Noções de Econometria

Tutorial com aplicação do software Gretl.

Carlos Antônio Soares de Andrade

Professor de Economia da
Universidade Federal de Campina Grande

Campina Grande
2012

Sumário

1 Introdução **1**
1.1 O que é econometria 1
1.2 Características dos dados econômicos 4
1.3 Fonte dos dados 4

2 Introdução ao Gretl **6**
2.1 Instalação 6
2.2 Entrada de dados 6
2.3 Importação de dados 9
2.4 Análise dos dados 10
2.4.1 Estatística descritiva 10
2.4.2 Correlação 13
2.4.3 Exercício prático 16

3 Regressão linear simples **18**
3.1 Uma aplicação no Gretl 18
3.1.1 Diagnóstico da regressão 19
3.1.2 Exercício prático 24

4 Regressão linear múltipla **26**
4.1 Estimação de mínimos quadrados ordinários utilizando Gretl 27
4.2 Análise dos resultados 27
4.2.1 Teste de significância do modelo 29

Referências Bibliográficas **31**

Lista de Figuras

2.1 Tela inicial do Gretl 7
2.2 Menu Arquivo do Gretl 7
2.3 Selecionando a entrada manual de dados 8
2.4 Janela para informar número de observações 8
2.5 Janela para informar a estrutura dos dados 8
2.6 Tela para confirmação da estrutura de dados. 9
2.7 Tela para informar o nome da primeira variável 9
2.8 Tela de inserção dos dados 9
2.9 Seqüencia para importação de dados 10
2.10 Dados da Companhia MB importados para o Gretl 11
2.11 Estatísticas descritivas da variável salário 11
2.12 Histograma da variável salário 14
2.13 Diagrama de caixa com as variáveis salário e idade. 14

3.1 Resultados do Modelo de Mínimos Quadrados Ordinários. 19
3.2 Intervalos de confiança para os coeficientes. 20

Lista de Tabelas

2.1 Dados hipotéticos. 15
2.2 Valores do coeficiente de correlação e qualificação da correlação. 16
2.3 Dados para exercício prático 17

3.1 Média de notas escolares e renda familiar 19
3.2 Esquema de asteriscos e níveis de significância no Gretl. 21

4.1 Observações semanais sobre receitas, preço e gastos com propaganda para a cadeia de lanchonetes 28
4.2 Resultado da regressão do modelo da cadeia de lanchonetes 29

1 Introdução

O presente texto reúne notas de aula da disciplina Econometria I, ministrada no curso de Ciências Econômicas da Universidade Federal de Campina Grande. Voltado não só para os alunos que se matriculam nesta disciplina, mas também para qualquer pessoa interessada em aprender a interpretar dados estatísticos sobre a realidade econômica do Brasil e internacional. A ferramenta básica é um modelo econométrico que alia fundamentos da teoria econômica com a técnica estatística da análise de regressão. O modelo explica o comportamento de uma ou diversas variáveis econômicas.

O curso tem um caráter eminentemente aplicado, onde exemplos práticos servem para introduzir conceitos de análise empírica e econométrica. Um aspecto importante do curso é que os alunos são treinados no uso de um *software* econométrico: o pacote econométrico Gretl, que usa licença *open source* e que vem sendo adotado nos principais centros de ensino e laboratórios de econometria do mundo.

Não há pretensão de substituir os manuais consagrados de econometria, mas servir de auxiliar ao processo de aprendizagem. Os exemplos serão resolvidos passo-a-passo com o pacote econométrico referido acima.

1.1 O que é econometria

Segundo o prêmio Nobel de Economia, Lawrence Klein (1978)

> O principal objetivo da econometria é dar conteúdo empírico ao raciocínio econômico apriorístico. Este raciocínio apriorístico é composto, principalmente, pelo que chamamos de teoria econômica. Mas podem servir também como estrutura e objeto da análise econômica as descrições gerais não quantitativas das instituições econômicas e seu funcionamento inter-relacionado, desde que as proposições apriorísticas possam ser colocadas em forma matemática.

Conforme o economista Jeffrey M. Wooldridge(2006),

> A econometria é baseada no desenvolvimento de métodos estatísticos para estimar relações econômicas, testar teorias, avaliar e implementar políticas de governo e de negócios. A aplicação mais comum da econometria é a previsão de importantes variáveis macroeconômicas, tais como taxas de juros, taxas de inflação e produto interno bruto (PIB). Ainda que as previsões de indicadores econômicos sejam bastante visíveis e, muitas vezes, extensamente publicadas, os métodos econométricos podem ser usados em áreas econômicas que não têm nada a ver com previsões macroeconômicas.

A econometria tem sido usada em campos de estudo da ciência política (efeitos de gastos em campanhas políticas sobre os resultados de eleições), da educação (efeito de gastos públicos com escolas sobre o desempenho de estudantes) e outros. É,portanto, a combinação de teoria econômica, estatística e, nos dias atuais, ciência da computação.

A teoria econômica proporciona a base para identificar as variáveis econômicas importantes. Para se estruturar uma análise econômica empírica, o primeiro passo "é a formulação cuidadosa da questão de interesse", segundo Wooldridge (*op. cit*).

A questão de interesse ou problema pode se referir a um aspecto da teoria para ser testado ou os efeitos de uma política governamental. Os métodos econométricos podem ser aplicados para responder a uma gama de questões.

Um estudo de economia aplicada segue um fluxo de tarefas constituídas das seguintes etapas: formulação do problema, coleta de dados, especificação do modelo econométrico, estimação dos parâmetros desconhecidos do modelo, diagnóstico do modelo estimado, re-especificação do modelo se necessário e, por fim, aplicação do modelo.

- *Formulação do problema.* A formulação do problema é muito importante para a especificação do modelo econométrico: definição das variáveis dependente e independente, forma funcional do relacionamento entre estas variáveis e quantidade e disponibilidade dos dados para a análise.

  Na opinião do professor Philip Hans Franses (2002), a econometria não se restringe a validar ou testar teorias. O econometrista também pode descobrir novas regularidades estatísticas de dados econômicos. Na tarefa de validar ou não uma teoria, os dados disponíveis podem não permitir rejeitar a hipótese. A teoria econômica frequentemente se refere ao longo prazo, contudo focar no curto-prazo pode ser mais frutífero. Para a questão "são os carros médios igualmente caros no Japão e nos Estados Unidos, após o ajuste da taxa de câmbio ente esses dois países?", parece ser uma formulação melhor do que "a hipótese da paridade do poder de compra permanece para o Japão e os EUA?".

- *Coleta dos dados.* Nesta etapa alguns cuidados são necessários. Concentrar-se em coletar dados estatísticos relevantes para a solução da questão de interesse. Atentar para o fato de que diferentes definições embasam a coleta e publicação de dados por diferentes instituições públicas e privadas como IBGE, FIESP, FGV, IPEA, e outras. Há problemas também concernentes à disponibilidade dos dados: dados faltantes, interrupção da série de coleta, credibilidade da fonte.

- *Especificação do modelo econométrico.* Esta etapa depende das duas anteriores. Após a revisão da teoria e da coleta dos dados relevantes à questão de interesse, passa-se à declaração da hipótese de pesquisa e especificação matemática e estatística do modelo. Segundo Hill (1999), a teoria econômica não objetiva prever o comportamento específico de um indivíduo ou firma, mas descreve o comportamento médio ou sistemático de muitos indivíduos. Um exemplo é a clássica formulação de Keynes (1985) da relação entre consumo e renda:

> A lei psicológica fundamental em que podemos basear-nos com inteira confiança, tanto a priori, partindo do nosso conhecimento da natureza humana, como a partir dos detalhes dos ensinamentos da experiência, consiste em que os homens estão dispostos, de modo geral e em média, a aumentar o seu consumo à medida que a sua renda cresce, embora não em quantia igual ao aumento de sua renda

Embora Keynes não tenha especificado a relação funcional entre renda e consumo, ficou estabelecida nos estudos econométricos a relação linear entre as duas variáveis, como na equação

$$Y = \beta_1 + \beta_2 X + \epsilon \tag{1.1}$$

Onde Y é a despesa de consumo, X é a renda e β_1 e β_2 são os parâmetros a serem estimados, respectivamente, coeficiente de intercepto e coeficiente de inclinação. O parâmetro ϵ é o componente aleatório, chamado também erro aleatório. Esse componente incorpora as demais variáveis não consideradas na função e outros fatores.

- *Estimação dos parâmetros do modelo.*

  A equação 1.1 acima é um modelo econométrico cujos parâmetros β_1 e β_2, desconhecidos serão estimados por algum método da estatística matemática. Dentre os mais utilizados tem-se o método da análise de regressão e o método dos mínimos quadrados ordinários. Dada a evolução recente da econometria e a disponibilidade de computadores para cada pesquisador individual, esta tarefa vem sendo realizada com o auxílio dos *softwares* chamados pacotes estatísticos. Dentre eles o Gretl.

- *Diagnóstico do modelo.*

  Uma vez especificado o modelo, e tendo os valores estimados dos parâmetros é o momento de avaliar se o modelo é coerente com os pressupostos teóricos e a percepção do analista. Consiste esta etapa na aplicação de vários testes de hipóteses sobre os parâmetros estimados e no modelo completo, objetivando verificar se foi bem especificado. Enfim, busca-se responder às questões: as variáveis explicativas eram todas importantes? Utilizou-se a forma funcional correta? Qual o grau de explicação do modelo? Não se obtendo respostas satisfatórias, procede-se à mudança do modelo aproveitando o conhecimento obtido da primeira análise.

- *Aplicação do modelo.*

  Chegando-se ao modelo mais adequado, depois de todos os testes e confrontos com os pressupostos teóricos,passa-se à aplicação do modelo para tarefas de previsão ou avaliação dos efeitos de uma política.

1.2 Características dos dados econômicos

Diferentemente de estudos das áreas das ciências exatas e naturais, onde os dados para análise são gerados em experimentos controlados em laboratório, em economia e na análise empírica econométrica os dados são obtidos da observação dos fenômenos econômicos e sociais, são dados não experimentais. Segundo Hill . *et. al.*(1999), a maioria dos dados econômicos é coletada para fins administrativos e não como objeto de pesquisa.

Os dados econômicos podem ser quanto ao tipo **dados quantitativos**, quando assumem valores numéricos sejam monetários, volume, extensão e proporções; ou **dados qualitativos** quando assumirem valores expressos por atributos como sexo (masculino, feminino), respostas binárias (sim e não ou 0 e 1).

Conforme Wooldridge(2006), quanto à apresentação os dados podem corresponder às seguintes estruturas:

Dados de corte transversal Consiste em uma amostra de indivíduos, consumidores, empresas, cidades, estados, países ou uma variedade de outras unidades, tomada em um determinado ponto no tempo. Ignoram-se quaisquer diferenças de tempo.

Dados de séries de tempo consiste em observações sobre uma ou muitas variáveis ao longo do tempo. Normalmente referindo-se a cronologias como anos, trimestres, meses e dias. São exemplos de séries temporais produto interno bruto, índice de preços ao consumidor e volume de vendas de automóveis.

Cortes transversais agrupados É uma combinação de dados de corte transversal e de séries temporais. Um exemplo é quando se utilizam duas amostra de dados da PNAD (Pesquisa Nacional por Amostra de Domicílios),efetuada anualmente pelo IBGE, uma para o ano de 2005 e outra para o ano de 2009. As mesmas variáveis são estudadas como renda, anos de estudo, condição do domicílio, etc. O detalhe é que os indivíduos amostrados não são os mesmos.

Dados de painel ou longitudinais Consiste em uma série de tempo para cada membro do corte transversal do conjunto de dados. Como exemplo temos quando coletamos dados de investimento e financeiros sobre um conjunto de empresas ao longo de um período de cinco anos. Aqui as mesmas empresas são acompanhadas ao longo de um determinado período.

1.3 Fonte dos dados

A compilação dos dados para um estudo econométrico é uma tarefa crítica. Depende da acessibilidade, da periodicidade em que são colocados à disposição do público e do formato de apresentação (impressos em relatórios ou boletins, gravados em meio digital como arquivo de planilhas ou texto não formatado).

No Brasil, fontes de dados econômicos e sociais podem ser obtidos das páginas de *internet* de instituições como o IBGE (`http://www.ibge.gov.br`), Banco Central do

Brasil (`www.bcb.gov.br/`) e do IPEA (Instituto de Pesquisa Econômica Aplicada) que mantém uma página de dados econômicos intitulada IPEADATA (`www.ipeadata.gov.br`).

Alguns livros e manuais de economia (micro, macro, economia brasileira e economia internacional) trazem tabelas com dados, mas para utilizá-los há o inconveniente de manualmente inseri-los nas planilhas.

2 Introdução ao Gretl

2.1 Instalação

O software Gretl [1] é gratuito e está disponível no site oficial em português `http://gretl.sourceforge.net/gretl_portugues.html`. A versão para instalação no sistema operacional Windows está no link `http://gretl.sourceforge.net/win32/index_pt.html`. Baixar o arquivo `gretl_install.exe`. Para executar o arquivo de instalação basta clicar duas vezes com o botão esquerdo do mouse sobre este arquivo.

Grande parte do software está em português do Brasil, mas ainda há muitas páginas de ajuda em inglês. É um projeto colaborativo e aos poucos essas lacunas vão sendo preenchidas pela colaboração voluntária [2].

Depois de instalado o programa aparecerá na área de trabalho um ícone com o desenho de uma figura feminina, uma camponesa. Clicando duas vezes neste ícone com o mouse ou seguindo a seqüencia na barra de comandos do Windows `Iniciar > Todos os programas > gretl > gretl`, será iniciado o programa apresentando a janela reproduzida na Figura 2.1. Como ainda não foi carregado nenhum arquivo de dados apenas as opções de menu estão disponíveis `Arquivo` e `Ferramentas`, as demais opções aparecem em cinza indicando que não estão disponíveis.

2.2 Entrada de dados

A entrada manual de dados no gretl é realizada através do comando `Novo conjunto de dados` do menu `Arquivo`, como mostram a Figura 2.2 e a Figura 2.3.

Surgirá uma pequena janela Figura 2.4 solicitando a informação de quantas observações serão inseridas.

Após ser informada a quantidade de observações, deve-se selecionar a estrutura dos dados, conforme a Figura 2.5: `Dados de corte` referindo-se a dados de corte transversal, `Série temporal` indicando se os dados são de série temporal e `Painel` se os dados estão na estrutura Painel.

Em seguida, pede-se para confirmar a estrutura do conjunto de dados conforme a Figura 2.6. Antes de clicar no botão `Aplicar`, clicar com o mouse no quadrado, assinalando-o, onde se lê `Inicie a introdução de valores`.

[1] Gretl é um acrônimo para Gnu Regression, Econometrics and Time-series Library. Trata-se de uma biblioteca de funções estatísticas e econométricas para análise de regressão e de séries temporais. Segundo Adkins(2007), é de fácil uso e razoavelmente poderoso.

[2] Para se informar sobre os aspectos de colaboração e voluntariado na comunidade de software aberto o leitor poderá começar pelo link `http://pt.wikipedia.org/wiki/Software_livre`.

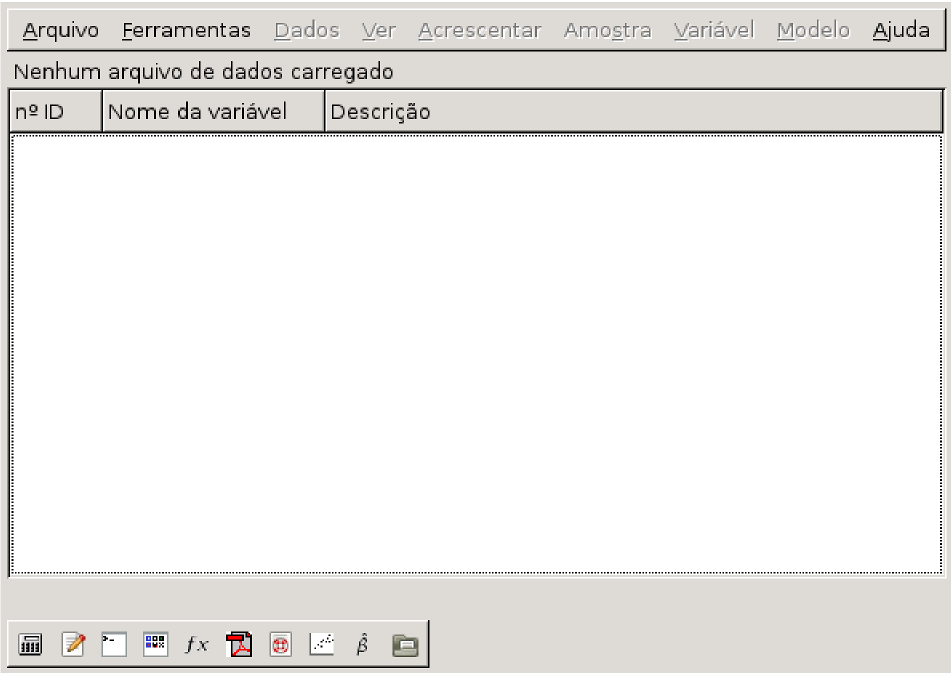

Figura 2.1: Tela inicial do Gretl

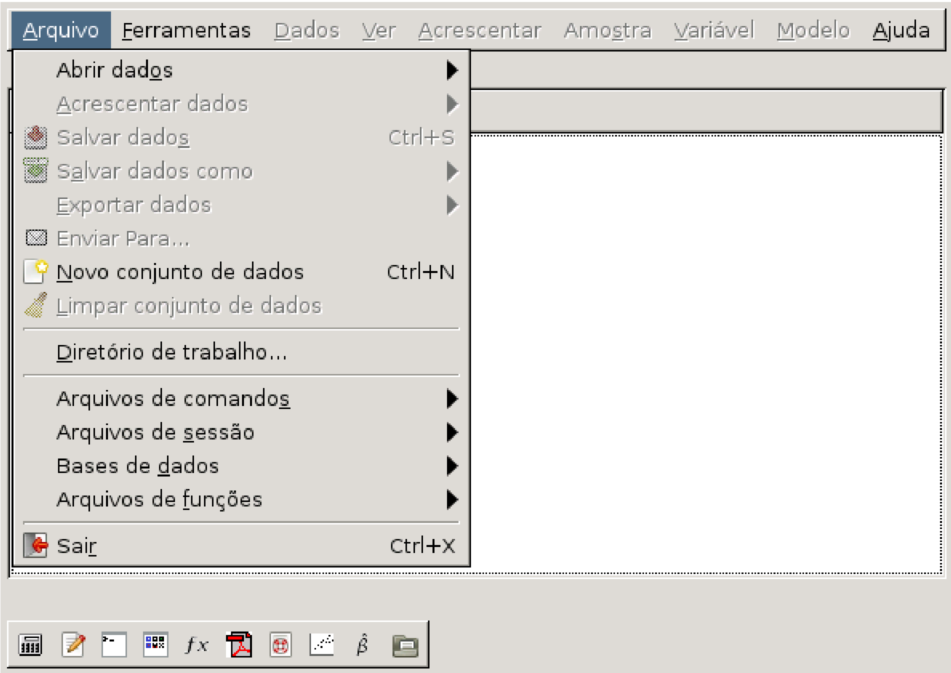

Figura 2.2: Menu Arquivo do Gretl

Como mostra a Figura 2.7, pede-se um nome para a primeira variável. No gretl as variáveis serão nomeadas com um máximo de 15 caracteres; iniciar com uma letra e depois números e letras e o caractere sublinhado _ ; não usar acentos como em "salário";

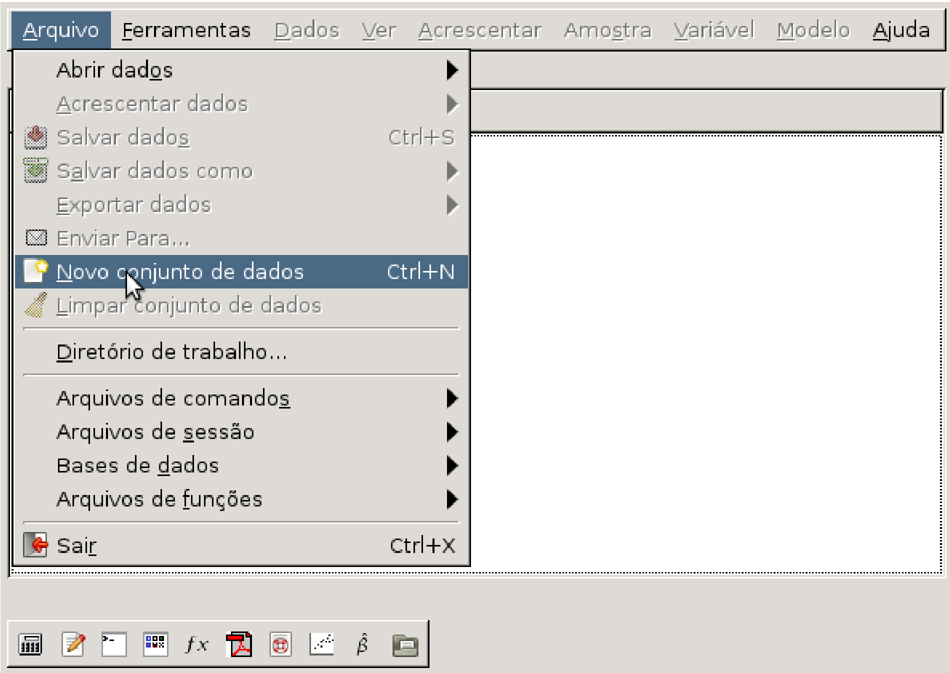

Figura 2.3: Selecionando a entrada manual de dados

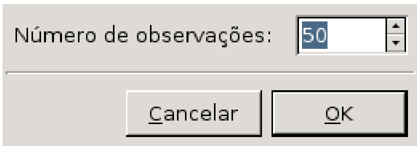

Figura 2.4: Janela para informar número de observações

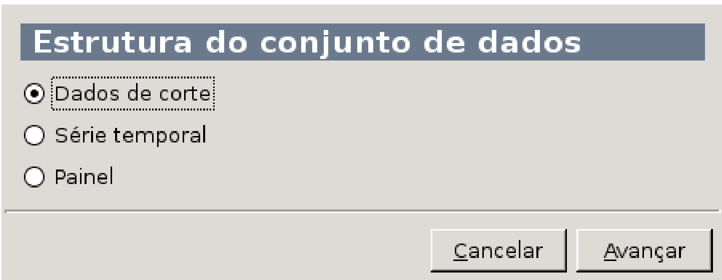

Figura 2.5: Janela para informar a estrutura dos dados

não ter espaços como em “Segundo Grau completo”;

A janela de inserção de dados na Figura 2.8 tem no canto superior esquerdo três ícones:$+$,$\surd$,$\times$. Correspondem respectivamente às ações inserir mais uma variável, confirmar os dados inseridos e fechar a janela de inserção de dados.

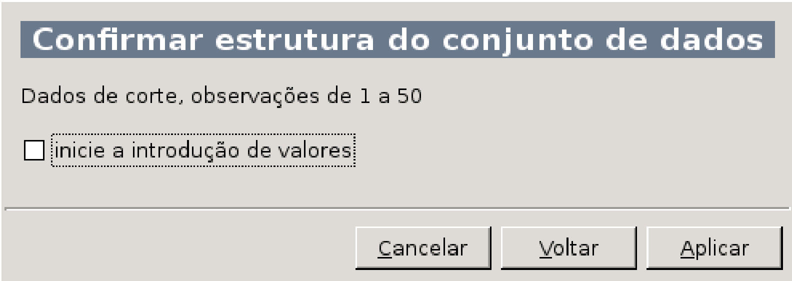

Figura 2.6: Tela para confirmação da estrutura de dados.

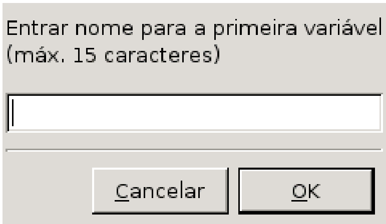

Figura 2.7: Tela para informar o nome da primeira variável

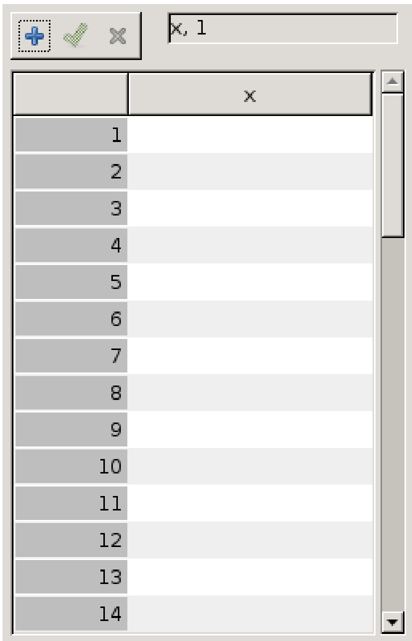

Figura 2.8: Tela de inserção dos dados

2.3 Importação de dados

A outra maneira de inserir dados no gretl é através da seqüencia `Arquivo - Abrir dados - Importar`. Como se vê na Figura 2.9. Nota-se que há uma variedade de formatos de dados, mas neste texto faremos referência apenas às opções "texto/CSV" e "Excel".

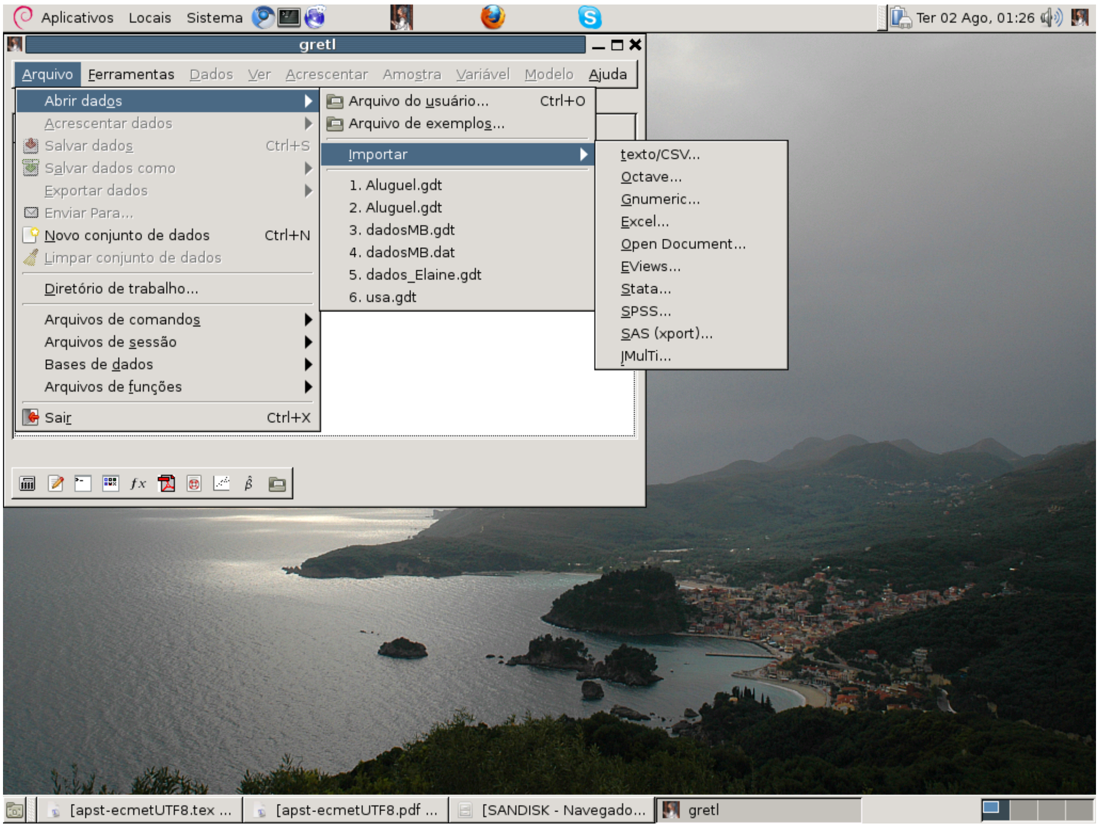

Figura 2.9: Seqüencia para importação de dados

2.4 Análise dos dados

O *software* Gretl tem como foco a análise de regressão e de séries temporais, mas tem ferramentas que auxiliam na descrição estatística sumária de variáveis. Na barra de menu superior há duas formas de se obter o resumo estatístico de uma variável ou um conjunto selecionado: o comando `Ver` com a opção `Estatísticas descritivas` e o comando `Variável` também com a opção `Estatísticas descritivas`. A resposta ao comando é uma janela de resultados informando média, mediana, mínimo, máximo, desvio padrão, coeficiente de variação, enviesamento(assimetria) e excesso de curtose.

2.4.1 Estatística descritiva

O objetivo da análise descritiva é resumir um conjunto de dados, extraindo as características e informações mais relevantes para o estudo.

Sejam os dados da empresa fictícia Companhia MB, apresentados no livro Estatística básica, de Bussab e Morettin(2005), página 11, Tabela 2.1. Para serem lidos pelo Gretl, os dados foram digitados em uma planilha adaptando os títulos das colunas extraindo os caracteres til, circunflexo e cedilha.

Salva a planilha, então procedeu-se à importação dos dados pelo Gretl usando a seqüência de comandos utilizando o mouse: `Arquivo > Abrir dados > Importar > Excel`. Abre-se uma janela onde o usuário deve informar o caminho da planilha que contém os dados para análise. O *software* solicita a confirmação da estrutura de dados, no caso Corte transversal.

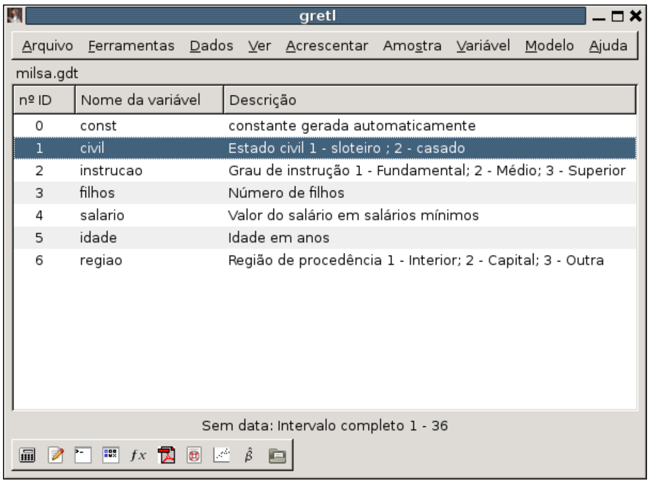

Figura 2.10: Dados da Companhia MB importados para o Gretl

Para se obter uma descrição numérica da distribuição da variável salário, procede-se à seqüência : selecionar a variável salário clicando com o mouse deixando-a destacada. Clicar com o mouse no comando `Ver` ou `Variável` e clicar em `Estatísticas descritivas` e uma janela com os resultados surgirá na tela do computador, como se vê na Figura 2.11.

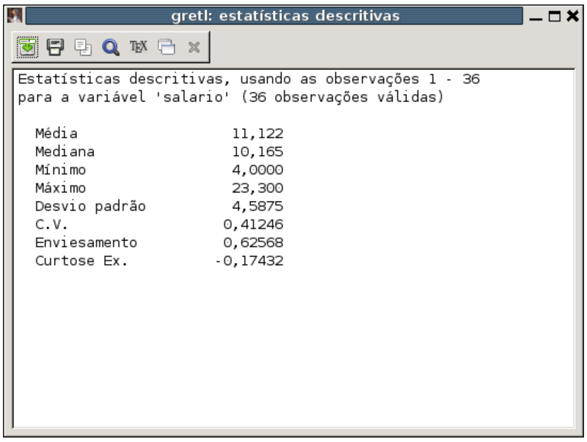

Figura 2.11: Estatísticas descritivas da variável salário

Medidas descritivas da distribuição

A média e a mediana são medidas de posição. O desvio padrão e o coeficiente de variação (C.V.) são medidas de dispersão. Com o nome de Enviesamento tem-se o coeficiente de assimetria da distribuição e com o nome de Curtose Ex. o excesso de curtose.

Segundo Merrill e Fox(1980) [3], define-se a média aritmética de um conjunto de dados como

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

No cálculo da média, leva-se em conta cada número do conjunto de observações, com igual peso. Cada conjunto de números tem uma única média.

A média tem como desvantagem sofrer alteração em seu valor com a presença de observações de valores extremamente grandes ou extremamente pequenos.

A mediana é o valor central de um conjunto de observações dispostas por ordem de grandeza. A mediana não é afetada por valores extremos.

A estas medidas de posição deve-se acrescentar a informação sobre a variabilidade em trono da média. São as medidas de dispersão. Com os valores `Mínimo` e `Máximo`, fornecidos pelo Gretl, obtemos a Amplitude Total, ou seja,

$$Amplitude\ Total = Máximo - Mínimo$$

As medidas de dispersão mais utilizadas são desvio padrão e variância. O desvio padrão é definido como a raiz quadrada positiva da variância. A variância de um conjunto de dados se define como a média dos quadrados dos desvios dos dados com relação à média. Gretl calcula a variância como

$$var(X) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Portanto o desvio padrão será

$$dp(X) = \sqrt{var(X)}$$

A variância e o desvio padrão terão valor zero quando todos os dados de uma variável assumirem o mesmo valor. Assim, quanto mais próximo de zero indica que o conjunto de valores é bastante concentrado e homogêneo.

A outra medida de dispersão informada pelo Gretl é o coeficiente de variação (C.V.). Relaciona o desvio padrão com a média e se define como

$$C.V. = \frac{dp(X)}{\bar{X}}$$

O coeficiente de variação é comumente expresso como uma percentagem, assim deve-se multiplicar por 100 o valor informado pelo Gretl. A dispersão dos dados será tanto maior quanto o valor do C.V. se aproxime de 1 ou de 100%.

As duas últimas medidas fornecidas pelo Gretl informam a respeito da forma da distribuição: Enviesamento (assimetria) e Curtose Ex (excesso de curtose). Uma distribuição é dita assimétrica se não é simétrica em relação à média. Segundo Merrill e Fox (*op.*

[3]Fonte das definições seguintes.

cit.), o coeficiente de assimetria α_3 é a medida de assimetria mais usada, e se define como

$$\alpha_3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3}{s^3}$$

onde s^3 é o cubo do desvio padrão. Para uma distribuição simétrica, α_3 é zero. Valores positivos de α_3 indicam que a distribuição é positivamente assimétrica, ou seja, seu gráfico (histograma) apresenta uma longa cauda à direita. Para valores negativos, apresentará uma cauda longa à esquerda.

O coeficiente de curtose é a medida de "achatamento" da distribuição. A medida da curtose é fornecida pelo cálculo de α_4, definido como

$$\alpha_4 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4}{s^4}$$

onde s^4 é a quarta potência do desvio padrão. A distribuição normal é muito utilizada como referência para medida do achatamento. Para a distribuição normal, α_4 é igual a 3. Desta forma, se α_4 excede 3 (Curtose Ex. é positiva) , a distribuição é dita *leptocúrtica* (mais apontada do que a distribuição normal), e se α_4 é menor do que 3 (Curtose Ex. é negativa), a distribuição é dita *platicurtica* (menos apontada do que a distribuição normal).

Representação gráfica da distribuição

A representação gráfica de uma distribuição de freqüência chama-se histograma e no Gretl pode ser obtido pela seqüência, usando o mouse: `Variável > Distribuição de freqüência`, informar o número de classes que se deseja (o Gretl sugere 7), clicar na caixa `mostrar gráfico` e clicar em `OK`. O gráfico será exibido em uma janela como a da Figura 2.12. Clicando com o botão direito do mouse sobre a janela do gráfico surgirá um menu com as opções de salvamento do gráfico para inserção no texto de um relatório ou no corpo de um artigo para publicação.

Outro gráfico utilizado para descrever uma distribuição de freqüência é conhecido como gráfico de caixa, também chamado de *desenho esquemático* em Bussab e Morettin(2005). Pode ser obtido no Gretl pela seqüência: clicar em `Variável`, `Gráfico das variáveis`, `Diagramas de caixa`, informar o nome de uma variável ou mais de uma variável (neste caso separadas por espaço) e `OK`. O resultado aparece em uma janela como a Figura 2.13. Clicando com o botão direito do mouse sobre a janela do gráfico surgirá um menu com as opções de salvamento do gráfico para inserção no texto de um relatório ou no corpo de um artigo para publicação.

2.4.2 Correlação

Um tópico importante na análise de variáveis sócio-econômicas diz respeito ao grau de associação entre elas. A resposta é fornecida pelo cálculo do coeficiente de correlação que informa o grau de associação, mas não pode ser usado para determinar a existência

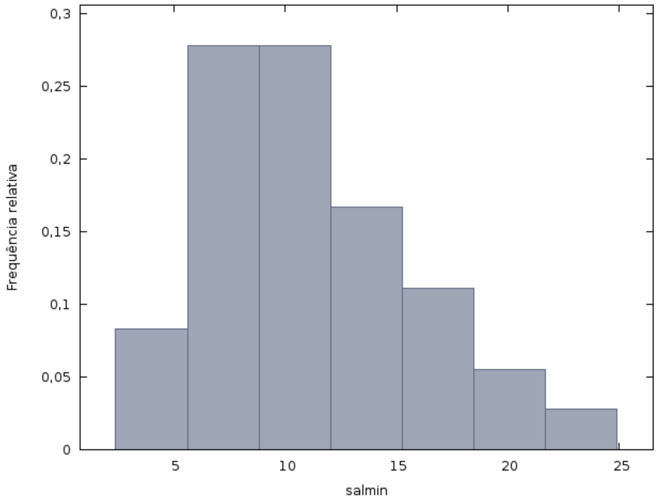

Figura 2.12: Histograma da variável salário

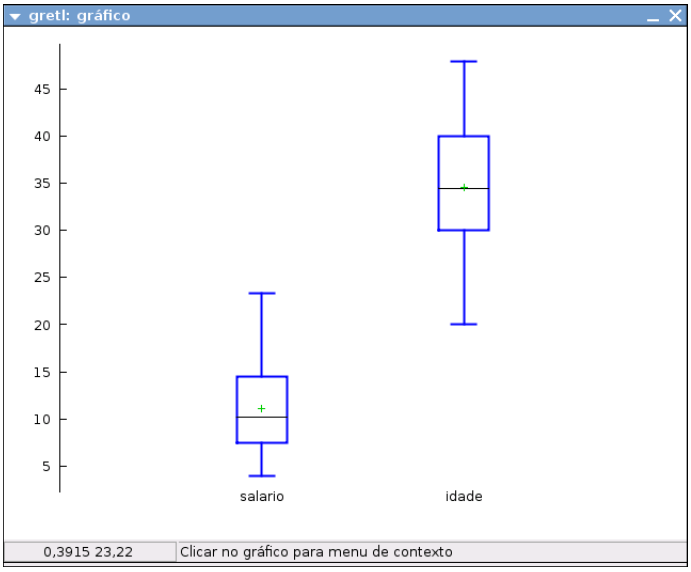

Figura 2.13: Diagrama de caixa com as variáveis salário e idade.

de um vínculo de causalidade entre as variáveis. Segundo Hoffman(1980), para uma amostra de n pares de valores X_i, Y_i (com $i = 1, ..., n$) a estimativa de correlação linear entre X e Y é dada por

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

onde $x_i = X_i - \bar{X}$ e $y_i = Y_i - \bar{Y}$, com $\bar{X} = \frac{1}{n} \sum X_i$ e $\bar{Y} = \frac{1}{n} \sum Y_i$.

Considere os dados da tabela 2.1

Para calcular o coeficiente de correlação entre X e Y no Gretl utiliza-se a seqüência de menus `Ver - Matriz de correlação - Variáveis selecionadas- OK`. A saída será

Tabela 2.1: Dados hipotéticos.

X	Y
1	1
2	9
3	2
4	7
5	6

igual ao conteúdo do quadro abaixo

Coeficientes de correlação, usando as observações 1 – 5
5% valor crítico (bilateral) = 0,8783 para n = 5

X	Y	
$1,0000$	$0,3730$	X
	$1,0000$	Y

Segundo Levin(1987), o coeficiente de correlação oscila entre os valores $-1,00$ e $+1,00$ e sendo interpretados segundo a Tabela 2.2.

Tabela 2.2: Valores do coeficiente de correlação e qualificação da correlação.

Valor do coeficiente	Interpretação
$-1,00$	correlação negativa perfeita
$\vdots$	
$-0,95$	correlação negativa forte
$\vdots$	
$-0,50$	correlação negativa moderada
$\vdots$	
$-0,10$	correlação negativa fraca
$\vdots$	
$0,00$	ausência de correlação
$\vdots$	
$+0,10$	correlação positiva fraca
$\vdots$	
$+0,50$	correlação positiva moderada
$\vdots$	
$+0,95$	correlação positiva forte
$\vdots$	
$+1,00$	correlação positiva perfeita

Logo, o coeficiente de correlação entre X e Y sendo igual a $0,03730$ pode ser interpretado como uma correlação positiva entre fraca e moderada.

2.4.3 Exercício prático

1. A Tabela 2.3 traz informações sobre 30 alunos distribuídas nas variáveis Idade (em anos), Altura (em metros), Peso (em quilogramas), TV (horas gastas assistindo TV , por semana) e IMC (Índice de Massa Corpórea, calculado pela fórmula: $IMC = \frac{Peso}{Altura^2}$). Usando o Gretl, realize as tarefas a seguir.

   a) Obtenha as estatísticas descritivas das variáveis e relate sobre o formato de cada distribuição e a dispersão dos dados.

   b) Obtenha para cada variável os gráficos histograma e diagrama de caixa. O que informam estes gráficos?

   c) Obtenha a matriz de correlação entre as variáveis e comente o resultado.

Tabela 2.3: Dados para exercício prático

Aluno	Idade	Altura	Peso	TV	IMC
1	17	1.60	60.5	16	23.630
2	18	1.69	55.0	7	19.255
3	18	1.85	72.8	15	21.275
4	25	1.85	80.9	20	23.638
5	19	1.58	55.0	5	22.030
6	19	1.76	60.0	2	19.367
7	20	1.60	58.0	7	22.650
8	18	1.64	47.0	10	17.471
9	18	1.62	57.8	12	22.029
10	17	1.64	58.0	10	21.537
11	18	1.72	70.0	8	23.655
12	18	1.66	54.0	0	19.516
13	21	1.70	58.0	30	20.022
14	19	1.78	68.5	2	21.617
15	18	1.65	63.5	10	23.369
16	19	1.63	47.4	18	17.862
17	17	1.82	66.0	10	19.952
18	18	1.80	85.2	10	26.263
19	20	1.60	54.5	5	21.200
20	18	1.68	52.5	14	18.662
21	21	1.70	60.0	5	20.747
22	18	1.65	58.5	5	21.458
23	18	1.57	49.2	10	19.948
24	20	1.55	48.0	28	19.955
25	20	1.69	51.6	4	18.069
26	19	1.54	57.0	5	24.013
27	23	1.62	63.0	5	24.084
28	18	1.62	52.0	10	19.859
29	18	1.57	49.0	12	19.805
30	25	1.65	59.0	2	21.649

3 Regressão linear simples

Como ensina Kmenta(1988), "a teoria econômica preocupa-se sobretudo com relações entre variáveis. Relações de oferta e procura, função custo, função produção e muitas outras (...), a econometria se preocupa em testar as proposições teóricas incorporadas nestas relações e em estimar os parâmetros nelas envolvidos".

Sendo x e y duas variáveis relacionadas em alguma proposição teórica ou empírica, a tarefa é estudar como y varia com variações em x? Conforme Wooldridge (*op.cit*),

> Ao escrever um modelo que "explicará y em termos de x", defrontamo-nos com três questões. Primeira, como nunca há uma relação exata entre duas variáveis, como consideramos outros fatores que afetam y ? Segunda, qual é a relação funcional entre y e x? E terceira, como podemos estar certos de que estamos capturando uma relação *ceteris paribus* entre y e x (se esse for um objetivo desejado)?

As ambigüidade resolve-se com a especificação do modelo que relaciona y a x. Uma equação tal como

$$y = \beta_0 + \beta_1 x + u \tag{3.1}$$

A equação 3.1 é chamada de modelo de regressão ou equação de regressão. As variáveis x e y têm vários nomes mas os mais descritivos são variável explicada para y e variável explicativa para x. A variável u , que em alguns livros é representada por ϵ, é chamada de termo de erro ou perturbação da relação, representa outros fatores além de x que afetam y como não observados. O parâmetro β_1 é chamado de parâmetro de inclinação da relação entre x e y e é muito importante na análise em economia aplicada. O parâmetro de intercepto β_0 tem seus usos mas é raramente central para uma análise (Wooldridge, 2006).

3.1 Uma aplicação no Gretl

Sejam os dados da tabela 3.1, retirados do livro de Pindyck e Rubinfeld (*op. cit.*). Entrando esses dados no Gretl, tal como descrito na seção 2.4.1, procede-se à seqüência de menus `Modelo -> Mínimos Quadrados Ordinários` e na janela que se abre informar como variável dependente a variável Y selecionando-a com o mouse e clicando na seta azul. Depois selecionar a variável X e clicar na seta verde inserindo-a como variável independente. Notar que o Gretl já colocou a constante como `const`. Por fim, clicar no botão `OK` e aparecerá a janela com os resultados, indicados a seguir na Figura 3.1.

Tabela 3.1: Média de notas escolares e renda familiar

Y (média das notas)	X (renda dos pais em US$ 1000,00)
4,0	21,0
3,0	15,0
3,5	15,0
2,0	9,0
3,0	12,0
3,5	18,0
2,5	6,0
2,5	12,0

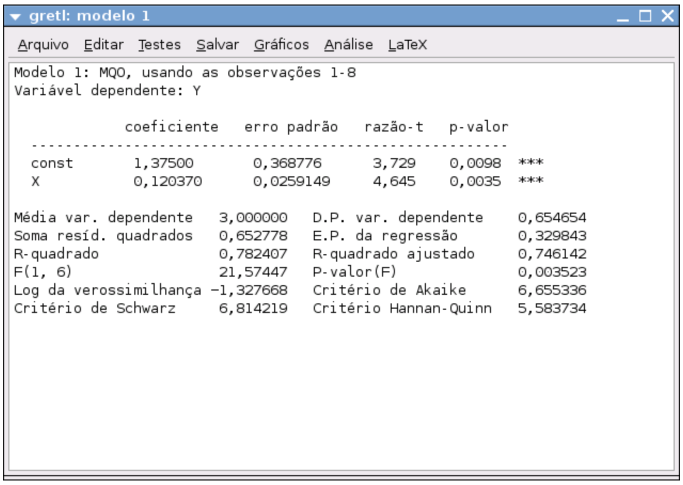

Figura 3.1: Resultados do Modelo de Mínimos Quadrados Ordinários.

Lendo os resultados do Gretl, temos na primeira coluna o termo `const` que se refere ao intercepto da equação e X a variável independente. Na segunda coluna os valores estimados dos coeficientes β_0 e β_1, respectivamente 1,37500 e 0,120370, desta forma o modelo estimado fica como na equação 3.2

$$\widehat{Y} = 1,375 + 0,12X \tag{3.2}$$

Segundo a equação 3.2, para cada unidade de X a mais, $\hat{Y}$ aumenta em 0,12. Os demais itens da janela de resultados servem para diagnosticar se o modelo é significativo nos parâmetros e se a equação tem um bom ajuste com relação aos dados da amostra.

3.1.1 Diagnóstico da regressão

O pressuposto do Modelo 1 é que a renda dos pais explica a média de notas dos alunos. Tendo o modelo estimado, procede-se à verificação da significância dos parâmetros, em especial de $\beta 1$, o coeficiente de X, renda dos pais. Assim, para verificar se a renda dos pais é significativa para explicar a média de notas dos alunos efetua-se a verificação do intervalo de confiança dos parâmetros e o teste da nulidade $H_0 : \beta_1 = 0$.

Significância dos parâmetros

Para se obter os intervalos de confiança no Gretl, ativar a janela de resultados `Modelo 1` e clicar no menu `Análise -> Intervalos de confiança para os coeficientes`. Surgirá a janela da figura 3.2.

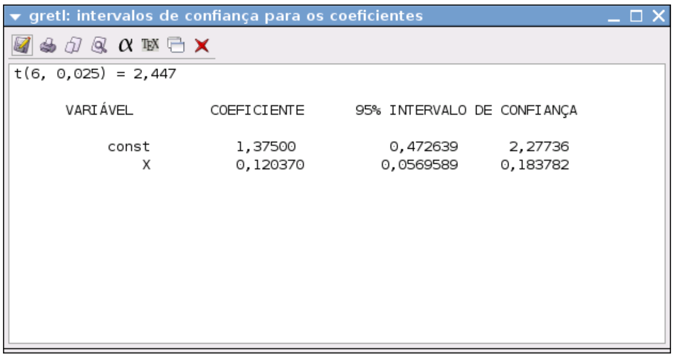

Figura 3.2: Intervalos de confiança para os coeficientes.

O nível de confiança padrão é $\alpha = 0,05$, que pode ser modificado clicando no botão com o α na barra de comandos. Observando os resultados, o intervalo de confiança para β_0 é $[0,472639\ ;\ 2,27736]$ e para β_1 é $[0,0569589\ ;\ 0,183782]$. O zero não está incluído, então para esta amostra os parâmetros não são zero com 95% de probabilidade.

Para o teste de hipótese Hill *et. al.*(*op.cit.*) sugerem as quatro etapas a seguir:

1. Determine as hipóteses nula (H_0) e alternativa (H_1).

2. Especifique a estatística de teste e sua distribuição se a hipótese nula é verdadeira.

3. Escolha α e determine a região de rejeição.

4. Calcule o valor amostral da estatística de teste.

Aplicando ao estudo das médias de notas e renda dos pais.

1. A hipótese nula é $H_0 : \beta_1 = 0$. A hipótese alternativa é $H_1 : \beta_1 \neq 0$. Se a hipótese nula é verdadeira, não há relação entre a renda dos pais e as médias de notas. Se a hipótese alternativa é verdadeira, então há relação entre a renda dos pais e as médias de notas.

2. Estatística de teste $t_c = \frac{\beta_k}{dp(\beta_k)} \sim t_{(T-2)}$ se a hipótese nula é verdadeira. Onde $dp(\beta_k)$ é o erro-padrão do coeficiente e T o número de observações. Os graus de liberdade da distribuição t_c é igual a T menos o número de parâmetros envolvidos.

3. O α é um valor de probabilidade, em geral utiliza-se $\alpha = 0,01$ ou $\alpha = 0,05$. Neste caso $\alpha = 0,05$. A região de rejeição é dada pelos valores de α e graus de liberdade (8 observações menos dois parâmetros = 6). Assim o valor crítico de t_c será dado em uma tabela estatística da distribuição t de Student para $t_{gl=6,\alpha=0,05/2}$, pois o teste é bicaudal se a hipótese alternativa é $H_1 : \beta_1 \neq 0$. No Gretl o valor crítico de t_c pode ser obtido de forma automática pela seqüência de menus da janela principal: `Ferramentas -> Tabelas estatísticas -> t` e preencher as janelas `gl` com os graus de liberdade e `probabilidade da cauda direita` com o valor de α escolhido. O valor de t_c crítico para a amostra de médias de notas é igual a 2,447. A regra de rejeição da hipótese nula em favor da hipótese alternativa é dada pela condição do valor amostral de $t \geq 2,447$ ou $t \leq -2,447$, ou, equivalente, se $|t| \geq 2,447$.

4. Por fim, o valor amostral da estatística t de cada parâmetro é dado na janela de resultados do Modelo 1 na coluna `razão - t`, que é o resultado da divisão do valor do coeficiente pelo `erro padrão`. Verifica-se que a `razão - t` de cada um dos coeficientes supera o valor crítico da estatística t da amostra (2,447). Portanto para cada um dos parâmetros a hipótese de nulidade foi rejeitada ao nível de $\alpha = 0,05$. Mais importante, no caso de β_1 ter sido rejeitada a hipótese de nulidade, significa que a variável X, rendimento dos pais, é significativa no modelo.

   Outra forma de avaliar a significância dos parâmetros é observar a coluna `p-valor`. Segundo Hill *et. al.*(*op.cit*), podemos decidir se rejeitamos uma hipótese nula comparando-o com o nível de significância α. A regra é:

> **Regra de rejeição para um teste de hipótese**: Quando o **P-valor** de um teste de hipótese é *menor* do que o valor escolhido α, o procedimento de teste conduz à *rejeição* da hipótese nula.

A consequência prática é que conhecendo o P-valor do teste, a decisão de rejeição da hipótese nula vai resultar da comparação do P-valor com o nível de significância α escolhido.

Outra facilidade fornecida pelo Gretl são os asteriscos ($*$) ao lado dos p-valores. O esquema da Tabela 3.2 demonstra que quanto mais asteriscos melhor ou mais significante a estimativa.

Tabela 3.2: Esquema de asteriscos e níveis de significância no Gretl.

Asteriscos	Valor do α	Interpretação
$***$	0,01	Muito significativo
$**$	0,05	Significativo
$*$	0,10	Pouco significativo

Qualidade do ajustamento

Conforme Pindyck e Rubinfeld(*op.cit*), "uma boa regressão é aquela que ajuda a explicar uma grande proporção da variância de Y". A medida desse grau de explicação é dada pelo R^2 ou R-quadrado. O R^2 dá a proporção da variação total de Y explicada pela regressão de Y contra X. Varia entre 0 e 1 e normalmente se interpreta em porcentagem. Se $R^2 \simeq 0$ indica um ajuste fraco, mas ao contrário, quanto mais próximo de 1 melhor será o ajuste. O caso de $R^2 = 1$ ocorre apenas quando todos os pontos estão sobre a linha de regressão. Na Figura 3.1, o valor de `R-quadrado` é 0,782407 ou 78,24% o que indica um bom ajuste, ou seja que a renda dos pais ajuda a explicar 78% da variação nas médias de notas da amostra de 8 pessoas. Mas Wooldridge (*op.cit*) adverte "que um R-quadrado aparentemente baixo não significa, necessariamente, que uma equação de regressão de MQO é inútil. (...) usar o R-quadrado como o principal padrão de medida de sucesso de uma análise econométrica pode levar a confusões".

Resumindo os resultados da regressão

Os resultados da estimação do modelo de regressão podem ser apresentados de forma resumida na forma de uma equação de regressão ajustada

$$\widehat{\mathrm{Y}} = \underset{(0{,}36878)}{1,37500} + \underset{(0{,}025915)}{0,120370}\,\mathrm{X}$$

$$T = 8 \quad \bar{R}^2 = 0,7461$$

(erros padrão entre parênteses)

Intervalo de confiança da previsão

Após a obtenção do modelo estimado e da verificação do seu alcance explicativo através dos testes de hipótese, pode-se utilizar a equação estimada para realizar previsões de Y para dado X. Porém assim como os valores estimados de β_0 e β_1 têm um intervalo de confiança, o valor previsto $\hat{y}_0$ para um dado x_0 não observado, mas com valor não muito distante de $\overline{X}$, também tem um intervalo de confiança. Para isto será necessário conhecer o desvio padrão da previsão dado pela equação 3.3(Hill *et.al.*)

$$dp(f) = \sqrt{\hat{\sigma}^2\left[1 + \frac{1}{T} + \frac{(x_0 - \bar{x})^2}{\sum(x_t - \bar{x})^2}\right]} \tag{3.3}$$

Assim, o intervalo de confiança da previsão será dado por

$$\hat{y}_0 \pm t_c dp(f) \tag{3.4}$$

Para o modelo de média de notas em função da renda dos pais, considere-se a equação estimada

$$\widehat{Y} = 1,375 + 0,12X$$

e desejamos prever a média de nota y_0 para uma renda de $x_0 = 20\,mil$. Temos

$$y_0 = 1,375 + 0,12 * (20) = 3,7824$$

A obtenção do intervalo de confiança da previsão no Gretl não é feita de forma direta. No livro de Adkins(2010) há uma sugestão de código com base na fórmula 3.5

$$v\hat{a}r(f) = \hat{\sigma}^2 + \frac{\hat{\sigma}^2}{T} + (x_0 - \bar{x})^2 v\hat{a}r(b_2) \tag{3.5}$$

Antes de apresentar o código sugerido deve-se informar o leitor que o software Gretl tem mais de um ambiente de operação. O ambiente até agora comentado é chamado de ambiente gráfico ou interface gráfica com o usuário (sigla GUI em inglês), onde a interação é realizada através de menus acionados com cliques do mouse. Outros dois ambientes estão disponíveis. Um é chamado de console e é acionado clicando com o mouse no terceiro ícone , da esquerda para a direita, na barra inferior de comandos na janela principal do Gretl. O console também é acionado digitando o comando `gretlcli` seguido de `Enter` na janela do DOS do Windows, que está em `Iniciar - Programas - Acessórios`.

No console os comandos são digitados informando parâmetros e variáveis de saída. O outro ambiente não tem um nome mas é acionado clicando com o mouse no segundo ícone, antes do ícone do console. Seu ícone é parecido com o Bloco de Notas do Windows, uma folha e um lápis. Neste ambiente os comandos são digitados em uma seqüência e depois são todos executados de uma vez clicando-se no ícone com o desenho de uma roda dentada, na barra de menus da janela.

Ao executar um modelo, em qualquer dos ambientes, o Gretl salva os resultados de alguns cálculos em variáveis do sistema. Assim, os valores estimados dos coeficientes da regressão são salvos na variável `$coeff`. Então para acessar o valor estimado do coeficiente da constante digita-se no console o comando adequado informando o parâmetro `$coeff(const)`. A soma dos quadrados dos resíduos é salva na variável `$ess`. Os graus de liberdade e o número de observações estão nas variáveis `$df` e `$nobs`. Para computar $\bar{x}$ utiliza-se a função interna `mean(x)` e para se obter o valor crítico de uma distribuição a função `critical`.

Abaixo o código sugerido seguido de comentários.

```
ols Y const X
genr yhat0 = $coeff(const)+$coeff(X)*20
genr sig2 = $ess/$df
genr f = sig2 + sig2/$nobs + ((20 - mean(X))^2)*($stderr(X)^2)
genr l_inf = yhat0 - critical(t,$df,0.025)*sqrt(f)
genr l_sup = yhat0 + critical(t,$df,0.025)*sqrt(f)
```

Comentário do código:

1. Na primeira linha o comando `ols` informa que o método dos mínimos quadrados ordinários será utilizado para estimar a equação de regressão de Y em X [1]. A partícula `const` instrui que a reta tem um intercepto.

2. Na segunda linha é feito o cálculo do Y previsto com o valor de $X = 20$ utilizando-se dos valores estimados dos coeficientes salvos nas variáveis `$coeff(const)` para β_0 e `$coeff(X)` para β_1. O valor previsto será salvo na variável `yhat0`.

3. Na terceira linha é gerado o $\hat{\sigma}^2$ da equação 3.5. Que é o resultado da divisão da `Soma resid. quadrados` informada na janela de resultados do modelo (salvo na variável `$ess`) pelos graus de liberdade do modelo — número de observações menos o número de parâmetros — salvo na variável `$df`.

4. Na quarta linha é calculada a $v\hat{a}r(f)$. Descrevendo: `$nobs` é o número de observações da amostra; `mean(X)` calcula a média da variável X;e `($stderr(X)^2)` calcula o erro padrão de β_1, o coeficiente de X.

5. Na quinta linha é calculado o limite inferior do intervalo de previsão,salvo na variável `l_inf`. Aqui é usado o comando `critical`: `critical(t, $ df,0.025)` dá o valor crítico da distribuição t com graus de liberdade salvo na variável `$df` com o $\alpha = 0,05/2 = 0,025$; `sqrt(f)` é a raiz quadrada do f calculado na quarta linha.

6. Na última linha do código é calculado o limite superior do intervalo de previsão, salvo an variável `l_sup`.

Então o valor previsto da média para $X = 20$ é `yhat0=3,7824` com o seguinte intervalo de predição, considerando o nível $\alpha = 0,05$ de confiança: valor mínimo da média igual `l_inf=2,832` e valor máximo da média igual a `l_sup=4,723` ou $[2.832, 4.723]$.

3.1.2 Exercício prático

1. Use o Gretl e os dados seguintes sobre preço e consumo de arroz para estimar uma função procura da forma $\hat{Q} = a + bP$, onde Q é o consumo de arroz em milhares de toneladas , e P é o preço de varejo em dólares por tonelada .

   Agora, responda ao seguintes quesitos:

   a) Qual a interpretação econômica dos coeficientes a e b na função procura $\hat{Q} = a + bP$?

   b) Os coeficientes estimados de a e de b são consistentes com aquilo que se esperaria com base na teoria econômica?

   c) Teste as hipóteses: $H_0 : a = 0$ contra $H_A : a \neq 0$ e $H_0 : b = 0$ contra $H_A : b < 0$ fixando o nível de confiança em5%.

[1] O Gretl faz distinção entre maiúsculas e minúsculas. Assim X é diferente de x.

Ano	Consumo	Preço
1957	178	105
1958	224	105
1959	160	130
1960	315	130
1961	229	130
1962	250	150
1963	181	150
1964	306	170
1965	237	170
1966	300	180

d) Qual o grau de bondade do ajuste, R^2?

e) Usando a função procura estimada, projete o consumo para o preço de $ 200 por tonelada.

f) Qual a sua conclusão sobre o modelo estimado?

4 Regressão linear múltipla

O modelo de regressão linear múltipla é uma extensão do modelo de regressão linear simples. A principal diferença é que o modelo contém mais de uma variável explicativa. Isto muda a interpretação dos coeficientes estimados e requer mais algumas suposições. A forma geral do modelo é mostrada na equação 4.1

$$Y = \beta_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon \tag{4.1}$$

onde Y é a variável dependente, X_k é k-ésima variável explicativa, $k = 2, 3, \ldots, K$, ϵ é o erro aleatório, e $\beta_1, \beta_2, \ldots, \beta_k$ são os parâmetros a serem estimados. Tal como na regressão linear simples, ϵ tem média zero, variância constante e não é auto-correlacionado. Os parâmetros β são independentes e não podem estar perfeitamente correlacionados.

Os parâmetros $\beta_2, \beta_3, \ldots, \beta_k$ medem o efeito de uma umidade em cada uma das variáveis X_k sobre o valor esperado de Y,mantidas constantes todas as outras variáveis explicativas. O parâmero β_1 é o termo de intercepto.

Como exemplo prático utilizaremos o modelo aplicado para explicar a receita total em uma cadeia de lanchonetes publicado em Hill *et. al.* (*op. cit.*, p. 147). O gerente de uma cadeia de lanchonetes deve decidir quanto gastar com propaganda e que atrativos, como preços mais baixos, deve introduzir para aquela semana. As questões de interesse para a gerência estão descritas a seguir:

> (...) o quanto a receita total varia em função do nível de despesas com a propaganda. Um aumento nessas despesas gera um aumento da receita total? Em caso afirmativo, o aumento da receita justifica o aumento das despesas com propaganda? A gerência tem interesse também na estratégia de preços. A redução dos preços conduz a um aumento ou a diminuição da receita total? Se uma redução dos preços implica apenas um pequeno aumento da quantidade vendida, a receita total cairá (a demanda é inelástica em relação ao preço); uma redução de preço que conduza a um grande aumento na quantidade vendida produzirá um aumento na receita total (a demanda é elástica em relação ao preço).

A tarefa é construir um modelo econômico em que a receita total dependa de uma ou mais variáveis explicativas. Com base na hipótese de que a receita total, RT, esteja relacionada linearmente com o preço, p, e com as despesas com anúncios/propagandas, a, especificamos o modelo econômico

$$RT = \beta_1 + \beta_2 p + \beta_3 a \tag{4.2}$$

O modelo econômico (equação 4.2) descreve o comportamento esperado de muitos pontos de venda distintos. Para considera a diferença entre a receita total observável e o valor esperado receita total, adicona-se ao modelo um erro aleatório, $\epsilon = RT - E(RT)$. Esse erro aleatório contém todos os fatores que fazem a receita total semanal difira de seu valor esperado. Entre estes fatores não considerados explicitamente na equação do modelo estão as condições meteorológicas, o comportamento dos concorrentes e outros. Assim o modelo econométrico para a receita total da cadeia de lanchonetes será

$$RT = \beta_1 + \beta_2 p + \beta_3 a + \epsilon \tag{4.3}$$

4.1 Estimação de mínimos quadrados ordinários utilizando Gretl

Os dados do exemplo estão na tabela 4.1. Depois de inseridos os dados no Gretl, clicamos no menu `Modelo`, seguido de `Mínimos Quadrados Ordinários`. Surgirá uma janela onde colocamos a variável `RT` como variável dependente e as variáveis `p` e `a` como variáveis independentes. Após clicar no botão OK, a janela com os resultados será mostrada na tela do computador, tendo na barra de títulos `gretl: modelo 1`. Durante uma sessão do Gretl a cada vez que executamos um modelo este será momeado em sequência modelo 1, modelo 2, etc.

O resultado dos cálculos efetuados pelo Gretl aparecem, no modo tabular, como mostrados na tabela 4.2 e na forma de equação como na equação 4.4, com erros-padrão entre parênteses.

$$\begin{aligned} \widehat{RT} &= 104,786 - 6,642p + 2,984a \\ &\quad (6,483) \qquad (3,1912) \qquad (0,167) \\ R^2 &= 0,8617 \qquad F(2,49) = 159,83 \\ T = 52 & \qquad\qquad \hat{\sigma} = 6,0696 \end{aligned} \tag{4.4}$$

4.2 Análise dos resultados

Comparando-se à janela de resultados da regressão linear simples, a novidade para a regressão múltipla é a adição da linha referente a mais uma variável independente. Agora temos `const`, `p` e `a`. As colunas são iguais: `coeficiente`, `erro-padrão`, `razão-t` e `p-valor`. Para se obter o intervalo de confiança dos coeficientes estimados segue-se o menu da janela `gretl:modelo 1`: `Análise` e `Intervalos de confiança para os coeficientes`.

O teste de significância de cada um dos coeficientes individualmente segue o mesmo procedimento: comparação da `razão-t` do coeficiente com o valor crítico da estatística t para a amostra sob análise, ou comparação do `p-valor` com o nível de significância α especificado para a análise.

Tabela 4.1: Observações semanais sobre receitas, preço e gastos com propaganda para a cadeia de lanchonetes

Semana	RT	p	a	Semana	RT	p	a
1	123.10	1.92	12.40	27	124.20	2.12	8.80
2	124.30	2.15	9.90	28	98.40	2.13	3.20
3	89.30	1.67	2.40	29	114.80	1.89	5.40
4	141.30	1.68	13.80	30	142.50	1.50	17.30
5	112.80	1.75	3.50	31	122.60	1.93	11.20
6	108.10	1.55	1.80	32	127.70	2.27	11.20
7	143.90	1.54	17.80	33	113.00	1.66	7.90
8	124.20	2.10	9.80	34	144.20	1.73	17.00
9	110.10	2.44	8.30	35	109.20	1.59	3.30
10	111.70	2.47	9.80	36	106.80	2.29	7.10
11	123.80	1.86	12.60	37	145.00	1.86	15.30
12	123.50	1.93	11.50	38	124.00	1.91	12.70
13	110.20	2.47	7.40	39	106.70	2.34	6.10
14	100.90	2.11	6.10	40	153.20	2.13	19.60
15	123.30	2.10	9.50	41	120.10	2.05	6.30
16	115.70	1.73	8.80	42	119.30	1.89	9.00
17	116.60	1.86	4.90	43	150.60	2.12	18.70
18	153.50	2.19	18.80	44	92.20	1.87	2.20
19	149.20	1.90	18.90	45	130.50	2.09	16.00
20	89.00	1.67	2.30	46	112.50	1.76	4.50
21	132.60	2.43	14.10	47	111.80	1.77	4.30
22	97.50	2.13	2.90	48	120.10	1.94	9.30
23	106.10	2.33	5.90	49	107.40	2.37	8.30
24	115.30	1.75	7.60	50	128.60	2.10	15.40
25	98.50	2.05	5.30	51	124.60	2.29	9.20
26	135.10	2.35	16.80	52	127.20	2.36	10.20

Fonte: Hill *et.al.* (*op. cit*)

A medida do grau de ajuste da regressão informada pelo `R-quadrado` (R^2) agora tem de ser complementada pelo `R-quadrado ajustado`. A explicação, segundo Hill *et.al.*(*op.cit.*)

> Uma dificuldade com R^2 é que seu valor pode ser aumentado adicionando-se cada vez mais variáveis, mesmo que essas variáveis não tenham qualquer justificativa econômica.(...) Os programas de regressão para computador frequentemente apresentam uma medida alternativa de aderência, chamada R^2 ajustado, em geral simbolizada por $\bar{R}^2$

Tabela 4.2: Resultado da regressão do modelo da cadeia de lanchonetes

Modelo 1: MQO, usando as observações 1–52
Variável dependente: RT

	Coeficiente	Erro Padrão	razão-t	p-valor
const	104,786	6,48272	16,1638	0,0000
p	−6,64193	3,19119	−2,0813	0,0427
a	2,98430	0,166936	17,8769	0,0000

Média var. dependente	120,3231	D.P. var. dependente	16,31873
Soma resíd. quadrados	1805,168	E.P. da regressão	6,069611
R^2	0,867085	R^2 ajustado	0,861660
$F(2, 49)$	159,8280	P-valor(F)	3,37e–22
Log da verossimilhança	−166,0111	Critério de Akaike	338,0222
Critério de Schwarz	343,8759	Hannan–Quinn	340,2664

4.2.1 Teste de significância do modelo

Com a possibilidade de integrar mais variáveis ao modelo, surge a necessidade de se efetuar um teste conjunto da relevância de todas as variáveis incluídas.

A hipótese nula a ser testada é que conjuntamente todos os coeficientes, exceto o associado ao termo constante — ou seja o coeficiente estimado de `const` — são zero. Isto é, a formulação das hipóteses nula e alternativa seria

$$\begin{aligned} H_0 &: \beta_2 = 0, \beta_3 = 0, \ldots, \beta_k = 0 \\ H_1 &: ao\, menos\, um\, dos\, \beta_k\, é\, diferente\, de\, zero \end{aligned}$$

Caso a hipótese nula não seja rejeitada, nenhuma das variáveis explicativas tem influência sobre a variável dependente. A estatística de teste é a estatística F. A hipótese nula é rejeitada se o valor de F do modelo calculado pelo Gretl for maior ou igual ao valor critico da estatística F tabelada.

O valor crítico da estatística F tabelado é encontrado com base na informação dos graus de liberdade, v_1 e v_2, e do nível de confiança α.

Observando o resultado do modelo 1 na tabela 4.2, verificamos que a a estatística F calculada para o modelo pelo Gretl é $F(2, 49) = 159,8280$. O termo antes do sinal de igualdade $F(2, 49)$ informa que os graus de liberdade no numerador da estatística F é igual a 2 — resultado de $3 - 1$, três parâmetros $(\beta_1, \beta_2\, e\, \beta_3)$ menos o parâmetro da constante (β_1); os graus de liberdade no denominador da estatística F é igual a 49 — resultado de $52-3$, o número de observações (52) menos o número de total de parâmetros no modelo(3).

O valor crítico da estatística $F(2, 49)$, considerando um nível de confiança $\alpha = 0,05$, pode ser obtido consultando uma tabela de F ou utilizando o menu `Ferramentas`, na janela principal do Gretl. Assim, clicando em `Ferramentas`, `Tabelas estatísticas` e selecionando a aba F preenchemos as informações solicitadas: `gl(numerador)`, digitamos 2; `gl(denominador)`, digitamos 49; e `probabilidade da cauda direita`, digitamos 0,05. Então clicando em OK, aparece a janela com o resultado: `Valor crítico = 3,18658`. Desta forma, a hipótese nula é rejeitada no modelo 1, concluindo-se que a relação estimada é significativa. À mesma conclusão chegamos observando o valor da estatística `P-valor`(F) =3,37e-22. O número 3,37 vezes 10^{-22}, ou 0.000000000000000000000337, muito menor do que $0,05$, implicando na rejeição da hipótese nula.

Referências

ADKINS, L. C. *Using gretl for Principles of Econometrics.* 3rd Edition, Version 1.3131, 2010. Disponível em : `http://www.learneconometrics.com/gretl/`. Acessado em :28 de setembro de 2011.

BUSSAB, W. O.; MORETTIN, P. A. *Estatística básica.* 5. ed. São Paulo: Saraiva, 2005.

COTTRELL, A.; LUCCHETTI, R. *Gretl User's Guide.* Disponível em: `http://gretl.sourceforge.net/gretl_portugues.html`. Acessado em: 15/02/2011.

FRANSES, P. H. *A concise introduction to econometrics: an intuitive guide.* Cambridge-UK : Cambridge University Press, 2002.

HILL, R. C.; GRIFFITHS, W. E.; JUDGE, G. G. *Econometria.* São Paulo:Saraiva, 1999.

HOFFMAN, R. *Estatístca para economistas.* São Paulo: Pioneira, 1980.

LEVIN, J. *Estatística aplicada a ciências humanas.* 2. ed. São Paulo: Harbra, 1987.

KEYNES, J. M. *A teoria geral do emprego, do juro e da moeda.* 2. ed. — São Paulo: Nova Cultural, 1985.

KLEIN, L. R. *Introdução à econometria.* São Paulo: Atlas, 1978.

MERRIL, W. C.; FOX, K. A. *Estatística econômica: uma introdução.* São Paulo: Atlas, 1980.

PINDYCK, R. S.; RUBINFELD, D. L. *Econometria: modelos e previsões.* Rio de Janeiro: Elsevier, 2004.

WOOLDRIDGE, J. M. *Introdução à Econometria: uma abordagem moderna.* São Paulo: Thomson Pioneira, 2006.