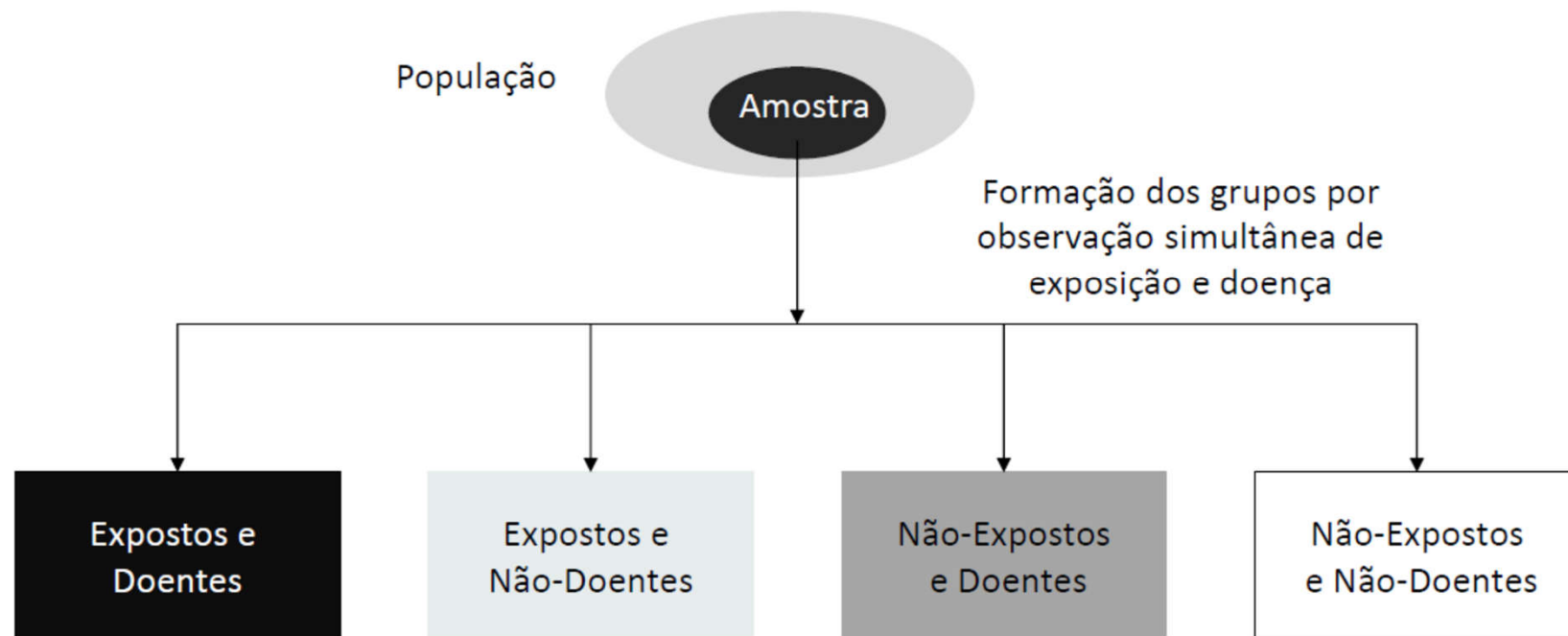
1. Transversais

2. Estudos Caso-Controlo

3. Estudos Coorte

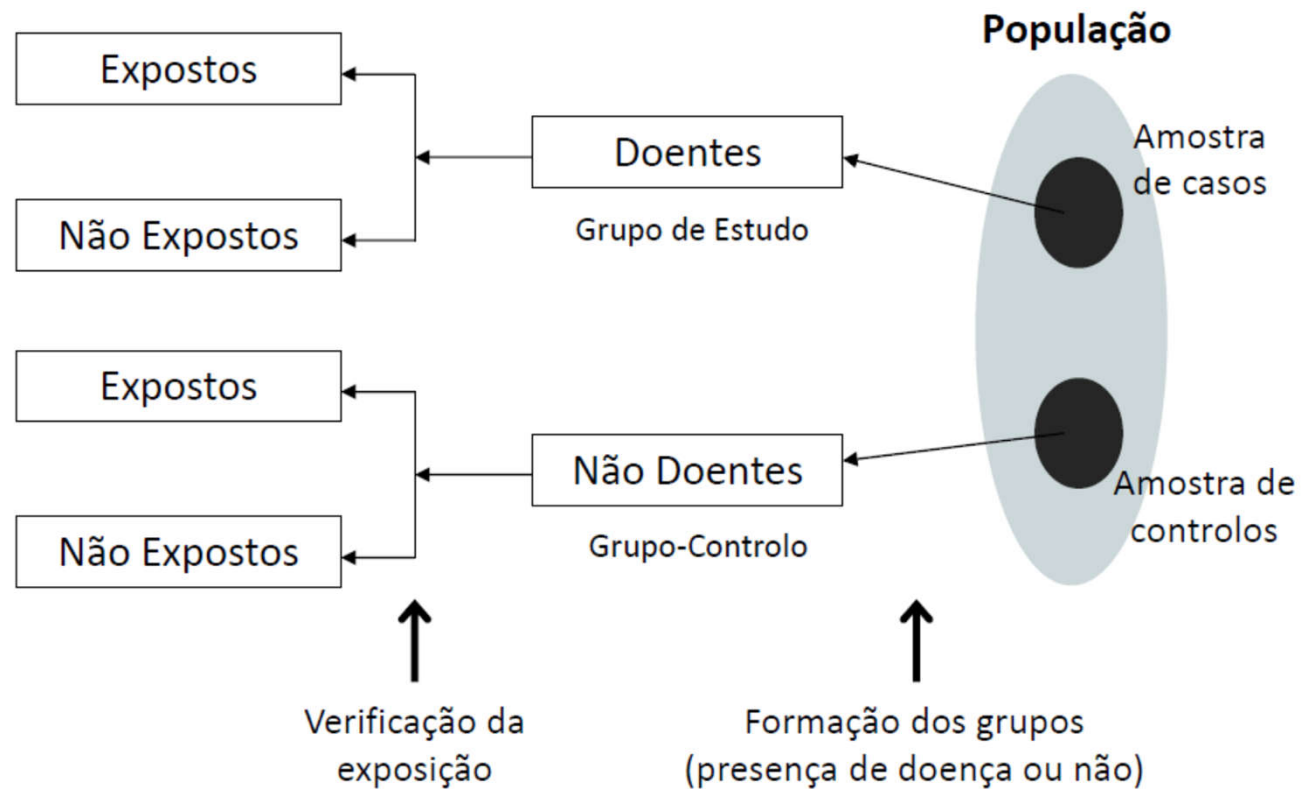
Tipos de Estudos

B. Estudos Analíticos | 1. Transversais



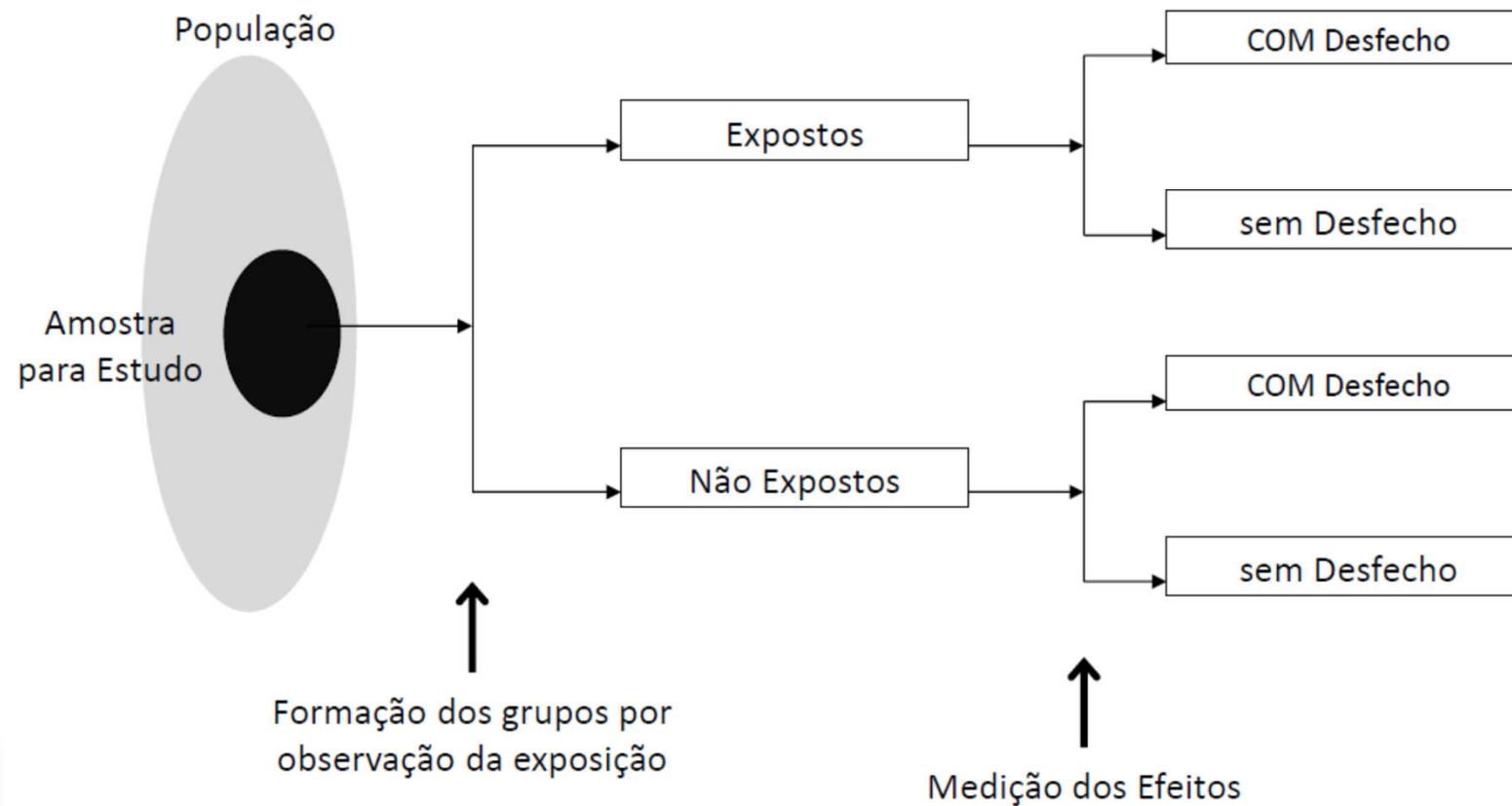
Tipos de Estudos

B. Estudos Analíticos | 2. Estudos Caso-Controllo



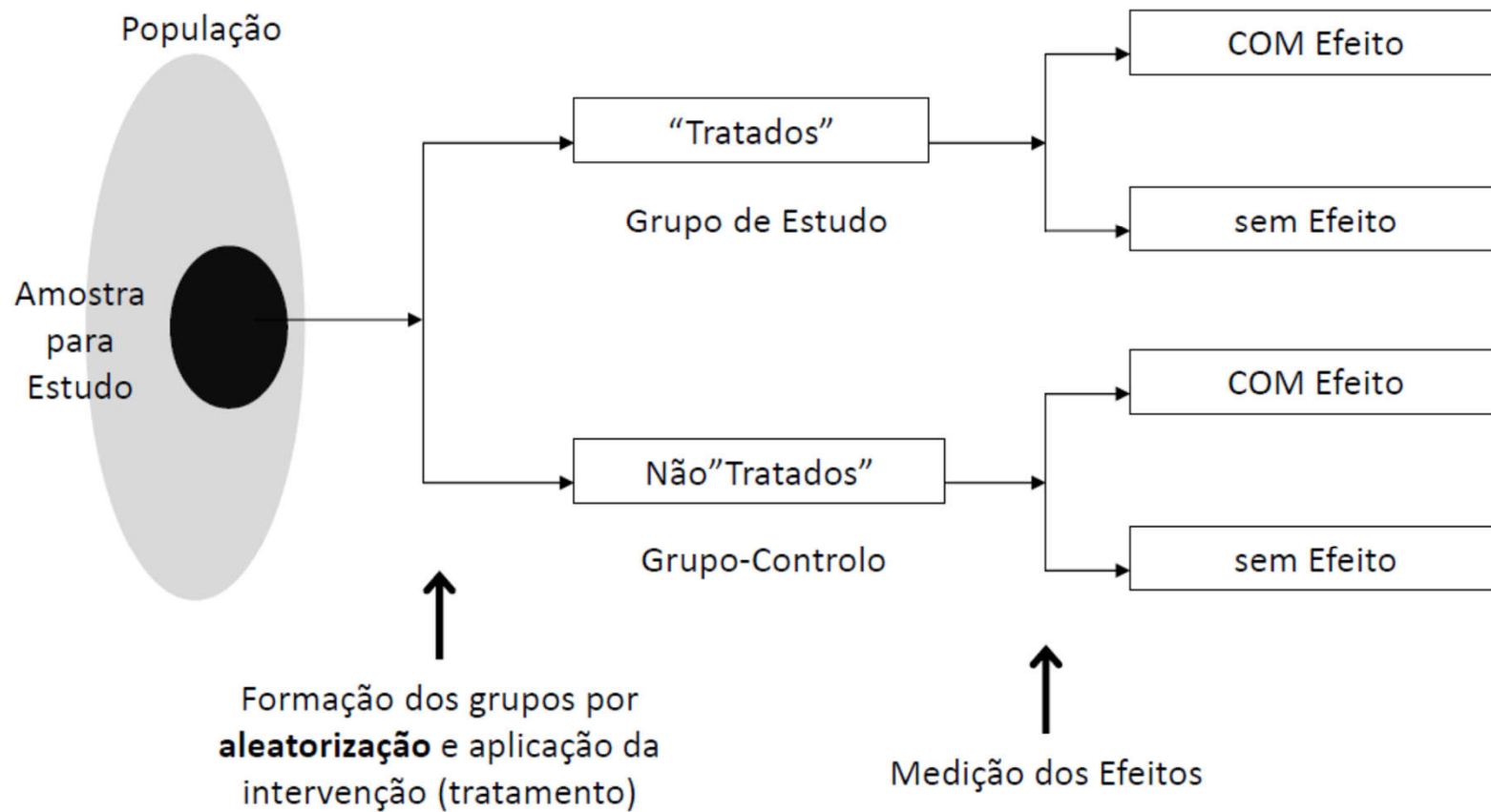
Tipos de Estudos

B. Estudos Analíticos | 3. Estudos Coorte

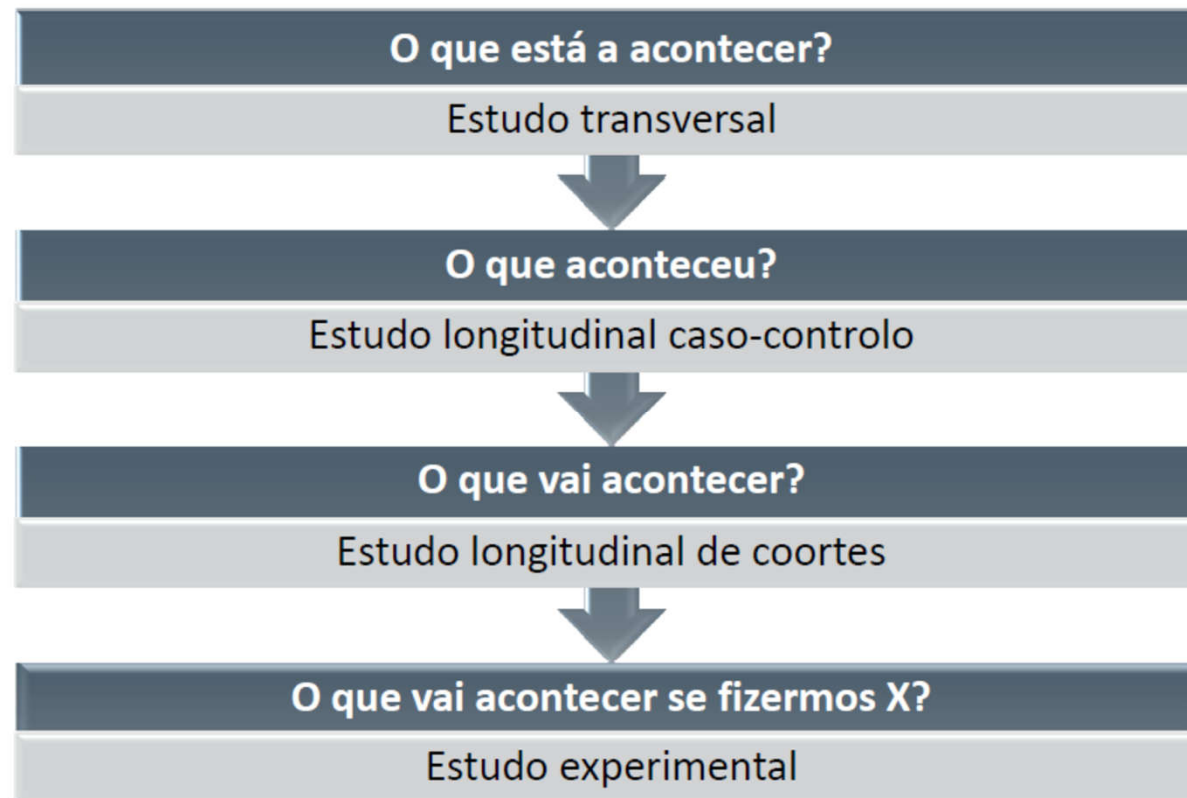


Tipos de Estudos

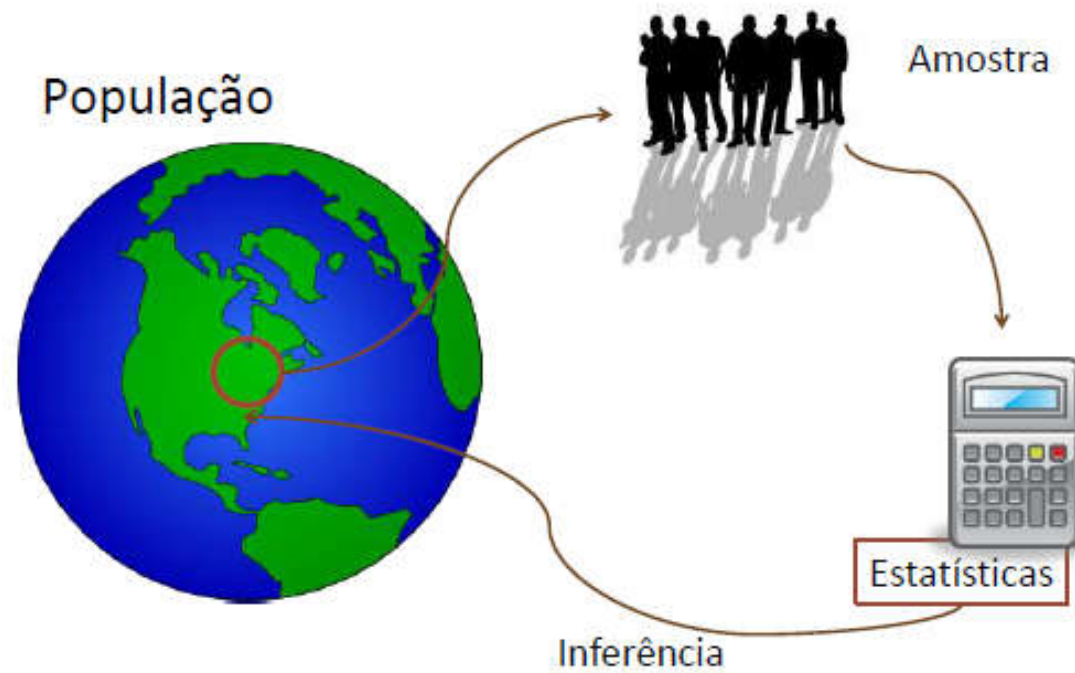
C. Estudos Experimentais | ensaio clínico aleatorizado



Tipos de Estudos



Alguns conceitos gerais–Inferência Estatística

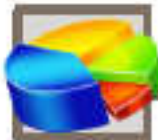


Alguns conceitos gerais



Síntese

- Resumir informações para melhorar compreensão dos dados
- Média
- Mediana
- Desvio Padrão
- Outras Estatísticas



Visualização

- Interpretações e análises baseadas em elementos visuais
- Gráficos
- Diagramas



Inferência

- Inferir o comportamento de uma população usando como base uma amostra
- Testes de hipóteses

Alguns conceitos gerais



População

É o conjunto de todos os elementos sob investigação.

Em geral NÃO temos acesso aos dados de toda a população. Temos uma amostra.



Exemplo

Desejamos saber se um novo tratamento para hipertensão é ou não eficaz.

Gostaríamos que o tratamento pudesse ser usado por TODOS aqueles que sofrem de hipertensão.

Nossa população é o conjunto de todas as pessoas que sofrem de hipertensão

Alguns conceitos gerais



Amostra e População

- Uma amostra refere-se a uma certa população, e as conclusões sobre esta amostra inferem sobre esta população.
- EX: Amostra de ratos suíços recém nascidos irão trazer informações sobre ratos suíços recém nascidos. Nada pode-se dizer em relação a outros tipos de ratos.



Aleatoriedade

- Uma vez determinada a população, a escolha dos elementos não pode ter viés.
- EX: Se vamos analisar coelhos de um certo tipo, não podemos, por exemplo, pegar só os coelhos mais “calminhos”



Tamanho da amostra

- Devido ao custo de obtenção de amostras vivas, o tamanho é baixo.
- Cuidado para não ser tão baixo!
- EX: Não é possível realizar alguns testes estatísticos com menos de 4 elementos na amostra

Por onde começar

- ▶ Podemos classificar as variáveis estatísticas em:

Qualitativas → sempre que se refere a um atributo da unidade estatística, tais como a cor, a profissão, entre outros;

Quantitativas → são mensuráveis, como por exemplo, o peso, o nº de filhos, a temperatura, o nível de ferro no sangue, etc...

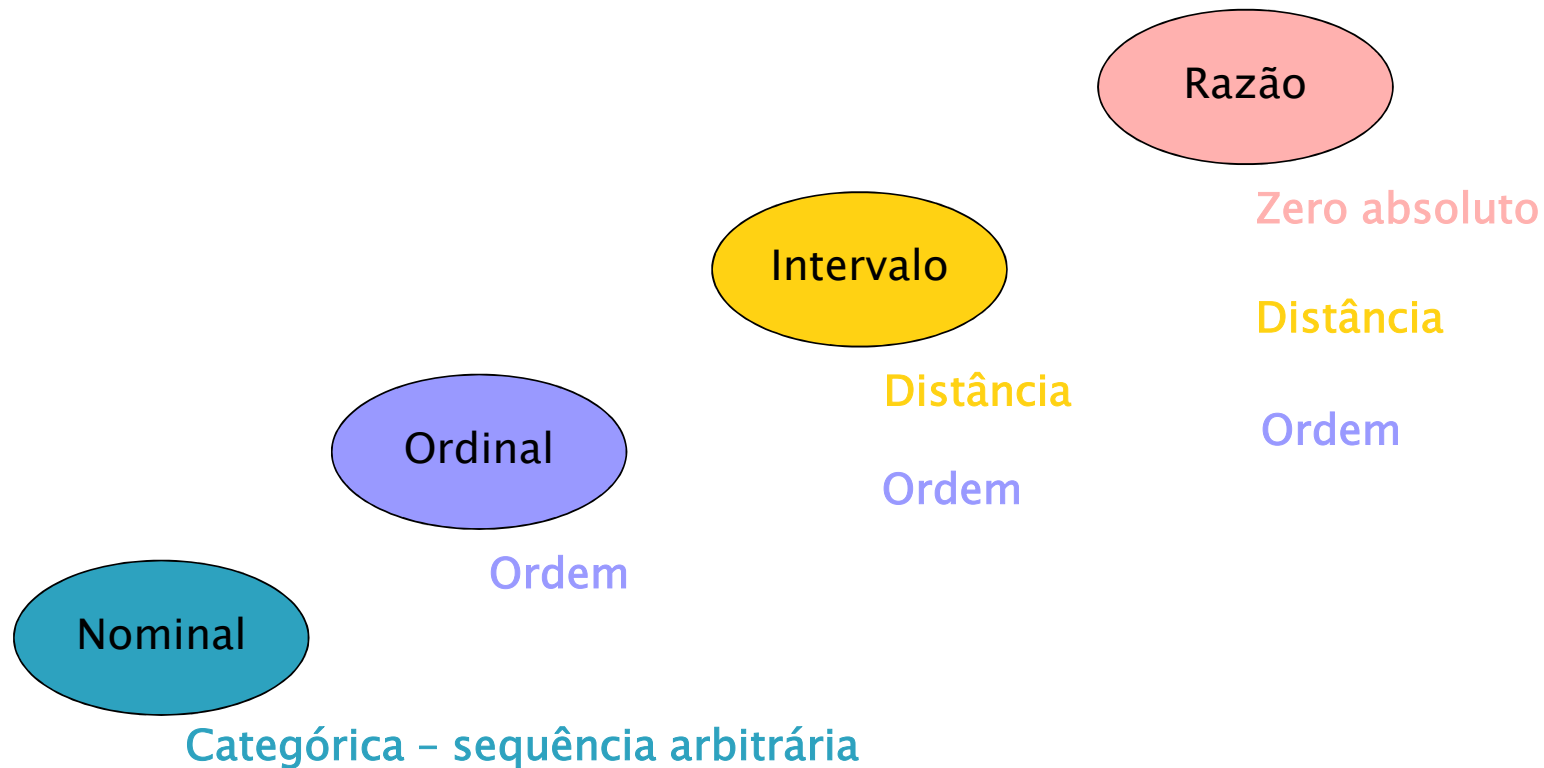
- ▶ As variáveis quantitativas podem ainda classificar-se em:

Contínuas → assumem um número infinito de valores dentro de um determinado intervalo;

Discretas → assumem um número finito de valores dentro de um determinado intervalo

Por onde começar

- Escalas de Steven (1946)



Alguns exemplos



Variáveis Qualitativas

- **Nominal**
 - EX: Órgão afetado por um tratamento
 - Moda, proporções
- **Ordinal**
 - EX: Pouca, muita dor
 - Mediana, proporções



Variáveis Quantitativas

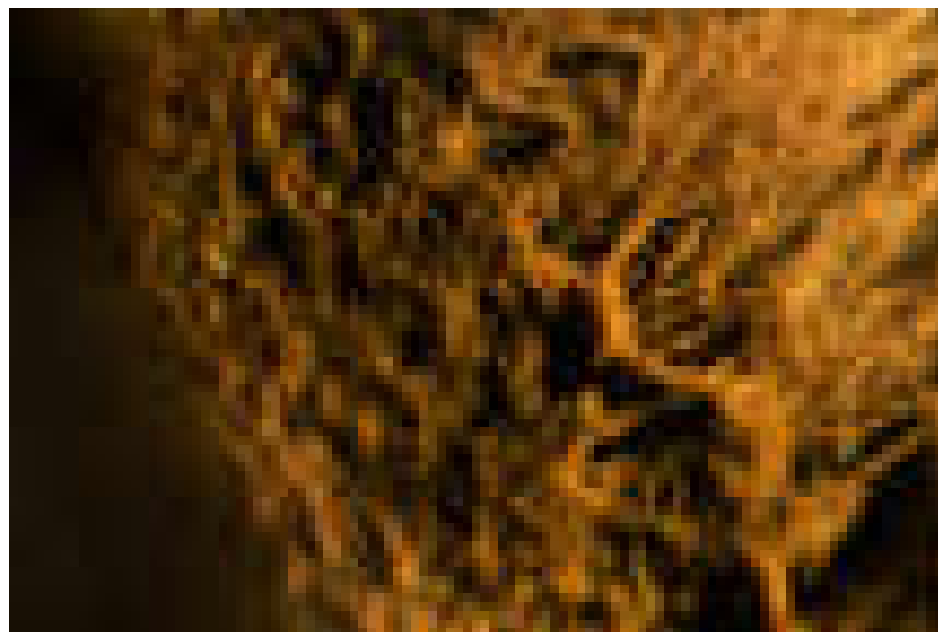
- **Intervalar ou de razão**
 - EX: Absorbância, pressão, comprimento, volume, %
 - Média, DP, Mediana, quartis, %, ...
- **Intervalar ou de razão**
 - EX: Absorbância, pressão, comprimento, volume, %
 - Média, DP, Mediana, quartis, %, ...

Etapas de um plano amostral

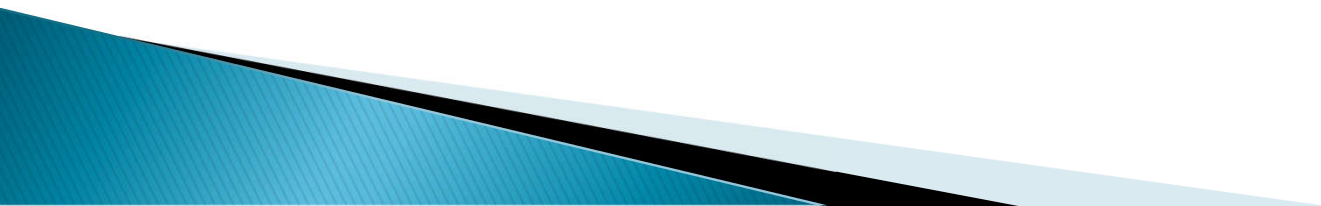
- ▶ Definir a população alvo
- ▶ Identificar a base de sondagem
- ▶ Escolher a técnica amostral
- ▶ Determinar a dimensão da amostra
- ▶ Selecionar os elementos da amostra
- ▶ Recolher a informação necessária dos elementos da amostra

População alvo

- ▶ **População alvo** é a totalidade dos elementos sobre os quais incide a nossa análise e dos quais se pretende obter informação.
- ▶ A definição da **população alvo** é uma das fases mais importantes na realização de um estudo.
- ▶ Para definir corretamente a **população alvo**, primeiro temos de ter a certeza quais são os objetivos, e depois, podemos perguntar: sobre quem incide o **estudo**?









Amostragem

- ▶ A escolha da técnica amostral depende essencialmente dos objetivos do estudo, das condições particulares do estudo, custos, tempo, etc e da avaliação das vantagens e desvantagens inerentes a cada técnica.
- ▶ As técnicas de amostragem dividem-se essencialmente em duas grandes categorias: aleatórias (ou probabilísticas, ou casuais) e não aleatórias (ou não probabilísticas, ou dirigidas).

Amostragem

Técnicas Aleatórias	
Vantagens	Desvantagens
<ul style="list-style-type: none">• Não existe subjetividade na escolha das unidades amostrais, eliminando-se o enviesamento por vezes introduzido pelos entrevistadores.• Determinação prévia da dimensão da amostra com base no grau de confiança e precisão desejadas.• Permite a utilização dos princípios utilizados na Teoria das Probabilidades, que garante à partida a melhor forma de diminuir o erro amostral.	<ul style="list-style-type: none">• Necessidade de ter uma base de sondagem.• Por vezes as características da base de sondagem levam a despende muito tempo e elevados custos na recolha das unidades amostrais.• Problema das não-respostas.

Amostragem

Técnicas Não Aleatórias	
Vantagens	Desvantagens
<ul style="list-style-type: none">• Não é necessário uma base de sondagem• Custos mais baixos• Recolha de informação mais rápida	<ul style="list-style-type: none">• Impossibilidade da aplicação da Teoria das Probabilidades• A qualidade dos resultados só pode ser avaliada subjetivamente.• Estão sempre presentes enviesamentos.

Amostra enviesada

- ▶ Uma amostra que não seja representativa da População diz-se **enviesada** e a sua utilização pode dar origem a interpretação erradas, como o exemplo que a seguir se exemplifica:
- Utilizar uma amostra constituída pelos 10 pacientes que revelaram ter bons hábitos alimentares para verificar se um determinado complemento alimentar contribui para um aumento do bem estar geral de saúde.
- Amostragem por conveniência.
- Amostragem por resposta voluntária.

Erros que podem surgir num processo de amostragem

- ▶ Quando se recolhe uma amostra, mesmo que cuidadosamente planeada e recolhida, trata-se de uma “amostra” da população, portanto alguma coisa se pode perder.
- ▶ Existem fundamentalmente dois tipos de erros:
- ▶ - **Erro de amostragem.** Resulta da não observação da totalidade da população. Numa amostragem aleatória é controlável aumentando a dimensão da amostra.
- ▶ - **Erro de medida.** Erros que são introduzidos durante o processo de amostragem: impossibilidade de observar certas unidades, valores mal registados, erros de codificação. Dificilmente é quantificado.

Métodos de amostragem

Aleatória

Simples
Estratificada
Sistemática
Clusters
Multi-etapas
Multi-fases

Não Aleatória

Intencional
Bola de Neve (*Snowball*)
Conveniência
Quotas
Sequencial

Determinação da dimensão da amostra

Este é um assunto que merece destaque afim de se esclarecerem alguns “boatos” sobre este tema.

Qualquer que seja a técnica de amostragem adoptada há um fator em comum entre elas: qual a dimensão da amostra a recolher.

Existem sempre dois efeitos opostos: a precisão que aumenta com a dimensão da amostra e o custo que aumenta no mesmo sentido. Não existe o melhor de dois mundos, há que escolher um processo que alie da melhor forma estas características. O cálculo da dimensão da amostra pode ser feita matematicamente desde que a técnica de amostragem seja aleatória.

Vão ser abordadas duas possíveis formas para estabelecer uma dimensão para a amostra:

- O caminho do esforço mínimo
- A estimação por meio das “Regras do polegar” (*Rules of thumb*)

Determinação da dimensão da amostra

▶ “Caminho do esforço mínimo”

- ▶ É talvez o pior dos métodos e no entanto o mais vulgarmente utilizado.
- ▶ O procedimento consiste em obter a dimensão o maior quanto possível. Este processo vai depender do tempo, custos, recursos, motivação do investigador, etc. Mas o que geralmente mais condiciona são os custos sempre inerentes a um processo de amostragem. Este tipo de técnica resulta sempre numa dimensão da amostra mais pequena do que o desejável.
- ▶ Há que ter em consideração que por vezes é preferível dedicar mais tempo e recursos nesta fase para que depois toda a análise estatística tenha qualidade.

Determinação da dimensão da amostra

- ▶ **A estimação por meio das “Regras do polegar”**
- ▶ Esta técnica também não tem por base nenhum procedimento matemático, mas resulta de uma aproximação feita a muitas experiências por investigadores e é preferível à anterior.
- ▶ Esta técnica consiste em estimar um valor mínimo para a dimensão da amostra de acordo com a técnica estatística que se vai utilizar. Basicamente o investigador tem de prever quais as técnicas que tem de aplicar e seleccionar a dimensão que corresponde ao máximo valor de entre as técnicas que tem que aplicar. Vamos supor que para a técnica 1 tem valor mínimo de 50 casos, técnica 2, 75 casos, técnica 3 150 casos. Então o valor deve ser 150 casos.

Determinação da dimensão da amostra

Vão ser enunciadas algumas regras do polegar para algumas técnicas univariadas e multivariadas. É de referir que vão ser apresentados os valores mínimos das dimensões das amostras para cada técnica estatística e que se pretende utilizar um $\alpha=0.05$

Teste t para duas amostras independentes

60 casos no total. 30 casos em cada amostra. Se não houver dimensões iguais nas duas amostras, a diferença não deve ser muito grande e o mínimo de uma das amostra não deve ser inferior a 15 casos.

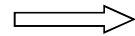
Análise de variância simples (one way ANOVA)

O tamanho mínimo da amostra depende do número de grupos (k), i.e., o número de níveis da variável independente.

Determinação da dimensão da amostra

Análise de variância simples (one way ANOVA) (cont.)

k	N (total)
3	80
4	95
5	105
6	115



A dimensão de cada nível (k) da variável independente deve ser k/N

Coeficiente de correlação paramétrica (tipo Pearson)

A amostra deve ter uma dimensão mínima de 40 casos. Mas se utilizar um nível de significância de 0.01, o mínimo pode ser de 30 casos.

Qui-quadrado

Para uma tabela de contingência do tipo $r \times c$, onde r é o número de linhas, ou seja o número de categorias da variável 1 e c o número de colunas, i.e., o número de categorias da variável 2.

O quadro abaixo indica o número da dimensão da amostra para cada tipo de tabela de contingência:

Determinação da dimensão da amostra

Qui-quadrado (cont)

		Número de colunas c				
		2	3	4	5	6
Número de Linhas r	2	45	55	65	70	80
	3	55	70	85	95	100
	4	65	65	100	110	120
	5	70	95	110	125	135
	6	80	100	120	135	150

Determinação da dimensão da amostra

—algumas considerações

- ▶ Em muitos casos é possível determinar o tamanho mínimo de uma amostra para estimar um parâmetro estatístico, como por exemplo, a MÉDIA POPULACIONAL (μ) .
- ▶ A fórmula para cálculo do tamanho da amostra para uma estimativa confiável da MÉDIA POPULACIONAL (μ) é dada por:

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

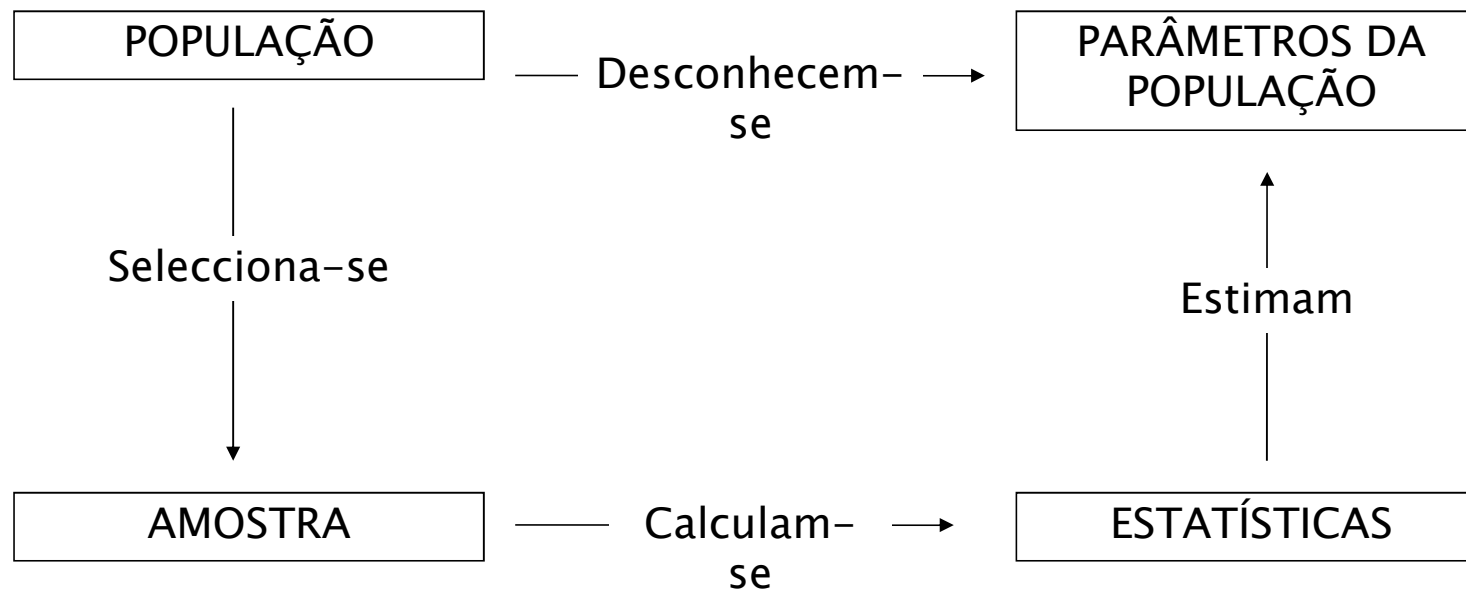
n = Número de indivíduos na amostra

$Z_{\alpha/2}$ = Valor crítico que corresponde ao grau de confiança desejado.

σ = Desvio-padrão populacional da variável estudada

E = Margem de erro ou ERRO MÁXIMO DE ESTIMATIVA. Identifica a diferença máxima entre a MÉDIA AMOSTRAL (\bar{X}) e a verdadeira MÉDIA POPULACIONAL.

Algumas considerações



Algumas considerações

- ▶ Enquanto que o objetivo de um estudo observacional é de recolher informação acerca de uma população selecionando e observando uma amostra da população tal qual ela se apresenta, pelo contrário, uma experimentação impõe um tratamento às unidades experimentais com o fim de observar a resposta. O princípio base de uma experimentação é o método da comparação, em que se comparam os resultados obtidos na variável resposta de um grupo de tratamento com um grupo de controlo.

Algumas considerações

Unidades experimentais são os objectos sobre os quais incide a experimentação e a quem é aplicado uma condição experimental específica, a que chamamos **tratamento**. **Variável resposta** é a variável cujo comportamento pretendemos estudar. **As variáveis explanatórias** são as variáveis que explicam ou causam mudanças na variável resposta.

▶ Exemplo

- ▶ Será que a aspirina reduz o perigo de um ataque cardíaco?
- ▶ O estudo conhecido por Physicians' Health Study, foi uma experimentação médica levada a cabo com o objetivo de responder a esta questão específica. Metade de um grupo de 22000 médicos (homens) foram escolhidos aleatoriamente para tomar aspirina todos os dias. A outra metade dos médicos tomou um placebo, que tinha o mesmo sabor e aspecto da aspirina. Depois de vários anos 239 médicos do grupo que tomou placebo, contra 139 do grupo que tomou aspirina, tiveram ataques cardíacos. Esta diferença é suficientemente grande para evidenciar o efeito da aspirina na prevenção dos ataques cardíacos.

Algumas considerações

- ▶ No estudo considerado anteriormente temos:
- ▶ Unidades experimentais – 22000 médicos
- ▶ Tratamentos - aspirina ou placebo
- ▶ Variável explanatória - se o indivíduo tomou aspirina ou placebo
- ▶ Variável resposta – se o indivíduo teve ou não ataque cardíaco.

Organização dos dados



Software



Cálculos das Estatísticas Básicas

- **Microsoft Excel**
 - Software comercial, mas presente em muitos computadores



Área de Biológicas

- **BioEstat (Livre) – Simple mas completo**
 - www.mamiraua.org.br/download/
- **OpenEpi (Livre) - Online – Só testes paramétricos**
 - www.openepi.com
- **BioStat (comercial) - Completo**
 - www.analystsoft.com/br/products/biostat/
- **Sigmaplot+SigmaStat (comercial) – Completo**



Gerais

- **Minitab (comercial) – Completo e fácil de usar**
- **SPSS (comercial) – Completo e fácil de usar**
- **S-Plus (comercial) – Completo mas menos amigável**
 - Possui versão livre – R-Statistics – exige programação
- **Statistica (comercial)**

Organização dos dados

Variáveis Qualitativas

	Grupo A	Grupo B	Grupo C	Total
Infectado	13	9	5	27
Não infectado	7	11	15	33
Total	20	20	20	60

Variáveis Quantitativas

Amostra	Grupo A	Grupo B	Grupo C
1	2,5	3,5	4
2	3	2,7	3,5
3	1,3	2,4	3,1
4	2,6	-	-
5	-	-	-
•	-	-	-
•			
•			

Grupos independentes ou emparelhados (dependentes)



Medidas em vários órgãos de um mesmo elemento amostral

Medidas em vários tempos de um mesmo elemento amostral

DEPENDENTES

Pergunte-se: Se eu tirar essa medida deste grupo, e logo tirar o elemento amostral, ele irá também ser tirado de outros grupos? Se sim, então há dependência!

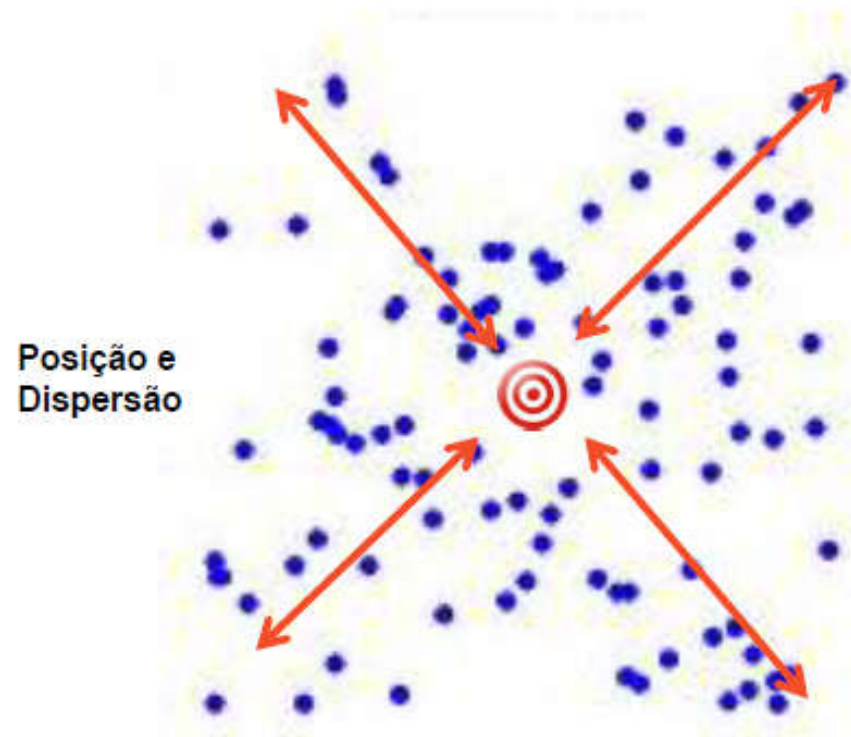


Cada medição provém de elementos amostrais distintos

Ao tirar uma medida, ou um elemento, os outros grupos não são afetados

INDEPENDENTES

Medidas resumo



Medidas de posição

MÉDIA

- É a medida centralizadora mais usada.
- Para calcular faça some os elementos e divida pelo número de elementos
- Ex: $(3 + 5 + 9 + 4 + 8 + 2) / 6 = 5.16$.

3	5	9	4	8	2
---	---	---	---	---	---

MEDIANA

- Leva em conta não a grandeza dos números mas sua disposição
- Ordena-se os elementos e toma-se o central (ou média dos centrais)
- Para calcular ordene os valores: 2, 3, 4, 5, 7, 8, 9.
- A mediana é vale 5

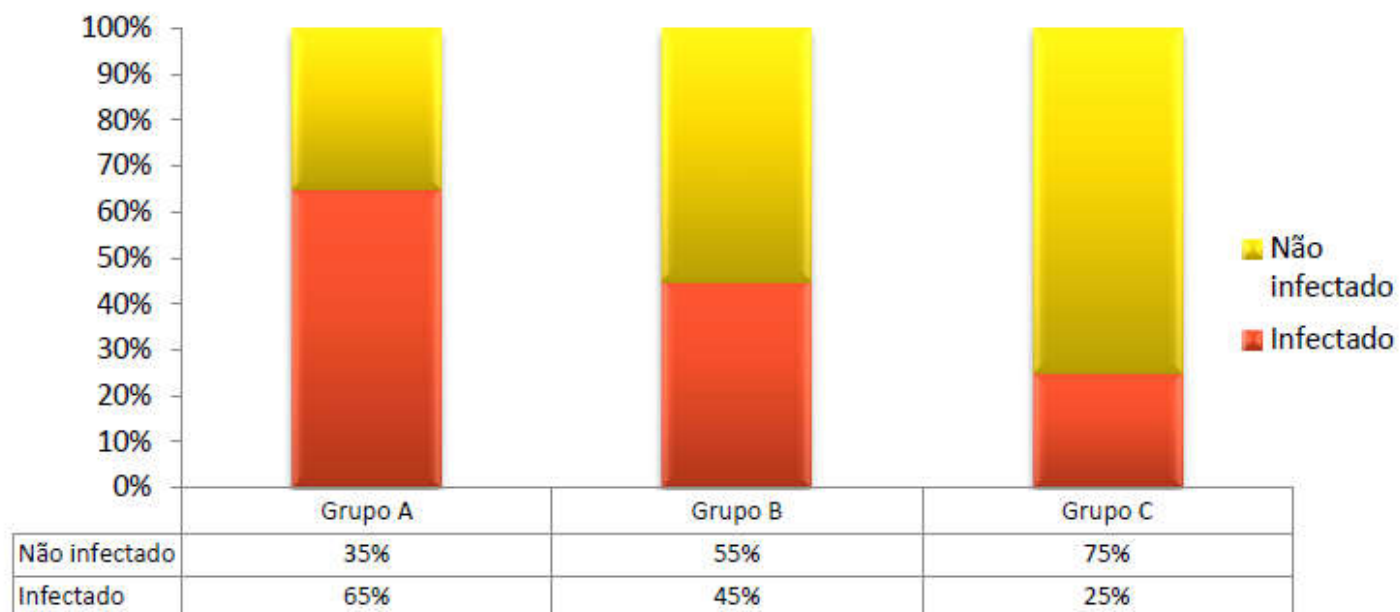
3	5	9	7	4	8	2
---	---	---	---	---	---	---

Medidas de dispersão

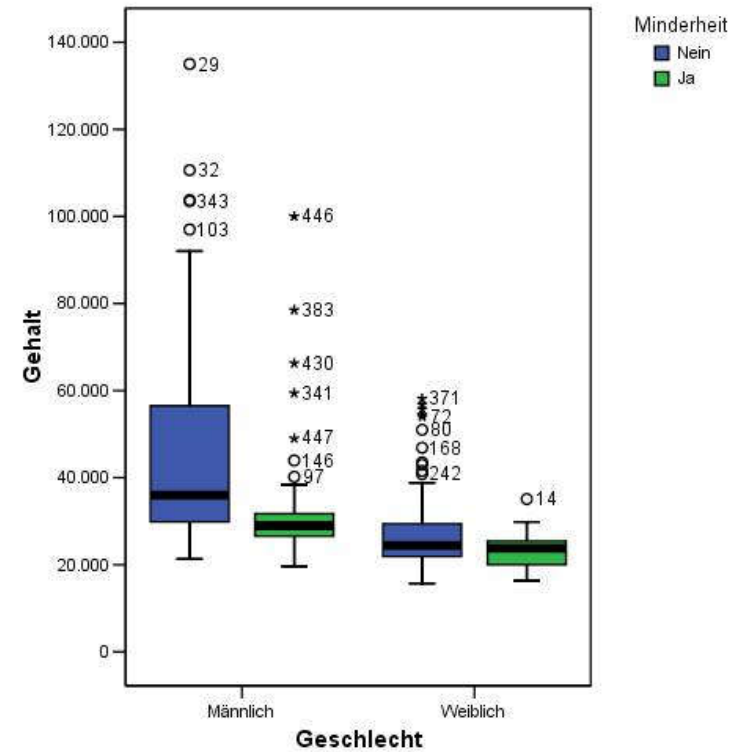
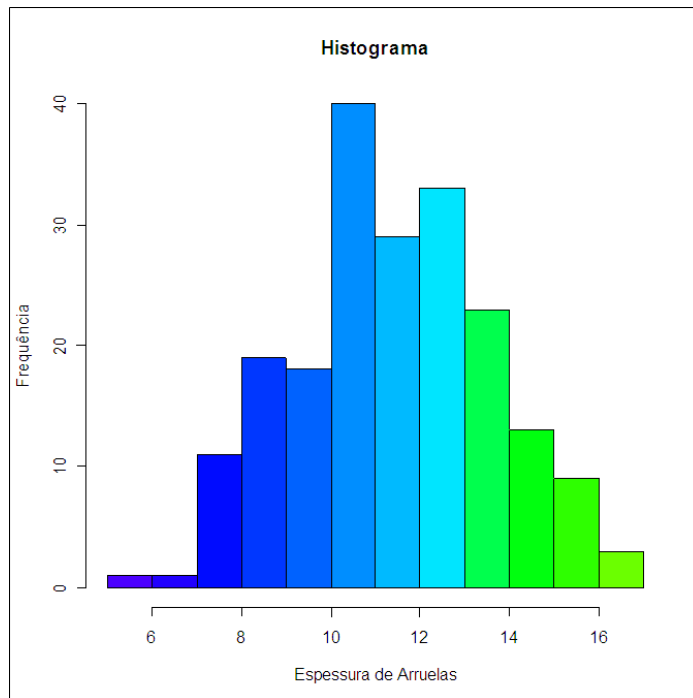
DESVIO PADRÃO AMOSTRAL

- É a medida de dispersão mais usada.
- Soma-se as diferenças com relação às médias ao quadrado, e depois divide-se por $N-1$ (e não N , Por quê?). Depois toma-se a raiz.

Visualização de variáveis qualitativas



Gráficos-dados quantitativos



Grupos INDEPENDENTES

	Grupo A	Grupo B	Grupo C
1	3,07	5,68	2,46
2	3,72	7,02	2,21
3	3,23	7,57	2,44
4	3,74	7,15	3,82
5	3,86	5,49	3,49
6	3,32	5,36	2,25
7	3,48	5,29	2,27
8	3,27	5,34	3,86
9	3,91	5,17	2,67
10	3,59	6,00	2,04
Média	3,52	6,01	2,75
Desvio Padrão	0,29	0,90	0,70

Gráfico de pontos

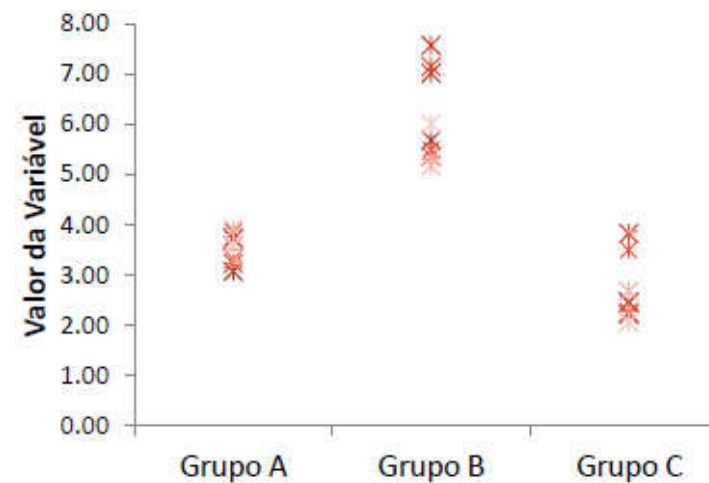
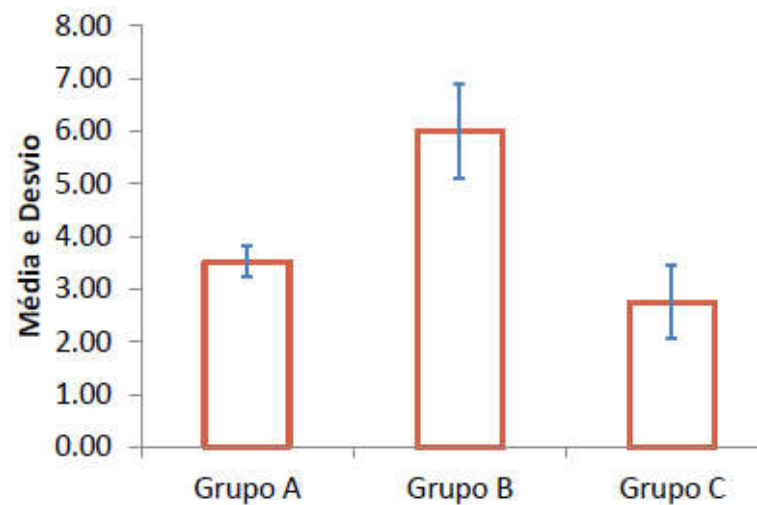


Gráfico Resumo



- ▶ Resolução de exercícios com recurso ao SPSS