

ANÁLISE DA CORRELAÇÃO E REGRESSÃO DA EXPANSÃO AÇUCAREIRA DA REGIÃO OESTE DE SÃO PAULO

Teucle MANNARELLI FILHO¹

Resumo: O objetivo deste artigo é fazer uma análise da Correlação Estatística e das Análises de Regressão linear da expansão açucareira da Região Oeste do Estado de São Paulo; com base na evolução açucareira desta importante fronteira agrícola. O presente estudo tem caráter exploratório para melhor entender de que maneira esta se processando a expansão da indústria açucareira na Região Oeste do Estado de São Paulo, ainda correlacionar a evolução da produção em relação ao Estado, a nível Brasil e em especial os volumes mundiais de açúcar. Os dados primários foram coletados com metodologia descrita no presente estudo; com tratamento estatístico e utilização de planilhas eletrônicas tipo Excel. A justificativa do presente trabalho se dá em razão da importância do setor sucroalcooleiro em especial a expansão açucareira da região voltado basicamente para exportação e a carência de estudos neste sentido.

Palavras-chave: Agricultura; açúcar; sucroalcooleiro; cana-de-açúcar; correlação

1 Introdução

O complexo agroindustrial canavieiro é segundo Ramos (1999) a mais antiga atividade econômica do Brasil e tem sua história marcada por uma forte intervenção estatal, que tem como marco fundamental a criação do Instituto do Açúcar e do Alcool (IAA) em 1933; que teve entre suas atribuições fundamentais: fixação dos limites de produção de açúcar e álcool, através de cotas individuais de produção, quando da definição do Plano Anual de Safra; além da própria fixação de preços tanto da matéria prima cana-de-açúcar, como dos produtos finais açúcar e álcool; bem como todo o controle e operacionalização dos excedentes exportáveis de açúcar.

No início da década de 1990, com intuito de reduzir, modernizar e dar mais competitividade ao estado brasileiro o governo Collor promoveu a extinção do IAA em março

¹ Diretor Financeiro UDOP e Doutorando em Finanças na Universidade Extremadura-Espanha. Professor Universitário na FAC-FEA-CEP 16015-280-Araçatuba (SP)

de 1990; vindo este a ser o marco inicial da desregulamentação do setor sucroalcooleiro e sua entrada em um regime de livre mercado.

A própria permanência deste importante setor da economia brasileira sob tutela governamental por mais de três décadas e sua desregulamentação para uma situação de condições mais competitivas com regras de livre mercado; trouxe mudanças estruturais já a partir da extinção do IAA; em especial para o mercado de açúcar.

Uma passagem sem regras claras de transição de um modelo com controle de produção, preço e cotas para um sistema de livre mercado; em um primeiro momento acarreta uma natural instabilidade nos preços do açúcar, principalmente no mercado interno, que passa a refletir a volatilidade dos preços internacionais do produto; haja vista que o açúcar é uma commodity internacional com preços formados nas Bolsas de Mercadorias de Nova York (açúcar demerara) e na Bolsa de Londres (açúcar branco).

Ainda segundo Ramos (1999) os preços internos do açúcar no mercado interno passaram a ter correlação muito forte com os preços do mercado externo; principalmente com o aumento acentuado das exportações brasileiras, como podemos constatar ao longo deste artigo. Os mecanismos de arbitragem entre os preços domésticos do açúcar e os do mercado externo são de fundamental importância na formação do preço do açúcar já a partir da desregulamentação ocorrida com a extinção do IAA.

A Região Oeste de São Paulo tem se caracterizado como uma nova fronteira para produção de cana-de-açúcar no Estado de São Paulo e em especial com um aumento na produção de açúcar em relação à produção de álcool (anidro e hidratado).

Nosso estudo tem como objetivo utilizando-se das técnicas Estatísticas para estimar as Correlações existentes na evolução da produção açucareira da Região Oeste de São Paulo com os dados primários de produção de açúcar do Estado de São Paulo, o Brasil e a própria produção mundial de açúcar. Ainda suplementando vamos investigar utilizando as técnicas de Regressão Linear pelo Método dos Mínimos Quadrados, para verificação da existência de relações significativas da evolução de produção da produção de açúcar da Região Oeste como uma variável dependente; buscando as variáveis explicativas na própria evolução de produção de outras regiões.

Toda a metodologia da pesquisa e preparação das planilhas eletrônicas tipo Excel, seguiram a modernas técnicas estatísticas, com um a completa revisão bibliográfica ao longo deste artigo.

Finalmente realizamos ampla discussão quanto à qualidade dos dados coletados, quer sejam eles por amostragem ou da população total estudada; bem como dos resultados obtidos.

Tendo como objetivo maior entender a sustentabilidade da expansão açucareira na Região Oeste de São Paulo.

2 O Setor Sucroalcooleiro no Oeste Paulista

A Região Oeste Paulista teve grande impulso na produção de cana-de-açúcar com a implantação da segunda fase do Proalcool, a partir de 1980 com a implantação das Destilarias Autônomas, o que consolidou a região como grande produtora de cana-de-açúcar, em uma fase inicial até meados da década de 90 com o objetivo de produzir álcool hidratado, mas já a partir da década de 90, segundo Mannarelli (2002), a região inicia sua diversificação no mix de produção com incremento da produção de álcool anidro e açúcar.

As terras de boa qualidade da região, segundo Pinheiro (1997), favoreceram a expansão canavieira da região, e ainda a preponderância de uma pecuária extensiva que estava em uma fase de esgotamento, vindo deste modo à cana-de-açúcar adentrar na região como um novo ciclo econômico.

Podemos caracterizá-lo em duas fases distintas: na década de oitenta temos a implantação dos canaviais e consolidação da cultura com objetivo de produzir apenas álcool hidratado e na década de noventa, expansão dos canaviais com diversificação na produção das unidades agroindustriais, com incremento significativo de álcool anidro e açúcar Very High Polarization (VHP) de exportação, não abandonando a produção de álcool hidratado, mas sim, ajustando um mix de produção, que de um lado melhora a rentabilidade da atividade e com a diversificação, reduz o risco da atividade econômica, que a partir de 2000 consolida definitivamente a região como pólo de produção canavieiro no Brasil.

Esta consolidação do pólo canavieiro do Oeste Paulista, centrado na cidade de Araçatuba, que além de ser a sede da região administrativa, é a maior cidade da região e atrai a atenção dos grandes grupos nacionais produtores de açúcar e álcool. Deste modo uma série de aquisições e fusões se iniciam a partir do final de 1999 e início de 2000, onde podemos citar a vinda do Grupo Cosan e José Pessoa para a região de Araçatuba.

A criação da União das Destilarias do Oeste Paulista (UDOP)² em 1985, uma entidade com objetivo de congregar e representar as empresas produtoras de álcool do Oeste Paulista, que vai tratar dos interesses específicos da região nas questões relacionadas ao trabalho, políticas regionais, formação e aperfeiçoamento de profissionais através de cursos de

² Em 02 de janeiro de 1986 toma posse a primeira diretoria da entidade

qualificação, inovações tecnológicas; vem contribuir de forma decisiva para a consolidação da região de Araçatuba como novo pólo sucroalcooleiro brasileiro.

A UDOP iniciou seu trabalho com 16 unidades associadas cresceu, chegando a 19 e, atualmente, está com 27 associadas, sendo que, 17 delas são produtoras de açúcar. Tendo iniciado como uma entidade de defesa dos interesses dos produtores de álcool, e agora, estes mesmos produtores diversificam sua produção, passando também a produzir açúcar, a própria entidade adaptou-se a esta nova realidade, e a sigla UDOP passou a ter novo significado, Usinas e Destilarias do Oeste Paulista e não mais União das Destilarias do Oeste Paulista, quando da sua instituição. A partir de 2000 a entidade passa a oferecer também cursos de pós-graduação *latu sensu* em convênio com a Esalq-USP e UFSCar-USP nas áreas de comercialização industrial e agrícola, atraindo além dos profissionais da região, outros vindos de outras regiões do Brasil.

Outras entidades representativas do setor sucroalcooleiro no Estado de São Paulo, tem importância no setor, segundo Belik (2002), a União da Indústria Canavieira do Estado de São Paulo (ÚNICA) foi criada em 1997 em substituição a Associação das Indústrias de Açúcar e Alcool (AIAA) e a Sociedade dos Produtores de Açúcar e Alcool (Sopral) com objetivo de congregar em uma única entidade as empresas do setor sucroalcooleiro do Estado de São Paulo, em suas demandas comuns junto ao Governo e outras demandas comuns ao longo da cadeia produtiva. Ainda segundo o autor, a consolidação da ÚNICA como entidade exclusiva de representação dos interesses paulistas, em seu início foi confusa e, ao longo do tempo vem se consolidando, quer pelas ações concretas de atuação, quer pela profissionalização da entidade.

Deste modo, tanto a UDOP, como entidade representativa regional e a ÚNICA, como entidade estadual, podem ser definidas, segundo Caves e Porter (1977), como sendo Grupos Estratégicos que levam em consideração apenas à concorrência empresarial, com uma atuação estratégica semelhantes em um sistema agroindustrial coordenado.

3 Econometria

Vamos utilizar em nosso estudo os conceitos e técnicas econométricas; segundo Gujarati (2000) Econometria, significa literalmente a “medida econômica” com a aplicação da estatística matemática aos dados econômicos para dar suporte empírico aos modelos construídos pela economia matemática e para obter resultados numéricos.

Deste modo com técnicas econométricas iremos formular o conjunto de hipóteses que seja suficientemente específicas quanto realistas, para tirar o máximo proveito possível dos dados a nossa disposição.

A metodologia básica em nossa investigação seguirá a metodologia clássica econométrica, como segue:

- Formulação da Hipótese ;
- Especificação do modelo matemático da teoria;
- Especificação do modelo econométrico da teoria;
- Obtenção dos dados;
- Estimativa dos parâmetros do modelo econométrico;
- Teste de Hipótese;
- Previsão;
- Utilização do modelo para fins de controle ou políticas.

Desde modo dentro deste modelo básico, conduziremos nossa investigação para analisar a evolução da produção de açúcar da Região Oeste de São Paulo, buscando na teoria econométrica modelos matemáticas que possam explicá-la a luz das variáveis explicativas desta evolução, com consonância com os dados coletados dentro da pesquisa aqui desenvolvida.

Formulação da Hipótese

Nosso estudo está focado na hipótese que a evolução da produção de açúcar na Região Oeste de São Paulo, pode ser explicada por modelos econométricos de correlação com variáveis explicativas. Isto posta nossa hipótese básica de trabalho consiste na premissa que existe correlação por técnicas econométricas, entre a variável dependente, em nosso caso a produção de açúcar da região com variáveis explicativas; desde modo vamos selecionar as

principais variáveis explicativas de nossa hipótese e testar o modelo econométrico, buscando quais são as variáveis explicativas se apresentam mais consistentes nas diversas alternativas de regressão que vamos analisar ao longo deste estudo.

Especificação do Modelo Matemático da Teoria

O modelo matemático proposto será o modelo de regressão por uma ou mais variáveis, de onde temos de acordo com Gujarati (2000) que este tipo de análise se preocupa em relacionar a variável dependente com variáveis explicativas; sendo:

$$E(Y/X_i) = f(X_i)$$

Sendo que esta equação básica expressa as relações entre a variável dependente Y com as variáveis explicativas X. Ainda que mais especificamente iremos tratar das relações lineares existentes entre as variáveis Y e X.

Segundo Gujarati (2000) esta relação $E(Y/X_i) = f(X_i)$ é conhecida como Função de Regressão Populacional de duas variáveis que expressa a distribuição de Y dado X; sendo importante que vamos tratar apenas das funções lineares, ou das correlações lineares entre X e Y, tanto nas regressões simples, ou de uma variável explicativa X como buscando nas regressões múltiplas de mais uma variável explicativa X.

O significado do Termo Linear ainda Segundo Gujarati (2000) é no sentido de que a curva de regressão entre as variáveis assume uma reta, como uma função de regressão como:

$$E(Y/X_i) = \beta_1 + \beta_2 X_i$$

Deste modo o termo regressão linear significa que vamos utilizar uma regressão linear dos parâmetros $\beta(s)$, ou seja somente elevados a primeira potencia e ainda não é multiplicado nem dividido por qualquer outro parâmetro; podendo ou não ser linear nas variáveis explicativas X(s); e que deste modo a linearidade deve estar presente tanto nos parâmetros como nas variáveis.

Muitas vezes o modelo explicativo de apenas duas variáveis Y e X pode não ser plenamente explicativo, por existir outra variável que ajuda ajustar o modelo da regressão. Deste modo ainda segundo Gujarati (2000) precisamos ampliar nosso modelo de regressão simples de duas variáveis para abranger modelos que envolvam mais de duas variáveis; sendo que a adição de mais variáveis nos levará ao exame de modelos de Regressão Múltipla, nos quais a variável dependente Y depende de duas ou mais variáveis explicativas X; sendo o modelo mais simples de regressão de três variáveis, com uma variável dependente e duas variáveis explicativas, onde podemos ter:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Onde Y é a variável dependente, X2 e X3 são as variáveis explicativas e u é o termo de perturbação estocástico e u_i é a i-ésima observação. Ainda que na equação acima temos que β_1 é o termo de intercepto.

Obtenção dos Dados

O sucesso de qualquer análise econométrica, segundo Gujarati (2000) depende basicamente da disponibilidade de dados apropriados e confiáveis, com respeito à natureza, fontes adequadas. Ainda que podemos classificar os diferentes tipos de dados disponíveis em: série temporal, de corte e combinados.

Dados de Serie Temporal: representam um conjunto de observações coletados em intervalos de tempo regular, com sendo: diariamente, semanalmente, mensalmente, anualmente, etc; podendo ser quantitativos ou qualitativos. Estes dados de serie temporal apresentam devem ser estacionários, ou seja que seu valor médio e sua variância não se alterem sistematicamente com o tempo; deste modo quando estamos usando series temporais, devemos testar a estacionaridade destes dados.

Dados de Corte (Cross-Section): são dados de uma ou mais variáveis coletados no mesmo tempo ponto do tempo, a exemplo de dados sobre censo populacional, e que apresentam também problemas quanto a heterogeneidade.

Dados Combinados: na verdade representam elementos de series temporais, combinados com dados de corte; sendo que sempre devemos estar atentos aos problemas inerentes a estacionaridade e a heterogeneidade dos mesmos

Alem da classificação dos dados quanto a seu tipo, devemos considerar aspectos relevantes quanto a Fonte dos mesmos; que podem ser de agencias governamentais, organizações privadas, ou mesmo individuais; sendo que a confiabilidade da fonte vai ser de fundamental importância para o trabalho de investigação e pesquisa. Outro aspecto muito importante quanto aos dados diz respeito a sua qualidade e precisão e que todo pesquisador deve estar ciente que os resultados de sua pesquisa são tão bons quanto a qualidade dos dados e sempre que possível proceder à checagem deles por mais de uma fonte, e é medida de precaução e cautela.

Segundo Gujarati (2000) os dados ainda podem ser classificados como experimentais, como sendo aqueles em que o pesquisador tem controle sobre eles, podendo repeti-los em um novo experimento; e os dados denominados não experimentais, que em sua maioria são dados

econômicos que são coletados para fim da pesquisa; e que podem ainda ser qualitativos, quando são eminentemente numéricos e os quantitativos, que traduzem situações ou estado.

Teste de Hipótese

Uma vez supondo-se que o modelo esteja ajustado por metodologia adequada de regressão linear, como em nosso estudo, precisamos desenvolver critérios adequados para descobrir se nossas estimativas obtidas satisfazem as expectativas da teoria que esta sendo testada; segundo Hill (2003) uma teoria ou hipótese que não seja verificável por meio da evidencia empírica não pode ser admitida com parte da investigação científica.

Estes aspectos da confirmação ou rejeição da teoria com base na evidencia da amostra se baseia no ramo da estatística chamado de inferência estatística ou teste de hipótese

Ainda que segundo Gujarati (2000) se um modelo escolhido confirmar a hipótese podemos usá-lo para prever os valores futuros da variável dependente Y com base nos valores futuros conhecidos ou esperados da variável explicativa X

Correlação Entre Variáveis

Quando vamos estudar o relacionamento entre duas variáveis, o conceito de Correlação é de fundamental importância para o entendimento das variações existentes; segundo Triola (1999) podemos definir Correlação como sendo alguma forma de relacionamento entre duas ou mais variáveis.

Quando estudamos apenas duas variáveis, o Coeficiente de Correlação Linear r mede o grau de relacionamento linear entre os valores emparelhados de x e y em uma amostra; sendo que seu valor pode ser calculado segundo Triola (1999) pela seguinte formula :

$$r = \frac{n \sum xy - (\sum x) (\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

Ainda segundo o mesmo autor, este Coeficiente também pode ser chamado de coeficiente de correlação momento produto de Person³, que é calculado com base em dados amostrais e segue a estatística básica amostral.

A interpretação do Coeficiente de Correlação Linear é bastante importante; sendo que podemos utilizar uma gama de softwares estatísticos no mercado; mas que em nosso estudo vamos nos valer da Planilha Eletrônica tipo Excel, na função Análise de Dados. Os valores de r deve estar sempre entre -1 e +1 sendo que valores próximos de -1 e de +1 demonstram maior correlação e valores próximos de zero podemos concluir pela ausência de correlação linear entre as duas variáveis x e y analisadas.

Nosso estudo vai utilizar-se da Tabela de Valores Críticos do Coeficiente de Correlação r de Person; que foi desenvolvida por Karl Person como instrumento de análise para valores críticos e de uso prático nas análises de correlação linear.

Tabela Valores Críticos do Coeficiente r

No Observações	$\alpha = 0,05$	$\alpha = 0,01$
4	0,950	0,999
5	0,878	0,959
6	0,811	0,917
7	0,754	0,875
8	0,707	0,834
9	0,666	0,798
10	0,632	0,765
11	0,602	0,735
12	0,576	0,708
13	0,553	0,684
14	0,532	0,661
15	0,514	0,641
16	0,497	0,623
17	0,482	0,606
18	0,468	0,590
19	0,456	0,575
20	0,444	0,561

³ Person em homenagem a Karl Person (1857-1936) quem estabeleceu este coeficiente

Esta tabela serve como modelo de decisão quando vamos testar a hipótese $H_0 : \rho = 0$ contra $H_1: \rho \neq 0$; sendo ρ o coeficiente de correlação da população. Deste modo vamos rejeitar H_0 se o valor absoluto de r é maior que o valor crítico da tabela. Ainda nesta tabela temos tabulados valores α de 5% (0,05) e 1% (0,01), o que ajuda quando desejamos maior precisão da correlação.

Finalmente a interpretação que vamos utilizar em nosso estudo nas correlações lineares que vamos realizar é do tipo:

“verificar se o valor calculado r excede o valor na tabela, e se assim o for concluímos que existe correlação linear significativa; em caso contrário (valor r calculado for menor que tabela, não temos evidências suficientes para apoiar a existência de correlação linear significativa.”

Devemos sempre lembrar que ao utilizarmos a tabela podemos usar e fazer referência ao nível de chance que desejamos, no nosso caso 5% ($\alpha = 0,05$) ou 1% ($\alpha = 0,01$).

Segundo Triola (1999) o Coeficiente de Correlação Linear r tem as seguintes propriedades:

O valor de r está sempre entre -1 e $+1$ sendo: $-1 < r < +1$

O valor de r não varia se todos os valores de qualquer uma das variáveis são convertidas para uma escala diferente

O valor de r não é afetado pela escolha de x ou y ; sendo que permutando-se todos os valores de x e y o valor de r permanecerá inalterado.

O cálculo de r serve para medir o grau de intensidade de um relacionamento linear entre duas variáveis; mas é inadequado para relacionamentos não lineares.

Ainda segundo o mesmo autor, alguns erros comuns podem envolver a utilização da Correlação de Person; como evitarmos conclusões que impliquem em causalidade, evitarmos utilização de dados que são médias e finalmente no que diz respeito à propriedade da linearidade, ou seja conclusões que nos dizem que não existem correlações lineares, não significa que x e y não estejam de alguma forma relacionados de outra maneira.

Teste Formal de Hipótese

O teste forma de Hipótese segunda Triola (1999) serve para formulação de hipóteses para testar se existe realmente correlação linear significativa entre as duas variáveis x e y . Nosso estudo vai sempre testar a hipótese nula ($H_0 : \rho = 0$) contra a hipótese alternativa ($H_1: \rho \neq 0$), que podemos expressar da seguinte maneira:

$H_0 : \rho = 0$ (Não há correlação linear significativa)

$H_1: \rho \neq 0$ (Existe correlação linear significativa)

Deste modo testando a hipótese nula ($H_0 : \rho = 0$) e esta sendo rejeitada confirmamos a hipótese alternativa ($H_1: \rho \neq 0$) e conseqüentemente temos regressão linear significativa; e sempre de outro lado toda vez que a hipótese nula for aceita não existe correlação linear entre as variáveis x e y .

A estatística do Teste Estatístico vamos realizar de duas maneiras diferentes: Distribuição t Student e pela Tabela dos Valores Críticos de Person (r), descritos neste Capítulo.

Devemos ao longo de nosso estudo então realizar sempre os dois testes estatísticos: Distribuição t Student e Tabela Valores Críticos r de Person; de modo que em ambos os testes estatísticos vamos testar a hipótese nula ($H_0 : \rho = 0$) que é ausência de correlação linear contra a hipótese alternativa ($H_1: \rho \neq 0$) que nos indica forte correlação linear; sempre com a utilização de cálculos pelo Excel.

4 Regressão Linear

4.1 Regressão Linear Simples

A análise de regressão linear simples segundo Neufeld (2003) nos fornece uma maneira para investigar a relação entre duas variáveis; de outro lado Gujarati (2000) diz que este tipo de análise se preocupa em estimar ou prever a média ou valor médio das variáveis dependentes a partir de valores conhecidos ou fixados de uma variável explicativa, por meio da relação.

$$E(Y/X_i) = f(X_i)$$

Temos que esta equação segundo Gujarati (2000) é conhecida como Função de Regressão Populacional (FRP) de duas variáveis, expressando simplesmente que a média da distribuição de Y , dado um X_i relaciona-se funcionalmente com X_i , e que a média de Y varia com X ; e deste modo a FRP: $E(Y/X_i)$ é uma função linear do tipo:

$$E(Y/X_i) = \beta_1 + \beta_2 X_i$$

onde β_1 e β_2 são parâmetros desconhecidos, porém fixos, e que são chamados de coeficientes de regressão; ou também denominados de intercepto (β_1) e coeficiente de inclinação (β_2).

Segundo Neufeld (2003) quando usamos a regressão linear simples para analisar uma amostra de dados, estamos fazendo algumas suposições sobre estes dados, presumindo que existe uma relação linear entre as duas variáveis, e que poderíamos reescrever a equação de Gujarati (2000) como sendo:

$$E(Y/X_i) = \beta_1 + \beta_2 X_i + \epsilon$$

A letra ϵ (epípsilon grega) inserida na fórmula é um número aleatório, que significa que a relação não pode ser representada de maneira perfeita por uma linha reta; mesmo se conhecêssemos β_1 e β_2 não teríamos como prever o valor real de Y . Ainda segundo Neufeld (2003) existe uma diferença entre o valor real de Y e aquele que seria se a relação fosse perfeitamente linear, esta diferença é ϵ e seu valor difere aleatoriamente para cada observação. Ainda que a regressão exige certos pressupostos acerca da distribuição de ϵ : que seja normal com média zero, variância constante e independente uns dos outros.

Os ajustes de uma reta a uma base de dados amostrais, ainda segundo o mesmo autor; nos dão conta que os valores de β_1 e β_2 são estimados selecionando-se uma amostra de dados e encontrando a reta que melhor ajusta os dados de acordo com o critério específico da minimização das somas dos quadrados dos erros; sendo que esta reta estimada é representada pela função:

$$\hat{Y} = \beta_1 + \beta_2 x$$

O valor \hat{Y} é chamado de valor previsto de y e fornecido pela equação linear e que para cada observação na população existe um valor para o ϵ associado a ela. Quando uma amostra é então selecionada os valores amostrais não incidem precisamente sobre uma reta e conseqüentemente, haverá ainda segundo Neufeld (2003) uma diferença para cada observação entre o valor de \hat{Y} e o valor Y .

Segundo Downing (1998) o Diagrama de Dispersão é uma maneira de plotar graficamente as relações entre a variável explicativa e a variável dependente; onde a visualização torna-se um importante fator de auxílio nas análises de regressão linear.

Deste modo à medida que temos um diagrama de dispersão com valores de X dispostos ao longo do eixo horizontal e os valores de Y ao longo do eixo vertical podemos obter o coeficiente angular e o intercepto da reta que melhor se ajusta aos pontos, pelo método dos mínimos quadrados das diferenças entre os valores reais e os valores obtidos na reta de regressão, partindo das fórmulas:

$$\text{Coeficiente angular } (\beta^1) = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$\text{Intercepto } (\beta^2) = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

Sendo que as barras sobre as letras indicam valores médios e que esta reta minimiza a soma dos quadrados dos erros (distância vertical de cada ponto em relação à reta de regressão).

Ainda segundo o mesmo autor temos sempre que verificar a precisão da reta de regressão e como já dissemos simplesmente conhecer o coeficiente angular e do intercepto da reta de regressão, não podemos afirmar sobre o ajuste dos dados, e deste modo temos a necessidade de estabelecer uma medida do grau desse ajuste entre a reta de regressão obtida com os dados amostrais.

Segundo Gujarati (2000) no Modelo Clássico de Regressão Linear pelo Método dos Mínimos Quadrados, que é a pedra angular da maior parte da teoria econométrica; deve-se formular 10 hipóteses⁴ básicas:

Hipótese 1: O modelo de regressão é linear nos parâmetros

Hipótese 2: Valores dos regressores Xs são fixados amostragem repetida

Hipótese 3: O valor médio da perturbação é zero

Hipótese 4: As variâncias condicionais de ϵ são idênticas

Hipótese 5: Não existe autocorrelação entre as perturbações

Hipótese 6: A covariância entre X_i e ϵ é zero

Hipótese 7: O número de observações n deve ser sempre maior que o número de parâmetros estimados

Hipótese 8: Os valores de X em uma dada amostra não podem ser todos iguais

Hipótese 9: O modelo de regressão está corretamente especificado

Hipótese 10: Não existem relações lineares perfeitas entre as variáveis explicativas.

⁴ Este modelo clássico foi inicialmente desenvolvido por Gauss em 1821 e desde então tem servido como um padrão com o qual podem ser comparados os modelos de regressão que não satisfazem as hipóteses gaussianas.

Deste modo, segundo Gujarati (2000), uma vez satisfeitas as hipóteses, Método Clássico de Regressão Linear apresenta propriedades estatísticas relevantes para sua ampla utilização. O Teorema de Gauss-Marlov⁵ nos diz: “*dadas às hipóteses do modelo clássico de regressão linear os estimadores por mínimos quadrados, na classe dos estimadores lineares não viesados, tem mínima variância.*”

Ainda segundo Gujarati (2000) este Teorema tem tanto importância prática como teórica, uma vez que as propriedades estatísticas dos estimadores se apresentam como lineares e não viesado e deste modo tem uma menor variância quando são colhidas diferentes amostras para estimação linear.

Resíduos e Propriedade de Mínimos Quadrados

Como já dissemos a equação da regressão representa a reta que melhor se ajusta aos dados; pelo critério que utiliza a menor distância vertical entre os pontos que representam os dados originais e a própria reta de regressão; sendo que tais distâncias são denominadas resíduo. Assim segundo Triola (1999) “um resíduo é a diferença ($y - \hat{y}$) entre um valor amostral observado y e o valor \hat{y} predito com base na reta de regressão calculada e deste modo uma reta verifica a propriedade dos mínimos quadrados, se a soma dos quadrados dos resíduos é a menor possível.”

Em nosso estudo vamos também apresentar a tabela dos Resíduos para análise da reta de regressão calculada.

Intervalos de Variação e Predição

O conceito de correlação linear r serve para explicar a correlação entre duas variáveis x e y ; e à medida que calculamos uma reta de regressão do tipo $y = ax + b$ vamos ter uma série de pontos que não são explicados pela reta de regressão e deste modo precisamos dos conceitos, que segundo Triola (1999) são:

Desvio Total = Variação Explicada – Variação Não Explicada

$$\sum (y - \hat{y})^2 = \sum (\hat{y} - \tilde{y})^2 - \sum (y - \hat{y})^2$$

⁵ Conhecido como Teorema de Gauss-Markov, na verdade a abordagem dos mínimos quadrados foi desenvolvida por Gauss em 1821 e a abordagem da mínima variância de Markov é posterior de 1900.

Deste modo podemos introduzir o Conceito de Coeficiente de Determinação R^2 que é segundo Gujarati (2000) o valor da variação de y que pode ser explicado pela reta de regressão; e que é calculado como:

$R^2 = r \times r$ (ou seja, o quadrado do coeficiente correlação person) ou

$$R^2 = \frac{\text{Variação explicada}}{\text{Variação total}}$$

A utilização prática do Coeficiente de Determinação R^2 é muito importante, pois segundo Triola (1999) uma vez que o conhecemos podemos afirmar sobre uma reta de regressão que possui $R^2 = 0,64$; que 64% da variação de y pode ser explicada pela sua reta de regressão e que conseqüentemente 36% da variação total de y permanecem não explicadas pela respectiva reta de regressão; o que realmente ajuda na análise objeto de nosso estudo.

Ainda segundo Downing (1998) e Triola (1999) o coeficiente de determinação múltiplo ajustado (R^2) é a medida do grau de ajustamento da equação de regressão múltipla aos dados amostrais, com valores variando de 0 a 1; sendo que para um ajustamento perfeito teríamos o valor 1. e serve também como uma medida de aderência da equação de regressão aos dados amostrais.

Segundo Downing (1998) o coeficiente R^2 ajustado pode ser interpretado analogamente ao coeficiente r^2 da regressão linear simples, pois mede a porcentagem da variação da variável dependente que pode ser explicada pela regressão. Ainda que em nosso trabalho de investigação vamos utilizar a planilha Excel para realizar os cálculos de R^2 e nos concentrarmos nas interpretações dos resultados obtidos.

Ainda segundo o mesmo autor, para que os resultados da regressão sejam confiáveis, o número de observações deve ser significativamente maior do que o número de coeficientes que estamos estimando e que o valor de R^2 não aumentara necessariamente se acrescentarmos outra variável; sendo que Gujarati (2000) melhor exemplifica as diferenças entre R^2 e R^2 ajustado; como sendo o termo ajustado para os graus de liberdade associados as somas de quadrados e que existe uma relação matemática entre R^2 e R^2 ajustado, como sendo:

$$\mathbf{R^2 \text{ ajustado} = (1 - k/n) R^2}$$

e que sempre teremos: R^2 ajustado $<$ R^2 e conforme aumenta o numero de variáveis X o R^2 ajustado aumentara menos que o R^2 .

Ainda sobre qual R^2 devemos utilizar na pratica, segundo Theil (1978) e' aconselhável usar R^2 ajustado em vez de R^2 pois R^2 tende a fornecer um quadro demasiadamente otimista do ajuste da regressão, particularmente quando o numero de variáveis explicativas não for muito pequeno comparado co m numero de observações. Finalmente segundo Gujarati (2000) e' importante notar que ao comparar dois modelos com base no coeficiente de determinação, que seja ajustado ou não, o tamanho da amostra e a variável dependente devem ser iguais; caso contrario não podemos compará-los.

À medida que utilizamos a regressão linear, estimado uma reta de regressão do tipo: $y = ax + b$, não sabemos a precisão exata da regressão e precisamos recorrer ao conceito de intervalo de predição, que segundo Triola (1999) é uma estimativa intervalar de confiança de um valor previsto de y.

De acordo que o estabelecimento de um intervalo de predição exige uma medida da dispersão dos pontos amostrais em torno da reta de regressão; sendo o resíduo, também denominado desvio não explicado como sendo a distância vertical entre um ponto amostral e a reta de regressão calculada.

Segundo Gujarati (2000) podemos definir o erro padrão da estimativa como sendo uma medida das diferenças entre os valores amostrais de y observados e os valores previstos de \hat{y} obtidos pela reta de regressão linear e que segundo Triola (199), podemos calcular como sendo:

$$Se = \sqrt{\frac{\sum (y-\hat{y})^2}{n-2}} \quad \text{onde: } \hat{y} \text{ e' o valor predito de } y .$$

Deste modo com o auxilio do conceito do erro padrão da estimativa, podemos construir estimativas de intervalos que dirão sobre a confiabilidade de nossas estimativas pontuais de y para um determinado nível de confiança, sendo:

$\hat{y} - E < y < \hat{y} + E$ onde a margem de erro E é:

$$E = t_{\alpha/2} \text{ Se } \sqrt{1 + 1/n + \frac{n(x_0 - \bar{x})^2}{n(\sum x^2) - (\sum x)^2}}$$

Onde x_0 representa o valor dado de x, $t_{\alpha/2}$.

Finalmente podemos com o erro padrão, construir um intervalo de predição para uma equação linear de regressão a um determinado intervalo nível de predição consultando as tabelas estatísticas e que em nosso estudo vamos nos situar ao nível de 95%.

Método Alternativo P

O método do valor P para teste de hipótese, está baseado no cálculo do valor de uma probabilidade P e segundo Triola (1999) dada uma hipótese nula e um conjunto de dados amostrais o valor P reflete a plausibilidade de se obter tais resultados não são da hipótese nula ser de fato verdadeira, e deste modo segundo o mesmo autor um valor P muito pequeno (menor que 0,05) sugere que os resultados amostrais são improváveis sob a hipótese nula, constituindo pois evidencia contra a hipótese nula.

Como definição podemos dizer que o valor P ou valor da probabilidade; é a probabilidade de obter um valor da estatística amostral de tese no mínimo tão extremo como o que resulta dos dados amostrais, na suposição de a hipótese nula ser verdadeira.

A utilização da abordagem de P, se faz necessária, segundo Triola (1999) pois, enquanto a abordagem tradicional resulta em uma conclusão do tipo rejeita ou não rejeitar; os valores P dão o grau de confiança ao rejeitarmos a hipótese nula e também pode sugerir que os resultados amostrais sejam extremamente incomuns.

A interpretação dos resultados obtidos de P; sintetizamos como:

Valor P	INTERPRETAÇÃO
Inferior a 0,01	elevada significância estatística Forte evidencia contra hipótese nula
0,01 a 0,05	estatisticamente significativa

Deste modo devemos rejeitar a hipótese nula se o valor de P é no Máximo igual ao nível de significância α e de outro lado não rejeitar a hipótese nula se o valor de P for maior que o nível de significância α .

Análise da Estatística F “Snedecor”

Segundo Fonseca (1985) a distribuição de probabilidade F de “Snedecor” é unicaudal à direita e oferece abscissa que deixa 5% à direita e podemos consultar tabelas estatísticas para determinação de seus valores críticos. A utilização da estatística F nas análises de regressão linear são importantes pois como resultado da Tabela ANOVA, temos o cálculo do valor F, que é a estatística F, seguindo pois a distribuição F (1, n-2) e deste modo segundo Bussab(2002) o F que aparece na tabela ANOVA, serve para testar a hipótese $H_0: \beta = 0$ e pode-se usar a estatística e tabelas F e ainda segundo o mesmo autor a comparação do valor F obtido na regressão serve para indicar a significância das estatísticas da reta ajustada.

Ainda a estatística F, segundo Downing (1998) tem utilidade para testarmos a existência de correlação entre as variáveis independentes com a variável dependente e sua regressão; na verdade testando a hipótese nula $H_0: B_1 = B_2 = \dots B_{m-1} = 0$ o que em outras palavras que o verdadeiro valor dos coeficientes é zero. Para testarmos esta hipótese vamos calcular a estatística F:

$$F = \frac{(\text{SQREG}) / (m - 1)}{(\text{SQE}) / (n - m)} \quad \text{Onde :}$$

SQREG é a somatória do quadrado de todos os desvios calculados pela diferença entre o valor y previsto pela reta regressão e o valor médio de y.

SQE é a somatória do quadrado de todos os desvios entre o valor efetivo e o valor ajustado pela reta de regressão

m – 1 é o número de variáveis independentes

n é o número de observações

Deste modo se a hipótese nula H_0 fosse verdadeira o valor calculado da estatística F será menor que o valor encontrado na tabela F de “Snedecor” para o valor crítico desejado e determinado grau de liberdade. Deste modo quando a tabela ANOVA nos fornece o valor F, devemos compará-lo com a tabela estatística F a um valor crítico e se o F calculado for maior que o valor da tabela estatística conseqüentemente vamos rejeitar a hipótese nula de que não existe correlação linear entre as variáveis analisadas. Finalmente segundo Downing (1998) se a hipótese nula é falsa e existe correlação linear, então, devemos esperar que SQREG seja maior que SQE de forma que a estatística F será maior do que seria se a hipótese nula fosse verdadeira.

5 Análise da Evolução da Produção de Açúcar

A tabela abaixo apresenta a evolução da produção de açúcar do Brasil, do Estado de São Paulo, que é o principal produtor brasileiro de açúcar com mais da metade da produção nacional e finalmente à evolução da produção da Região Oeste do estado de São Paulo.

Com base nestes dados primários e nos conceitos econométricos discutidos neste artigo, buscaremos as possíveis correlações existentes na evolução da produção de açúcar e também em relação à evolução dos preços de exportação do açúcar.

Considerando que o mercado brasileiro de açúcar funciona sem qualquer interferência governamental, e em plenas condições de livre mercado, e que os preços do mercado interno são historicamente os preços do mercado internacional; vamos considerar para nossas análises o preço de referencia como o preço do açúcar exportado pelo Brasil, expresso em Dólares por tonelada; partido-se da premissa verdadeira que na situação brasileira de livre mercado, sem qualquer tarifa de importação, o preço de exportação na verdade é o grande formador de preços do mercado interno.

A Tabela de Evolução da Produção de Açúcar e Preços do Açúcar de Exportação foi compilada e elaborada pelo autor, com base nas informações disponíveis em diferentes fontes primárias de dados disponíveis.

Tabela Evolução Produção Açúcar e Preços Exportação

Ano	Brasil (mil ton)	São Paulo (mil ton)	Região Oeste (mil ton)	Preço Exp. (Us\$/ton)
1990/91	7.365,35	3.471,14	0	341,11
1991/92	8.604,32	4.567,30	0	266,03
1992/93	9.261,26	4.940,85	83,77	248,23
1993/94	9.273,88	5.536,97	97,16	257,09
1994/95	11.726,54	6.705,61	112,25	268,84
1995/96	12.652,88	7.244,09	177,02	307,47
1996/97	13.663,22	7.929,52	244,93	297,29
1997/98	14.910,61	8.734,86	374,35	277,75
1998/99	17.961,65	11.806,28	624,08	231,61
1999/00	19.387,17	13.091,18	835,57	157,51
2000/01	16.244,61	9.671,39	579,31	184,30
2001/02	19.196,24	12.328,48	966,42	203,98
2002/03	22.381,27	14.272,86	nd	nd

Fonte: Única em <http://www.unica.com.br>; Udop em <http://www.udop.com.br> e USDA em <http://www.usda.gov>; Preço Açúcar Exportado Brasil em <http://www.bmf.com.br>

5.1 Regressão Linear: Produção Açúcar Brasil x Preço Exportação

A correlação da Produção com o Preço é sempre importante no sentido de verificarmos a sensibilidade do incremento de produção em relação aos preços praticados; de onde devemos esperar uma correlação positiva, ou seja que aumentos de preço no produto, acarretam aumento na produção.

<i>Estatística de regressão</i>	
R múltiplo	0,71660688
R-Quadrado	0,51352542
R-quadrado ajustado	0,46487796
Erro padrão	3100,41886
Observações	12

ANOVA

	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>
Regressão	1	101471137,6	1,01E+08	10,556058	0,008735138
Resíduo	10	96125971,38	9612597		
Total	11	197597109			

	<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>
Interseção	27981,2962	4590,188675	6,095892	0,0001163
Variável X 1	-57,7164432	17,76432317	-3,24901	0,0087351

Obtemos deste modo a seguinte reta de Regressão Linear Simples:

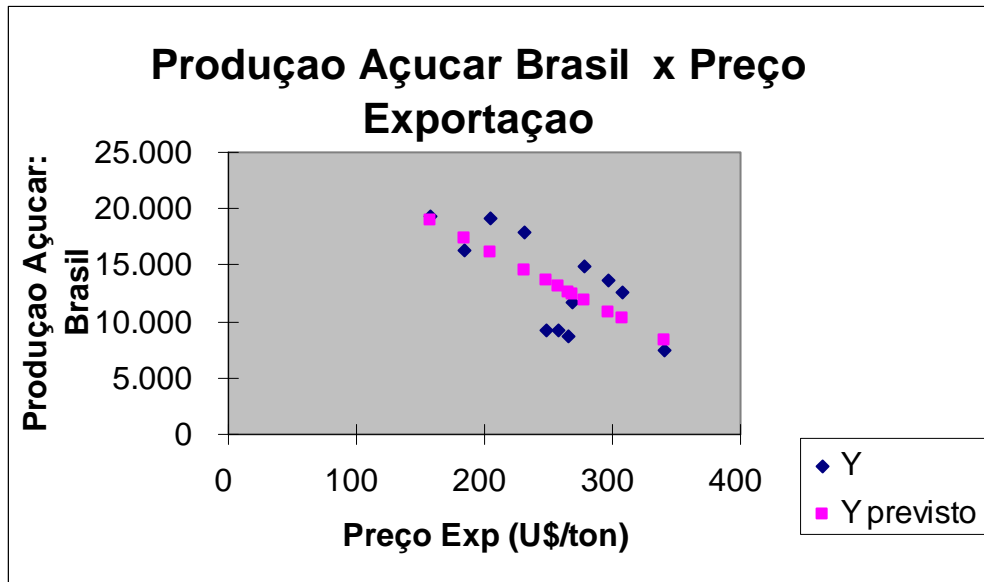
$$Y = - 57,72 X + 27.981$$

A reta de regressão tem um baixo grau de ajustamento sendo que somente 46,48% da variação na produção pode ser explicada pelo preço, medida pelo R² Ajustado (0,4648) e principalmente sinal negativo do coeficiente angular da reta de regressão (- 57,71) nos indica uma reta de regressão com inclinação invertida do que previamente deveríamos obter. Deste modo à medida que os preços do açúcar diminuem temos um aumento na produção de açúcar na Região Oeste São Paulo

As causas desta regressão invertida e com baixo coeficiente de ajustamento, pode ter varias causas, tais como: ganhos de produtividade, redução nos custos, maturação dos investimentos em canaviais e industriais, paridade cambial, ganhos marginais, etc; que no entanto não iremos discutir neste artigo.

Ainda importante constatar que os valores estatísticos P, F e t Student; obtidos nos indicam que o ajustamento da reta tem significância estatística com 95% de aceitação.

De onde podemos visualizar no seguinte gráfico de dispersão:



5.2 Regressão Linear: Produção Açúcar Brasil x Produção Açúcar Oeste São Paulo

Temos ao longo do período de observação que vai da safra 90/91 ate a safra 2002/03 um grande crescimento da produção de açúcar tanto no Brasil, como na Região Oeste de São Paulo; sendo que a Região Oeste não representa mais de 5% da produção brasileira de açúcar, o que nos dá a referencial econometrico para buscarmos uma correlação entre as produções de açúcar do Brasil com a Região Oeste de São Paulo.

<i>Estatística de regressão</i>	
R múltiplo	0,953764118
R-Quadrado	0,909665993
R-quadrado ajustado	0,898374242
Erro padrão	103,0777928
Observações	10

ANOVA

	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>
Regressão	1	855954,9197	855954,9	80,560225	1,89057E-05
Resíduo	8	85000,25099	10625,03		
Total	9	940955,1706			

	<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>
	-			
Interseção	771,9957648	135,6093731	-5,69279	0,0004582
Variável X 1	0,081889219	0,009123603	8,975535	1,891E-05

Obtemos deste modo a seguinte reta de Regressão Linear Simples:

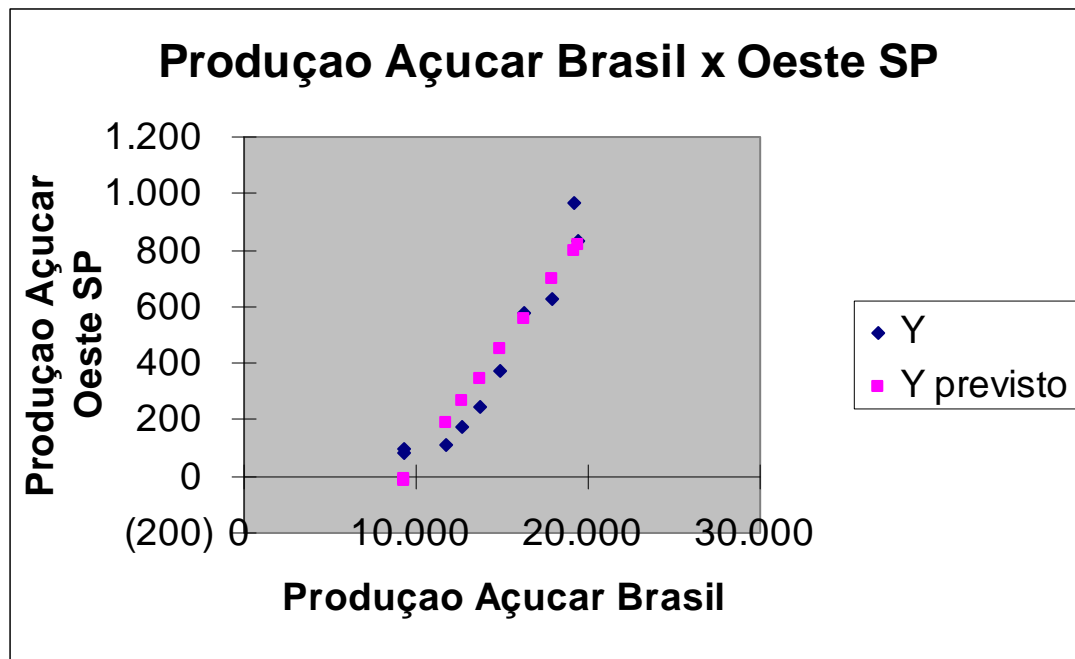
$$Y = 0,0818 X + 771,99$$

A reta de regressão tem um alto grau de ajustamento; sendo que 89,83% da variação na produção da Região Oeste SP pode ser explicada pela Produção do Brasil, medida pelo R² Ajustado (0,8983). Deste modo à medida que aumenta a produção de açúcar no Brasil, aumenta a produção de açúcar na Região Oeste de São Paulo, com um bom ajuste de correlação linear, expresso pela equação acima.

Este artigo não tem por objetivo buscar as causas do aumento da produção de açúcar, no Brasil e na Região Oeste SP; mas que podemos constatar a alta correlação de suas produções, e que deste modo os mesmos fatores desencadeantes e determinantes para o aumento da produção de açúcar do Brasil se dão também para a Região Oeste SP.

Ainda importante constatar que os valores P e F obtidos nos indicam que o ajustamento da reta tem significância estatística com 95% de aceitação.

De onde podemos visualizar no seguinte gráfico de dispersão:



5 Conclusões Finais

O presente artigo teve objetivo exploratório, no sentido de fazer uma discussão conceitual econométrica e da correlação linear, e com base nos dados primários buscar identificar os mecanismos de correlação existentes na produção de açúcar da Região Oeste SP, dentro de um contexto mais abrangente, que diz respeito à própria produção de açúcar do Brasil, que é o maior produtor e maior exportador mundial de açúcar e também no contexto globalizado de preços internacionais de açúcar na medida da própria competitividade do açúcar brasileiro e da Região Oeste SP.

Podemos deste modo concluir que a produção brasileira de açúcar a partir da década de noventa com o processo de liberação de preços e produção, com o fim do Instituto do Açúcar e do Alcool (IAA); teve um aumento sustentado e continuado e que o fator preço, que seria o fator determinante para este aumento na produção; na verdade não representa importância no aumento da produção de açúcar, sendo que pelo contrario temos um ajustamento linear de regressão invertido com coeficiente angular negativo; nos mostrando

que à medida que os preços do açúcar diminuíram no mercado internacional a produção de açúcar brasileiro e na Região Oeste de SP aumentou.

No tocante a análise da correlação linear simples da produção de açúcar brasileiro com a produção da Região Oeste SP, iremos obter um forte ajuste de correlação; e alto coeficiente de explicação de 89,83%; de modo que podemos inferir que produção de açúcar da Região Oeste SP, que representa aproximadamente 5% da produção brasileira, esta de fato integrada economicamente com a produção brasileira de açúcar, e à medida que produção de açúcar no Brasil aumentou a partir da década de 90, a produção da Região Oeste respondeu com aumentos com alta correlação e ajustamento estatístico significativo; demonstrando que a Região Oeste responde aos mesmos fatores que impulsionaram a produção de açúcar do Brasil.

Finalmente que não esta sendo objeto deste nosso estudo investigativo a determinação dos fatores que impulsionaram e que serão determinantes na expansão açucareira do Brasil e da Região Oeste de São Paulo; mas que em Mannarelli (2002) temos que a maior rentabilidade do açúcar de exportação e a melhor liquidez do açúcar são alguns dos fatores explicativos desta expansão, medidos pela Análise das Demonstrações Financeiras das empresas produtoras.

MANNARELLI FILHO, Teucle. Analysis of correlation and regression of sugar expansion in the west region of São Paulo. **Economia & Pesquisa**, Araçatuba, v. 7, n. 7, p. , mar. 2005.

Abstract: The objective of this article is to make an analysis of the Statistical correlation and of the analysis of linear regression of sugar expansion in the West Region of São Paulo; with basis in the sugar evolution of this important agricultural frontier. This study has an explanatory nature to better understand the process of the expansion of the sugar industry in the West Region of São Paulo; also correlate the evolution of production as to the state, concerning to Brazil and specially the r world sugar volume.

The primary data were collected with methodology described in this study; with statistical treatment and utilization of electronic plan excell type. The justification of this work is due to the importance of the sugar-alcohol sector, specially the sugar expansion of the region turned basically to the exportation and the lack of study in this sense.

Key words: Agriculture; sugar; sugar alcohol; correlation

Referências Bibliográficas

- ASSAF NETO, A., MARTINS, E. **Administração financeira**. São Paulo: Atlas, 1986.
- BELIK, W., et al. **Agroindústria canavieira no Brasil**. São Paulo: Atlas, 2002.
- BOLSA DE MERCADORIAS & FUTUROS (BM&F). Disponível em: <<http://www.bmf.com.br/>>. Acesso em: 22 maio 2004.
- BUSSAB, W., MORETTIN, P.A. **Estatística básica**. São Paulo: Saraiva, 2002.
- CAVES, R., PORTER, M. From entry barriers to mobility barriers: conjectural decisions and contrived deterrence to new competition. **Journal of Economics**, Cambridge, v.91, n.2, p-241-261, 1977.
- DOWNING, D., WESTERNFIELD, R.W. y BRADFORD, D.J. **Princípios da administração financeira**. São Paulo: Saraiva, 1998.
- FOSECA, J.S. et al. **Estatística aplicada**. São Paulo: Atlas, 1976.
- GITMAN, L. **Princípios de administração financeira**. São Paulo: Harbra, 1997.
- GUJARATI, D.N. **Econometria básica**. São Paulo: Makron Books, 2000.
- HILL, R.C.; GRIFFITHS, W.E.; HUDGE, G.G. **Econometria**. São Paulo: Saraiva, 2003.
- MANNARELLI, T.F. **Revista Economia e Pesquisa**. Araçatuba: Fundação Educacional, 2002. ?
- MARTINS, E. **Administração financeira**. São Paulo: Atlas, 1987.
- MORAES, M.A.F.D. **A Desregulamentação do setor sucroalcooleiro do Brasil**. Americana: Caminho Editorial, 2000.
- MORAES, M.A.F.D. et al. **Agroindústria canavieira no Brasil: evolução, desenvolvimento e desafios**. São Paulo: Atlas, 2002.
- NEUFELD, J.L. **Estatística aplicada à administração usando Excel**. São Paulo: Person Education do Brasil, 2003.
- PINHEIRO, C., BODSTEIN, O.C. **História de Araçatuba**. Araçatuba: Academia Araçatubense de Letras, 1997.
- RAMOS, P. **Agroindústria canavieira e propriedades fundiárias no Brasil**. Tese (Doutorado) - Escola de Administração de Empresas de São Paulo, Fundação Getulio Vargas, São Paulo, 1991.
- THEIL, H. **Principles of econometrics**. New York: John Wiley & Sons, 1978.
- TRIOLA, M.F. **Introdução à estatística**. Rio de Janeiro: Livros Técnicos Científicos, 1999.
- UNIÃO DA INDÚSTRIA CANAVIEIRA DO ESTADO DE SÃO PAULO (UNICA). Disponível em: <<http://www.unica.com.br/>>. Acesso em: 22 jan. 2005

USINAS E DESTILARIAS DO OESTE PAULISTA (UDOP). Disponível em:
<<http://www.udop.com.br/>>. Acesso em: 20 jan. 2005