OBSERVATORIO DE LA ECONOMÍA LATINOAMERICANA

revista eumednet.

Octubre 2019 - ISSN: 1696-8352

OS DADOS ESTATÍSTICOS E SEU PROCESSAMENTO INICIAL COM AS TECNICAS E SOFTWARES MAIS USUAIS NA ECONOMETRIA (*)

Pedro Hubertus Vivas Agüero (**)

Para citar este artículo puede utilizar el siguiente formato:

Pedro Hubertus Vivas Agüero (2019): "Os dados estatísticos e seu processamento inicial com as tecnicas e softwares mais usuais na econometria", Revista Observatorio de la Economía Latinoamericana (octubre 2019). En línea:

https://www.eumed.net/rev/oel/2019/10/tecnicas-softwares-econometria.html

Resumen: El propósito de este trabajo es presentar las exigencias mínimas para los datos estadísticos que se usan en la econometría y las principales formas y prácticas para deducir rápidamente los parámetros de una ecuación de regresión y sus indicadores de eficiencia o fuerza, utilizando sucesivamente el álgebra y la estadística elemental, una calculadora HP-12C y softwares como el Excel, el Eviews y la Stata. Todo esto se espera conseguir mostrando un caso específico de la economía brasileña, la relación entre el índice Ibovespa y la tasa de interés Selic. Todo lector, al leer y seguir esta ayuda memoria, podrá entender de manera práctica, las diferentes alternativas para calcular los parámetros β_1 y β_2 de una ecuación de regresión, así como la covarianza y el coeficiente de regresión R y el coeficiente de determinación R^2, todo respaldado por los cálculos correspondientes. Al final, se esperaría que cada lector pueda comprender y aplicar estos cálculos para sus propios casos del mundo real y disciplinas econométricas, superando así los temores y dificultades iniciales.

Palabra-clave: Los datos estadísticos y su procesamiento en la econometría

Resumo: O propósito deste artigo é apresentar as exigências mínimas para os dados estatísticos a utilizar na econometria e as principais formas e práticas para deduzir rapidamente os parâmetros de uma equação de regressão e dos indicadores da eficiência ou força destes mesmos, utilizando sucessivamente a álgebra e a estatística elementar, uma calculadora HP-12C, e os softwares como o Excel, Eviews e o Stata. Tudo isto aplicado para um caso específico da economia brasileira, da relação entre o índice Ibovespa e a taxa de juros Selic. A leitura e acompanhamento desta ajuda memoria lhe permitirá ao leitor, examinar de forma prática, as diferentes alternativas para calcular os parâmetros β_1 e β_2 de uma equação de regressão, assim como da covariância e do coeficiente de correlação R e do coeficiente de determinação R^2, tudo sustentado nos cálculos e referências bibliográficas correspondentes. No final se esperaria que cada leitor possa entender e aplicar estes cálculos para seus próprios casos do mundo real e nas disciplinas de econometria, superando o receio e dificuldades usuais para entender e aplicar estes indicadores.

Palavra-chave: Os dados estatísticos e seu processamento na econometria

1. Apresentação

Próximos a cumprir os 90 anos da fundação da Sociedade de Econometria, criada em Cleveland EUA em dezembro de 1930, entre outros por Irving Fisher; vê-se como esta disciplina conseguiu grandes progressos e avanços em sua formulação e aplicação, graças à contribuição de muitos economistas, matemáticos e estatísticos, e aos avanços no processamento dos dados estatísticos. Justamente este processamento dos dados, hoje em dia adquire elevados níveis de complexidade e sofisticação, pelo processamento computacional e aos vários softwares que vão aparecendo para este propósito, cada vez mais complexos e difíceis de acompanhar

Os economistas temos a obrigação intrínseca de dominar e aplicar esta disciplina em nossas atividades e nesse sentido, ela forma parte da grade curricular das escolas de economia. Entretanto, entender plenamente esta disciplina e acompanhar seus avanços permanentes não é uma tarefa fácil e nesse sentido, urge ter uma ajuda memória para recordar e aplicar os itens iniciais e mais importantes desta disciplina. Isto último é o propósito deste artigo, facilitar o entendimento destas categorias, apresentando as exigências mínimas para os dados estatísticos a utilizar e as principais formas e práticas para deduzir os parâmetros de uma equação de regressão e dos indicadores da eficiência ou força destes mesmos, utilizando sucessivamente a álgebra e a estatística elementar, uma calculadora HP-12C, e os softwares mais usuais, como o Excel, Eviews e o Stata. Tudo isto aplicado com detalhe e parcimônia para um caso específico da economia brasileira, da relação entre o índice Ibovespa (Bolsa de Valores) e a taxa de juros do Selic (política fiscal)

2. O que é econometria?

Literalmente, econometria significa "medição económica" (Gujarati & Porter, 2012 p. 25), no entanto ela é mais abrangente. Segundo Paulo Sandroni a econometria se define como um:

> "Ramo da economia que cuida do estabelecimento de leis quantitativas para os fenômenos econômicos. Partindo da teoria econômica geral, analisa os dados fornecidos pela estatística, mediante a aplicação de métodos matemáticos. Com isso, prepara o quadro de variáveis concretas que poderá servir de base a uma programação econômica" (Sandroni 1999 p. 189)

Nestes termos deve-se entender à econometria, como uma combinação eficaz da teoria econômica, as matemáticas e as estatísticas em prol de confirmar hipóteses e deduzir ensinamentos para a microeconomia (indivíduos e empresas) e na macroeconomia (governos e economia internacional)

3. Por que é uma disciplina autônoma?

A econometria dá conteúdo empírico à grande parcela da teoria econômica. Quantifica o que foi estabelecido por determinada relação de conteúdo econômico. Portanto a econometria existe de forma autônoma para, principalmente fazer a verificação empírica dos diferentes postulados, teoremas e hipóteses da economia. Nesse sentido a econometria é uma simbiose da teoria económica, das matemáticas e das estatísticas e assim ela assume vida própria.

4. Metodologia da Econometria

Seguindo o pensamento formal ou tradicional, a econometria observa o seguinte roteiro:

1) Formulação da teoria ou da hipótese
2) Especificação do modelo matemático
3) Especificação do modelo econométrico
4) Obtenção de dados
5) Estimativa dos parâmetros do modelo econométrico
6) Teste de hipóteses
7) Previsão ou predição

8) Utilização do modelo para fins de controle ou política

a. Formulação da teoria ou da hipótese

A formulação da teoria ou da hipótese deve ser a base para a criação do modelo. Como exemplo pode-se citar a teoria Keynesiana do consumo que relaciona positivamente consumo (C) e renda (Y), assim:

> “Definiremos, portanto, aquilo a que chamaremos propensão a consumir como a relação funcional entre Y (determinado nível de renda medido em unidades de salário) e C (o gasto que, para o consumo, se toma de dito nível de rendimento), de modo que C = f(Y) (Keynes 1988 p. 74)

b. Especificação do modelo matemático

Para o exemplo anterior, poder-se-ia exemplificar o modelo matemático como uma reta, que represente a função consumo, onde C é o consumo; Y é a renda e o coeficiente β_2 que seria a inclinação da reta ou a propensão marginal a consumir, onde com base no modelo anterior, ele deve ser positivo, mas menor que a unidade (1), pois o consumo não aumentaria na mesma escala da renda, segundo Keynes:

$C = \beta_1 + \beta_2 Y$

Em que C= despesas de consumo
Y = renda
β_1 = intercepto da reta, no eixo das ordenadas
β_2 = coeficiente angular, $0 < \beta_2 < 1$; ou inclinação da reta

c. Especificação do modelo econométrico

A equação acima demonstra uma relação determinística e exata entre consumo e renda, no entanto sabe-se que as relações econômicas não são exatas, até porque não apenas a renda altera o consumo, pelo menos não é a única variável.
Para isso o econometrista insere no modelo o termo de perturbação ou erro (μ); esta μ é a variável que representa todos os outros fatores que afetam o consumo, mas que não são considerados explicitamente como a renda. O modelo seria:

$$C = \beta_1 + \beta_2 Y + \mu$$

Neste modelo econométrico C = (Y) alguém poderia pensar, e porque não Y = f(C’)? Este raciocínio chama-se “especificação” e sobre isso Ben Bernanke, ex-presidente do Banco Central dos EUA, apresenta este mesmo dilema, numa outra situação, assim:

> “Um estatístico que estuda os dados da Grande Depressão notaria o fato básico de que o estoque de dinheiro, a produção e os preços nos Estados Unidos caíram juntos em 1929 a 1933 e subiram juntos nos anos subsequentes. Mas essas correlações não podem responder às perguntas cruciais: O que causou o quê? As mudanças no estoque monetário (M) estão causando mudanças nos preços (P) e na produção (Y), como Friedman e Schwartz deveriam concluir? Ou, pelo contrário, o estoque de dinheiro está reagindo passivamente às mudanças no estado de economia?” (Bernanke, 2002 p. 2)

Quer dizer, a função seria M = f(PY) ou PY = f(M)? Nestes e outros casos só o domínio da teoria econômico e da natureza dos eventos, ajudariam a defini-lo

d. Apresentação dos dados

Os dados estatísticos que se vão utilizar correspondem por um lado ao índice Ibovespa, indicador de rentabilidade ou nível dos preços dos títulos (ações) da Bolsa de Valores de São Paulo (B3) e por outro lado, a taxa Selic, indicador da taxa de juros oficiais, utilizados pelo governo para remunerar seus títulos públicos, entre outros usos. Com estes dados vai-se construir um modelo econométrico assumindo que o nível de preços ou de rentabilidade da Bolsa de Valores (Y) é influenciado, entre outros, pelo nível de remuneração dos títulos públicos (X), existindo uma relação inversa entre eles, já que na medida que este último se eleva, os investidores buscaram aplicar seus recursos nos títulos públicos; à inversa, quando ele diminui, os investidores buscaram aplicar seus recursos na Bolsa. Por outro lado, em vários estudos (Levine & Zervos 1996, e Vivas 2019; entre outros) se prova que o crescimento das economias (PIB) é influenciado diretamente pelo nível de rentabilidade das bolsas de valores.

O índice Ibovespa representa neste caso o valor ou preço ponderado das ações de uma amostra de quase 70 empresas que operam na Bolsa de Valores de São Paulo, valores vigentes no último dia útil do mês indicado. A taxa Selic (Overnight) e a média dos juros que o governo paga aos bancos que lhe emprestam dinheiro. São os juros efetivos mensais, vigentes em cada mês e devidamente capitalizados para um ano.

Tabela Nº 1

Relações entre a taxa Ibovespa e a taxa Selic Overnight

n	Período	Índice Ibovespa	Taxa % Selic	n	Período	Índice Ibovespa	Taxa % Selic
1	mar/17	64.984	13,35	17	jul/18	79.220	6,68
2	abr/17	65.403	9,9	18	ago/18	76.677	7,06
3	mai/17	62.711	11,75	19	set/18	79.342	5,79
4	jun/17	62.899	10,16	20	out/18	87.423	6,68
5	jul/17	65.920	10,03	21	nov/18	89.504	6,04
6	ago/17	70.835	10,03	22	dez/18	87.887	6,04
7	set/17	74.293	7,96	23	jan/19	97.393	6,68
8	out/17	74.308	7,96	24	fev/19	95.584	6,04
9	nov/17	71.970	7,06	25	mar/19	95.414	5,79
10	dez/17	76.402	6,68	26	abr/19	96.353	6,42
11	jan/18	84.912	7,19	27	mai/19	97.457	6,68
12	fev/18	85.353	5,79	28	jun/19	100.967	5,79
13	mar/18	85.365	6,55	29	jul/19	101.812	7,02
14	abr/18	86.115	6,42	30	ago/19	101.134	6,17
15	mai/18	76.753	6,42	31	set/19	104.745	5,66
16	jun/18	72.762	6,42				

Fonte: IPEADATA 2019 s/p

e. Estimativa dos parâmetros do modelo econométrico

Com os dados anteriores pode-se deduzir os valores de um modelo econométrico utilizando a técnica da análise de Regressão dos Mínimos Quadrados (RMQ), que se vai mostrar mais na frente, do qual obteve-se assim as seguintes estimativas de β_1 e β_2:

$$\hat{Y}_i = 118.980,99 - 4.892,48X_i$$

O acento circunflexo em cima de Y indica que se trata de uma estimativa.

f. Teste de hipótese

O teste de hipótese corresponde ao campo da estatística conhecido como inferência estatística. Seu objetivo no modelo é assegurar que os resultados são estatisticamente confiáveis. Por exemplo, dar resposta à pergunta se existem probabilidades de que $\beta_2 = 0$, na qual o modelo se anularia

g. Previsão ou predição

Após estes passos pode-se utilizar o modelo para prever os valores futuros da variável dependente Y (previsão) com base nos valores futuros conhecidos ou esperados da variável explicativa X (previsor). Por exemplo, qual seria o nível de rentabilidade (preços) da Bolsa de Valores, caso o governo decida diminuir a taxa Selic para 3%?

h. Controle ou elaboração da política econômica

Evidentemente, em função dos planos e metas de crescimento da economia, poder-se-ia definir uma política monetária (fixar a taxa Selic, por exemplo), que permita atingir determinados patamares de rentabilidade (preço) da Bolsa de Valores, que como é dito, influenciaria também o crescimento da economia

5. Mais sobre organização e apresentação dos dados estatísticos

Definido o campo, assunto e tema da pesquisa e identificados os dados estatísticos que lhe devem dar suporte, devem-se proceder a fazer um julgamento sobre a qualidade e forma da matéria prima em mãos (os dados estatísticos) para tirar o melhor proveito destes e evitar cair em erros que possam desnaturalizar o conteúdo e resultados da pesquisa, além de buscar sua melhor apresentação e leitura. Neste julgamento devem-se usar critérios sobre o tamanho da amostra, apresentação dos dados, homogeneização das unidades de medida e simplificação dos números, assim:

a. Tamanho da amostra

A estatística formal recomenda que o número de observações a ser considerado numa pesquisa deve ser o suficientemente grande como para que estes dados se qualifiquem como próprios de uma distribuição normal, já que segundo Jan Kmenta na distribuição normal:

> "1. A distribuição é continua e simétrica em torno da média. Isto tem as seguintes implicações:
> a) A média, mediana e moda são iguais, e
> b) A média divide a área sob a curva normal em duas metades exatas
> 2. O domínio da distribuição estende entre $-\infty$ e $+\infty$, isto é; a distribuição é ilimitada.
> 3. A altura máxima da curva normal é atingida no ponto $x = \mu$, e os pontos de inflexão (isto é, os pontos em que a distribuição se torna mais achatada ou em que a curva muda de concavidade) ocorre em $x = \mu \pm \sigma$. O que significa que o desvio padrão mede a distância do centro da distribuição ao ponto de inflexão (Kmenta, 1988 p. 95)

Por outro lado, o mesmo Kmenta cita o fato de que numa amostra de distribuição normal, existem condições para que tanto a média aritmética como o desvio padrão da amostra sejam iguais aos seus equivalentes da população:

> "Teorema do Limite Central: Se X tem uma distribuição qualquer com média μ e variância σ^2, então a distribuição de X aproxima-se da distribuição normal com média μ e variância σ^2/n quando o tamanho da amostra aumenta" (Kmenta, 1988 p. 121).

Porém, qual deveria ser o tamanho mínimo de uma amostra, para que ele mereça ter uma distribuição normal? A estatística em geral tem uma serie de técnicas para definir apropriadamente o tamanho de uma amostra e eles requerem de uma série de exigências e cuidados complexos, no entanto, e no dia a dia, afirma-se que um tamanho de 30 ou mais observações seria suficiente:

> "Uma regra prática muito usada é que a amostra deve consistir de 30 ou mais observações" (Stevenson, 1981 p. 181).

Logo, toda pesquisa que se prece de séria e rigorosa deveria ter pelo menos estas 30 informações ou dados.

b. Como se devem apresentar os dados

Todos os dados devem ser apresentados em tabelas, quadros e gráficos, devidamente enumerados e rotulados, com estes cuidados adicionais:

1) No início de qualquer trabalho de pesquisa, sempre se devem apresentar os dados originais, como eles foram encontrados na fonte, identificando o período e as unidades de medida consideradas, assim como informações sobre a fonte, (autor, ano de publicação e a página especifica onde aparecem estes dados), guardando o restante da informação para as referências bibliográficas, no final do capítulo ou estudo.

2) Especificamente, no caso de tabelas, devem-se seguir estes conselhos de Joanne Smailes e Ângela McGrane (2002 p. 46):

> "Ser claro sobre o que se deseja mostrar na tabela
> Dar proeminência à variável de interesse, listando-a primeiro
> Toda tabela deve possuir um título claro e explicativo
> Toda linha/coluna deve receber um título claro
> Sempre que útil, deve-se incluir os totais das linhas e das colunas"

3) Conforme se vão usando e transformando estes dados, devem-se indicar nas tabelas, quadros e gráficos subsequentes, (seja no título, no cabeçalho ou em notas de rodapé destas tabelas e gráficos), o procedimento seguido para apresentar os cálculos correspondentes, para que o leitor possa se orientar na leitura dos mesmos; quer dizer, estas tabelas, quadros e gráficos devem ser, na medida do possível, autofalantes.

4) Segundo a ABNT:

<u>Tabelas</u> significam números, e números com os quais se fazem cálculos, deduções ou transformações.

<u>Quadros</u> são utilizados para mostrar informações qualitativas ou até quantitativas, porem que não sofrem maiores transformações no interior da mesma.

c. Homogeneização das unidades de medida

Todos os dados apresentados devem ser identificados com suas unidades de medida correspondentes e sempre se deve usar a mesma unidade de medida ao longo de todo o período histórico considerado, em cada tabela e coluna. Estas unidades podem ser:

- Indivíduos ou habitantes (população humana)
- Toneladas (t)
- Litros (l)
- Metros (m)
- Kilowatts (KW)
- Anos
- Percentuais (%)
- Unidades monetárias (R$, US$, €, £, ¥, etc.)
- Hectares (Ha)

Quando se trata de unidades monetárias, além deles serem apresentados numa única unidade monetária, eles devem estar em valores constantes ou valores deflacionados, utilizando-se um ano base determinado. Este ano base deve ser claramente especificado (por exemplo: base: 2000 =100 ou reais de 2013) e na medida do possível deve-se usar um ano base recente, para dar-lhe valor real aos resultados conseguidos. Quando não seja possível ter os dados deflacionados, é melhor não usar os dados nominais, porque nas comparações entre anos, a inflação desnaturaliza a solidez de qualquer cálculo. No pior dos casos é melhor usar os valores traduzidos a dólares americanos (US$).

d. Simplificação dos números

Na medida do possível todos os números devem ser expressos em valores resumidos, utilizando-se para esse efeito, as maiores unidades de medida de cada sistema. Por exemplo:

1.545.820 kg	por 1.545,8 toneladas	ou	1,5 mil toneladas
180.000.000 habitantes	por 180.000 mil habitantes	ou	180 milhões de habitantes

Na simplificação de valores devem-se utilizar os critérios de arredondamento, onde valores iguais ou maiores de cinco justificam adicionar uma unidade a mais à unidade anterior e valores inferiores a cinco, devem ser desconsiderados. Exemplo:

1.565 por 1,6 mil
1.433 por 1,4 mil

Quando se trabalha com o Excel, em princípio não seria necessário o arredondamento dos dados, visto que ele facilmente considera todos os valores (quando se transporta os valores com o mouse ou se escrevem as células correspondentes) embora isto seja um problema na apresentação, já que cifras muito extensas significam menor poder de impacto ou elegância e também para a elaboração dos gráficos, já que as grandes quantidades "incham" os valores das coordenadas.

Sobre o uso de casas decimais o Excel nos permite selecionar alternativas de apresentação dos cálculos, considerando decimais ou não; inclusive no caso em que se decida mostrar casas decimais, ele nos permite escolhas sobre o número destes, assim:

a) Clicar no cabeçalho ou início da coluna (ou linha) de nosso interesse
b) Com o lado direito de o mouse selecionar "formatar células"
c) Clicar em "número"
d) Em "casas decimais", selecionar o número de casas de interesse.
e) Aqui se pode aproveitar para clicar em "usar separador de 1000 (.)".
f) OK

6. Cálculo dos parâmetros (β_1 e β_2) de regressão:

6.1 Dedução das "equações normais" de regressão:

O cálculo estatístico nos oferece um jogo de "equações normais" para encontrar os parâmetros indicados, inicialmente de uma amostra e posteriormente da população. Estas equações normais aparecem assim:

a) Assumindo que $Y_i = \beta_1 + \beta_2 X_i + U_i$; uma função aleatória, em que o subíndice i mostra todos os valores considerados
b) Também se poderia chegar a ele, utilizando uma amostra, fazendo esta construção alternativa $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{U}_i$ em que os parâmetros β_i levam um chapéu encima
c) Ambos caminhos levam aos mesmos resultados, quer dizer. $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{U}_i$ em que $\hat{\beta}_1 + \hat{\beta}_2 X_i = \hat{Y}_i$
d) Neste último caso $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ é uma função linear ou uma função determinística
e) Logo, $Y_i = \hat{Y}_i + U_i$ ou $U_i = Y_i - \hat{Y}_i$ ou também $\Sigma U_i = \Sigma(Y_i - \hat{Y}_i)$
f) Neste último caso, elevando ao quadrado ambos os lados, sem as somatórias e os i, consegue-se: $U_i^2 = (Y_i - \hat{Y}_i)^2$
g) $U_i^2 = (Y_i - \hat{Y}_i)^2$
h) Efetuando as operações algébricas implícitos nesta função e logo efetuando um processo de derivação parcial com respeito a $\hat{\beta}_1$ e $\hat{\beta}_2$; consegue-se as "equações normais"

Estas "equações normais" aparecem assim (Gujarati & Porter 2012 p. 80)

(I) $\Sigma Y_i = n\beta_1 + \beta_2 \Sigma X_i$
(II) $\Sigma Y_i X_i = \beta_1 \Sigma X_i + \beta_2 \Sigma X_i^2$

6.2 Organizando os dados para viabilizar as "equações normais"

Tabela Nº 2

Os dados do Ibovespa e Selic para atender às equações normais

n Nº Obs..	Período mês / ano	Ibovespa Y	Selic X	XY	X²
1	mar/17	64.984	13,35	867.536,40	178,22
2	abr/17	65.403	9,90	647.489,70	98,01
3	mai/17	62.711	11,75	736.854,25	138,06
4	jun/17	62.899	10,16	639.053,84	103,23
5	jul/17	65.920	10,03	661.177,60	100,60
6	ago/17	70.835	10,03	710.475,05	100,60
7	set/17	74.293	7,96	591.372,28	63,36
8	out/17	74.308	7,96	591.491,68	63,36
9	nov/17	71.970	7,06	508.108,20	49,84
10	dez/17	76.402	6,68	510.365,36	44,62
11	jan/18	84.912	7,19	610.517,28	51,70
12	fev/18	85.353	5,79	494.193,87	33,52
13	mar/18	85.365	6,55	559.140,75	42,90
14	abr/18	86.115	6,42	552.858,30	41,22
15	mai/18	76.753	6,42	492.754,26	41,22
16	jun/18	72.762	6,42	467.132,04	41,22
17	jul/18	79.220	6,68	529.189,60	44,62
18	ago/18	76.677	7,06	541.339,62	49,84
19	set/18	79.342	5,79	459.390,18	33,52
20	out/18	87.423	6,68	583.985,64	44,62
21	nov/18	89.504	6,04	540.604,16	36,48
22	dez/18	87.887	6,04	530.837,48	36,48
23	jan/19	97.393	6,68	650.585,24	44,62
24	fev/19	95.584	6,04	577.327,36	36,48
25	mar/19	95.414	5,79	552.447,06	33,52
26	abr/19	96.353	6,42	618.586,26	41,22
27	mai/19	97.457	6,68	651.012,76	44,62
28	jun/19	100.967	5,79	584.598,93	33,52
29	jul/19	101.812	7,02	714.720,24	49,28
30	ago/19	101.134	6,17	623.996,78	38,07
31	set/19	104.745	5,66	592.856,70	32,04
	Total	2.571.897	228,21	18.391.998,87	1.790,64

Fonte: Elaborado com os dados da Tabela Nº 1

Com estes resultados já se pode dar valor as variáveis consideradas nas "equações normais" indicadas; quer dizer, vai-se resolver estas equações:

I) $2.571.897 = 31\beta_1 + 228{,}21\beta_2$
II) $18.391{,}998{,}87 = 228{,}21\beta_1 + 1.790{,}64\beta_2$

Este resultado oferece equações do 1º grau com dois parâmetros (β_1 e β_2)

6.3 Cálculo das β_i pelo método da adição (Bezerra et al 1982 p. 146-147)

Já temos estas equações:

(I)	2.571.897	=	31	β1 +	228,21	β2
(II)	18.391.998,87	=	228,21	β1 +	1.790,64	β2

Multiplico (II) por um número que permita anular uma coluna: (31/228,21) **-0,13584**

(I)	2.571.897	=	31	β1 +	228,21	β2
(II)	-2498365,387	=	-31	β1 +	-243,2395	β2

	73.532	=	0		-15,03	β2
	β2	=	-4892,48			

Substituindo este valor na primeira equação:

(I)	2.571.897	=	31	β1 +	228,21	-4892,48

Efetuando as operações:

(I)	2.571.897	=	31	β1 +	-1116514
	3.688.411	=	31	β1	
	β1	=	118981		

Equação de regressão:

$\hat{Y}_i =$ 118.980,99 + -4.892,48 X_i

6.4 Cálculo das β_i com uma calculadora HP (HP-12C 1994 p. 97-102)

<u>Quadro Nº 1</u>

Operando com uma calculadora HP-12C

Primeiro deve-se limpar a memória da calculadora, assim f reg e g end

A seguir deve-se introduzir os valores de Y e X, incluindo todos os decimais:

64.984	enter	13,35	Σ+	Na tela aparece	1
65.403	enter	9,90	Σ+	Na tela aparece	2
62.711	enter	11,75	Σ+	Na tela aparece	3
62.899	enter	10,16	Σ+	Na tela aparece	4
65.920	enter	10,03	Σ+	Na tela aparece	5
70.835	enter	10,03	Σ+	Na tela aparece	6
74.293	enter	7,96	Σ+	Na tela aparece	7
74.308	enter	7,96	Σ+	Na tela aparece	8
71.970	enter	7,06	Σ+	Na tela aparece	9
76.402	enter	6,68	Σ+	Na tela aparece	10
84.912	enter	7,19	Σ+	Na tela aparece	11
85.353	enter	5,79	Σ+	Na tela aparece	12
85.365	enter	6,55	Σ+	Na tela aparece	13
86.115	enter	6,42	Σ+	Na tela aparece	14
76.753	enter	6,42	Σ+	Na tela aparece	15
72.762	enter	6,42	Σ+	Na tela aparece	16
79.220	enter	6,68	Σ+	Na tela aparece	17
76.677	enter	7,06	Σ+	Na tela aparece	18
79.342	enter	5,79	Σ+	Na tela aparece	19
87.423	enter	6,68	Σ+	Na tela aparece	20
89.504	enter	6,04	Σ+	Na tela aparece	21
87.887	enter	6,04	Σ+	Na tela aparece	22
97.393	enter	6,68	Σ+	Na tela aparece	23
95.584	enter	6,04	Σ+	Na tela aparece	24
95.414	enter	5,79	Σ+	Na tela aparece	25
96.353	enter	6,42	Σ+	Na tela aparece	26
97.457	enter	6,68	Σ+	Na tela aparece	27
100.967	enter	5,79	Σ+	Na tela aparece	28
101.812	enter	7,02	Σ+	Na tela aparece	29
101.134	enter	6,17	Σ+	Na tela aparece	30
104.745	enter	5,66	Σ+	Na tela aparece	31
Clicando	RCL1	aparece	31	n tamanho da amostra	
Clicando	RCL2	aparece	228,21	que é a	ΣX
Clicando	RCL3	aparece	1.790,64	que é a	ΣX^2
Clicando	RCL4	aparece	2.571.897	que é a	ΣY
Clicando	RCL6	aparece	18.391.998,70	que é a	ΣXY

Para encontrar o valor de β1 se clica nestes três botões:

0 g ỹ,r	aparece	118980,99	que é valor de $\beta 1$

Para encontrar o valor de $\beta 2$ também se clica assim:

1 g ỹ,r	aparece	114089,50	

Subtraindo deste último valor o $\beta 1$ já calculado:

114.088,50	-	118980,99	=	-4.892,48	que é o valor de $\beta 2$

Assim consegue-se também a equação linear:

$$\hat{Y}_i = 118.980,99 + -4.892,48\ X_i$$

6.5 Cálculo das β_i utilizando o Excel 16 (Mynbaev & Lemos, 2004 p. 34-38)

Inicialmente este seria o procedimento a seguir:

1) Tenha na tela do computador a tabela de dados (Tabela Nº 2 anterior), clicar no botão superior “Dados”. Deve aparecer “Análise de Dados” no canto superior direito
2) Se clicando em “Dados” não aparece nada então se deve clicar sucessivamente em: ARQUIVO, OPÇÕES, SUPLEMENTOS, Ir, FERRAMENTAS DE ANALISE e OK. Abrindo novamente o Excel, agora já deve aparecer no canto superior direito “Analise de Dados”

Para deduzir a regressão linear, utilizando os dados da Tabela Nº 2 procede-se assim:

a) Clicar em “dados” e logo em “análise de dados” (botões que aparecem na parte superior da tela do Excel).
b) Na janela que aparecerá, selecionar “regressão” e logo “ok”.
c) Em “intervalo Y de entrada” clicar no interior da janela (limpeza) e logo com o mouse carregar todos os valores da coluna Y (sem o total), do indicador Ibovespa, de abaixo para cima.
d) Idem, no “intervalo X de entrada”, clicar no interior da janela (limpeza) e logo com o mouse carregar todos os valores da coluna X, do índice Selic, de abaixo para cima (sem o total)
e) Em “Intervalo de Saída” escrever a identificação da célula de saída, onde deve começar a tabela de resultados. Por exemplo, escrever G3
f) Clicar em OK

Na parte final desta planilha que se acaba de conseguir procede-se a escrever manualmente a equação de regressão linear, como aparece a seguir:

Quadro Nº 2

Operando com o Excel

RESUMO DOS RESULTADOS

Estatística de regressão	
R múltiplo	0,73418538
R-Quadrado	0,53902817
R-quadrado Aj.	0,52313259
Erro padrão	8837,3217
Observações	31

ANOVA

	gl	SQ	MQ	F	F Signif.			
Regressão	1	2648356228	2,65E+09	33,91057	2,59E-06			
Resíduo	29	2264849391	78098255					
Total	30	4913205620						
	Coeficientes	Erro padrão	Stat t	valor-P	95% Inf.	95% Sup.	Inferior 95%	Superior 95%

Interseção	118980,986	6385,34688	18,63344	1,11E-17	105921,5	132040,5	105921,49	132040,487
Variável X 1	-4892,48307	840,159603	-5,82328	2,59E-06	-6610,8	-3174,16	-6610,802	-3174,16375

Nestes resultados o β_1 é o valor que aparece como "interseção" e o β_2 aquele da "Variável X 1"
Quer dizer, a equação procurada seria esta:

$\hat{Y}i$ =	118.980,99	+	-4.892,48 Xi

6.6 Cálculo das β_i através do Eviews10 (Braga 2014 s/p)

O procedimento a seguir deve obedecer a este roteiro:

1) Preparar uma tabela especial para ser carregado ao Eviews, na base da Tabela Nº 2. Ele deve ser copiado numa nova Tabela do Excel, na qual se devem apagar tudo aquilo que não sejam os dados mesmos, incluindo os totais, deixando só os títulos Ibovespa e Selic, Esta Tabela deve ser salva em Meus Documentos com este título: **Ibovespa_Selic.xls** (tudo seguido, sem deixar espaços)
2) Abrir o EViews
3) Clicar em *Create a nex Eviews Worfile* (ou clicar sucessivamente em *FILE>NEW>WORKFILE*)
4) Em *Workfile Struture Type* selecionar *Dated_regular frequency*
5) Selecionar *Monthly*
6) Escrever em *Start Date*: 2017M03 (para março 2017)
7) Idem, em *End Date*: 2019M09 (para setembro 2019).
8) Idem, em *Work File Nome, WF*: Ibovespa_Selic
9) Idem, em *Page*: 1
10) *OK*
11) Clicar sucessivamente em *PROC>IMPORT>IMPORT FROM FILE*
12) Buscar em Meus Documentos o arquivo Ibovespa_Selic.xls e clicar encima dele
13) Abrir
14) Em *Excel Read Step 1 of 3*, clicar em Avançar
15) Em *Excel Read step 2 of 3, Header Type*, selecionar *Names in firt line* e logo Avançar
16) Em *Excel Read step 3 of 3, Start Date*, escrever 2017M03, *Finish*
17) Procurar um botão na parte centro superior da tela que diz *QUICK*
18) Clicar sucessivamente em *QUICK>ESTIMATE DQUATION>LEAST SQUARE*
19) Na janela aberta escrever: Ibovespa Selic C (deixando espaços em branco)
20) *OK*
21) Assim aparecerá a planilha que vem a seguir:
22) Para imprimir: Clicar em *PRINT* e logo *OK*

Quadro Nº 3

Relações Ibovespa – Selic conseguida pelo Eviews

Dependent Variable: IBOVESPA
Method: Least Squares
Date: 10/14/19 Time: 17:57
Sample: 2017M03 2019M09
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SELIC	-4892.483	840.1596	-5.823278	0.0000
C	118981.0	6385.347	18.63344	0.0000

R-squared	0.539028	Mean dependent var	82964.42
Adjusted R-squared	0.523133	S.D. dependent var	12797.40
S.E. of regression	8837.322	Akaike info criterion	21.07370
Sum squared resid	2.26E+09	Schwarz criterion	21.16621
Log likelihood	-324.6423	Hannan-Quinn criter.	21.10385
F-statistic	33.91057	Durbin-Watson stat	0.568050
Prob(F-statistic)	0.000003		

Fonte: Com os dados da Tabela Nº 2

Deste Quadro pode-se deduzir a equação de regressão linear:

$\hat{Y}$ =	118.981 -	4.892,483 X

6.7 Cálculo dos β_i-utilizando o Stata (Monteiro 2011)

O procedimento é parecido àquele seguido com o Eviews:

1) Tenha em Meus Documentos a Tabela Ibovespa_Selic.xls preparado anteriormente
2) Clicar no arquivo Stata, que já deve estar carregado no computador
3) Clicar em *StatSE-64*, quando se deve abrir uma planilha
4) Clicar sucessivamente em *File, Import, Excelspreadsheet.xls*
5) Clicar em *Brouse* e logo localizar em Meus Documentos o arquivo **Ibovespa_Selix.xls**
6) Selecionar *First row as variável names*
7) Na parte inferior da planilha, em *Commands* escrever ***reg Ibovespa Selic*** e logo, *enter*
8) Deve aparecer uma planilha como o Quadro a seguir
9) Para sair e limpar o software, escrever embaixo de Commands **clear**

Quadro Nº 4

Relações Ibovespa Selic conseguido através do Stata

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   14.1   Copyright 1985-2015 StataCorp LP
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     Special Edition                  College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)
Single-user Stata perpetual license:
       Serial number:  10699393
         Licensed to:  Andrey
Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.
```

```
. import excel "E:\Ibovespa-Selic.xls.xlsx", sheet("Planilha1") firstrow
. reg Ibovespa Selic
     Source |       SS           df       MS      Number of obs   =        31
-------------+----------------------------------   F(1, 29)        =     33.91
      Model |  2.6484e+09         1  2.6484e+09   Prob > F        =    0.0000
   Residual |  2.2648e+09        29  78098254.9   R-squared       =    0.5390
-------------+----------------------------------   Adj R-squared   =    0.5231
      Total |  4.9132e+09        30   163773521   Root MSE        =    8837.3
------------------------------------------------------------------------------
   Ibovespa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Selic |  -4892.483   840.1596    -5.82   0.000    -6610.802   -3174.164
      _cons |     118981   6385.347    18.63   0.000     105921.5    132040.5
------------------------------------------------------------------------------
Command
```

Fonte: Com os dados da Tabela Nº 2, organizado como Ibovespa_Selic.xls

Desta planilha pode-se deduzir a equação da reta, da mesma forma e conteúdo, como nos casos anteriores; quer dizer:

$$\hat{Y}_i = 118.981 - 4.892,48X_i$$

6.8 Balanço parcial com os resultados conseguidos até aqui

Aplicando a função de regressão $\hat{Y} = f(X)$ conseguida nos 5 métodos anteriores aos dados inicialmente apresentados, consegue-se estes resultados:

Tabela Nª 3

Comparações entre o Y (real) e o $\hat{Y}$ (teórico)

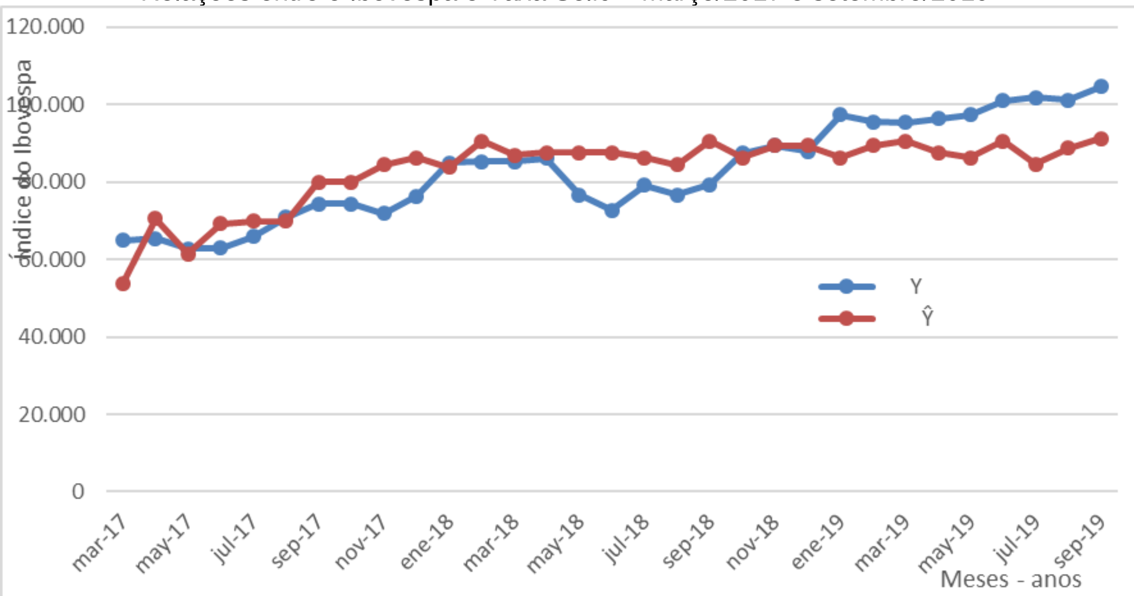
n	Período mês ano	Ibovespa Y	Selic X	Deduções $\hat{Y}$	$\hat{u} = Y - \hat{Y}$	$\hat{u}^2 = (Y - \hat{Y})^2$
1	mar/17	64.984	13,35	53.666,34	11.317,66	128.089.498,70
2	abr/17	65.403	9,90	70.545,40	-5.142,40	26.444.313,35
3	mai/17	62.711	11,75	61.494,31	1.216,69	1.480.335,09
4	jun/17	62.899	10,16	69.273,36	-6.374,36	40.632.438,16
5	jul/17	65.920	10,03	69.909,38	-3.989,38	15.915.158,06
6	ago/17	70.835	10,03	69.909,38	925,62	856.771,16
7	set/17	74.293	7,96	80.036,82	-5.743,82	32.991.475,26
8	out/17	74.308	7,96	80.036,82	-5.728,82	32.819.385,64
9	nov/17	71.970	7,06	84.440,06	-12.470,06	155.502.281,11
10	dez/17	76.402	6,68	86.299,20	-9.897,20	97.954.546,93
11	jan/18	84.912	7,19	83.804,03	1.107,97	1.227.591,81
12	fev/18	85.353	5,79	90.653,51	-5.300,51	28.095.394,34
13	mar/18	85.365	6,55	86.935,22	-1.570,22	2.465.596,32
14	abr/18	86.115	6,42	87.571,24	-1.456,24	2.120.648,17
15	mai/18	76.753	6,42	87.571,24	-10.818,24	117.034.414,97
16	jun/18	72.762	6,42	87.571,24	-14.809,24	219.313.723,90
17	jul/18	79.220	6,68	86.299,20	-7.079,20	50.115.057,69
18	ago/18	76.677	7,06	84.440,06	-7.763,06	60.265.028,79
19	set/18	79.342	5,79	90.653,51	-11.311,51	127.950.233,04
20	out/18	87.423	6,68	86.299,20	1.123,80	1.262.928,81
21	nov/18	89.504	6,04	89.430,39	73,61	5.418,71
22	dez/18	87.887	6,04	89.430,39	-1.543,39	2.382.046,85
23	jan/19	97.393	6,68	86.299,20	11.093,80	123.072.421,88
24	fev/19	95.584	6,04	89.430,39	6.153,61	37.866.939,32

25	mar/19	95.414	5,79	90.653,51	4.760,49	22.662.275,75
26	abr/19	96.353	6,42	87.571,24	8.781,76	77.119.228,93
27	mai/19	97.457	6,68	86.299,20	11.157,80	124.496.524,41
28	jun/19	100.967	5,79	90.653,51	10.313,49	106.368.099,17
29	jul/19	101.812	7,02	84.635,75	17.176,25	295.023.402,60
30	ago/19	101.134	6,17	88.794,37	12.339,63	152.266.584,30
31	set/19	104.745	5,66	91.289,53	13.455,47	181.049.627,86
	Total	2.571.897	228,21	2.571.897,00	0,00	2.264.849.391,06

Nota 1: A coluna Ŷ se consegue aplicando a fórmula de RMQ já conseguida
Nota 2: Observe como a somatória da coluna û é zero
Fonte: Com os dados da Tabela Nº 2

O desenho dos valores Y e Ŷ ao longo do período indicado pode ser representado na seguinte Figura:

Gráfico Nº 1
Relações entre o Ibovespa e Taxa Selic – março/2017 e setembro/2019



Fonte: Com os dados da Tabela Nº 3

Este Gráfico mostra o comportamento volátil (ciclos de alta e baixa) do Ibovespa (Y), suavizado pela regressão (Ŷ).

7. Dedução da covariância entre X e Y (Lapponi 1997 p. 130-140)

A covariância é um artificio estatístico que mostra as interrelações entre as vaiáveis X e Y, que pode ser uma relação direta ou positiva (quando um cresce o outro também cresce, e vice-versa) ou inversa ou negativa (quando um cresce o outro diminui e vice-versa) e para isso se utilizam os desvios de cada variável, com relação à sua média aritmética correspondente. A covariância é representada pelo símbolo Cov(Y, X) e ela pode ser populacional ou amostral. Para o caso amostral sua fórmula ´é a seguinte:

$$Cov(Y,X) = \frac{\Sigma(Y - \bar{Y})(X - \bar{X})}{n - 1}$$

Em que $\bar{Y}$ é a média aritmética de Y e $\bar{X}$ é a média aritmética de X

No caso de trabalhar com dados populacionais, o denominador desta fórmula é simplesmente n. Uma grande limitação da covariância é que ele pode ter valores ilimitados, entre menos infinito e mais infinito [$- \infty \leq Cov(Y,X) \leq + \infty$]

7.1 Dedução da covariância manualmente ou pela estatística elementar

Tabela Nº 4

Cálculo da covariância Cov(Y,X)				$\bar{Y}$ =	82964,42	$\bar{X}$ = 7,361613
n	Período	Y	X	$Y - \bar{Y}$	$X - \bar{X}$	$(Y-\bar{Y})(X-\bar{X})$
1	mar/17	64.984	13,35	-17.980	5,99	-107.673,71
2	abr/17	65.403	9,90	-17.561	2,54	-44.577,68
3	mai/17	62.711	11,75	-20.253	4,39	-88.879,84
4	jun/17	62.899	10,16	-20.065	2,80	-56.150,81
5	jul/17	65.920	10,03	-17.044	2,67	-45.481,11
6	ago/17	70.835	10,03	-12.129	2,67	-32.365,99
7	set/17	74.293	7,96	-8.671	0,60	-5.188,87
8	out/17	74.308	7,96	-8.656	0,60	-5.179,89
9	nov/17	71.970	7,06	-10.994	-0,30	3.316,06
10	dez/17	76.402	6,68	-6.562	-0,68	4.473,03
11	jan/18	84.912	7,19	1.948	-0,17	-334,23
12	fev/18	85.353	5,79	2.389	-1,57	-3.753,92
13	mar/18	85.365	6,55	2.401	-0,81	-1.948,34
14	abr/18	86.115	6,42	3.151	-0,94	-2.966,63
15	mai/18	76.753	6,42	-6.211	-0,94	5.848,75
16	jun/18	72.762	6,42	-10.202	-0,94	9.606,73
17	jul/18	79.220	6,68	-3.744	-0,68	2.552,24
18	ago/18	76.677	7,06	-6.287	-0,30	1.896,37
19	set/18	79.342	5,79	-3.622	-1,57	5.693,04
20	out/18	87.423	6,68	4.459	-0,68	-3.039,03
21	nov/18	89.504	6,04	6.540	-1,32	-8.642,79
22	dez/18	87.887	6,04	4.923	-1,32	-6.505,75
23	jan/19	97.393	6,68	14.429	-0,68	-9.834,71
24	fev/19	95.584	6,04	12.620	-1,32	-16.678,20
25	mar/19	95.414	5,79	12.450	-1,57	-19.565,92
26	abr/19	96.353	6,42	13.389	-0,94	-12.606,86
27	mai/19	97.457	6,68	14.493	-0,68	-9.878,33
28	jun/19	100.967	5,79	18.003	-1,57	-28.293,09
29	jul/19	101.812	7,02	18.848	-0,34	-6.438,58
30	ago/19	101.134	6,17	18.170	-1,19	-21.651,11
31	set/19	104.745	5,66	21.781	-1,70	-37.062,12
	Total	2.571.897	228,21	0	0,00	-541.311,27

Fonte: Com os dados da Tabela Nº 2

$$Cov(Y,X) = \frac{-541.311}{30} = -18.043,7$$

Isto quer dizer que a Cov(Y,X) = -18.043,7; um valor negativo, confirmando com isso que as relações entre o Ibovespa e a taxa Selic são inversos.

7.2 Dedução da covariância diretamente através do Excel

Todo o grande esforço que se acaba de apresentar para deduzir manualmente a covariância se pode simplificar utilizando diretamente os artifícios que nos facilita o Excel. O procedimento é o seguinte:

a) Tenha em mãos a Tabela Nº 4, aberto numa planilha em Excel
b) Clicar o mouse numa célula qualquer, livre, à direita da Tabela anterior
c) Clicar no botão $\mathbf{f_x}$ que está localizado na parte superior esquerdo da tela
d) Em “Selecionar uma categoria” marcar ESTATISTICA
e) Em “Selecionar uma função” marcar COVARIAÇÃO.S
f) Ok
g) Em “Matriz 1” introduzir com o mouse todos os valores da coluna Y (sem o total)
h) Em “Matriz 2” introduzir com o mouse todos os valores da coluna X (sem o total
i) Ok
j) Deve aparecer na célula reservada inicialmente o valor de -18.043,7; que é o mesmo valor conseguido manualmente

7.3 Dedução da covariância utilizando o Eviews

O Eviews nos permite conseguir a covariância entre Y e X, seguindo inicialmente o mesmo caminho indicado anteriormente para deduzir as β_i, isto é:

a) File, New, Workfile
b) Selecionar *Dated – regular frequency*
c) Em *Frequency* selecionar *Monthly*
d) Em *Start date* escreva 2017M3 (para o mês de março 2017)
e) Em *End date* escreva 2019M09 (para o mês de setembro de 2019)
f) Em *WF* escreva Ibovespa_Selic
g) Em *Page* escreva 1
h) OK
i) Proc, Import, Import from file
j) Clicar em Meus documentos o arquivo Ibovespa-Selic.xls
k) Avançar
l) Em *Header type* selecionar *Names in first line*
m) Em *Start date* escrever 2017M1
n) Finish
o) Quick, Group statistics, Covariances
p) Na janela aberta escrever **Ibovespa Selic**
q) OK
r) Em *Covariance* aparecerá -18.043,7 que é a covariância

7.4 Dedução da covariância utilizando o Stata

Igualmente, o Stata também nos permite conseguir a covariância entre o Ibovespa e o Selic, seguindo inicialmente o mesmo caminho para deduzir as β_i; que já se explicou anteriormente, isto é:

a) Clicar em *StatSE-64*, quando se deve abrir uma planilha
b) Clicar sucessivamente em *File, Import, Excelspreadsheet.xls*
c) Em Meus Documentos clicar encima de Ibovspa_Selix.xls
d) Abrir
e) Marcar *Import first name as variable name*
f) Ok
g) Em *Commands* escrever manualmente isto: **corr Ibovespa Selic, cov**
h) Na janela correspondente aparecerá o valor -18.043,7 que é o mesmo valor conseguido com o Excel

8. Dedução do coeficiente de correlação R

O coeficiente de correlação R tem o mesmo significado que a covariância, com algumas vantagens, como que ele tem valores limitados, entre um mínimo de -1 e um máximo de + 1 ($-1 \leq R \leq 1$) e por outro lado, ele é o mesmo tanto para uma amostra como para uma população

8.1 Cálculo do coeficiente de correlação R manualmente (Lapponi 1997 p. 144-160)

A fórmula de seu cálculo deriva da covariância, apresentada anteriormente:

$$R = \frac{Cov(Y,X)}{SxSy} \quad \text{ou também } R = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{n - 1\ (Sx\ Sy)}$$

Onde Sx = desvio padrão de X e Sy = desvio padrão de Y

Isto quer dizer que para conseguir os desvios padrões indicados, previamente deve-se calcular as variâncias de Y e X, o que se vai fazer a seguir:

Tabela Nº 5

Dados para calcular o coeficiente de correlação R

					Media Y =	Media X =
					82964,41935	7,361613
n	Y	X	$Y - \bar{Y}$	$X - \bar{X}$	$(Y - \bar{Y})^2$	$(X - \bar{X})^2$
1	64.984	13,35	-17980,42	5,99	323.295.480	35,86
2	65.403	9,90	-17561,42	2,54	308.403.450	6,44
3	62.711	11,75	-20253,42	4,39	410.200.996	19,26
4	62.899	10,16	-20065,42	2,80	402.621.054	7,83
5	65.920	10,03	-17044,42	2,67	290.512.231	7,12
6	70.835	10,03	-12129,42	2,67	147.122.814	7,12
7	74.293	7,96	-8671,42	0,60	75.193.514	0,36
8	74.308	7,96	-8656,42	0,60	74.933.596	0,36
9	71.970	7,06	-10994,42	-0,30	120.877.257	0,09
10	76.402	6,68	-6562,42	-0,68	43.065.348	0,46
11	84.912	7,19	1947,58	-0,17	3.793.070	0,03
12	85.353	5,79	2388,58	-1,57	5.705.317	2,47
13	85.365	6,55	2400,58	-0,81	5.762.787	0,66
14	86.115	6,42	3150,58	-0,94	9.926.158	0,89
15	76.753	6,42	-6211,42	-0,94	38.581.730	0,89
16	72.762	6,42	-10202,42	-0,94	104.089.361	0,89
17	79.220	6,68	-3744,42	-0,68	14.020.676	0,46
18	76.677	7,06	-6287,42	-0,30	39.531.642	0,09
19	79.342	5,79	-3622,42	-1,57	13.121.922	2,47
20	87.423	6,68	4458,58	-0,68	19.878.941	0,46
21	89.504	6,04	6539,58	-1,32	42.766.115	1,75
22	87.887	6,04	4922,58	-1,32	24.231.800	1,75
23	97.393	6,68	14428,58	-0,68	208.183.939	0,46
24	95.584	6,04	12619,58	-1,32	159.253.816	1,75
25	95.414	5,79	12449,58	-1,57	154.992.058	2,47
26	96.353	6,42	13388,58	-0,94	179.254.092	0,89
27	97.457	6,68	14492,58	-0,68	210.034.894	0,46
28	100.967	5,79	18002,58	-1,57	324.092.910	2,47
29	101.812	7,02	18847,58	-0,34	355.231.296	0,12
30	101.134	6,17	18169,58	-1,19	330.133.661	1,42
31	104.745	5,66	21780,58	-1,70	474.393.693	2,90
Total	2.571.897	228,21	0,00	0,00	4.913.205.620	110,64

Fonte: A partir dos dados da Tabela Nº 4

Cálculo das variâncias (Var):

Var X ===> $Sx^2 = \Sigma(X - \bar{X})^2 = 110,64 = 3,69$

n - 1 30

$$\text{Var Y} \Longrightarrow S_y^2 = \frac{\Sigma(Y - \bar{Y})^2}{n - 1} = \frac{4.913.205.620}{30} = 163.773.521$$

Cálculo dos desvios padrões:

Desvio padrão de X ==> $S_x = \sqrt{S_x^2} =$ 1,9204289

Desvio padrão de Y ==> $S_y = \sqrt{S_y^2} =$ 12797,403

$S_xS_y =$ 24576,503

Cálculo dos coeficientes de correlação R

$$R = \frac{Cov(Y,X)}{S_x\,S_y} = \frac{-18043{,}709}{24576{,}503} = -0{,}7341854$$

Isto quer dizer que o coeficiente de correlação R é igual a -0,734

8.2 Cálculo do coeficiente de correlação com uma calculadora HP-12C (HP, 1981 p.100-102)

Logo de fazer os cálculos assinalados anteriormente em 6.4, para deduzir os β_i da equação de regressão, pode-se conseguir os desvios padrões de X e Y assim:

a) Clicando em **g s** aparece o desvio padrão de X = 1,9204289
b) Clicando em **><** aparece o desvio padrão de Y = 12.797,403

E agora sim se pode deduzir o valor do coeficiente de correlação, assim:

a) Clicando em **0 g ŷ,r** aparecerá o valor de β_1 = 118.980,99
b) Clicando em **><** aparecerá o valor do coeficiente de correlação R = - 0,73418538

8.3 Cálculo do coeficiente de correlação R utilizando o Excel

a) Ter aberta a Tabela Nº 2, numa planilha em Excel
b) Clicar numa célula qualquer, que este livre
c) Clicar no botão superior da tela, em f**x**
d) Em Selecionar uma categoria, escolher ESTATISTICA
e) Em Escolher uma função, selecione CORREL
f) Em Matriz 1 carregar com o mouse todos os valores da coluna Y, sem o total
g) Em Matriz 2 carregar com o mouse todos os valores da coluna X, sem o total
h) Clicando em Ok aparecerá o valor -0,72419 que é o coeficiente de correlação

Este último valor também aparece no Quadro Nº 2 anterior, na parte superior, com o nome de R Múltiplo; só que ele aparece com valor positivo, quando na verdade ele é negativo. Daí o cuidado e alerta que se deve ter em tomar ligeiramente este valor como coeficiente de correlação. Nestes casos seria aconselhável ver o sinal do coeficiente $\beta_{2;}$ se ele for negativo, então o R também será negativo e vice-versa

8.4 Cálculo do coeficiente de correlação utilizando o Eviews

Neste caso deve-se seguir inicialmente tudo aquilo que já foi indicado no item 6.6 anterior, para deduzir os valores dos parâmetros βi; quer dizer:

a) Abrir o Eviews
b) Clicar sucessivamente em *File, New, Work file*
c) Selecionar *Date-regular frequency* e *Monthly*
d) Em *Start date* escrever 2017M3

e) Em *End date* escrever 2019M9
f) Em WF escrever Ibovespa_Selic
g) Em Page escrever 1 e clicar em Ok
h) Clicar sucessivamente em *Proc, Import, Import from file*
i) Clicar encima do arquivo **Ibovespa_Selic.xls** e Abrir
j) *Avançar* e em *Header type* selecionar *Names in first line*
k) *Avançar*, em *Start date* escrever 2017M3, *Finish*
l) Clicar sucessivamente em *Quick, Group statistics, Correlation*
m) Em *Series list* escrever **Ibovespa Selic** e logo clicar em Ok
n) Na janela correspondente aparecera -0,73419, que é o coeficiente de correlação

8.5 Cálculo do coeficiente de correlação utilizando o Stata

Igual, neste caso deve-se seguir aquilo que já se indicou no item 6.7 anterior, para deduzir os valores dos β_i, com estes detalhes:

a) Abrir o Stata
b) Clicar em *Stata SE-64*, File, *Import, Excel spreadshet, Brouse*
c) 0k
d) Selecionar *Import from first row as variable names*
e) Ok
f) Escrever embaixo de *Commands* **corr Ibovespa Selic**
g) Clicar em Enter
h) Na tela aparecera *corr Ibovespa Selic* = -0,7342; que é o coeficiente de correlação

9. Dedução do coeficiente de determinação R² (Matos, 2000 p. 68)

O coeficiente de determinação R^2 mede a bondade de ajuste do modelo de regressão; quer dizer, o grau relativo ou % de explicação, da variável independente (X = Selic), com respeito à variável dependente (Y = Ibovespa), quando se considera um universo de explicação de 1% a 100%. Evidentemente, quanto maior este valor, maior será o significado ou relevância da variável independente e vice-versa. Ele tem estas propriedades

a) Os valores de R^2 vão de 0 a 1: $0 \leq R^2 \leq 1$
b) Quanto maior seja o valor de R^2 mais representativo será o modelo estatístico, do problema em estudo
c) Um $R^2 = 0{,}90$ significa que um 90% da variação de Y (Ibovespa) seria explicado pelas variações de X (taxa Selic)
d) Ele também pode ser deduzido elevando ao quadrado o Coeficiente de correlação R

9.1 Dedução do coeficiente de determinação R^2 utilizando a estatística e geometria elementar (Wonnacott & Wonnacott, 1976 p. 110-114)

Para deduzir o valor de r^2 seguem-se estes passos:

a) Elabora-se um gráfico (Gráfico Nº 3), com os valores de X e Y nos eixos das abscissas (horizontal) e das ordenadas (vertical) respectivamente.
b) Identifica-se o ponto médio ou média aritmética de X ($\overline{X} = 7{,}36$); que aparece no cabeçalho da Tabela 5) e a partir dele traça-se uma paralela ao eixo das Y; igual, identifica-se a média aritmética de Y ($\bar{Y} = 82.964$) e traça-se uma paralela ao eixo das X, partindo deste último ponto.

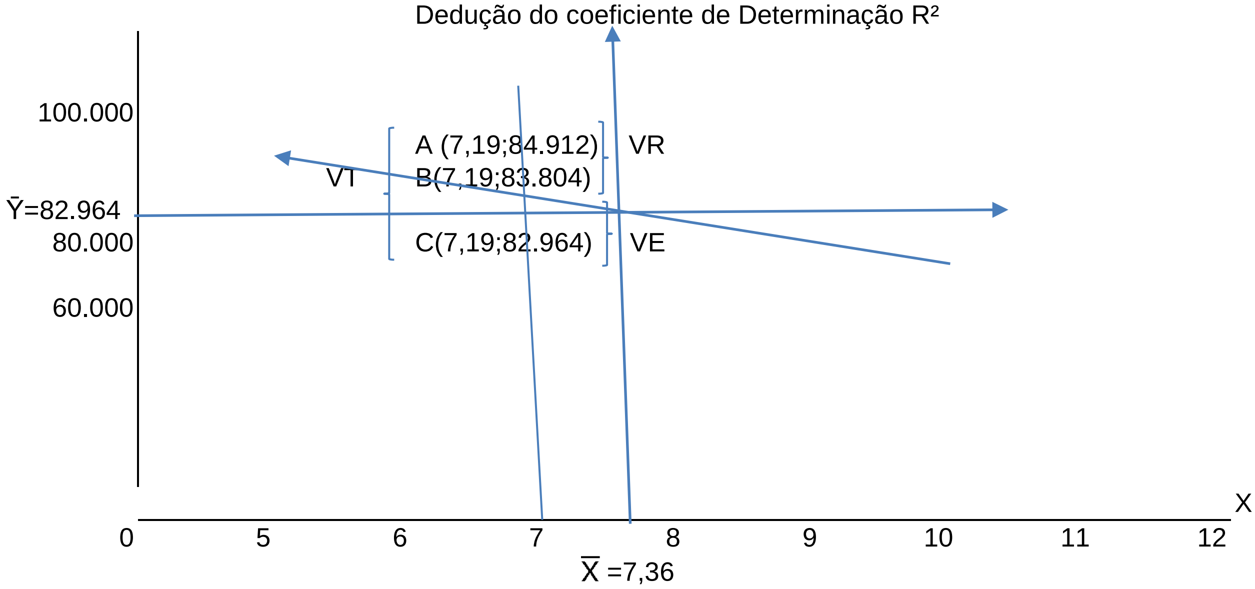
c) Desenha-se uma reta utilizando os valores de $\hat{Y}$ (já calculados na Tabela 3). Esta reta passará forçadamente pelo eixo das paralelas vertical e horizontal último

d) Na Tabela 3 identifica-se um ponto onde $Y > \hat{Y}$ (na realidade poderia ser qualquer ponto, porem só para facilitar o raciocinio aceita-se esta proposta). A escolha cai na informação 11 que lhe corresponde a janeiro de 2018. Nesse ponto $Y = 84.912$; $\hat{Y} = 83.84$ e $\bar{Y} = 82.964$. O valor de X seria 7,19

e) Tudo isto último pode-se ver no seguinte Gráfico

Gráfico Nº 3

Dedução do coeficiente de Determinação R^2



Fazendo as deduções seguintes deste Gráfico: com os valores de Y

VT	Distancia AC =	84.912	-	82.964	=	1.948
VR	Distancia AB =	84.912	-	83.804	=	1.108
VE	Distancia BC =	83.804	-	82.964	=	840

Sendo que VT = VR+VE ou AC = AB+BC ou

1.948	=	840	+	1.108

Ou VE = VT - VR

Dividindo tudo por VT

$$\frac{VE}{VT} = \frac{VT}{VT} - \frac{VR}{VT}$$

Introduzindo somatórias ∑

$$R^2 = 1 - \frac{\sum(Y - \hat{Y})^2}{\sum(Y - \bar{Y})^2}$$

$$R^2 = 1 - \frac{\sum(Y - \hat{Y})^2}{\sum(Y - \bar{Y})^2} \quad \begin{array}{l} \Leftarrow\!= \text{Tab. 3} \\ \Leftarrow\!= \text{Tab. 5} \end{array}$$

Substituindo valores, temos:

$$R^2 = 1 - \frac{2264849391}{4913205260}$$

$$R^2 = 1 - 0{,}46097186$$

$$R^2 = 0{,}539028135$$

9.2 Dedução do coeficiente de determinação R^2 através da HP-12C

Anteriormente em 8.2, operando a calculadora HPO-12C, conseguiu-se o valor de R = -0,73418538, do qual pode-se deduzir o valor de R^2 assim:

- 0,73418538 enter Y^X 2 enter ➔ aparecerá 0,53902817 = R^2

9.3 Dedução do coeficiente de correlação R^2 através de Excel

Idem, anteriormente em 6,5 conseguiu-se deduzir os parâmetros da equação de regressão, que aparece no Quadro Nº 2. Neste Quadro pode-se ler:

R-Quadrado = 0,53902817 ➔ Isto último é o R^2

9.4 Dedução do coeficiente de correlação R^2 através de Eviews

Idem, anteriormente em 6.6 conseguiu-se deduzir os parâmetros da equação de regressão, que aparece no Quadro Nº 3. Neste Quadro pode-se ler:

R-squared = 0,539028 ➔ Isto último é o R^2

9.5 Dedução do coeficiente de correlação R^2 através de Stata

Idem, anteriormente em 6.7 conseguiu-se deduzir os parâmetros da equação de regressão, que aparece no Quadro Nº 4. Neste Quadro pode-se ler:

R-squared = 0,5390 ➔ Isto último é o R^2

10. Balanço final

Ao longo deste artigo se apresentaram sugestões do como apresentar os dados que se utilizam na econometria, assim como as técnicas mais fáceis e entendíveis para calcular os parâmetros de uma equação de regressão e os valores da covariância, do coeficiente de correlação R e coeficiente de determinação R^2. Tudo isto deve ser a parte inicial de qualquer curso de econometria. Em cursos mais avançados da econometria deve-se abordar os fundamentos teóricos e práticos de tudo isto, além de definir os intervalos de confiança e testes de hipóteses para estes parâmetros, funções de regressão múltipla, exames de situações nos quais se violam os princípios fundamentais da econometria como da existência de homocedastícidade e da ausência de autocorrelaçao e da multicolinearidade; inclusive a campos mais avançados como as técnicas de dados em painel e series de tempo. Isto último se coloca para reconhecer que o conteúdo deste artigo é uma pequena parte do amplo mundo da econometria, porem ao qual se deveria facilitar seu acesso.

Notas de pé de página:

(*) Este material forma parte da disciplina Econometria I, lecionado no Departamento de Economia da Faculdade de Economia, Administração, Contabilidade e Atuariais da Pontifícia Universidade Católica de São Paulo Brasil (FEA-PUCSP).

(**) Professor associado do Departamento de Economia da FEA-PUCSP de São Paulo Brasil desde 1994 e ex-docente da FCE-UNMSM Lima Peru

REFERENCIAS BIBLIOGRAFICAS:

BERNANKE, Ben S. *On Milton Friedman's Ninetieth Birthday*. In: **The Conference to Honor Milton Friedman**, University of Chicago, Chicago, Illinois, November 8, 2002. Disponível em https://fraser.stlouisfed.org/title/453/item/8873, accessed on October 3, 2019

BEZERRA, Manoel Jairo; BEZERRA, Roberto Zaremba. **Álgebra 1**. 2ª edição revista. Rio de Janeiro: Fename, 1982

BRAGA, Carlos de Souza. **Eviews 7. Manual de referência rápida**. São Paulo: Mimeografado Departamento de Economia FEA-PUCSP 2014.

GUJARATI, Damodar N; PORTER, Dawn C. **Econometria básica**. 5ª edição. Porto Alegre: AMGH Editora, 2012

HP-12C. **Manual del propietario y guía para la solución de problemas.** Singapura, 1994.

IPEADATA. **Índice de ações, Ibovespa fechamento, Diário**. Rio de Janeiro: Ipeadata, 2019. Disponível em http://www.ipeadata.gov.br/Default.aspx

IPEADATA. **Selic, Taxa de juros over/Selic, (% a.m.), Mensal** Rio de Janeiro: Ipeadata, 2019. Disponível em http://www.ipeadata.gov.br/Default.aspx

KEYNES, John Maynard. **A teoria geral do emprego, do juro e da moeda.** São Paulo: Nova Cultural, 1988 (Coleção “Os Economistas”)

KMENTA, jan. **Elementos de econometria. Teoria estatística básica**. Volume 1. São Paulo: Atlas, 1988

LAPPONI, Juan Carlos. **Estatísticas usando Excel 5 e 7.** São Paulo: Mimeografado, s/d Poli/USP, 1997

MATOS, Orlando Carneiro de. **Econometria básica**. 3ª edição. São Paulo: Atlas 2000.

MYNBAEV, Kairat T; LEMOS, Alan. **Manual de econometria**. Rio de Janeiro: Editora FGV, 2004

MONTEIRO, Orlando. **Funções básicas no stata**. Viçosa MG: Economia & TV. Universidade Federal de Viçosa, 18/09/2011. Disponível em https://www.youtube.com/watch?v=61J2UAPMuXk&list=PL135C39D5EF47F78F

LEVINE, Ross; ZERVOS, Sara. *Stock market development and long run growth.* **The World Bank Economic Review**, Vol. 10 Nº 2, p. 323-339, may 1996. Disponível em http://faculty.haas.berkeley.edu/ross_levine/papers/1996_WBER_LRGrowth.pdf

SANDRONI, Paulo. **Novíssimo dicionário de economia**. São Paulo: Editora Best Seller, 1999. Disponível em http://sinus.org.br/2014/wp-content/uploads/2013/11/FMI.BMNov%C3%ADssimo-Dicion%C3%A1rio-de-Economia.pdf

SMAILES, Joanne e McGRANE, Ângela. **Estatística aplicada à Administração com Excel**. São Paulo: Atlas, 2002

STEVENSON, William J. **Estatística aplicada à administração**. São Paulo: Harbra, 1981.

VIVAS, Agüero Pedro Hubertus. *Bolsas de valores e crescimento econômico. Uma análise econométrica com dados em painel*. **Observatório de la Economia Latinoamericana**. Universidad de Málaga, España, noviembre de 2018. Disponível em https://www.eumed.net/rev/oel/2018/11/crescimento-economico.html?fbclid=IwAR3_WPl6QE22rfyK9rSlEcx5hIfkqnFXaHwz2IxJ2kvfJwTPr0CFGyGb-04 Este trabalho foi apresentado previamente no XXXIII Congresso da ANGE e XVI Seminário de Economia da PUC-SP, realizado em São Paulo entre os dias 01 a 05 de outubro de 2018, com o título de *Mercados mobiliários e crescimento econômico dos países: Uma análise econométrica com dados em painel*.

WONNACOTT, Ronald J; WONNACCOTT, Thomas H. **Econometria.** Rio de Janeiro: Livros Técnicos e Científicos, 1976