

# AULAS 01 E 02

# Introdução

**Ernesto F. L. Amaral**

**06 e 20 de agosto de 2013**  
**Metodologia de Pesquisa (DCP 854B)**

**Fonte:**

**Babbie, Earl. 1999. “Métodos de Pesquisas de *Survey*”. Belo Horizonte: Editora UFMG. Capítulos 3-5 (pp.77-158).**

**Triola, Mario F. 2008. “Introdução à estatística”. 10<sup>a</sup> ed. Rio de Janeiro: LTC. Capítulo 1 (pp.2-31).**

**Wooldridge, Jeffrey M. 2008. “Introdução à econometria: uma abordagem moderna”. São Paulo: Cengage Learning. Capítulo 1 (pp.1-17).**

# ESTRUTURA DA AULA

- 1) **Pesquisa de survey como método das ciências sociais:**  
Babbie (cap.3)
- 2) **Tipos de desenhos de pesquisa:** Babbie (cap.4)
- 3) **A lógica da amostragem do survey:** Babbie (cap.5)
- 4) **Tipos de dados:** Triola (cap.1)
- 5) **Tipos de bancos de dados:** Babbie (cap.4); Wooldridge (cap.1)
- 6) **Sejamos críticos:** Triola (cap.1)

# **1) Pesquisa de *survey* como método das ciências sociais**

**Babbie, 1999: 77-91 (capítulo 3)**

## PESQUISA DE *SURVEY*

- Babbie (1999) aborda o método de pesquisa com aplicação de questionários (pesquisa de *survey*).
- Esse método pode ser aplicado a uma grande quantidade de desenhos de pesquisa.
- O método consiste em aplicar um questionário a uma amostra da população de interesse.
- As entrevistas podem ser pessoais, por telefone ou por correio.
- As respostas são codificadas de forma padronizada e registradas de forma quantitativa em programas computacionais.
- Os dados são analisados, correlacionados e generalizados para a população.

## CARACTERÍSTICAS CIENTÍFICAS DO *SURVEY*

- É um tipo de ferramenta disponível aos pesquisadores sociais.
- Os métodos de *survey* podem não ser apropriados para determinados tópicos de pesquisa.
- De todo modo, o *survey* pode ser usado em combinação com outros métodos.
- Esse tipo de pesquisa tem função pedagógica, porque suas deficiências ficam mais claras do que em outros métodos, permitindo avaliações mais conscientes de suas implicações.

## PESQUISA DE *SURVEY* É LÓGICA

- Mesmo que haja comportamentos racionais e irracionais, os cientistas sociais devem procurar entender o comportamento social racionalmente.
- Os dados de *survey* facilitam a aplicação cuidadosa do pensamento lógico.
- O formato da pesquisa de *survey* permite desenvolvimento e teste rigorosos de explicações lógicas.
- O exame de uma grande quantidade de entrevistas permite testar proposições complexas que envolvam diversas variáveis em interação simultânea.

## PESQUISA DE *SURVEY* É DETERMINÍSTICA

- Todo evento ou situação (resultado) tem razões determinadas que o antecederam (causa).
- Sempre que o pesquisador de *survey* procura explicar causas e razões de eventos, características e correlações observadas, a investigação deve assumir postura determinística.
- O sistema determinístico de causa e efeito é clarificado no *survey*, porque é possível elaborar um modelo lógico claro.
- A disponibilidade de numerosos casos e variáveis permite ao analista explicar uma variável (dependente) com base em muitas informações (variáveis independentes).
- É possível ainda analisar efeito de variáveis intervenientes.

## PESQUISA DE *SURVEY* É GERAL

- Ao buscar observar e entender padrões gerais de eventos e correlações entre acontecimentos, a teoria social busca a generalização de suas conclusões.
- Pesquisas amostrais são realizadas para entender a população maior da qual a amostra foi selecionada.
- Análises descritivas e explicativas em pesquisas de *survey* visam desenvolver proposições gerais sobre o comportamento humano.
- Grande número de casos permite que achados possam ser replicados entre subconjuntos da amostra.
- Relato cuidadoso da metodologia de um *survey* facilita réplicas posteriores por outros pesquisadores e/ou para outros subgrupos, permitindo testes e retestes.



## PESQUISA DE *SURVEY* É PARCIMONIOSA

- O cientista social busca explicar o maior número de fenômenos com o menor número de informações.
- Como um *survey* tem grande número de variáveis, é possível analisar importância relativa de cada uma.
- Ou seja, busca-se obter o máximo de compreensão com menor número de variáveis.
- As variáveis explicativas não precisam ser definidas a priori.
- É possível coletar uma grande quantidade de informações que serão usadas em diferentes modelos explicativos, com o objetivo de selecionar o mais adequado.

## PESQUISA DE *SURVEY* É ESPECÍFICA

- É preciso definir exatamente quais são os métodos e conceitos utilizados na pesquisa social.
- A medição de cada variável no *survey* é construída a partir de respostas específicas a itens específicos de questionário, o qual é codificado e quantificado de forma específica.
- Ao explicitar como métodos, medidas e conceitos foram elaborados, o *survey* permite ser avaliado e criticado.
- Se variáveis não são conceituadas e medidas corretamente, correlações observadas podem não fazer sentido.
- A superficialidade e aproximações de todo tipo de pesquisa científica são mais aparentes e explícitas no *survey*.
- É mais difícil saber como conceitos são construídos e medidos na pesquisa qualitativa.

# OUTRAS CARACTERÍSTICAS DA PESQUISA DE *SURVEY*

- *Survey* é empiricamente verificável: proposições, teorias e hipóteses podem ser testadas na prática (no mundo real).
- *Survey* é intersubjetivo: detalhamento de métodos e técnicas utilizados permite que outro cientista replique estudo e alcance resultados semelhantes.
- *Survey* é aberto a modificações: coleta e quantificação permite fonte permanente de informações e nova análise dos dados quando há modificações de teorias.

## COMPARAÇÃO DO *SURVEY* COM OUTROS MÉTODOS

- Uma investigação abrangente será beneficiada se utilizar métodos diferentes focados em um tópico.
- A pesquisa de *survey* pode se beneficiar de conhecimentos de outros métodos:
  - Experimentação.
  - Análise de conteúdo.
  - Observação participante.

## PESQUISA DE *SURVEY* E EXPERIMENTAÇÃO

- Na lógica do experimento controlado (grupos de tratamento e controle) é realizado emparelhamento sócio-econômico de indivíduos, de forma que única diferença entre grupos é o recebimento da política.
- Na pesquisa de *survey*, o pesquisador também realiza o controle de variáveis, permitindo análises de subgrupos.
- Nos dois casos, o objetivo é de excluir influência de variáveis secundárias.

## PESQUISA DE *SURVEY* E ANÁLISE DE CONTEÚDO

- A codificação de respostas de questionários é uma forma de análise de conteúdo.
- Se forem coletadas perguntas abertas, o pesquisador de *survey* precisará categorizar as respostas posteriormente.
- Por isso, os pesquisadores de *survey* devem utilizar as experiências e métodos de analistas de conteúdo.

# PESQUISA DE *SURVEY* E OBSERVAÇÃO PARTICIPANTE

- Pesquisas de *survey* podem utilizar as técnicas desenvolvidas por pesquisadores que utilizam observação participante:
  - Obtenção de informações sobre interações inconscientes entre pessoas (*rappport*).
  - Manutenção da neutralidade quando a atividade de campo é realizada.
  - Realização de observações precisas ao coletar os dados.

# CIÊNCIAS SOCIAIS SÃO CIÊNCIAS DA COMPLEXIDADE

- Problemas de amostragem e de generalização são maiores no estudo do comportamento social.
- A pesquisa de *survey* é excelente para desenvolvimento de métodos úteis e de entendimento amplo.
- A medição de conceitos polêmicos (preconceito, por exemplo) pode ser de difícil aplicabilidade.
- A simples presença de pesquisadores no campo pode afetar entrevistados.
- Alguns participantes podem formar opiniões no momento em que são entrevistados.
- É preciso ter compreensão mais sofisticada sobre medição, conceituação, probabilidade e complexidade de correlações e causalidades nas ciências sociais do que nas naturais.



## **2) Tipos de desenhos de pesquisa**

**Babbie, 1999: 93-111 (capítulo 4)**

## TIPOS DE DESENHOS DE PESQUISA

- Há uma tendência a considerar a análise de dados de *survey* como mais desafiadora e interessante do que o desenho do *survey* e a coleta de dados.
- Durante a análise, começa-se a ganhar entendimento do assunto estudado e pode-se compartilhar as descobertas com colegas.
- No entanto, é preciso que haja a correta elaboração do desenho de pesquisa e sua execução, para que não haja problemas na etapa da análise.
- Portanto, o desenho é igualmente desafiador e importante.
- Há vários tipos de pesquisa de *survey*, com objetivos, custos, tempo e escopo diferenciados.

## FINALIDADES DA PESQUISA DE *SURVEY*

- Uma pesquisa de *survey* pode ser realizada para atender diferentes objetivos:
  - Intenção de voto.
  - Venda de produtos.
  - Projetar infra-estrutura.
  - Modificar programa social.
- De uma forma geral, três objetivos gerais permeiam os interesses de pesquisas de *survey*:
  - Descrição.
  - Explicação.
  - Exploração.

# DESCRIÇÃO

- *Surveys* são frequentemente realizados para descobrir a distribuição de certos traços e atributos da população, permitindo análise descritiva.
- Há preocupação em **como** é a distribuição de determinada variável (descrição), e não no **porquê** (explicação).
- Exemplos de dados descritivos:
  - Distribuições de candidatos por idade e sexo.
  - Taxas de aprovação do governo.
  - Percentual de políticos na escala esquerda/direita.
- Além de descrever a amostra total (e inferir para a população total), pesquisadores descrevem subgrupos.

# EXPLICAÇÃO

- Muitas pesquisas de *survey* têm o objetivo adicional de fazer análises explicativas sobre a população, procurando relações de causa e efeito.
- Explicar quase sempre requer análise multivariada, que é a análise do impacto simultâneo de duas ou mais variáveis (independentes) sobre outra variável (dependente).
- Há uma preocupação em explicar o porquê de determinada distribuição da variável, geralmente utilizando modelos estatísticos.
- Por exemplo, podemos querer explicar escala esquerda/direita de deputados por partido de pertencimento, Estado de residência, partido do governador...

# EXPLORAÇÃO

- Pesquisadores podem fazer pesquisa sobre um tema que está começando a ser investigado, e não se sabe exatamente suas causas e consequências.
- Primeiramente, pode ser realizada entrevista em profundidade (qualitativa) para captar o máximo de informações possíveis do público em estudo:
  - Não é necessário ter amostra representativa.
  - Não é preciso coletar dados com questionário padronizado.
- Posteriormente, as informações são analisadas para melhor entender os indivíduos, sem preocupação explicativa, mas visando a elaboração de questionário padronizado.

## UNIDADES DE ANÁLISE

- Em uma pesquisa de *survey*, dados são coletados para descrever cada unidade que está sendo analisada.
- As unidades de análise são tipicamente pessoas, mas podem ser domicílios, bairros, municípios, Estados, países, empresas, universidades...
- Os dados são agregados e manipulados, permitindo descrever a amostra estudada, além de realizar análises explicativas.
- Um mesmo *survey* pode envolver mais de uma unidade de análise, como é o caso da Pesquisa Nacional por Amostra de Domicílios (PNAD) do Instituto Brasileiro de Geografia e Estatística (IBGE), que coleta dados de pessoas e domicílios.

# FALÁCIA ECOLÓGICA

- Unidades de análise num *survey* podem ser descritas com base em seus componentes.
- Se o objeto da pesquisa for descrever os municípios e agregar as várias descrições para descrever todos municípios, a unidade básica de análise é o município.
- Pesquisador pode selecionar unidade de análise incorreta, incorrendo em falácia ecológica. Por exemplo:
  - Ao analisar taxas de criminalidade por raça, o correto seria ter dados por indivíduos.
  - Se utilizarmos dados municipais, poderíamos concluir que negros possuem maiores taxas de criminalidade.
  - No entanto, os crimes podem estar sendo realizados por brancos que residem em municípios com maioria negra.



## IDENTIFICAÇÃO DAS UNIDADES DE ANÁLISE

- É importante que as unidades de análise sejam identificadas no momento da montagem dos bancos de dados.
  
- O mais seguro é que em pesquisas que possuam mais de uma unidade de análise, sejam criados arquivos separados de dados para cada unidade.

# **3) A lógica da amostragem do *survey***

**Triola, 2008: 2-31 (capítulo 1)**

**Babbie, 1999: 113-158 (capítulo 5)**

# CONCEITOS BÁSICOS

- **Estatística** (como ciência) é um conjunto de métodos para:
  - Planejamento de estudos e experimentos.
  - Obtenção de dados.
  - Organização, resumo, apresentação, análise, interpretação e elaboração de conclusões baseadas nos dados.
- **Censo** é um conjunto de dados obtidos de todos os membros da população.
- **Amostra** é um subconjunto de membros selecionados de uma população.

## POR QUE FAZER AMOSTRAGEM?

- Há duas razões para selecionar uma parcela (amostra) da população e fazer estimativas sobre toda a população: tempo e custo.
- Muitas vezes *surveys* por amostragem são mais precisos do que entrevistar todos componentes da população, devido à logística do *survey*.
- Um grande número de entrevistas requer muitos pesquisadores, os quais podem não ter boa qualidade.
- A qualidade dos dados coletados se reduz com a menor qualidade dos entrevistadores.
- Uma pesquisa em menor escala permite procedimentos de acompanhamento mais severos.

## POR QUE FAZER AMOSTRAGEM? (cont.)

- Se forem realizadas uma grande quantidade de entrevistas, o tempo aumenta, dificultando especificar a data de referência da entrevista.
- O ponto médio da entrevista pode ser utilizado como data de referência.
- Também é possível pedir aos entrevistados para responder sobre uma data específica no passado, mas isso introduz problema de recordação imprecisa.
- Geralmente é recomendado que pesquisas de *survey* tenham período de coleta de dados de no máximo 3 meses.
- A supervisão, relatórios e treinamento são muito mais difíceis em um grande *survey*, e a qualidade dos dados pode ser menor do que aquela obtida com menos entrevistas.

# DADOS POR AMOSTRAGEM SÃO REALMENTE EXATOS?

- Se forem tomados procedimentos de amostragem bem estabelecidos, *surveys* por amostragem podem permitir estimativas muito precisas sobre as populações.
- Raramente é possível determinar com exatidão o grau de precisão dos achados de uma amostra.
- Pesquisadores de opinião política e eleitoral representam um grupo de analistas a quem é dada a oportunidade de verificar a precisão de seus dados (resultados reais da eleição).

## 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.

# HOMOGENEIDADE E HETEROGENEIDADE

- **Homogeneidade:** se todos os membros de uma população fossem idênticos uns aos outros em todos os sentidos, qualquer tamanho de amostra seria suficiente. Na realidade, bastaria um caso para estudar as características de toda população.
- **Heterogeneidade:** se há variação na população estudada, o pesquisador deve usar procedimentos de amostragem mais controlados.
  - Uma amostra de indivíduos de uma população deve conter a mesma variação existente na população, para permitir descrições úteis dela.
  - A amostragem probabilística é um método eficiente para extrair uma amostra que reflita corretamente a variação existente na população como um todo.



# VIÉS CONSCIENTE E INCONSCIENTE NA AMOSTRAGEM

- É preciso entrevistar pessoas com perfis diferenciados para representarem a população total.
- Porém, sem método de amostragem robusto, não saberemos as proporções adequadas dos tipos diversos de indivíduos que devem ser entrevistados.
- Há então o perigo dos vieses pessoais do pesquisador afetarem a amostra selecionada.
- Esse viés pode acontecer de maneira consciente ou inconsciente na coleta de dados.

# REPRESENTATIVIDADE E PROBABILIDADE DE SELEÇÃO

- Como vimos, as amostras de *surveys* devem representar as populações das quais são retiradas.
- Além disso, amostras necessitam ser representativas para os interesses substantivos da pesquisa.
- A representatividade vai depender se queremos fazer análises específicas por:
  - Áreas geográficas.
  - Nível de escolaridade.
  - Nível sócio-econômico.
  - Faixa etária.
  - Sexo...

# CONCEITOS E TERMINOLOGIA DE AMOSTRAGEM

- **Elemento** (unidade de análise) é a unidade sobre a qual a informação é coletada e que serve de base para a análise (pessoas, famílias, corporações, países...).
- **Universo** é a agregação teórica e hipotética de todos elementos definidos num *survey*. O universo não é especificado quanto a tempo e lugar.
- **População** é a agregação teoricamente especificada de elementos do *survey*. É a coleção completa de todos os elementos a serem estudados. É preciso definir o elemento e o referencial de tempo da pesquisa. É a especificação teórica do universo.
- **População do *survey*** é a agregação de elementos da qual a amostra do *survey* é de fato selecionada. Nem todos elementos da população terão chance de ser selecionados para a amostra.

## CONCEITOS E TERMINOLOGIA DE AMOSTRAGEM (cont.)

- **Unidade de amostra** é o elemento ou conjunto de elementos considerados para seleção em alguma etapa da amostragem.
  - Numa amostra simples, as unidades de amostra são o mesmo que os elementos.
  - Em outros casos, uma amostra pode ter diferentes unidades de amostra (setores censitários, domicílios, pessoas), mas só as pessoas são elementos.
- **Moldura de amostragem** é a lista de unidades de amostra da qual a amostra é selecionada.
  - Na amostra de etapa única, a moldura de amostragem é a lista dos elementos compondo a população de *survey*.
  - Muitas vezes, são as molduras que definem as populações de *survey*, e não o contrário.

## CONCEITOS E TERMINOLOGIA DE AMOSTRAGEM (cont.)

- **Unidade de observação** (unidade de coleta de dados) é um elemento (ou agregação de elementos) do qual se coleta dados.
  - Unidade de análise (elemento) e unidade de observação podem ser a mesma coisa (indivíduo, por exemplo).
  - Mas em uma pesquisa podemos entrevistar chefes de domicílio (unidades de observação) para coletar informação sobre todos membros (unidades de análise).
- **Variável** é um conjunto de características mutuamente excludentes, como sexo, idade, emprego...
  - Podemos descrever os elementos de uma população com base nas características individuais das variáveis.

## CONCEITOS E TERMINOLOGIA DE AMOSTRAGEM (cont.)

- **Parâmetro** é uma medida numérica que descreve alguma característica de uma população.
- **Estatística** é uma medida numérica que descreve alguma característica de uma amostra.
- **Erro amostral**: pesquisa de *survey* busca estimar parâmetros com base em amostras, o que gera erros.
  - Teoria da probabilidade permite estimar o grau de erro.
- **Níveis de confiança** indicam o percentual de confiança de que sua estatística (amostral) está dentro de um **intervalo de confiança** do parâmetro (populacional).
  - Nível e intervalo são estimados usando o erro amostral.

# **4) Tipos de dados**

**Triola, 2008: 2-31 (capítulo 1)**

# TIPOS DE DADOS

- **Dados** são observações coletadas de um determinado grupo de interesse.
- **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 MEDIÇÃO

- Variáveis sociais possuem diferentes níveis de medição.
- **Nominal:** distingue as categorias que compõem uma variável (sexo, religião, região de residência...). As categorias da variável nominal são mutuamente excludentes.
- **Ordinal:** as categorias de uma variável são ordenadas em uma escala (classe social, religiosidade, alienação...). Os números têm significado somente de indicação de ordem.
- **Intervalo:** usa números para descrever uma variável e distâncias entre pontos têm significado real. Diferença entre 20 e 40 graus Fahrenheit é a mesma que entre 60 e 80. Mas 40 não é necessariamente duas vezes mais quente que 20.
- **Razão:** é o mesmo que a medição de intervalo, mas tem zero real. Uma pessoa de 20 anos tem dobro de idade de uma pessoa de 10 anos.

# 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 (cont.)

- 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.

# RESUMO DOS 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.

# **5) Tipos de bancos de datos**

**Babbie, 1999: 93-111 (capítulo 4)**

**Wooldridge, 2008: 1-17 (capítulo 1)**

# DESENHOS BÁSICOS DE *SURVEY*: BANCOS DE DADOS

- Após especificar os objetivos e unidades de análise da pesquisa, é preciso escolher entre diversos desenhos diferentes:
  - *Surveys* interseccionais (*cross-sectional*).
  - *Surveys* longitudinais (tendências, coortes ou painel).
  - *Surveys* interseccionais servindo como longitudinais.
- Wooldridge (2008) classifica os dados econômicos em:
  - Dados de corte transversal = *surveys* interseccionais.
  - Cortes transversais agrupados = estudos de tendências.
  - Dados de séries de tempo = estudos de coortes.
  - Dados de painel ou longitudinais = estudos de painel.

# **DADOS DE CORTE TRANSVERSAL (Wooldridge)**

## ***SURVEYS INTERSECCIONAIS (Babbie)***

- Um conjunto de dados de corte transversal consiste em uma amostra de uma unidade de análise, tomada em um determinado ponto no tempo.
- Esses dados são muito utilizados em economia e em outras ciências sociais.
- Dados em um determinado ponto do tempo são importantes para testar hipóteses e avaliar políticas.
- Dados podem ter problemas de seleção amostral, no caso de determinados indivíduos não revelarem informações acuradas.
- Amostragem deve ser realizada de forma acurada para evitar que coleta se concentre em unidades com características semelhantes.

## EXEMPLO DE DADOS DE CORTE TRANSVERSAL

– Conjunto de dados de corte transversal para o ano de 1976 de 526 trabalhadores (Wooldridge 2008):

Número da observação	Salário por hora	Anos de escolaridade	Anos de experiência no mercado de trabalho	Feminino	Estado civil (casado)
1	3,10	11	2	1	0
2	3,24	12	22	1	1
3	3,00	11	2	0	0
4	6,00	8	44	0	1
5	5,30	12	7	0	1
...	...	...	...	...	...
525	11,56	16	5	0	1
526	3,50	14	5	1	0



# CORTES TRANSVERSAIS AGRUPADOS (Wooldridge)

## ESTUDOS DE TENDÊNCIAS (Babbie)

- Uma população pode ser amostrada e estudada em ocasiões diferentes.
- Um mesmo conjunto de variáveis é coletado em diferentes períodos do tempo, em **distintas** amostras aleatórias de uma mesma população (Censo Demográfico, Pesquisa Nacional por Amostra de Domicílios – PNAD).
- Agrupar cortes transversais de diferentes anos é eficaz para analisar os efeitos de uma política pública.
- O ideal é coletar dados de anos anteriores e posteriores a uma importante mudança de política governamental.
- Além de aumentar o tamanho da amostra, a análise de corte transversal agrupada é importante para estimar como uma relação fundamental mudou ao longo do tempo.
- Geralmente são utilizados dados secundários, coletados por outros pesquisadores ou instituições.

# EXEMPLO DE CORTES TRANSVERSAIS AGRUPADOS

– Conjunto de dados sobre os preços da moradia em 1993 e 1995 nos Estados Unidos (Wooldridge 2008):

Número da observação	Ano	Preço comercializado	Impro	Arquad	Quantidade de dormitórios	Quantidade de banheiros
1	1993	85.500	42	1.600	3	2,0
2	1993	67.300	36	1.440	3	2,5
3	1993	134.000	38	2.000	4	2,5
...	...	...	...	...	...	...
250	1993	243.600	41	2.600	4	3,0
251	1995	65.000	16	1.250	2	1,0
252	1995	182.400	20	2.200	4	2,0
253	1995	97.500	15	1.540	3	2,0
...	...	...	...	...	...	...
520	1995	57.200	16	1.100	2	1,5

# DADOS DE SÉRIES DE TEMPO (Wooldridge)

## ESTUDOS DE COORTES (Babbie)

- Um conjunto de dados de séries de tempo consiste em observações sobre variáveis ao longo do tempo.
- Como eventos passados podem influenciar eventos futuros, o tempo é uma dimensão importante em um conjunto de dados de séries de tempo.
- A análise desses dados pode ser dificultada, porque observações econômicas não são independentes ao longo do tempo (variáveis possuem padrões sazonais).
- Há uma série de frequências possíveis: diárias, semanais, mensais, trimestrais, anuais, decenais...
- Estes dados são também chamados de estudos de coorte, em que mesma população é analisada, mas amostras estudadas podem ser diferentes:
  - Pessoas com 10 anos em 2000, 20 anos em 2010, 30 anos em 2020, 40 anos em 2030...

## EXEMPLO DE DADOS DE SÉRIES DE TEMPO

– Conjunto de dados de séries de tempo sobre efeitos do salário mínimo em Porto Rico (apud Wooldridge 2008):

Número da observação	Ano	Salário mínimo médio no ano	Taxa de trabalhadores cobertos pela lei de salário mínimo	Taxa de desemprego	Produto Nacional Bruto (PNB)
1	1950	0,20	20,1	15,4	878,7
2	1951	0,21	20,7	16,0	925,0
3	1952	0,23	22,6	14,8	1.015,9
...	...	...	...	...	...
37	1986	3,35	58,1	18,9	4.281,6
38	1987	3,35	58,2	16,8	4.496,7

## DADOS DE PAINEL OU LONGITUDINAIS (Wooldridge) ESTUDOS DE PAINEL (Babbie)

- Um conjunto de dados de painel consiste em uma série de tempo para **cada** membro do corte transversal.
- Os dados de painel são distintos dos dados de corte transversal agrupados (tendências) e de séries de tempo (coortes), porque as **mesmas** unidades são acompanhadas ao longo de um determinado período.
- Dados de painel podem ser coletados para indivíduos, domicílios, instituições ou unidades geográficas.
- Esses dados são os mais sofisticados para fins explicativos, mas são mais difíceis e caros de se obter.
- Pode haver problema de grande número de não respostas nas últimas ondas de entrevistas.
- A análise dos dados pode se tornar complicada quando se tentar avaliar as mudanças dos indivíduos no tempo.

## EXEMPLO DE DADOS DE PAINEL OU LONGITUDINAIS

- Conjunto de dados de painel sobre crime e estatísticas relacionadas em 1986 e 1990 em 150 cidades nos Estados Unidos (Wooldridge 2008):

Número da observação	Cidade	Ano	Homicídios	População	Desemprego	Polícia
1	1	1986	5	350.000	8,7	440
2	1	1990	8	359.200	7,2	471
3	2	1986	2	64.300	5,4	75
4	2	1990	1	65.100	5,5	75
...	...	...	...	...	...	...
297	149	1986	10	260.700	9,6	286
298	149	1990	6	245.000	9,8	334
299	150	1986	25	543.000	4,3	520
300	150	1990	32	546.200	5,2	493

# CORTE TRANSVERSAL USADO COMO LONGITUDINAL

- Alguns mecanismos podem ser utilizados num *survey* interseccional (corte transversal) para aproximar o estudo de processo ou mudança (longitudinal).
- Podem ser realizadas perguntas referentes ao passado (renda no ano anterior, local de residência anterior):
  - Há problemas de erro de memória.
  - Os dados devem ser interpretados como amostra da população atual, e não de população passada.
- Por exemplo, é possível utilizar um único banco de dados de corte transversal para comparar pessoas de diferentes idades (jovens e idosos) e coortes (calouros e veteranos).

# VARIAÇÕES DOS DESENHOS BÁSICOS

- Os desenhos básicos de pesquisa apresentados anteriormente podem ser modificados para se enquadrarem aos objetivos de um estudo:
  - **Amostras paralelas:** amostras separadas de populações diferentes, utilizando mesmo questionário (exemplo é a pesquisa sobre preconceito na UFMG).
  - **Estudos contextuais:** uso de dados sobre o ambiente ou meio da pessoa para descrever o contexto do indivíduo.
  - **Estudos sociométricos:** intenção é de observar as inter-relações entre membros da população estudada (redes de amizades, por exemplo).



## ESCOLHENDO O DESENHO APROPRIADO

- **Dados de corte transversal** são mais apropriados se objetivo é descrição de tempo único.
- **Mudanças ao longo do tempo** são mais difíceis de realizar, porque dados de painel exigem tempo e recursos:
  - É possível utilizar dados de corte transversal e comparar pessoas que passaram por uma experiência no passado, com aqueles que não passaram.
- **Estudos de painel** são mais viáveis economicamente quando o fenômeno estudado tem duração curta (por exemplo, opinião de voto durante uma campanha eleitoral).
- **Estudos de tendências** podem ser realizados quando dados antigos são complementados com dados coletados pelo pesquisador.

# **6) Sejam os críticos**

**Triola, 2008: 2-31 (capítulo 1)**

## 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*)