

# **Principais Conceitos em Estatística**

**Ernesto F. L. Amaral**

**08 de outubro de 2009**

**[www.ernestoamaral.com/met20092.html](http://www.ernestoamaral.com/met20092.html)**

**Fonte:**

**Triola, Mario F. “Introdução à estatística”. 10<sup>a</sup> ed., Rio de Janeiro: LTC, 2008.  
pp. 2-31.**

# AULAS DE METODOLOGIA QUANTITATIVA

1. Procedimentos de Pesquisa em Ciências Sociais

21/09/2009 - segunda-feira

2. Principais Conceitos de Indicadores Sociais

(Jannuzzi 2006, 13-36)

05/10/2009 - segunda-feira

**3. Principais Conceitos em Estatística**

**(Triola 2008, 2-31)**

**08/10/2009 - quinta-feira**

4. Principais Conceitos em Econometria

(Wooldridge 2008, 1-17)

15/10/2009 - quinta-feira

5. Fontes de Dados de Indicadores Sociais

05/11/2009 - quinta-feira

# CONCEITOS BÁSICOS

- **Dados** são observações coletadas de um determinado grupo de interesse.
- **Estatística** (como ciência) é um conjunto de métodos para o planejamento de estudos e experimentos, obtenção de dados e consequente organização, resumo, apresentação, análise, interpretação e elaboração de conclusões baseadas nos dados.
- **População** é a coleção completa de todos os elementos a serem estudados, já que inclui todos os sujeitos a serem estudados.
- **Censo** é um conjunto de dados obtidos de todos os membros da população.
- **Amostra** é um subconjunto de membros selecionados de uma população.

# TIPOS DE DADOS

- **Parâmetro** é uma medida numérica que descreve alguma característica de uma população.
- **Estatística** é uma medida numérica que descreve alguma característica de uma amostra.
- **Dados quantitativos** são números que representam contagens ou medidas (renda, anos de escolaridade).
  - **Dados discretos** são aqueles em que o número de valores possíveis são finitos ou “enumeráveis” (número de cômodos em um domicílio).
  - **Dados contínuos** resultam de infinitos valores possíveis em uma escala contínua (renda per capita).
- **Dados qualitativos** (ou categóricos ou de atributos) podem ser separados em diferentes categorias que se distinguem por alguma característica não-numérica (sexo, ideologia política).

# NÍVEIS DE MENSURAÇÃO DE DADOS

- Nível **nominal** de mensuração possui dados que informam nomes, rótulos ou categorias:
  - Os dados não são ordenados e não devem ser usados para cálculos de médias.
  - Raça e código postal, por exemplo.
- Nível **ordinal** de mensuração engloba dados que podem ser organizados em alguma ordem:
  - Sabemos que há diferenças relativas entre os valores dos dados, mas não sabemos as magnitudes das diferenças.
  - Na escala de frequência (pouco/médio/muito), é possível ordenar os dados, mas não sabemos se a diferença entre “pouco” e “médio” é o mesmo que “médio” e “muito”.

# NÍVEIS DE MENSURAÇÃO DE DADOS

- Nível **intervalar** de mensuração é similar ao ordinal, mas sabemos as magnitudes das diferenças entre dois valores:
  - Os dados não possuem um ponto inicial zero natural.
  - Sabemos as magnitudes das diferenças entre os anos censitários (1970, 1980, 1991 e 2000), mas o tempo não começou em zero.
- Nível de mensuração de **razão** é similar ao intervalar, mas há um ponto inicial zero natural:
  - Como há um zero que indica nenhuma quantidade, é possível dizer que uma quantidade é maior que outra em  $X$  vezes (razões significativas).
  - 30 anos de idade é 6 vezes maior do que 5 anos de idade, por exemplo.

# NÍVEIS DE MENSURAÇÃO DE DADOS

Nível	Resumo	Exemplo
Nominal	Apenas categorias. Os dados não podem ser arranjados em um esquema de ordem. Há categorias ou nomes apenas.	Município de residência.
Ordinal	As categorias são ordenadas, mas as diferenças não podem ser encontradas ou não têm significado.	Frequência à igreja: pouco, médio, muito.
Intervalar	As diferenças são significativas, mas não existe ponto inicial zero natural e as razões não têm sentido	Ano censitário (não há tempo zero).
Razão	Há um ponto inicial zero natural e as razões são significativas.	Taxa de desemprego.

## PENSAMENTO CRÍTICO

- “O sucesso em um curso introdutório de estatística normalmente requer mais *senso comum* do que habilidade matemática.” (Triola 2008, p.9)
- É importante interpretar os dados e resultados com métodos científicos, e saber usar o senso comum.
- “Há três tipos de mentiras: mentiras, mentiras horríveis e estatística.” (Benjamin Disraeli)
- “Números não mentem, mas mentirosos manipulam números.” (Autor desconhecido)
- “Algumas pessoas usam a estatística como os bêbados usam os postes - para apoio mais do que para iluminação.” (Andrew Lang)
- “Estatístico é um especialista que coleta números e depois os desvirtua.” (*Esar's Comic Dictionary*)



## PROBLEMAS EM AMOSTRAGEM

- Em grande parte, o objeto da estatística é o uso de dados amostrais para se fazer generalizações sobre uma população.
- Os dados amostrais devem ser coletados de modo apropriado, senão a amostra não será representativa para a população da qual foi obtida.
- Amostra de resposta voluntária (ou amostra auto-selecionada) é aquela em que os respondentes decidem, eles mesmos, se serão ou não incluídos.
  - É uma amostra tendenciosa e não deve ser usada para fazer afirmações de uma população maior.
  - Censo americano tenta corrigir esse problema com visita de recenseadores após envio dos questionários.
- Pequenas amostras podem gerar resultados não confiáveis.

## OUTROS PROBLEMAS NO USO DE DADOS

- Gráficos: pesquisas eleitorais de candidatos políticos.
- Pictogramas: desenhos de objetos podem distorcer dados.
- Questões intencionalmente fraseadas para provocar uma resposta desejada.
- Ordem das questões e dos itens podem enviesar dados coletados.
- A não-resposta ocorre com mais frequência em determinados grupos da população.
- Dados ausentes: pessoa fora do domicílio para trabalho, ou pessoa não possui telefone para responder tais pesquisas.
- Correlação indica que duas variáveis estão relacionadas. No entanto, correlação não implica causalidade.
- Estudos de interesse próprio: objetivos mercadológicos.
- Número preciso: ao invés de 191.856.686, use 191 milhões.

# MÉTODOS ESTATÍSTICOS

- Em um **estudo observacional**, as características específicas são observadas e medidas, mas não se procura modificar os sujeitos objeto do estudo.
  - Em um **estudo transversal**, os dados são observados, medidos e coletados em um ponto no tempo.
  - Em um **estudo retrospectivo** (ou de controle de caso), os dados são coletados do passado.
  - Em um **estudo prospectivo** (ou longitudinal ou de coorte), os dados são coletados no futuro de grupos (coortes) que compartilham fatores comuns.
- Em um **experimento**, é aplicado algum tratamento aos sujeitos, e os efeitos são observados:
  - Esses sujeitos são **unidades experimentais**.
  - Ocorre **confundimento** quando não é possível distinguir entre efeitos de diferentes fatores.

# PLANEJANDO EXPERIMENTOS

- É preciso controlar os efeitos das variáveis.
- **Experimento cego:** sujeitos e agentes não sabem se estão utilizando placebo ou medicamento.
- **Planejamento em blocos aleatorizado:** bloco é grupo de sujeitos semelhantes, mas os blocos são diferentes entre si (grupos de homens e mulheres). É preciso fazer amostra aleatória em cada bloco.
- **Planejamento experimental completamente aleatorizado.**
- **Planejamento experimental rigorosamente controlado:** sujeitos são escolhidos cuidadosamente de modo que aqueles que recebem cada um dos tratamentos são similares em relação ao que é importante no experimento.
- É importante ter um tamanho de amostra grande o bastante para se obter dados representativos. É preciso ainda utilizar métodos aleatórios de seleção dos sujeitos.

# TÉCNICAS DE AMOSTRAGEM

- Na **amostra aleatória**, os membros de uma população têm chances iguais de serem selecionados.
- Na **amostragem aleatória simples (AAS)**, toda amostra possível de ser obtida de tamanho  $n$  tem a mesma chance de ser escolhida.
- Na **amostragem probabilística**, os indivíduos do universo têm probabilidade conhecida e diferente de zero de serem selecionados para a amostra (mas não necessariamente a mesma probabilidade).
- Na **amostragem não-probabilística**, não são estimadas as probabilidades de seleção dos elementos da amostra.

## OUTRAS TÉCNICAS DE AMOSTRAGEM

- **Amostragem sistemática:** há a escolha de algum ponto inicial e a seguir é selecionado cada  $k$ -ésimo elemento da população.
- **Amostragem de conveniência:** usa resultados que são fáceis de coletar.
- **Amostragem estratificada:** a população é subdividida em pelo menos dois subgrupos diferentes (ou estratos), de modo que os sujeitos dentro do mesmo subgrupo tenham as mesmas características (sexo ou faixa etária). A seguir, é extraída uma amostra de cada subgrupo.
- **Amostragem por conglomerado:** a população é dividida em seções ou conglomerados (Estados da Federação). São então selecionados aleatoriamente alguns desses conglomerados e escolhidos todos seus membros.
- **Amostragem de múltiplos estágios:** técnicas combinadas.