

AULAS 01 E 02

Slides extras

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. Capítulo 5 (pp.113-158).

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

ESTRUTURA DA AULA

- 1) **Amostragem probabilística:** Babbie (cap.5)
- 2) **Amostragem não-probabilística:** Babbie (cap.5)
- 3) **Principais conceitos em econometria:** Wooldridge (cap.1)

1) Amostragem probabilística

Babbie, 1999: 113-158 (capítulo 5)

TEORIA DA AMOSTRAGEM PROBABILÍSTICA

- A **finalidade do *survey*** é selecionar um conjunto de elementos de uma população de modo que as estatísticas descrevam com precisão os parâmetros populacionais.
- A **amostragem probabilística** fornece métodos para ampliar a possibilidade de alcançar esta meta, bem como métodos para estimar o grau de sucesso provável.
- Em uma amostra com **seleção aleatória**, todos membros da população têm oportunidade igual de serem selecionados para a amostra.
- Os métodos de seleção aleatória:
 - Diminuem os vieses conscientes ou inconscientes do pesquisador.
 - Permitem utilizar a teoria da probabilidade para estimar parâmetros populacionais e erros amostrais.

DISTRIBUIÇÃO BINOMIAL DA AMOSTRAGEM

- Queremos avaliar uma **variável binomial** em nossa população, em que os valores possíveis são zero e um.
- A **distribuição amostral** é a distribuição das estatísticas das diferentes amostras selecionadas.
- Se **muitas amostras aleatórias** independentes são extraídas de uma população, as estatísticas amostrais fornecidas por estas amostras estarão distribuídas em torno do parâmetro populacional.
- A **teoria da probabilidade** possui uma fórmula para estimar o quão próximo as estatísticas da amostragem estão aglomeradas em torno do valor real.
- O **erro da amostragem** ou erro padrão (S) é medido com base nos parâmetros (P, Q) e tamanho da amostra (n):

$$S = \sqrt{\frac{PQ}{n}}$$

ERRO DA AMOSTRAGEM (ERRO PADRÃO)

- O **erro padrão** (S) indica em que medida as estimativas amostrais estão distribuídas em torno do parâmetro populacional.
- Aproximadamente 34% das estimativas amostrais estão até **um erro padrão** (desvio padrão) acima do parâmetro da população, e outros 34% até um desvio padrão abaixo do parâmetro.
 - Ou seja, 68% das amostras resultarão em estimativas dentro de (mais ou menos) um erro padrão.
- Cerca de 95% das amostras estarão dentro de (mais ou menos) **dois erros padrão** do valor verdadeiro.
- 99,9% das amostras estarão dentro de (mais ou menos) **três desvios padrão**.
- A proporção de amostras dentro de 1, 2 ou 3 erros padrão do parâmetro é constante em amostragens aleatórias.

TEORIA DA PROBABILIDADE NA PRÁTICA

- A teoria da probabilidade depende que se selecione **grandes números de amostras aleatórias**.
- Se **parâmetro for conhecido** e muitas amostras aleatórias forem selecionadas, podemos prever quantas amostras cairão dentro de intervalos específicos a partir do parâmetro.
- No entanto, geralmente **desconhecemos o parâmetro** e selecionamos **somente uma amostra**.
- Sabemos que uma amostra aleatória tem probabilidade de 68% de estar dentro da faixa de um erro padrão.
 - Ou seja, temos 68% de confiança que a estimativa da amostra está dentro de um erro padrão do parâmetro.
- Como não temos parâmetro, **usamos a estimativa da amostragem** na fórmula do erro padrão.
- Estamos x% confiante (**nível de confiança**) de que o parâmetro está entre dois valores (**intervalo de confiança**).

POPULAÇÕES E MOLDURAS DE AMOSTRAGEM

- É preciso discutir as relações entre as **condições de campo** e os pressupostos teóricos.
- A moldura de amostragem é a lista de elementos dos quais é selecionada uma amostra probabilística.
- O *survey* amostral é representativo dos membros desta lista.
- É importante saber até que ponto há semelhança ou diferença entre **populações e molduras de amostragem**.
- Se as molduras de amostragem são realmente representativas da população estudada, amostras corretamente selecionadas fornecem informações apropriadas para descrever a população.

SOBRE QUAL POPULAÇÃO ESTAMOS FALANDO?

- Os resultados dos *surveys* por amostragem só podem ser considerados representativos do conjunto de elementos que compõem a moldura de amostragem.
- As listagens não incluem de fato todos os elementos da população, por isso é preciso avaliar a extensão destas omissões e corrigi-las, se possível.
- No caso de existir omissões, é necessário informar nos relatórios de pesquisa, com o objetivo de expor corretamente qual população está sendo estudada.
- Para poder generalizar a população que compõe a moldura de amostragem, é necessário que todos elementos tenham representação igual (apareçam só uma vez).

MÉTODOS DE AMOSTRAGEM

- **Amostragem probabilística:** indivíduos do universo têm probabilidade conhecida e diferente de zero de serem selecionados para amostra (mas não necessariamente a mesma probabilidade).
 - Amostras probabilísticas são mais representativas do que outros tipos, porque são evitados os vieses de seleção.
 - Além disso, a teoria da probabilidade permite estimar a precisão ou representatividade da amostra.
- **Amostragem não-probabilística:** não são estimadas as probabilidades de seleção dos elementos da amostra.

AMOSTRAGEM ALEATÓRIA SIMPLES (AAS)

- O desenho da pesquisa é realizado com AAS quando toda amostra possível de ser obtida de tamanho n tem a mesma chance de ser escolhida.
- De posse da moldura de amostragem, você enumera cada elemento da lista, atribuindo só um número a cada um, sem saltos.
- Em seguida, é usada uma tabela de números aleatórios ou um programa computacional para selecionar os elementos da amostra.
- AAS é raramente realizada na prática, pois é necessário dispor de uma lista de elementos (moldura de amostragem) bem definida a priori.
- Mesmo de posse desta lista, geralmente os pesquisadores usam o método de amostragem sistemática.

AMOSTRAGEM SISTEMÁTICA

- Na amostragem sistemática, há a escolha de algum ponto inicial e a seguir é selecionado cada k -ésimo elemento da população.
- Por exemplo, se a lista tem 10.000 elementos e se deseja uma amostra de 1.000, é selecionado cada 10° elemento.
- **Amostra sistemática de início aleatório**: primeiro elemento é selecionado aleatoriamente.
- **Intervalo de amostragem** é a distância padrão entre os elementos selecionados na amostra (10 no exemplo acima).
- **Razão amostral** é a proporção de elementos selecionados da população: divisão de 1 pelo intervalo de amostragem.
- Há perigo da **periodicidade**, se lista de elementos estiver organizada com base em alguma informação da população.
- Resultados da AAS e da AS são idênticos, mas a AS é mais fácil e conveniente de ser realizada na prática.

AMOSTRAGEM ESTRATIFICADA

- AAS e AS são importantes para garantir grau de representatividade e permitir estimativa de erro.
- Na **amostragem estratificada**, intuito é de garantir representação das variáveis de estratificação para aumentar representação de outras variáveis.
- AE tem possibilidade de ser **mais representativa** em certas variáveis do que AAS, reduzindo erro amostral.
- Como erro é menor com população homogênea, idéia é de selecionar amostra dentro de **subconjuntos homogêneos** (com heterogeneidade entre subconjuntos).
- População é dividida em subgrupos (estratos) diferentes entre si. Sujeitos de um subgrupo possuem mesmas características. **Variáveis de estratificação** dependem do interesse de estudo e disponibilidade de informação.
- Com base na proporção relativa de cada subgrupo na população, **elementos dos subgrupos são selecionados** na mesma proporção do tamanho da amostra.

ESTRATIFICAÇÃO NA AMOSTRAGEM SISTEMÁTICA

- Se uma lista de elementos estiver agrupada por características da população, a amostragem sistemática realizará uma **estratificação implícita**.
- Por exemplo, se alunos são organizados por turma, sexo e notas em uma lista, uma amostra sistemática selecionaria números adequados de cada subgrupo.
- Uma lista ordenada de elementos pode ser mais útil do que uma lista não ordenada (aleatória).
- Nestas condições, a amostragem sistemática é mais precisa na representatividade dos subgrupos do que a amostragem aleatória simples.
- A lista só deve ser reordenada se o arranjo apresenta algum problema de periodicidade.

AMOSTRAGEM POR CONGLOMERADOS

- Muitas pesquisas não possuem listagens para fins de amostragem, tais como população de um município.
- Realizar amostragem de **grupos de elementos** e depois selecionar elementos de cada conglomerado.
- Idéia é de selecionar amostra dentro de **subconjuntos heterogêneos** (com homogeneidade entre subconjuntos).
- **Passos:** (1) população é dividida em conglomerados (seções); (2) são selecionados alguns dos conglomerados; e (3) seus membros são listados e amostrados.
- **Exemplo** de amostragem por conglomerados:
 - Ao pesquisar um município, pode ser feita lista de bairros (conglomerados) para amostragem AAS ou AS.
 - Após selecionar bairros, pode ser feita lista de quarteirões (conglomerados) para amostragem AAS ou AS.
 - Após selecionar quarteirões, é realizada listagem de domicílios para nova amostragem AAS ou AS.

PRECISÃO NA AMOSTRAGEM POR CONGLOMERADOS

- A amostragem por conglomerados é recomendada pela eficiência, embora não tenha muita precisão.
- Como a amostragem por conglomerados é geralmente realizada em múltiplas etapas, ela terá **um erro amostral para cada etapa**.
- A amostra inicial dos conglomerados representa a população de conglomerados em uma faixa de erro amostral.
- A amostra de elementos extraída de um conglomerado representa todos elementos, com um erro amostral.
- Há o **risco** de selecionar uma amostra de quarteirões desproporcionalmente ricos, assim como amostra de residências ricas dentro dos quarteirões.
- **Solução** é pensar: (1) na quantidade de conglomerados selecionados no início; e (2) no número de elementos selecionados em cada conglomerado.

AMOSTRA E HOMOGENEIDADE NOS CONGLOMERADOS¹⁷

- Sabemos que erro amostral é menor com o aumento do tamanho da amostra e com maior homogeneidade dos elementos a serem amostrados.
- Uma **amostra por conglomerados** representará melhor todos conglomerados se um **grande número** for selecionado e se todos eles forem bastante **semelhantes entre si**.
- Uma **amostra de elementos** representará melhor todos elementos de um conglomerado se um **grande número** de elementos for selecionado do conglomerado, e se todos os elementos forem bastante **semelhantes entre si**.
- Dado um tamanho máximo de amostra, se o número de conglomerados aumentar, o número de elementos num conglomerado deve diminuir.
- Ou seja, a representatividade dos conglomerados aumenta à custa da menor representação dos elementos de cada um, e vice-versa.

DIRETRIZ GERAL DE AMOSTRA DE CONGLOMERADOS

- Os elementos de um conglomerado natural dentro de uma população são tipicamente mais homogêneos do que todos os elementos da população total.
- Por exemplo, os moradores de um bairro são mais semelhantes entre si do que todos os moradores do município.
- Podem ser necessários **poucos elementos** para representar adequadamente um conglomerado natural, enquanto um **número maior de conglomerados** pode ser necessário para representar adequadamente a diversidade dos conglomerados.
- A **diretriz geral** (científica) no desenho de conglomerados é maximizar o número de conglomerados selecionados, diminuindo o número de elementos de cada um.
- Porém, um número menor de conglomerados pode ser listado com mais agilidade e economia.

REGRA SIMPLES DE AMOSTRA DE CONGLOMERADOS

- Pesquisadores populacionais **convencionalmente** selecionam cinco residências por zona ou quarteirão de recenseamento.
- Por exemplo, querendo entrevistar um total de 2.000 residências, selecionamos 400 quarteirões com cinco entrevistas de residências em cada um.
- De todo modo, amostragem por conglomerados envolve **perda de precisão**.
- Desenho amostral em etapas múltiplas está sujeito a erros amostrais em cada uma das etapas:
 - O erro amostral em cada etapa será maior do que no caso de uma amostra aleatória em etapa única, porque o tamanho da amostra é menor em cada etapa.
 - Isso está relacionado com a fórmula de cálculo do erro amostral: $s = \sqrt{\frac{PQ}{n}}$

AAS, AS E ESTRATIFICAÇÃO NOS CONGLOMERADOS

- Discutimos a amostragem por conglomerados como se fosse selecionada uma amostra aleatória simples a cada etapa do desenho.
- Informações podem ser usadas para **estratificar os conglomerados**:
 - Censo pode ser usado para estratificar quarteirões de recenseamento em termos de composição étnica, classe social, valores das propriedades, qualidade de estruturas, natureza da propriedade das construções e tamanho.
- Após **estratificação** das unidades primárias de amostragem de acordo com as variáveis disponíveis, pode-se usar tanto técnicas de amostragem **aleatória simples** ou amostragem **sistemática** para extrair a amostra.
- **Conglomerados e estratificação**: erro amostral será mais reduzido, quanto mais os conglomerados estiverem combinados em estratos homogêneos.

AMOSTRAGEM SIMPLES POR CONGLOMERADO

- Para garantir a seleção de uma amostra representativa de elementos, é preciso dar a cada elemento da população total uma oportunidade igual de ser selecionada.
- A forma mais simples de fazer isso numa amostra por conglomerados é dar a cada um a mesma chance de seleção e selecionar uma determinada proporção de elementos de cada conglomerado selecionado:
 - Temos 100.000 elementos agrupados em 1.000 conglomerados e queremos amostra de 1.000 elementos.
 - Selecionamos $1/10$ dos conglomerados (100) com igual probabilidade.
 - Depois selecionamos $1/10$ dos elementos de cada conglomerado escolhido.
 - Ou seja, cerca de 1.000 elementos foram selecionados e tiveram a mesma probabilidade de seleção.

AMOSTRAGEM ANTERIOR É SIMPLES E INEFICIENTE

- A técnica anterior é simples, mas não é a mais eficiente.
- A maioria das amostragens por conglomerados envolve conglomerados de tamanhos muito diversos.
- Ao selecionar conglomerados com igual probabilidade, retirando uma proporção fixa de elementos dos conglomerados selecionados:
 - Seleciona-se um número relativamente pequeno de conglomerados grandes na primeira etapa da amostragem.
 - Os elementos selecionados para representar todos os elementos dos grandes conglomerados são retirados de poucos desses conglomerados.
 - No limite, toda população da cidade residindo em dez grandes bairros poderia ser representada pelas pessoas vivendo em apenas um deles.

AMOSTRAGEM PPT

- Como vimos, a amostragem por conglomerados terá maior eficiência se houver a seleção de muitos conglomerados, com poucos elementos retirados de cada um.
- Esse princípio é utilizado pelo método de amostragem de probabilidade proporcional ao tamanho (PPT):
 - Há seleção de maior número de conglomerados.
 - Há garantia de representação dos elementos de conglomerados grandes.
 - Cada elemento na população recebe igual chance de seleção.

PROCEDIMENTOS DA AMOSTRAGEM PPT

- Na primeira etapa, é dada a cada conglomerado uma oportunidade de seleção proporcional ao seu tamanho.
- Na segunda etapa, o mesmo número de elementos é escolhido de cada conglomerado selecionado.
- Esses procedimentos igualam as probabilidades últimas de seleção de todos os elementos, já que:
 - Conglomerados com mais elementos têm maior probabilidade de serem selecionados.
 - Elementos em conglomerados maiores têm menor chance de seleção do que elementos em conglomerados menores.

FÓRMULA E EXEMPLO DE PPT

- A probabilidade de seleção de um elemento num desenho de amostragem PPT é:

$$[\textit{probabilidade do elemento}] = \left[\frac{\textit{número de conglomerados selecionados}}{\textit{conglomerados}} \right] \times \left[\frac{\textit{tamanho do conglomerado}}{\textit{tamanho da população}} \right] \times \left[\frac{\textit{elementos selecionados por conglomerado}}{\textit{tamanho do conglomerado}} \right]$$

- Se 100 conglomerados são selecionados e 10 elementos escolhidos de cada um, a partir de uma população total de 100.000, a probabilidade geral de seleção de cada elemento será de 1.000/100.000 ou 1/100.

- **Exemplo 1:** um conglomerado com 100 elementos:

$$\text{Prob}_1 = 100 \times (100/100.000) \times (10/100) = 1/100$$

- **Exemplo 2:** um conglomerado com 10 elementos:

$$\text{Prob}_2 = 100 \times (10/100.000) \times (10/10) = 1/100$$

- Independente do número de elementos num conglomerado, cada elemento tem a mesma probabilidade de seleção.

SIMPLIFICANDO A FÓRMULA DE PPT

- O tamanho do conglomerado é cancelado:

$$[probabilidade] = \left[\begin{array}{c} \text{número de} \\ \text{conglomerados} \\ \text{selecionados} \end{array} \right] X \left[\frac{\cancel{\text{tamanho do conglomerado}}}{\text{tamanho da população}} \right] X \left[\frac{\begin{array}{c} \text{elementos} \\ \text{selecionados} \\ \text{por conglomerado} \end{array}}{\cancel{\text{tamanho do} \\ \text{conglomerado}}} \right]$$

- Isso é o tamanho da amostra dividido pelo tamanho da população:

$$[probabilidade] = \frac{\left[\begin{array}{c} \text{número de} \\ \text{conglomerados} \\ \text{selecionados} \end{array} \right] X \left[\begin{array}{c} \text{elementos} \\ \text{selecionados} \\ \text{por conglomerado} \end{array} \right]}{\left[\begin{array}{c} \text{tamanho da} \\ \text{população} \end{array} \right]}$$

MODIFICAÇÕES NO DESENHO PPT

- Costuma-se fazer duas modificações no desenho de amostra PPT.
- **Primeiro**, podemos achar necessário representar conglomerados muito grandes na amostra.
 - Selecionamos todos esses conglomerados (probabilidade igual a 1) e seus elementos com a probabilidade geral.
- **Segundo**, se um número padrão de elementos for selecionado de cada conglomerado escolhido, conglomerados com menos elementos do que o número padrão representam um problema.
 - Solução é combinar pequenos conglomerados, de forma que cada combinação tenha pelo menos o número padrão a ser selecionado.
- Amostragem por conglomerados é difícil, mas é adequada quando for impossível obter uma lista de todos elementos da população.

AMOSTRAGEM PROBABILÍSTICA E PONDERAÇÃO

- Sabemos que uma amostra probabilística é aquela em que cada elemento da população tem uma probabilidade diferente de zero e conhecida de seleção, mesmo que elementos diferentes tenham probabilidades diferentes.
- A amostra será representativa da população se cada elemento da amostra receber um peso igual ao inverso da sua probabilidade de seleção.
- No caso de todos os elementos da amostra terem a mesma chance de seleção (amostra autoponderada), cada um recebe o mesmo peso (1).

AMOSTRAGEM E PONDERAÇÃO DESPROPORCIONAIS

- É ainda possível fazer amostragem e ponderação desproporcionais:
 - Ou seja, amostramos subpopulações desproporcionalmente para garantir número suficiente de casos de cada subpopulação para análise.
- Podemos sobre-amostrar uma primeira área para analisá-la detalhadamente, ao mesmo tempo em que a segunda área terá uma amostra proporcionalmente menor:
 - Devemos analisar as duas amostras separadamente ou comparativamente.
 - Se quisermos combinar as duas amostras para criar um quadro composto de toda região, devemos realizar um procedimento de ponderação da amostra.

GRAUS DE PRECISÃO NA PONDERAÇÃO

- Em um desenho complexo de amostra, podemos computar pesos para cada elemento até várias casas decimais, ou atribuir pesos aproximados para corrigir uma eventual amostragem desproporcional.
- A precisão que buscamos na ponderação deve ser comensurável com a precisão que queremos nos nossos resultados.
- Se nossos objetivos de pesquisa permitem tolerar erros de alguns pontos percentuais, não nos preocuparemos com cálculo de ponderações exatas.
- O grau de precisão da ponderação terá relação direta com o grau de erro amostral e não amostral de nossa pesquisa.
- Não há uma fórmula precisa para determinar qual a melhor ponderação.
- No entanto, devemos informar nos resultados se nossa ponderação foi aproximada ou precisa.

MÉTODOS DE PONDERAÇÃO

- Para a **ponderação aproximada**:
 - Primeiro criamos tabelas sem pesos de cada uma das amostras das subpopulações.
 - Depois multiplicamos as células da subpopulação que foi subamostrada (segundo grau de subamostragem), antes de adicioná-la à subpopulação que foi sobreamostrada.
- Para uma **ponderação mais extensa e rápida**:
 - Dados para certos casos podem ser copiados. Intenção é de aumentar número de casos, como se tivéssemos aplicado uma maior quantidade de questionários.
- Para **ponderações previamente calculadas**:
 - Podemos utilizar um banco de dados que já possui os pesos de cada elemento. Neste caso, somente informamos ao programa computacional qual variável contém a informação de peso.

PONDERAÇÃO E INFERÊNCIA ESTATÍSTICA

- Os procedimentos de ponderação têm efeitos sérios na maioria dos cálculos de inferência estatística.
- Se finalidade da pesquisa exige inferências estatísticas precisas, feitas com base em dados cuidadosamente ponderados, é preciso consultar um estatístico de amostragem antes da amostra ser desenhada.

PANORAMA DA AMOSTRAGEM PROBABILÍSTICA

- O método chave de amostragem usado na pesquisa de *survey* é a amostragem probabilística.
- Os elementos são escolhidos de uma população, com base na seleção aleatória e com probabilidades diferentes de zero e conhecidas.
- A amostragem probabilística pode ser simples ou complexa, dependendo da situação de campo.
- É o método mais eficiente para seleção de elementos:
 - Evita o viés consciente ou inconsciente na seleção de elementos por parte do pesquisador.
 - Utiliza métodos controlados de seleção que permitem estimar o grau de erro esperado (erro amostral).

2) Amostragem não-probabilística

Babbie, 1999: 113-158 (capítulo 5)

DESENHOS DE AMOSTRAGEM NÃO-PROBABILÍSTICA

- Apesar dos métodos de amostragem probabilísticos serem adequados à generalização, podemos decidir por utilizar amostragem não-probabilística:
 - Porque é mais barata.
 - Porque talvez não seja necessário ter representatividade exata.

- Há alguns métodos de amostragem não-probabilística:
 - Amostragem intencional ou por julgamento.
 - Amostragem por cotas.
 - Amostragem por confiança em sujeitos disponíveis.

AMOSTRAGEM INTENCIONAL OU POR JULGAMENTO

- Na amostragem intencional ou por julgamento, pode-se selecionar a amostra baseado no próprio conhecimento: (1) da população; (2) dos seus elementos; e (3) da natureza das metas de pesquisa.
- Pode-se realizar um pré-teste do questionário com a seleção de maior variedade de entrevistados para testar a aplicabilidade geral das perguntas.
- Este é um pré-teste do instrumento e não um *survey* propriamente dito.
- Ou seja, há a seleção de indivíduos-chave importantes para os interesses da pesquisa.

AMOSTRAGEM POR COTAS

- Na amostragem por cotas, sabemos qual a proporção da população que possui determinadas características (sexo, idade, raça, local de residência, ocupação):
 - São colhidos dados de pessoas com todas as combinações de características.
 - Todas pessoas recebem peso apropriado à sua porção na população total.
 - Quando todos elementos da amostra estiverem ponderados, os dados como um todo devem resultar numa representação razoável da população total.
- Há problemas na amostragem por cotas:
 - A moldura das cotas deve ser precisa.
 - Pode haver vieses na seleção de elementos com determinadas características.

CONFIANÇA EM SUJEITOS DISPONÍVEIS

- Na amostragem por confiança em sujeitos disponíveis, há a seleção de indivíduos que são fáceis de acessar.
- Também é chamada de amostragem de conveniência ou amostragem a esmo.
- Embora sirva à finalidade de pré-testar um questionário, este método de amostragem não deve ser usado para um estudo que pretende descrever uma população.

USOS NÃO-*SURVEY* DOS MÉTODOS DE AMOSTRAGEM

- Na discussão sobre a lógica e os métodos de amostragem, foi dada maior atenção à pesquisa de ***survey***, em que há a seleção de elementos de uma população para entrevista ou questionários auto-administrados.
- Métodos de amostragem podem ter aplicação **não-*survey***:
 - **Análise de conteúdo**: volume de documentos para analisar pode ser muito grande, o que exige uso de técnicas de amostragem de seleção de tais documentos.
 - **Experimentos de laboratórios**: sujeitos para experimentos são geralmente voluntários. Pode-se usar métodos de estratificação para satisfazer cotas.
 - **Observação participante**: ao utilizar métodos de amostragem de *survey*, um pesquisador pode obter um conjunto mais representativo de observações.

3) Principais conceitos em econometria

Wooldridge, 2008: 1-17 (capítulo 1)

ECONOMETRIA

- A econometria evoluiu como uma disciplina separada da estatística matemática, porque enfoca problemas inerentes à coleta e à análise de dados econômicos não-experimentais.
- **Dados não-experimentais** não são acumulados por meio de experimentos controlados de indivíduos, firmas ou segmentos da economia.
- Dados não-experimentais são também chamados de **dados observacionais** para enfatizar o fato de que o pesquisador é um coletor passivo de dados.
- **Dados experimentais** são frequentemente coletados em ambientes de laboratório nas ciências naturais, mas são muito mais difíceis de serem obtidos nas ciências sociais.
- O método de análise da **regressão múltipla** é utilizado por econometristas e estatísticos matemáticos, mas o foco e interpretação pode diferir significativamente.

ANÁLISE ECONÔMICA EMPÍRICA

- Os métodos econométricos são usados para testar uma teoria econômica ou para analisar relações que apresentam importância para análises de políticas públicas.
- Uma análise empírica usa dados para testar uma teoria ou estimar uma relação.
- O primeiro passo em qualquer análise empírica é a formulação cuidadosa da questão de interesse, a qual pode ser a de testar efeitos de uma política governamental ou, até mesmo, de testar hipóteses e teorias.
- O modelo econômico formal consiste em equações matemáticas que descrevem relações para testar teorias.

MICROECONOMIA

- Os indivíduos fazem escolhas para maximizar seu bem-estar (**maximização da utilidade**), sujeitas às restrições de recursos.
- Isso oferece um arcabouço para criar modelos econômicos para fazer previsões entre variáveis.
- A maximização da utilidade leva a um conjunto de **equações de demanda**, no contexto das decisões de consumo.
- Em uma equação de demanda, a quantidade demandada de cada produto depende do seu próprio preço, do preço dos bens substitutos e complementares, da renda do consumidor e das características individuais que influem no gosto.

MODELO ECONÔMICO

- O modelo econômico é a formulação teórica de uma relação entre variáveis econômicas.
- A quantidade de tempo gasto na atividade criminosa é uma função de vários fatores (Gary Becker 1968):

$$y=f(x_1, x_2, x_3, x_4, x_5, x_6, x_7),$$

y = horas gastas em atividades criminosas.

x_1 = “salário” por hora ocupada em atividade criminosa.

x_2 = salário-hora em emprego legal.

x_3 = renda de outras atividades que não o crime ou um emprego legal.

x_4 = probabilidade de ser capturado.

x_5 = probabilidade de ser condenado se capturado.

x_6 = sentença esperada se condenado.

x_7 = idade.

MODELO ECONOMÉTRICO

- Após elaborar o modelo econômico, é especificado um modelo econométrico, que será aplicado a dados existentes.
- A forma da função $f(.)$ deveria ser especificada antes de realizar uma análise econométrica.
- Se uma variável não pode ser obtida, é possível utilizar uma variável que se aproxima desta que se quer medir (**proxy**).
- Outros fatores são considerados no termo de erro u (ou termo de perturbação):
 - **Erro amostral** é a diferença entre o resultado amostral e o verdadeiro resultado da população (devidos ao acaso).
 - **Erro não-amostral** ocorre quando os dados amostrais são coletados, registrados ou analisados incorretamente.
- Modelo econométrico de Becker (1968):

$$\textit{crime} = \beta_0 + \beta_1 \textit{salário} + \beta_2 \textit{outrenda} + \beta_3 \textit{freqpris} + \beta_4 \textit{freqcond} + \beta_5 \textit{sentmed} + \beta_6 \textit{idade} + u$$

MODELO ECONOMÉTRICO NA PRÁTICA

- Na maioria dos casos, a análise econométrica começa pela especificação de um modelo econométrico, sem consideração de detalhes da criação do modelo econômico.
- É comum começar com um modelo econométrico e usar o raciocínio econômico e conhecimentos científicos como guias para escolher as variáveis.
- Após a especificação do modelo econométrico, várias hipóteses podem ser formuladas em termos das direções e influências dos parâmetros desconhecidos (independentes) sobre a variável de interesse (dependente).
- Após os dados terem sido coletados, os métodos econométricos são usados para estimar os parâmetros do modelo econométrico e para testar as hipóteses de interesse.