

ANÁLISE EXPLORATÓRIA

Introdução

2012

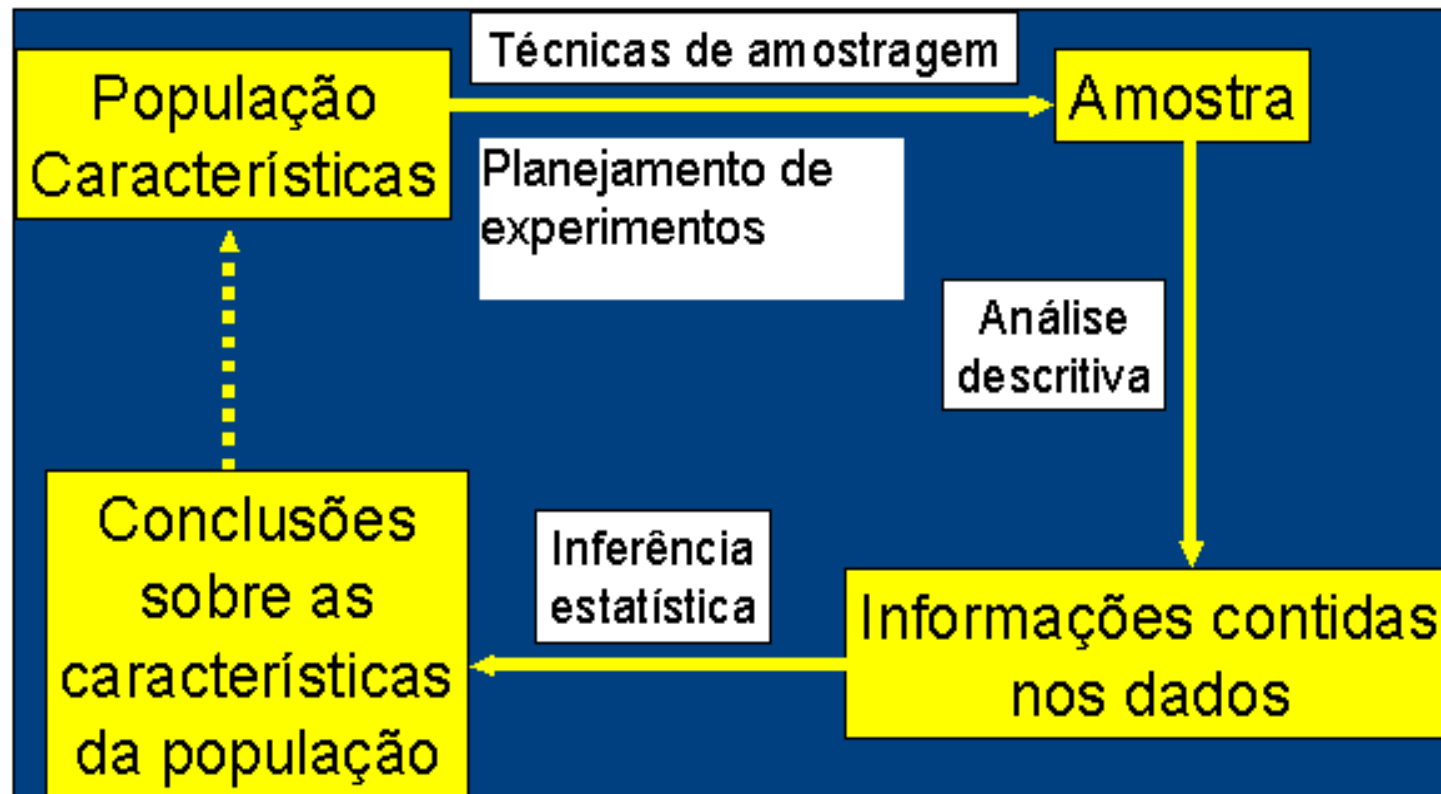
Estatística Descritiva e Análise Exploratória

Etapas iniciais. Utilizadas para descrever e resumir os dados. A disponibilidade de uma grande quantidade de dados e de métodos computacionais muito eficientes revigorou estas áreas da Estatística.

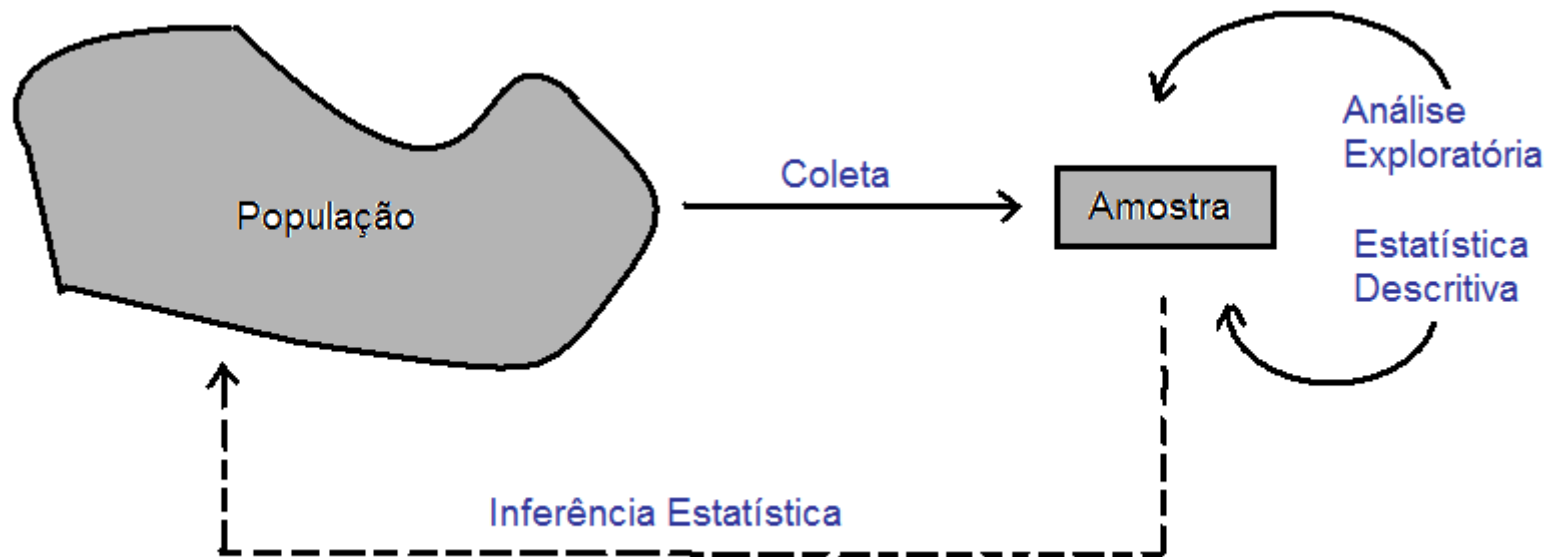
Probabilidade

Permite estudar os fenômenos aleatórios, ou seja, aqueles em que está presente a incerteza sobre os seus resultados.

Estatística



Estatística



O que é Estatística ?

Para muitos, Estatística não passa de conjuntos de tabelas de dados numéricos. Os estatísticos são pessoas que coletam esses dados.

A Estatística originou-se com a coleta de dados e a construção de tabelas para os governos.
A situação evoluiu e esta coleta de dados representa somente um dos aspectos da Estatística.

Definição de Estatística

A Estatística é uma ciência baseada na **Teoria da Probabilidade**, cujo objetivo principal é nos auxiliar a tomar decisões ou tirar conclusões em situações de incerteza, a partir de dados.

População: conjunto de todas as unidades que são de interesse em um certo estudo.

Amostra: qualquer subconjunto da população selecionado de acordo com certas regras.

Censo: estudo que inclui todos os elementos da população.

Coleta

Experimento planejado

Efeito de um ou mais fatores sobre outro(s).

Interferência do pesquisador.

Controle sobre fatores externos.

Levantamento observacional

Dados são coletados “como estão”.

Não há interferência do pesquisador.

Levantamento amostral (*survey*)

População bem definida.

Protocolo de coleta.

Amostragem

Uma área importante em muitas aplicações estatísticas é a da **Tecnologia de Amostragem**.

Exemplos:

- Pesquisa de mercado,
- Pesquisa de opinião,
- Avaliação do processo de produção.

Exemplo em R (www.r-project.org)

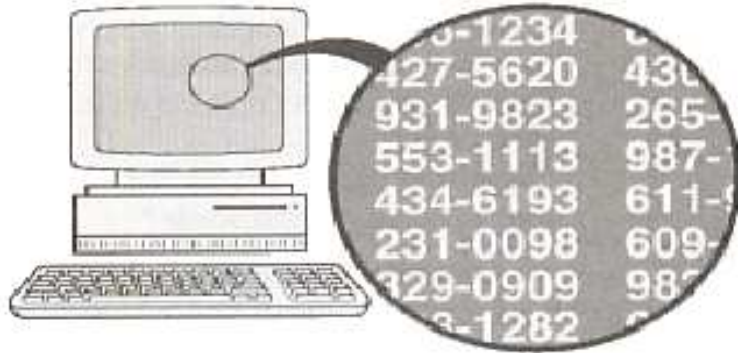
```
> alunos = read.csv("ListadePresença.csv", header = TRUE, sep = ";")

> names(alunos)
[1] "Codigo"    "Ingresso"  "Curso"     "Nome"

> nal = dim(alunos)[1]

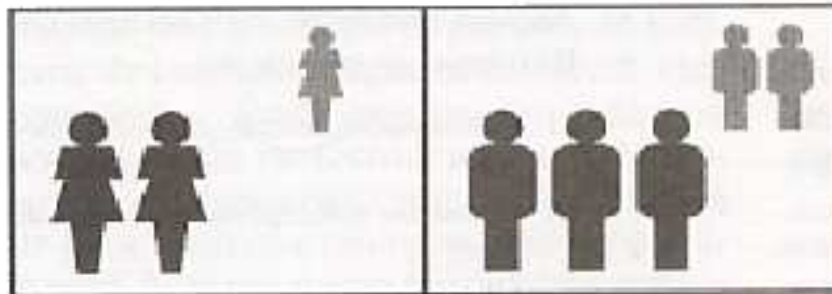
>(amostra = sample(1:nal, 5))
[1] 29  3 38 20 12

> alunos[amostra, -dim(alunos)[2]]
  Codigo Ingresso Curso
29 6426472 2008/1 55051
3  5910474 2007/1 97001
38 6811298 2009/1 18083
20 6811392 2009/1 18083
12 6882318 2009/1 18083
```



Amostragem Aleatória

Cada elemento da população tem uma chance conhecida de ser selecionado.



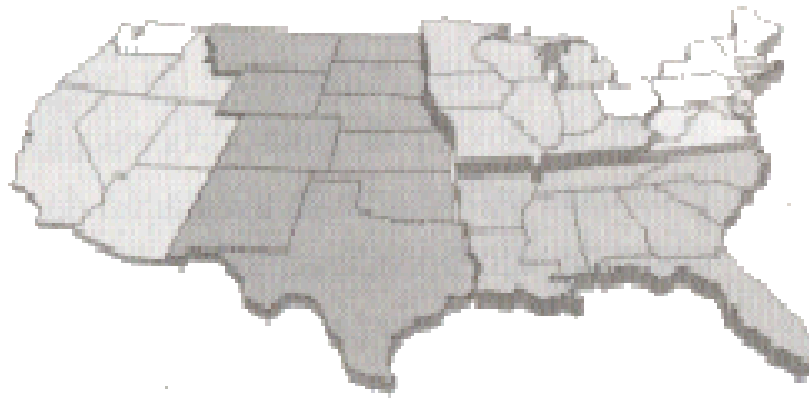
Amostragem Estratificada

Classificar a população em pelo menos dois estratos e selecionar uma amostra de cada um.



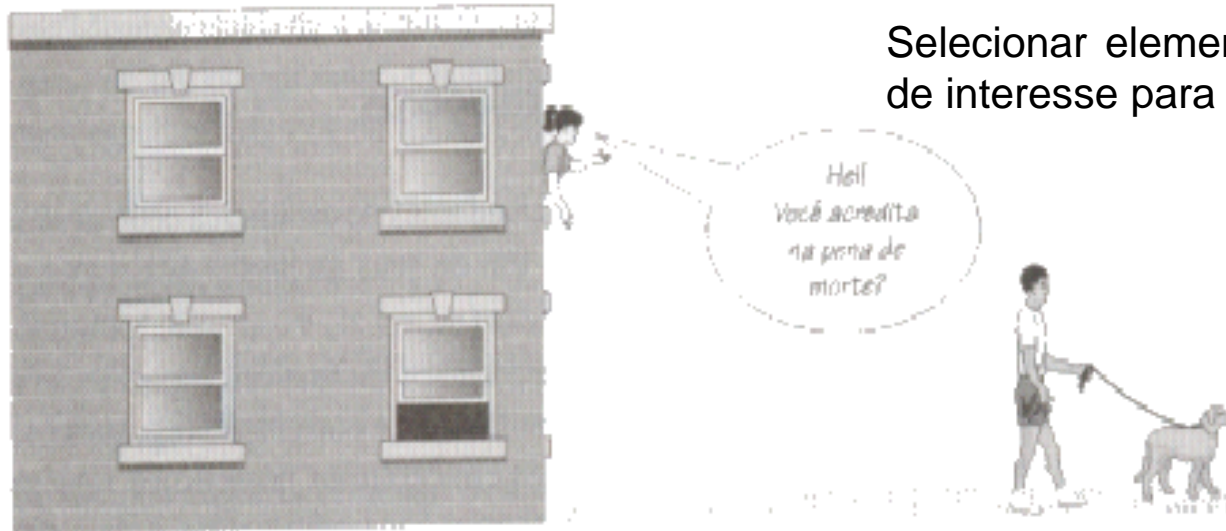
Amostragem Sistemática

Selecionar um elemento a cada k .



Amostragem por Conglomerados

Dividir a população em conjuntos homogêneos, mas com elementos heterogêneos. Selecionar aleatoriamente alguns destes conjuntos e tomar amostras deles.



Amostragem por Conveniência

Selecionar elementos de fácil acesso ou de interesse para o estudo.

Exemplo

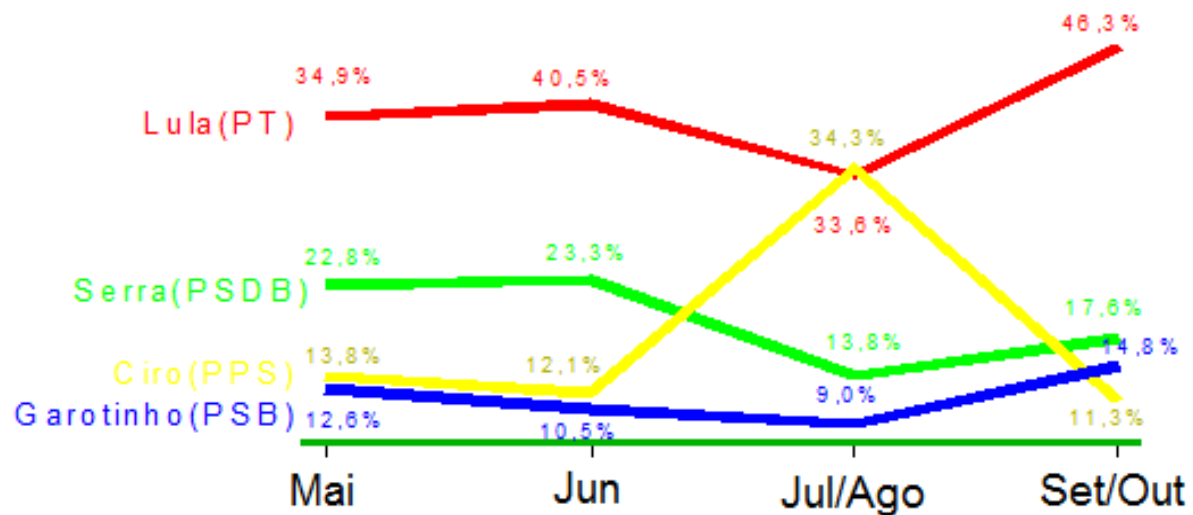
Numa pesquisa eleitoral um instituto de pesquisa procura, com base nos resultados de um levantamento aplicado a uma amostra da população, prever o resultado da eleição.

Eleição presidencial. Os institutos de pesquisa de opinião colhem periodicamente amostras de eleitores para obter as estimativas de intenção de voto da população. As estimativas são fornecidas com um valor e uma margem de erro.

A figura a seguir (**Instituto Toledo & Associados**) refere-se à intenção de voto no 1º turno das eleições para presidente em 2002.

Intenção de voto para presidente do Brasil-2002

Voto estimulado, em % do total de votos. A última pesquisa ouviu 2.202 eleitores. Margem de erro de 2,09%.



Fonte: Pesquisa Toledo & Associados.