# APLICAÇÃO DO MÉTODO DE REGRESSÃO DE CUMEEIRA ("RIDGE REGRESSION") NA ESTIMAÇÃO DE FUNÇÕES DE DEMANDA E DE PRODUÇÃO

MARIA NAÍMA KALIL

Orientador: RODOLFO HOFFMANN

Dissertação apresentada à Escola Superior de Agricultura "Luiz de Queiroz" da Universidade de São Paulo, para obtenção do Título de Mestre em Economia Agraria.

PIRACICABA
Estado de São Paulo - Brasil
agosto, 1977

*Aos meus pais*

# AGRADECIMENTOS

Ao Professor Rodolfo Hoffmann, pela valiosa orientação oferecida.

Aos Professores José Ferreira de Noronha e Geraldo Sant'Ana de Camargo Barros, pela leitura das elaborações finais deste trabalho, enriquecendo-o com importantes sugestões e comentários.

Ao Departamento de Ciências Sociais Aplicadas da ESALQ, e à Fundação Ford, pela bolsa de estudos, e também pelo financiamento da despesa com a publicação deste trabalho.

Ao Instituto de Pesquisas Econômicas e Sociais, pela bolsa concedida no período de elaboração da pesquisa.

Aos Professores Roberto Simionato de Moraes e Vivaldo Francisco Cruz, que me possibilitaram a utilização do Centro de Computação Eletrônica da ESALQ.

# ÍNDICE

Página

# LISTA DAS TABELAS

Página

Página

Página

Página

Página

# LISTA DAS FIGURAS

Página

Página

## RESUMO

Esta pesquisa teve como principal objetivo a aplicação do método de regressão de cumeeira a funções de demanda e de produção, comparando os resultados com os obtidos pelo método de mínimos quadrados ordinários. Foi verificada a possibilidade de se obter "melhores" resultados em estudos econômicos cujas variáveis são multicolineares.

Analisou-se as causas, efeitos e maneiras de detecção da multicolinearidade, como também os problemas existentes com relação aos métodos normalmente utilizados para contorná-la. As pesquisas com esse tipo de problema apresentam estimativas bastantes afetadas, de baixa precisão e portanto não confiáveis.

Utilizou-se, para confronto dos resultados, as pesquisas de SOBRAL (1973), referente ao estudo de demanda de alimentos (arroz, feijão, batatinha, carne e leite) e de BISERRA (1971), referente ao es-

tudo da função de produção de milho em Jardinópolis e Guaíra, municípios da DIRA de Ribeirão Preto. Os mesmos dados destes trabalhos foram aqui empregados. Os dados correspondentes à função de demanda são provenientes de diversas fontes, são anuais, e abrangem o período de 1950/70, para o Brasil. Os dados correspondentes à função de produção referem-se ao ano agrícola de 1969/70, e foram obtidos através de entrevistas diretas com agricultores.

No ajustamento das equações, utilizou-se o método de regressão de cumeeira ("ridge regression"), que fornece estimativas do vetor dos parâmetros com menor comprimento que aquelas obtidas pelo método de mínimos quadrados. Este método foi sugerido por HOERL e KENNARD (1970 a), e visa a obtenção de melhores resultados para análises de regressão onde o problema de multicolinearidade afeta seriamente as estimativas obtidas pelo método usual de mínimos quadrados. Êle baseia-se no acréscimo de pequenas quantidades positivas (no intervalo de 0 a 1), aos elementos da diagonal principal da matriz X'X quando na forma de matriz de correlações. Existem valores desse acréscimo para os quais o vetor dos estimadores de cumeeira ($\hat{\beta}^*$) está, em média, mais próximo de $\beta$ (o vetor dos parâmetros) do que o vetor dos estimadores de mínimos quadrados ($\hat{\beta}$), isto é, a esperança do quadrado do desvio é menor para $\hat{\beta}^*$ do que para $\hat{\beta}$. Uma apresentação deste método foi desenvolvida neste trabalho, visando maior clareza dos fundamentos teóricos em que o mesmo se apóia.

Através do gráfico de cumeeira e dos "fatores de inflação da variância das estimativas dos parâmetros", verificou-se a existência e importância da multicolinearidade nos dados de uma amostra.

No caso das funções de demanda analisadas, todas elas referentes a produtos alimentares, foram obtidos coeficientes de elasticidade-renda e elasticidade-preço que mostraram que essas demandas são inelásticas.

Os resultados correspondentes à função de produção de milho evidenciaram que todos os fatores estavam sendo utilizados no estágio racional de produção, e que as variáveis sementes e "despesas de custeio" estavam sendo sub-utilizadas.

Concluiu-se que o método de regressão de cumeeira é uma alternativa válida quando se tem dados seriamente afetados pelo problema de multicolinearidade. Obteve-se neste trabalho resultados melhores na aplicação do método às funções de produção do que para as funções de demanda.

# 1. INTRODUÇÃO

A teoria econômica pode ser aplicada e testada através dos modelos econométricos. Estes envolvem a representação matemática, em seu sentido qualitativo e quantitativo, dos problemas econômicos.

Recentemente novos métodos econométricos têm surgido, complementando os anteriores, ou colocando em segundo plano aqueles cuja validade era considerada incontestável. Entre estes está o método dos mínimos quadrados, que é quase sempre utilizado no ajuste de equações, devido às suas propriedades estatísticas. Porém existe a pressuposição de certas condições que na realidade podem não acontecer, e que conforme a gravidade do problema irão tornar o método dos mínimos quadrados impróprio para a obtenção de conclusões econômicas válidas. Neste caso as novas metodologias que forem surgindo e mostrarem-se mais adaptadas, vão contribuir com melhores resultados, maior grau de confiabilidade, e consequentemente para decisões mais corretas baseadas na análise econométrica de dados. Assim, medidas de política agrícola terão sua qualidade melhorada com a precisão de sua base de informações econométricas. Espera-se que este trabalho possa dar uma contribuição neste sentido, pois os diferentes métodos propostos devem ser experimentados com a finalidade não só de verificar sua validade como também para divulgá-los.

## 1.1. Importância das funções de demanda

O conhecimento da demanda de alimentos é particularmente importante na adequação da produção ao consumo dos gêneros alimentícios. Se o crescimento da demanda superar o da oferta de um alimento, pode-se esperar que ocorra "déficit" do mesmo, pelo menos temporariamente.

As medidas de política agrícola dirigidas à viabilização do abastecimento, bem como a melhorias no nível alimentar e nutricional da população, não podem prescindir de uma análise estrutural da demanda de alimentos básicos, entre os quais se destaca o arroz, o feijão, a batatinha, a carne e o leite, como os tradicionalmente consumidos no Brasil.

A quantidade demandada de alimentos é afetada pelo crescimento da população, pela urbanização, pela elevação da renda "per capita", pelas variações nos preços, etc.... Tal influência pode ser medida pelos coeficientes de elasticidade. Na prática, a obtenção de estimativas desses coeficientes pode ser muito útil para explicar e prever variações no consumo.

As evidências têm mostrado que os produtos agrícolas, em geral, têm demanda, a curto prazo, preço-inelástica . Pequenas variações nas quantidades resultariam portanto de modificações relativamente grandes nos preços. Por outro lado, um aumento no preço resultaria em maior dispêndio pelos consumidores. Isso devido principalmente ao fato desses produtos não possuirem substitutos próximos (JUNQUEIRA, 1964, p. 143). WAUGH (1973, p. v, parte preliminar) considera que a demanda de alimentos é inelástica não só quanto ao preço mas também com relação à renda. De fato, têm sido obtidos coeficientes de elasticidade da demanda inelásticos, em muitos trabalhos realizados. SOBRAL (1973, p. 37) apresenta, em uma análise de bibliografia sobre demanda de alimentos, uma tabela contendo vários coeficientes de demanda, estimados por diferentes autores, para locais distintos, e sendo quase todos inelásticos. Porém, sabe-se que nem sempre a demanda de alimentos é inelástica. Para determinados tipos de produtos, como é o caso da carne e outros produtos protéicos de origem animal, as elasticidades preço e renda serão maiores. No caso da elasticidade-renda, espera-se que para regiões de renda elevada ela seja baixa para produtos da agricultura. Mas com relação às populações de baixa renda as elasticidades-renda de alimentos tendem a ser altas. PEREZ (1973), estimou elasticidades-renda do consumo de alimentos, no município de Piracicaba, obtendo para os alimentos protéicos, com poucas exceções, elasticidades maiores do que 1. Além disso, os coeficientes, tanto para os produtos protéicos como para os energéticos e hortícolas, resultaram sempre maiores para o estrato de renda inferior, comparativamente ao superior.

## 1.2. Importância das funções de produção

A análise da produção agrícola tem sido fundamentada, basicamente, nas funções de produção, como é possível se observar em numerosos trabalhos realizados.

A utilização deste tipo de funções em economia agrícola visa determinar as relações econômicas entre a produção e os respectivos insumos que para ela contribuem.

Sabe-se que os estudos relacionados com análises de funções de produção somente possuem utilidade a nível de subsídio à adoção de medidas econômicas para determinada região estudada. Conhecer o modo como atuam os insumos revela-se bastante importante neste sentido. Portanto, existem certas limitações na aplicação deste tipo de funções, já que não é possível recomendar ao agricultor, a nível individual, mudanças nas utilizações dos fatores com base nos estudos de função de produção agregada.

Deste tipo de funções é possível se derivar curvas de oferta. Com relação à política agrícola, estas análises contribuem para a fixação de preços mínimos, estimativas de safras agrícolas, adequações da oferta à demanda[1/], previsão de variações decorrentes de mudanças nos preços, etc.... Isso, por sua vez, será útil no planejamento da comercialização dos produtos agropecuários, incluindo armazenamento e transporte.

## 1.3. Objetivos

Os objetivos deste trabalho são:

a. Verificar a aplicabilidade do método de regressão de cumeeira ("ridge regression"), comparando os resultados com aqueles provenientes da utilização do método de mínimos quadrados ordinários, nos casos em que existe multicolinearidade. Serão feitas comparações diretas com os resultados obtidos por SOBRAL (1973), para funções de demanda, e com aqueles obtidos por BISERRA (1971), com relação a funções de produção.

b. Analisar, no caso dos dois tipos de funções em que o método de regressão de cumeeira será aplicado, as características dos resultados obtidos, verificando a que

---

1/ Sabe-se que no Brasil e no mundo têm sido caracterizados "deficits" na produção e oferta de alimentos, que se espera que aumentarão com o crescimento da procura em função principalmente do aumento populacional.

tipo de modelos econômicos é mais conveniente o uso deste método.

c. Discutir as limitações que o problema de multicolinearidade vem causando em análises econômicas.

d. Estimar coeficientes de elasticidade da demanda em relação a preço, renda, urbanização, preço de outro produto (elasticidade-cruzada) e salário mínimo, pelo método de regressão de cumeeira, para arroz, batatinha, feijão, carne e leite, referentes ao período de 1950 a 1970, com dados relativos ao Brasil.

e. Estimar funções de produção de milho, através do mesmo método, para os municípios de Jardinópolis e Guaíra, separadamente, com dados de 1969/70, e analisar o nível de utilização dos fatores, aos preços do ano agrícola estudado.

# 2. A MULTICOLINEARIDADE EM ESTUDOS ECONÔMICOS

## 2.1. Análise do Problema

Segundo FARRAR e GLAUBER (1967, p. 92), quando surgem problemas de multicolinearidade, é freqüente existir certa tendência à eliminação de variáveis das equações de regressão. Este procedimento pode originar estimativas viesadas para os coeficientes, principalmente nos casos em que as variáveis que se apresentam correlacionadas são fundamentais à especificação do modelo.

É freqüente alguns autores comentarem as modificações nas propriedades dos parâmetros, em decorrência da não inclusão, ou inclusão em excesso, de variáveis num modelo. HEADY e DILLON (1966, p.202), afirmam que "se fossem omitidas quaisquer variáveis relevantes, o modelo ajustado seria viesado num sentido econômico; ... do mesmo modo, a inclusão não justificada de variáveis conduzirá ao viés". Esta conclusão não é verdadeira. Deve-se concordar que quando se retira uma ou mais variáveis importantes de um modelo, as estimativas dos parâmetros resultarão tendenciosas, mas se se inclui variáveis adicionais irrelevantes à especificação do mesmo, essas estimativas continuam não tendenciosas. JOHNSTON (1972, p. 168-169), prova esta afirmação, e cita que "... a exclusão de variáveis relevantes da regressão pode ser um erro muito sério: não somente conduz a estimativas de coeficientes seriamente viesadas, mas as inferências com base nestes coeficientes serão também imprecisas, desde que a estimativa da variância residual será viesada para cima. Se os dados e os graus de liberdade permitirem, será preferível errar por incluir variáveis na análise de regressão, antes de excluí-las". Deve-se observar no entanto que, conforme foi citado acima, apesar das estimativas dos parâmetros de um modelo com inclusão de variáveis irrelevantes continuarem não tendenciosas, há perda de eficiência das mesmas. KMENTA (1971, p. 399), deduz que "... se o erro de especificação consiste em incluir algumas variáveis independentes irrelevantes na equação de regressão, os estimadores de mínimos quadrados dos coeficientes de regressão permanecem não viesados mas tornam-se não eficientes". Portanto, o ideal consiste em nem deixar de incluir no modelo todas as as variáveis relevantes, nem incluir variáveis desnecessárias.

Mesmo no caso em que não forem obtidas estimativas dos parâmetros significativas, não se deve eliminar a respectiva variável do modelo. HEADY e DILLON (1966, p. 211), argumentam que não é recomendável a exclusão de variáveis de um modelo, nos casos de ajustamento de funções de produção, mesmo que os coeficientes estimados não se apresen tem estatisticamente significativos, se essas variáveis são importantes ao processo produtivo. Dever-se-ia então usar o critério proposto por Anderson, segundo o qual somente quando o desvio padrão de um coeficiente da variável fosse maior que o seu coeficiente em valor absoluto, é que se deveria retirá-la do modelo pré-determinado, devendo-se considerar ainda se ela não é essencial ao mesmo. No caso das variáveis se apresentarem altamente correlacionadas, o que se observa frequentemente é o procedimento de se "deixar de lado" algumas variáveis, para se obter parâmetros significativos, e dessa forma contornar o problema de multicolinearidade, ainda que conduzindo a estimativas viesadas e não eficientes.

A multicolinearidade ocorre quando alguma variável está correlacionada com outra ou com uma combinação linear de outras variáveis independentes no modelo. Quando essa correlação é, em valor absoluto, igual a um, diz-se que há multicolinearidade perfeita. Mas este tipo de multicolinearidade não é o mais importante, já que dificilmente ocorre na prática. De modo geral ocorre a multicolinearidade não perfeita.

Os efeitos deste tipo de problema refletem-se nas estimativas obtidas que serão de baixa precisão, e portanto não confiáveis. Os erros de estimação são grandes, podendo levar as estimativas com sinais incorretos. Se existirem dois fatores positivamente correlacionados ($r_{ij}$ positiva e elevada), há tendência de os respectivos coeficientes de regressão estimados apresentarem desvios grandes e opostos, ou seja, um deles terá grande valor positivo e outro elevado valor negativo (JOHNSTON, 1972, p. 159-163).

Devido à grande variância das estimativas, os testes "t" indicam que as respectivas variáveis não estão exercendo influência estatística significativa em Y, não obstante na realidade essa influência possa ser importante, estando apenas "ocultada" pela multicolinearidade (o poder do teste é fraco neste caso). Além disso, as estimativas resultantes são instáveis, pois com adição de poucas novas observações à amostra, seus valores podem se alterar bastante.

Portanto conclui-se que a imprecisão originada por um alto grau de multicolinearidade é prejudicial no sentido de que as estimativas obtidas pelo método de mínimos quadrados ordinários serão aparentemente não-significativas, quando o problema pode estar, neste caso, localizado na amostra (KMENTA, 1971, p. 389).

## 2.2. Medidas da Multicolinearidade

A preocupação com um grau elevado de multicolinearidade numa amostra, leva os pesquisadores a desejarem determinar uma medida para ela. Têm sido frequentes as pesquisas em que os autores realizam a observação dos coeficientes de correlação simples entre duas a duas das variáveis incluidas, o que não é correto. Pode existir multicolinearidade perfeita num modelo com mais de duas variáveis explicativas, sem que haja nenhum coeficiente de correlação simples igual à unidade, ou mesmo alto (podendo ser até de baixos valores). Por outro lado, se existir um coeficiente unitário ($r_{ij}$ = 1), então afirma-se que existe multicolinearidade perfeita. Portanto a existência de um coeficiente de correlação simples elevado entre duas variáveis independentes é uma condição suficiente mas não necessária para que surjam problemas devido à multicolinearidade, quando a regressão contém mais que duas variáveis independentes (KMENTA, 1971, p. 383-384).

KLEIN (1962, p. 101) propõe que se considere um valor relativo de $r_{ij}$, ou seja, se $r_{ij}$ for alto em comparação com o valor da correlação múltipla da equação ($r_{ij} > R$), o problema deve ser considerado grave.

A observação do valor do determinante de X'X é sugerida em vários textos. Este será baixo com altos graus de multicolinearidade. No limite, com multicolinearidade perfeita, será zero (a matriz X'X é singular). No entanto o determinante tem seu valor influenciado também pela dispersão das observações, a não ser que se utilize o valor do determinante correspondente à matriz X'X com as variáveis normalizadas.

O "$R^2$ delete" pode ser utilizado também como uma medida e se constitui no cômputo do $R_i^2$ quando se elimina uma variável independente da equação, alternadamente. Compara-se então o $R^2$ correspondente ao modelo completo com os diferentes $R_i^2$. Em caso de multicolinearidade, a diferença entre o $R^2$ e o mais alto dos $R_i^2$, será pequena. O "$R^2$ delete" é descrito por KMENTA (1971, p. 389).

O gráfico de cumeeira ("ridge trace") é uma figura traçada no espaço bi-dimensional onde se pode visualizar os parâmetros mais afetados pela multicolinearidade. É proposto por HOERL e KENNARD (1970 a) e será analisado nos próximos capítulos.

## 2.3. Estudos Relativos ao Problema

Analisando as implicações da multicolinearidade exata, no caso específico de estimação de funções de produção Cobb-Douglas, DOLL (1974), considera que neste tipo de função é coerente se esperar encontrar multicolinearidade elevada entre os insumos da amostra, como um problema que surge com o modelo. Isso é devido às pressuposições com relação à estimação dessas funções. Para modelos com dados de corte transversal, e com propriedades com uma atividade única, essas pressuposições seriam:

A - Todas as firmas possuem a mesma função de produção;

B - Cada firma maximiza seu lucro;

C - Os preços dos insumos e do produto são os mesmos para todas as firmas.

O autor afirma que a presença da multicolinearidade neste caso serve como uma verificação do modelo econômico, já que surge como conseqüência daquelas pressuposições básicas. Se houvesse retornos decrescentes à escala, com todas as propriedades no equilíbrio de longo prazo, todas elas usariam tanto a mesma quantidade de insumos, como na mesma proporção. Se os retornos à escala fossem crescentes ou unitários, as propriedades precisariam estar em qualquer ponto da trajetória de expansão, que será determinado em cada caso por fatores exógenos. Mas o autor considera que seria possível medir retornos à escala mesmo quando os insumos são perfeitamente correlacionados. Devido às pressuposições do modelo de Cobb-Douglas, este não pode ser estimado pelo método de mínimos quadrados ordinários, e mesmo quando ocorrer multicolinearidade quase exata, haveria obtenção de estimativas imprecisas. O método de mínimos quadrados com restrição deve ser utilizado então. DOLL sugere a estimação conjunta de $(\beta_1 + \beta_2)$, quando $X_1$ e $X_2$ são correlacionados. Partindo da hipótese da racionalidade do agricultor, obtém-se a trajetória de expansão da firma (que pode ser considerada como a restrição de "identificação"). Essa curva de otimização, juntamente com o valor estimado de $(\beta_1 + \beta_2)$, permite estimar $\beta_1$ e $\beta_2$. Essa estimação é distinta para firmas em equilíbrio de longo ou curto prazo, pois a maximização do lucro difere nestes casos. Porém os estudos de funções de produção visam basicamente verificar o estágio de produção em que os agricultores se encontram; já isto não é possível pelo método proposto por DOLL, que pressupõe que todos os agricultores sejam racionais, no sentido de objetivarem a maximização do lucro, para a dedução das estimativas.

Com relação à técnica de agregação de insumos dentro de categorias com insumos complementares, visando eliminar a correlação que geralmente aparece entre os insumos, DOLL cita que ela não se compatibiliza com a lógica do modelo Cobb-Douglas, o qual "impõe a propriedade de que todas as categorias de insumos precisam ser substitutos com elasticidade de substituição igual a um". Considera, então, que seria preferível estimar o modelo com restrição, da maneira como foi sugerida por êle.

Outro problema que DOLL julga importante, e que já foi considerado aqui, constitui-se na especificação incorreta do modelo de regressão, quando existe multicolinearidade. Êle afirma que existem dois tipos de erros de especificação. Primeiro, um tipo de erro que pode conduzir a interpretações falsas, que consiste em incluir variáveis desnecessárias. O outro seria a omissão de uma variável importante, quando correlacionada com outra. Seria possível omitir uma das variáveis correlaciondas no modelo estimado, porém dever-se-ia incluí-la na especificação teórica do modelo, para evitar erros de interpretação. Mas estimar um valor agregado do coeficiente de regressão correspondente aos insumos correlacionados não é conveniente, pois não possibilita analisar suas influências isoladas.

KMENTA (1971), sugere alguns métodos para resolver a multicolinearidade, ainda que de certa forma incompletamente. Uma sugestão é utilizar informação prévia sobre os coeficientes de regressão, de fontes diferentes daquelas da amostra. Por exemplo, num certo estudo com dados de série temporal, em que determinada variável esteja correlacionada com outra ou com uma combinação linear de outras variáveis, pode-se utilizar de estimativa de seu coeficiente a partir de outra amostra com dados de corte transversal, desde que haja tempo disponível e seu custo não seja elevado (KMENTA, 1971, p. 385).

Pode-se também aumentar o tamanho da amostra e com isto conseguir diminuição nas variâncias das estimativas, devido aos incrementos nos somatórios de quadrados ($\Sigma x_i^2$), estes por sua vez como consequência do aumento das observações (desde que estas sejam diferentes do valor médio). Mas os incrementos em $\Sigma x_i^2$ devem compensar as elevações nos coeficientes de determinação simples que poderão ocorrer(KMENTA, 1971, p.391).

Destinado à resolução de problemas econômicos cujos dados apresentam multicolinearidade, foi proposto o método de regressão de cumeeira ("ridge regression") por HOERL e KENNARD (1970 a). Equivale a uma estimação de $\beta$ com restrição. Algumas propriedades que tornam este método mais vantajoso que o de mínimos quadrados ordinários serão descritas no capítulo referente à metodologia. É introduzido também o gráfico de cumeeira ("ridge trace"), para representar os efeitos da não ortogonalidade. HOERL e KENNARD (1970 b) ilustram o uso deste gráfico com dois exemplos empíricos: um modelo de regressão com dez variáveis independentes, e outro com treze. No desenvolvimento dos modelos pelo método proposto, evidencia-se a instabilidade acentuada e superestimação dos parâmetros devido à existência de multicolinearidade.

BROWN e BEATTIE (1975), analisam as implicações da multicolinearidade, em especial para as funções de produção. Eles comentam que o procedimento errôneo de se cancelar variáveis no modelo original tem sido geralmente usado quando surgem problemas de multicolinearidade. Isto porém conduz ao viés de especificação. Uma alternativa melhor seria o uso do método de regressão de cumeeira ("ridge regression"), proposto por HOERL e KENNARD (1970 a). Porém os autores argumentam em seu trabalho que este método fornecerá melhores resultados quando aplicado a certas funções de produção do que a outro tipo de funções. Êles provam, através de um teorema desenvolvido pelos autores em colaboração com T.D. WALLACE, que no caso das funções de produção o viés esperado nas estimativas de cumeeira ("ridge estimates") será relativamente pequeno. Através da fórmula do viés, deduzida neste teorema, pode-se concluir que este será relativamente pequeno se as intercorrelações entre as variáveis são positivas e os valores verdadeiros dos coeficientes de regressão forem todos do mesmo sinal e de magnitude aproximadamente igual.

Portanto, o conhecimento anterior sobre os coeficientes de regressão permite inferir que o viés nas estimativas de funções de produção será provavelmente menor que naquelas de funções de demanda. Desse modo, é enfatizado o uso do método de regressão de cumeeira para estimar funções de produção.

Com relação à esperança do quadrado do desvio (mean square error), entende-se que os autores são favoráveis a um pequeno viés, se é possível haver diminuição da variância, até quando a esperança do quadrado do desvio para determinado valor de k[2/] for menor ou igual à de mínimos quadrados ordinários, já que existe sempre uma estimativa de cumeeira

---

2/ k é o valor do acréscimo (pequeno e positivo) aos elementos da diagonal da matriz X'X na forma de correlações, para obtenção do estimador de cumeeira.

com menor EQD (esperança do quadrado do desvio) que aquela corresponden te à de mínimos quadrados.

A grande redução na variância e EQD das estimativas de cumeeira em relação às de mínimos quadrados, é o argumento usado pelos autores para afirmar que para muitos problemas não ortogonais o método de mínimos quadrados estaria obsoleto.

Para escolha da melhor regressão pelo método de cumeeira (ou de um valor de k que determinaria essa regressão), é sugerida uma regra de decisão, que não é subjetiva como a do gráfico de cumeeira proposta por HOERL e KENNARD (1970 a, p. 65). No entanto os autores chegam à conclusão de que os resultados obtidos por esta regra são similares aos do gráfico de cumeeira. Após escolhidas as estimativas de acordo com o nível de k a partir do qual elas estabilizam, pode-se testar se elas possuem menor EQD que as de mínimos quadrados. Para isso utilizar-se-ia o teste F não-central de Wallace, como é sugerido por BROWN e BEATTIE (1975, p. 30). Êles ilustram o desenvolvimento desse teste aplicando-o a um exemplo com dados de irrigação na região do pacífico central (Estados Unidos), concluindo que não há razões para dizer que o método de regressão de cumeeira não deva ser aplicado nesse caso.

McDONALD e GALARNEAU (1975) propõem em seu trabalho dois métodos analíticos para especificar k, como alternativos ao método proposto por HOERL e KENNARD (1970 a), que é o gráfico de cumeeira. A escolha de k por estes métodos é objetiva, resultando num único valor. O método do gráfico de cumeeira visa encontrar uma melhor (quanto à esperança do quadrado do desvio) estimativa de $\beta_i^*$ (única), selecionando um único valor de k seguindo alguns critérios subjetivos, podendo portanto variar segundo o pesquisador. Já os métodos propostos por McDONALD e GALARNEAU visam estabelecer algumas regras matemáticas segundo as quais se determina um único estimador de cumeeira que seria o mesmo independentemente do pesquisador.

Os autores analisaram os seguintes métodos alternativos de escolha de k:

a. $k = 0$ sempre, o que corresponde a usar o método de mínimos quadrados.

b. Definindo:

$$Q = \hat{\beta}'\hat{\beta} - s^2(tr(X'X)^{-1}$$

escolha k tal que $(\hat{\beta}^*)'(\hat{\beta}^*) = Q$ se $Q > 0$; escolha $k = 0$ caso contrário[3/].

c. Escolha k tal que $(\hat{\beta}^*)'(\hat{\beta}^*) = Q$ se $Q > 0$; escolha $k = \infty$ caso contrário.

Dessa forma, para $Q > 0$, a especificação de k fica condicionada ao valor da estimativa do quadrado do comprimento do vetor de coeficientes de regressão.

---

3/ $\hat{\beta}^*$ é o estimador do vetor de coeficientes de regressão obtido pelo método de regressão de cumeeira.

## 2.4. Pesquisas que evidenciam a existência de multicolinearidade em dados de funções de demanda e produção

São bastante frequentes casos em que a estimação de funções em economia fica prejudicada pela presença de variáveis correlacionadas no modelo. Analisando algumas pesquisas sobre funções de demanda e produção, em que a multicolinearidade se mostrou bastante grave, foi constatado que as estimativas dos parâmetros se mostravam "instáveis", dificultando as conclusões. Em todos os trabalhos a presença de coeficientes altamente afetados é geral, chegando a impedir, em alguns casos, a obtenção de conclusões econômicas válidas.

Em funções de produção a variável área cultivada apresenta-se em geral correlacionada com as demais. Além disso, a homogeneidade na tecnologia adotada pelos agricultores numa determinada região é causa das dificuldades inerentes à multicolinearidade. O que tem sido feito para contornar o problema é uma transformação das variáveis para que o modelo se apresente na forma de "produtividade". Mas isto conduz geralmente a um baixo poder de "explicação" pelos fatores, isto é, um baixo valor do coeficiente de determinação.

Os estudos de demanda que se utilizam de séries cronológicas de dados também apresentam o problema. As variáveis que normalmente são utilizadas na estimação de funções desta natureza (renda, urbanização, preços) são em geral bastante correlacionadas.

Os comentários a seguir referem-se às pesquisas de SOBRAL

(1973) e BISERRA (1971), onde a multicolinearidade afetou bastante a estimativa dos parâmetros.

SOBRAL (1973), desenvolveu pesquisa sobre demanda de alimentos no Brasil, utilizando dados anuais secundários referentes ao período 1950/70. Para cada um dos produtos estudados (arroz, feijão, batatinha, carne e leite), foram ajustadas várias equações, através do método de mínimos quadrados ordinários. Destas, o autor selecionou as "melhores" e obteve coeficientes de elasticidade de demanda. Êle formulou seus modelos como uniequacionais, em primeiro lugar por facilidade de cálculo. Mas justifica através da indicação de WAUGH (1973, p. 16) de que FOX chegou aos mesmos resultados, na prática, com o uso do método de mínimos quadrados ordinários e equações simultâneas para produtos agrícolas.

Os modelos utilizados se apresentaram nas formas bilogarítmica, semilogarítmica, logarítmica-inversa, diferenças dos valores originais das variáveis e diferenças dos logaritmos dos valores das variáveis. A bilogarítmica apresenta o problema de possuir elasticidade constante em todos os pontos da curva, o que não se coaduna com a realidade no caso da elasticidade-renda da demanda de alimentos. Ao utilizar modelos com as primeiras diferenças, SOBRAL obteve baixos valores de $R^2$, auto-correlação nos resíduos e baixa significância dos coeficientes estimados. As formas semilogarítmica e logarítmica-inversa foram as que se ajustaram melhor, e portanto as que deram origem aos coe ficientes de elasticidade nos quais o autor se baseou para as conclusões econômicas. As variáveis incluidas na especificação dos modelos são o tempo (variável tendência), renda, urbanização, preços dos produtos e salário mínimo. O autor constatou existência de multicolinearidade elevada entre o tempo, a renda e a urbanização. Supondo estar o fator tempo incluido na influência da urbanização, cancelou-o, ajustando as equações com as demais variáveis. Apesar disso, não foram obtidos resultados satisfatórios, e as estimativas dos coeficientes de regressão mostraram-se "instáveis", algumas com sinais contrários aos esperados e com valores absolutos muito elevados. Principalmente no caso da carne e do leite, não foi possível obter conclusões válidas sobre os coeficientes de elasticidade da demanda.

BISERRA (1971), analisou as relações fator-produto na cultura do milho, com dados referentes ao ano agrícola 1969/70, obtidos por amostragem aleatória, diretamente de produtores dos municípios de Jardinópolis e Guaíra, compreendidos na Divisão Regional Agrícola (DIRA) de Ribeirão Preto. Foram analisadas 60 propriedades em Jardinópolis e 64 em Guaíra. Foram experimentados modelos nas formas linear e Cobb-Douglas, e consideradas oito variáveis na função estimada para Jardinópolis (área cultivada, trabalho humano, investimento em fertilizantes, investimento em máquinas e implementos, investimento em sementes, "despesas de custeio" anos de educação do produtor e contatos com técnico extensionista); em Guaíra, utilizou-se destas mesmas variáveis, exceto extensão rural.

O autor notou a existência de multicolinearidade entre as variáveis, principalmente do fator área cultivada com os demais. Êle interpreta o problema como resultante de tecnologia uniforme entre os produtores de milho nas regiões estudadas, e no caso da variável trabalho humano como conseqüência de dificuldades associadas ao cômputo da mesma. Visando obter melhores estimativas, a função de Cobb-Douglas foi ajustada aos dados de produção e quantidades dos fatores por unidade de área. Mas as variáveis assim definidas passaram a "explicar" muito pouco da variação no rendimento do milho, considerando-se os baixos coeficientes de determinação múltipla resultantes (os maiores coeficientes obtidos para Jardinópolis e Guaíra foram, respectivamente, 0,37 e 0,15).

A solução encontrada por BISERRA (1971, p. 42), foi a eliminação de algumas variáveis nas várias equações, estimadas pelo método de mínimos quadrados ordinários. O autor afirma ter se baseado em critério proposto por HEADY e DILLON (1966, p. 136), "segundo o qual quando o coeficiente de correlação simples é próximo da unidade (igual ou maior que 0,80), uma das variáveis independentes deve ser eliminada, pois, só assim "melhores" estimativas para os parâmetros das variáveis mais relevantes poderão ser obtidas." No entanto esta colocação de BISERRA reflete uma interpretação errada de HEADY e DILLON, que sugerem com muita cautela essa eliminação. Considerando-se que ambos os fatores correlacionados possam ser importantes à especificação do modelo, este não é um método correto para contornar a multicolinearidade existente.

# 3. METODOLOGIA

## 3.1. Material

No estudo relativo a funções de demanda, foram utilizados no presente trabalho, os mesmos dados de SOBRAL (1973), os quais referem-se ao Brasil e incluem informações anuais que abrangem o período de 1950 a 1970. O autor, além de apresentar tabela com os dados (ver SOBRAL, 1973, Apêndice 1, p. 108-113), fornece explicação detalhada sobre as fontes em que foram obtidos e o critério utilizado em seu processamento (ver SOBRAL, 1973, Metodologia, p. 53-59).

Na parte dedicada à análise de produção da cultura do milho, foram utilizados dados sobre o ano agrícola 1969/70, obtidos através de entrevistas diretas com agricultores, selecionados por amostragem, nos municípios de Jardinópolis (60 propriedades) e Guaíra (64 propriedades), pertencentes à Divisão Regional Agrícola (DIRA) de Ribeirão

Preto. O conjunto dos dados é apresentado em BISERRA (1971, Apêndices 1 e 2, p. 92-99).

## 3.2. Método

### 3.2.1. O Método de Regressão de Cumeeira[4/]

#### 3.2.1.1. Introdução

Considere-se o modelo de regressão múltipla

$$y = X\beta + u$$

onde y é o vetor de observações, X é a matriz de dimensões n x p, dos valores das variáveis independentes, β é o vetor dos parâmetros a serem estimados e u é o vetor dos erros. Admite-se que $E(u) = 0$ e que $E(uu') = I\sigma_u^2$, onde $\sigma_u^2$ é a variância dos erros.

Tendo em vista evitar a influência das unidades de medida das variáveis, utilizam-se, nos cálculos, as variáveis transformadas (normalizadas), de maneira que X'X e X'y sejam matrizes de correlações simples. Assim, se $Z_j$ é o valor da j-ésima observação da variável

---

4/ As deduções das fórmulas apresentadas neste capítulo, como também explicações complementares do método, encontram-se no Apêndice A.

dependente, o valor do j-ésimo elemento do vetor y é

$$Y_j = \frac{Z_j - \overline{Z}}{\sqrt{\Sigma(Z_j - \overline{Z})^2}}$$

onde $\overline{Z}$ é a média aritmética dos valores de $Z_j$ e $j = 1, \ldots, n$.

Analogamente, se $V_{ij}$ é a j-ésima observação da i-ésima variável independente, tem-se que

$$X_{ij} = \frac{V_{ij} - \overline{V}_i}{\sqrt{\Sigma\ (V_{ij} - \overline{V}_i)^2}}$$

com $i = 1, \ldots, p$ e $j = 1, \ldots, n$.

Deve-se distinguir dois tipos de coeficientes de regressão. Sejam $\beta_i$ os coeficientes para a equação com as variáveis normalizadas e $\theta_i$ os coeficientes para a equação com as variáveis na forma original.

Se $\hat{\theta}_i$ e $\hat{\beta}_i$ são os estimadores de $\theta_i$ e $\beta_i$, respectivamente, pode-se verificar que

$$\hat{\theta}_i = \hat{\beta}_i \frac{\sqrt{\Sigma(Z_j - \overline{Z})^2}}{\sqrt{\Sigma\ (V_{ij} - \overline{V}_i)^2}}$$

Sabe-se que o estimador de mínimos quadrados ordinários de $\beta$ é

$$\hat{\beta} = (X'X)^{-1}X'y$$

Então, como o método de regressão de cumeeira tem por objetivo diminuir a multicolinearidade, o estimador de cumeeira é obtido aumentando-se os elementos da diagonal principal da matriz X'X, na forma de matriz-correlação, por pequenas quantidades positivas (k), ou seja, o estimador de cumeeira é

$$\hat{\beta}^* = (X'X + kI_p)^{-1}X'y$$

Como X'X é uma matriz de correlações simples, só se consideram, geralmente, valores de k para o intervalo $0 \leq k \leq 1$.

Quando as variáveis estão na forma original, tem-se:

$$\hat{\theta}^* = (V'V + \Lambda)^{-1}V'z$$

onde z é o vetor das variáveis dependentes centradas, V a matriz das variáveis independentes também centradas e $\Lambda$ a matriz dos acréscimos a serem efetuadas aos elementos da diagonal de V'V.

O termo constante é:

$$\hat{\alpha}^* = \overline{Z} - \Sigma\hat{\theta}_i^* \overline{V}_i$$

Obviamente $\hat{\beta}^* = \hat{\beta}$ se $k = 0$.

O valor a ser escolhido para k é arbitrário, porém deve ser pequeno e positivo. O ideal consiste em se utilizar um k para o qual a esperança do quadrado do desvio seja mínima.

### 3.2.1.2. Análise das Estimativas pelo Gráfico de Cumeeira

O gráfico de cumeeira é uma representação geométrica mostrando como os valores de $\beta_i^*(k)$ variam em função dos valores de k, comumente no intervalo $(0,1)$. A utilização de k dentro deste intervalo decorre de se dispor das variáveis normalizadas, livres da influência das unidades de medida específicas para cada uma.

Utilizando-se este gráfico, é possível se analisar os efeitos da multicolinearidade sobre as estimativas dos parâmetros obtidas a partir de um grupo de dados, através da visualização das modificações nos valores dessas estimativas. Mas a principal finalidade da construção de tal gráfico é que êle é a base para a escolha do nível de k com o qual se obterá a regressão estimada. Essa escolha deve recair num k a partir do qual as estimativas dos parâmetros sejam relativamente estáveis.

### 3.2.1.3. Soma de Quadrados do Resíduo para o Estimador de Cumeeira

A equação correspondente à soma de quadrados do resíduo para o estimador de cumeeira é

$$S^*(k) = (y - X\hat{\beta}^*)'(y - X\hat{\beta}^*)$$

Para as variáveis ~~normalizadas~~ ela pode ser expressa da seguinte forma

$$S^*(k) = 1 - \hat{\beta}^{*\prime}X'y - k\left[\Sigma(\hat{\beta}_i^*)^2\right]$$

onde 1 corresponde à soma de quadrados total e $\hat{\beta}^{*\prime}X'y$ à soma de quadrados da regressão.

Quando as variáveis estão na forma original a expressão fica

$$S^*(\Lambda) = \Sigma z_j^2 - \Sigma \hat{z}_j^2 - (\hat{\theta}^*)'\Lambda\,(\hat{\theta}^*)$$

onde $z_j$ são os valores da variável dependente original centrada, $\Sigma z_j^2$ é a soma de quadrados total e $\Sigma\hat{z}_j^2$ é a soma de quadrados da regressão.

A relação entre as somas de quadrados do resíduo para as variáveis originais $[S^*(\Lambda)]$ e normalizadas $[S^*(k)]$ é

$$S^*(\Lambda) = \left[S^*(k)\right] z'z$$

O método de estimação de mínimos quadrados baseia-se na minimização da soma de quadrados do resíduo da regressão, para a obtenção das estimativas dos parâmetros. Estas são as melhores estimativas lineares não viesadas. Por outro lado, o método de regressão de cumeeira mostra a "sensibilidade" da solução $\hat{\beta}^* = (X'X + kI)^{-1}X'y$ para o critério de otimização $S^*(k) = (y - X\hat{\beta}^*)'(y - X\hat{\beta}^*)$, o que o método de mínimos quadrados não faz. Ou seja, este método se baseia em, fixado o valor da soma de quadrados do resíduo da regressão, minimizar o quadrado do comprimento do vetor das estimativas dos parâmetros. Este procedimento vai conduzir à matriz X'X com os elementos da diagonal aumentados de valores k, que é como foi definido o estimador de cumeeira.

Dessa forma $\hat{\beta}^*$ caracteriza-se em possuir comprimento menor do que $\hat{\beta}$. Define-se o comprimento de um vetor x como sendo a raiz quadrada do produto deste pela sua transposta.

Então[5/]

$$\hat{\beta}^{*\prime}\hat{\beta}^* < \hat{\beta}'\hat{\beta}$$

para $k \neq 0$.

#### 3.2.1.4. A distribuição de $\hat{\beta}^*$

Supondo que os erros possuem distribuição normal, tendo

---

5/ Ver HOERL e KENNARD (1970 a, p. 57).

$E(u) = 0$ e $E(uu') = I\sigma_u^2$ então $\hat{\beta}$ possui distribuição multinormal e a aplicação dos testes de significância comumente usados em análise de regressão é válida. Porém, no caso dos estimadores de cumeeira, estes são tendenciosos e os valores de t e F calculados da forma usual não podem ser utilizados aqui como testes de significância. Os valores de t devem ser interpretados apenas como medidas descritivas que mostram em quanto o valor da estimativa do parâmetro é maior que o valor do respectivo desvio padrão estimado. O $R^2$ como medida descritiva continua válido no método de regressão de cumeeira.

### 3.2.1.5. Propriedades dos Estimadores de Cumeeira

A extensa utilização do método de mínimos quadrados tem sua justificativa pelas propriedades de seus estimadores, as quais são bem conhecidas.

Define-se esperança do quadrado do desvio (mean square error) de um vetor de parâmetros estimados $(\hat{\beta})$, como sendo

$$EQD(\hat{\beta}) = E\left[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)\right]$$

Comparativamente àquela estimativa originada pelo método de mínimos quadrados, os estimadores de cumeeira irão sempre produzir uma estimativa, com um valor apropriado de k, com menor esperança do quadrado do desvio, não obstante êles se apresentarem viesados.[6/]

Para o estimador de cumeeira a esperança do quadrado do desvio equivale a

$$EQD(\hat{\beta}^*) = k^2\beta'(X'X + kI)^{-2}\beta + \sigma_u^2 tr \left[(X'X + kI)^{-1}(X'X)(X'X + kI)^{-1}\right]$$

Ou então a

$$EQD(\hat{\beta}^*) = \left[\Sigma \text{ viés}(\hat{\beta}_i^*)\right]^2 + \Sigma \text{ var}(\hat{\beta}_i^*)$$

Para o estimador de mínimos quadrados,

$$EQD(\hat{\beta}) = \Sigma \text{var}(\hat{\beta}_i)$$

pois neste caso o viés é nulo.

Quando existe multicolinearidade, as variâncias serão muito grandes e portanto $EQD(\hat{\beta})$ também tenderá a ser relativamente grande.

A matriz de variâncias e covariâncias dos coeficientes de regressão de cumeeira corresponde a

$$\text{var-cov}(\hat{\beta}^*) = \sigma_u^2(X'X + kI)^{-1}(X'X)(X'X + kI)^{-1}$$

---

6/ Ver HOERL e KENNARD (1970 a, p. 62).

Os elementos da diagonal principal da matriz $(X'X + kI)^{-1}(X'X)(X'X + kI)^{-1}$ são denominados "fatores de inflação da variância dos parâmetros" (variance inflation factors) e são êles responsáveis pelo aumento excessivo nas variâncias dos coeficientes de regressão quando a multicolinearidade é alta. Por esta razão, estes elementos se constituem numa avaliação dos parâmetros mais problemáticos, sendo que nestes casos êles diminuem sensivelmente com aumentos em $k$. Convém lembrar novamente que a matriz $(X'X + kI)^{-1}(X'X)(X'X + kI)^{-1}$, da maneira como foi definida aqui, contém as variáveis na forma normalizada.

O viés originado no estimador de cumeeira é

$$E(\hat{\beta}^*) - \beta = -k(X'X + kI)^{-1}\beta$$

Denominando

$$\Sigma \text{var}(\hat{\beta}_i^*) = \gamma_1(k)$$

$$\Sigma(\text{viés } \hat{\beta}_i^*)^2 = \gamma_2(k)$$

e analisando as variações em $\gamma_1(k)$ e $\gamma_2(k)$ através da Figura 1, nota-se que, à medida em que $k$ vai aumentando, $\gamma_1(k)$ diminui rapidamente, enquanto $\gamma_2(k)$ aumenta. Somando essas duas influências, EQD decresce até certo valor de $k$, aumentando depois continuamente. É possível encontrar um $k$ em que a esperança do quadrado do desvio seja menor para o estimador de cumeeira do que para o de mínimos quadrados, embora para isso seja introduzido um viés. Portanto, existe um $k$ para o qual

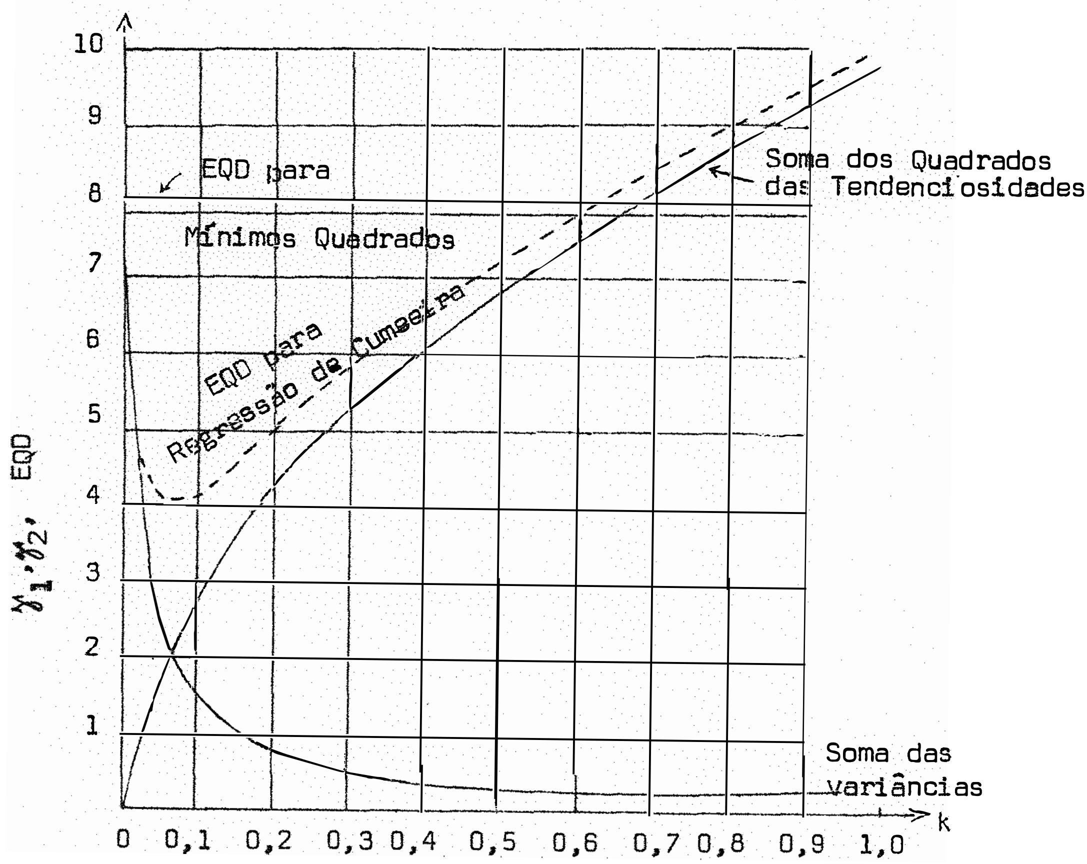


FIGURA 1. Representação das Relações entre a Soma das Variâncias, a Soma dos Quadrados das Tendenciosidades, a Esperança do Quadrado do Desvio e o valor de k.

Fonte: HOERL e KENNARD (1970 a, p.61).

$$EQD(\hat{\beta}^*) < EQD(\hat{\beta})$$

Considerando essa propriedade do método de regressão de cumeeira, pode-se dispor de um critério para sua utilização.

Percebe-se que o viés será zero para $k = 0$, e neste caso o estimador será o de mínimos quadrados, $\hat{\beta}$, com $EQD(\hat{\beta}) = \Sigma var(\hat{\beta})$. Como foi citado com relação aos "fatores de inflação da variância dos parâmetros", o método de regressão de cumeeira permite uma redução nas variâncias dos estimadores quando os dados são não ortogonais, porém será introduzido um viés.

A fórmula do viés pode ser escrita de outra forma[7/]:

$$E(\hat{\beta}_j^*) - \beta_j = \frac{kc_{jj}}{|W^{-1}|} \Big[\hat{b}^*_{j1}\beta_1 + \hat{b}^*_{j2}\beta_2 + \ldots + \hat{b}^*_{j.j-1}\beta_{j-1}$$

$$-1{,}0\beta_j + \hat{b}^*_{j.j+1}\beta_{j+1} + \ldots + \hat{b}^*_{jp}\beta_p\Big]$$

onde $\hat{b}^*_{ji}$ é a estimativa do coeficiente de regressão para a i-ésima variável na qual é feita uma regressão de $X_j$ sobre as $p - 1$ variáveis independentes remanescentes; $W^{-1} = (X'X + kI)$; $c_{jj}$ = o menor formado pelo cancelamento das j-ésima linha e j-ésima coluna da matriz $W^{-1}$.

---

7/ Esta fórmula é demonstrada por BROWN e BEATTIE (1975, p. 31).

É interessante verificar através da fórmula acima, sob que condições o viés será menor. Como em modelos econômicos é mais comum existirem intercorrelações positivas entre as variáveis, é mais útil analisar este caso. Com coeficientes de correlações positivos entre as variáveis, $\Sigma \hat{b}^*_{ji}$ usualmente é aproximadamente igual a um.[8/] Então, se todos os $\beta_i$ tiverem o mesmo sinal, bem como magnitude semelhante (considerando as variáveis normalizadas), então o viés será relativamente menor. Por outro lado, quando os $\beta_i$ possuem sinais e grandezas diferentes, o viés tenderá a ser maior.

Portanto espera-se obter, através do uso do método de regressão de cumeeira, resultados mais precisos para funções de produção, cujos coeficientes de regressão apresentam-se todos positivos (a menos que o uso dos insumos seja irracional), com magnitude aproximadamente igual. Já no caso das funções de demanda, os resultados originados por este método poderão não ser tão precisos.

Um teste proposto por TORO-VIZCARRONDO e WALLACE (1968, p. 565), pode ser usado para se avaliar quando as estimativas de cumeeira possuem EQO menor que as de mínimos quadrados ($H_0$). Calcula-se

$$\emptyset = \frac{SQR(\hat{\beta}^*) - SQR(\hat{\beta})}{SQR(\hat{\beta}) \;/\; (n-p)}$$

---

8/ Ver BROWN e BEATTIE (1975, p. 23).

onde $SQR(\hat{\beta}^*)$ = soma de quadrados do resíduo da regressão de cumeeira;

$SQR(\hat{\beta})$ = soma de quadrados do resíduo da regressão de mínimos quadrados;

$n - p$ = graus de liberdade da soma de quadrados do resíduo;

$\emptyset$ = valor do teste para EQD.

Ao nível de significância ($\alpha$) escolhido, compara-se $\emptyset$ com $\emptyset_\alpha$ da tabela (TORO-VIZCARRONDO e WALLACE, 1968, p. 566) para $n - p$ graus de liberdade. A regra de decisão consiste em obter $\emptyset < \emptyset_\alpha$, aceitando-se então a hipótese de que $\hat{\beta}^*$ possui menor EQD do que $\hat{\beta}$ aceita-se $H_o$).

Mas como os valores de $\emptyset_\alpha$ na tabela apresentada por TORO-VIZCARRONDO e WALLACE (1968) foram obtidos considerando uma restrição linear e o método de regressão de cumeeira implica numa restrição de segundo grau,[9/] a comparação indicada é apenas um indicador mais ou menos grosseiro para verificar se a estimativa de $\beta$ pelo método de cumeeira tem EQD menor do que a estimativa de mínimos quadrados.

---

9/ HOERL e KENNARD (1970 a, p. 59) mostram que o estimador de cumeeira para o vetor dos parâmetros é um vetor que, fixado o seu comprimento (o que é uma restrição de segundo grau), torna mínima a soma de quadrados dos desvios.

### 3.2.2. Modelos

Os símbolos utilizados na definição dos modelos são os seguintes:

$Y$ = variável dependente;

$a$ = intercepto da função;

$X_1, X_2, \ldots, X_n$ = variáveis independentes ou explicativas;

$b_1, b_2, \ldots, b_n$ = coeficientes de regressão.

#### 3.2.2.1. Formas da Função de Demanda

Julgou-se conveniente analisar os modelos nas formas semilogarítmica e logarítmica inversa, que se apresentaram mais adequadas, fornecendo os melhores resultados obtidos por SOBRAL (1973).

a) Função Semilogarítmica (Modelo I)

A função semilogarítmica assume a forma

$$\exp(Y) = \exp(a)\, X_1^{b_1}\, X_2^{b_2} \ldots X_n^{b_n}$$

que equivale à forma matemática linear:

$$Y = a + b_1 \ln X_1 + b_2 \ln X_2 + \ldots + b_n \ln X_n$$

A elasticidade de $Y$ em relação a $X_i$ é $\dfrac{b_i}{Y}$ .

b) Função Logarítmica inversa (Modelo II)

O modelo logarítmico inverso é

$$Y = \exp \left\{ a - \frac{b_1}{X_1} - \frac{b_2}{X_2} - \ldots - \frac{b_n}{X_n} \right\}$$

ou

$$\ln Y = a - b_1 \frac{1}{X_1} - b_2 \frac{1}{X_2} - \ldots - b_n \frac{1}{X_n}$$

Neste caso a elasticidade de Y em relação a $X_i$ é igual a $- \frac{b_i}{X_i}$.

Em todos os modelos a variável tendência é sempre mantida na forma original.

### 3.2.2.2. Forma da Função de Produção

A forma linear utilizada por BISERRA (1971) não será empregada em virtude de contrariar a lei dos rendimentos marginais decrescentes. Poder-se-ia experimentar a forma quadrática não usada por êle, porém os problemas de multicolinearidade que se está tentando eliminar seriam aumentados com a introdução de termos com produtos e quadrados. Além disso, não sendo pequeno o número de variáveis independentes, torna-se difícil a utilização do modelo nesta forma.

Ajustou-se a função de produção do milho na forma Cobb--Douglas, que é a seguinte

$$Y = a\, X_1^{b_1}\; X_2^{b_2} \ldots X_n^{b_n}$$

Por anamorfose, obtém-se

$$\ln Y = \ln a + b_1 \ln X_1 + b_2 \ln X_2 + \ldots + b_n \ln X_n$$

Deve-se notar que em todos os modelos são utilizados os logarítmos neperianos das variáveis.

### 3.2.3. Testes Estatísticos

A aplicação dos testes estatísticos às estimativas pelo método de regressão de cumeeira serve apenas como medida indicativa, como foi comentado na sub-seção 3.2.1.4., não se aplicando da maneira usual.

Além de utilizar o teste de Durbin-Watson, no caso das equações de demanda, para verificar a existência de autocorrelação nos resíduos, e o teste "F" de Snedecor, para testar a hipótese de nulidade de todos os coeficientes de regressão do modelo, será utilizado o teste "t" de Student para testar as seguintes hipóteses a respeito dos parâmetros:

a) Função de Demanda

O teste para as variáveis referentes aos preços dos produtos analisados é

$$H_o : b_i = 0$$
$$H_a : b_i < 0$$

Para os preços dos produtos substitutos e complementares o teste é

$$H_o : b_i = 0$$
$$H_a : b_i \neq 0$$

Teste para as variáveis renda, urbanização e tempo

$$H_o : b_i = 0$$
$$H_a : b_i > 0$$

b) Função de Produção

Neste caso será aplicado o teste

$$H_o : b_i = 0$$
$$H_a : b_i > 0$$

a todos os insumos incluidos na análise.

## 3.3. Definição das variáveis

A seguir são descritas as variáveis incluidas nos vários modelos.

### 3.3.1. Função de Demanda

A variável dependente é a quantidade anual aparente consumida em quilogramas por habitante (para arroz, batatinha, feijão e carne) ou em litros por habitante (para o leite), no Brasil.

As variáveis explicativas são:

a) variável tendência (tempo), correspondendo ao ano de 1950 o valor -10 e variando até 10 em 1970.

b) renda real anual disponível, em cruzeiros por habitante, no Brasil.

c) nível de urbanização, em porcentagem da população urbana em relação à total, anual, no Brasil.

d) preço médio anual real, a nível de varejo, em cruzeiros por quilograma (para arroz, batatinha, feijão, carne, carne seca e farinha) ou em cruzeiros por litro (para o leite), no Brasil.

e) o valor real do salário mínimo médio anual, em cruzeiros, no Estado da Guanabara.

### 3.3.2. Função de Produção

As variáveis são referentes ao ano agrícola de 1969/70, para a cultura do milho.

A variável dependente é o valor total da produção do milho, em cruzeiros.

As variáveis independentes são:

a) área cultivada com milho, em hectares.

b) trabalho humano, em dias-homens.

c) valor total gasto em fertilizantes, em cruzeiros.

d) valor total gasto com sementes, em cruzeiros.

e) valor total da depreciação, juros e conservação das máquinas e implementos, além do aluguel, quando este ocorreu, em cruzeiros; a parte da despesa total da propriedade, referente à cultura do milho, foi determinada de acordo com a participação dessa cultura no valor total da produção.

f) valor total das despesas de custeio, em cruzeiros (inclui gastos com combustíveis, lubrificantes, calcário e comercialização de insumos).

g) anos de educação formal completa do produtor.

h) número de contatos do produtor com o técnico extensionista.

# 4. RESULTADOS

Quase todos os resultados apresentados foram calculados para as variáveis normalizadas, seguindo a simbologia apresentada no capítulo 3, com exceção de alguns, para os quais utilizou-se as variáveis originais, cuja simbologia também é a mesma apresentada no capítulo 3.

Na escolha dos melhores coeficientes foram estabelecidos alguns critérios que determinaram, em conjunto, o valor de k (acréscimo) no gráfico de cumeeira. Pela ordem de importância, êles são:

a. o valor de k deve coincidir com o ponto a partir do qual as estimativas dos parâmetros são relativamente estáveis;

b. os sinais dos coeficientes, a partir deste ponto, deverão permanecer sem modificação, e de acordo com o estabelecido através da teoria econômica;

c. o valor da soma de quadrados dos resíduos da regressão não deverá ter sofrido aumento excessivo neste nível, em relação ao nível k = 0.

Esses critérios são recomendados por HOERL e KENNARD (1970 a, p. 64-65).

Foi experimentado o critério proposto por McDONALD e GALARNEAU (1975) para a escolha de k, mas verificou-se que este é inferior ao critério proposto por HOERL e KENNARD (1970 a).

## 4.1. Resultados para as Funções de Demanda

### 4.1.1. Multicolinearidade

As variáveis renda e urbanização se apresentaram correlacionadas entre si, e com as demais. É de se esperar que estas variáveis estejam relacionadas, pois sabe-se que o desenvolvimento econômico do Brasil esteve associado a um processo de industrialização que, por sua vez, levou a uma crescente urbanização.

A presença da multicolinearidade prejudicando as estimativas dos parâmetros é evidente em todos os produtos para os quais se estimaram os coeficientes de demanda, como se pode comprovar examinando os "fatores de inflação da variância das estimativas dos parâmetros" para o modelo II, nas tabelas 1 a 5. Os valores mais altos desses fatores foram encontrados para as estimativas dos coeficientes de regressão correspondentes ao tempo, à renda e à urbanização, considerando

Tabela 1. Estimativas dos "Fatores de Inflação da Variância das Estimativas dos Parâmetros" e Soma de Quadrados dos Resíduos, aos vários níveis de k, relativos à Função de Demanda do Arroz no Brasil, Modelo II.

| k | Fatores de Infl. da Var. das Est. dos Parâm. das variáveis | | | | | | | | $S^*(k)$ |
|---|---|---|---|---|---|---|---|---|---|
| | tempo | randa | urban. | preço arroz | preço feijão | preço batat. | preço far. | sal. mínimo | |
| 0,0 | 542,83 | 49,66 | 636,01 | 2,32 | 3,36 | 5,73 | 1,83 | 3,71 | 0,219 |
| 0,1 | 0,55 | 0,91 | 0,45 | 1,08 | 1,21 | 1,72 | 1,09 | 1,16 | 0,267 |
| 0,2 | 0,28 | 0,32 | 0,23 | 0,81 | 0,79 | 0,92 | 0,82 | 0,86 | 0,275 |
| 0,3 | 0,19 | 0,19 | 0,16 | 0,64 | 0,59 | 0,59 | 0,65 | 0,67 | 0,281 |
| 0,4 | 0,15 | 0,14 | 0,12 | 0,52 | 0,47 | 0,43 | 0,53 | 0,54 | 0,286 |
| 0,5 | 0,12 | 0,11 | 0,10 | 0,44 | 0,39 | 0,33 | 0,44 | 0,44 | 0,292 |
| 0,6 | 0,10 | 0,09 | 0,09 | 0,37 | 0,33 | 0,26 | 0,37 | 0,37 | 0,297 |
| 0,7 | 0,09 | 0,08 | 0,08 | 0,32 | 0,29 | 0,22 | 0,32 | 0,32 | 0,302 |
| 0,8 | 0,08 | 0,07 | 0,07 | 0,28 | 0,25 | 0,18 | 0,28 | 0,28 | 0,308 |
| 0,9 | 0,07 | 0,06 | 0,06 | 0,25 | 0,22 | 0,16 | 0,25 | 0,24 | 0,313 |
| 1,0 | 0,07 | 0,06 | 0,06 | 0,22 | 0,20 | 0,14 | 0,22 | 0,21 | 0,319 |

Tabela 2. Estimativas dos "Fatores de Inflação da Variância das Estimativas dos Parâmetros" e Soma de Quadrados dos Resíduos, aos vários níveis de k, relativos à Função de Demanda do Feijão no Brasil, Modelo II.

| k | Fatores de Infl. da Var. das Est. dos Par. das variáveis | | | | | | | | S*(k) |
|---|---|---|---|---|---|---|---|---|---|
| | tempo | renda | urban. | preço arroz | preço feijão | preço batat. | preço far. | salário mín. | |
| 0,0 | 542,83 | 49,66 | 636,01 | 2,32 | 3,36 | 5,73 | 1,83 | 3,71 | 0,396 |
| 0,1 | 0,55 | 0,91 | 0,45 | 1,08 | 1,21 | 1,72 | 1,09 | 1,16 | 0,449 |
| 0,2 | 0,28 | 0,32 | 0,23 | 0,81 | 0,79 | 0,92 | 0,82 | 0,86 | 0,476 |
| 0,3 | 0,19 | 0,19 | 0,16 | 0,64 | 0,59 | 0,59 | 0,65 | 0,67 | 0,499 |
| 0,4 | 0,15 | 0,14 | 0,12 | 0,52 | 0,47 | 0,43 | 0,53 | 0,54 | 0,519 |
| 0,5 | 0,12 | 0,11 | 0,10 | 0,44 | 0,39 | 0,33 | 0,44 | 0,44 | 0,536 |
| 0,6 | 0,10 | 0,09 | 0,09 | 0,37 | 0,33 | 0,26 | 0,37 | 0,37 | 0,551 |
| 0,7 | 0,09 | 0,08 | 0,08 | 0,32 | 0,29 | 0,22 | 0,32 | 0,32 | 0,565 |
| 0,8 | 0,08 | 0,07 | 0,07 | 0,28 | 0,25 | 0,18 | 0,28 | 0,28 | 0,577 |
| 0,9 | 0,07 | 0,06 | 0,06 | 0,25 | 0,22 | 0,16 | 0,25 | 0,24 | 0,589 |
| 1,0 | 0,07 | 0,06 | 0,06 | 0,22 | 0,20 | 0,14 | 0,22 | 0,21 | 0,599 |

Tabela 3. Estimativas dos "Fatores de Inflação da Variância das Estimativas dos Parâmetros" e Soma de Quadrados dos Resíduos, aos vários níveis de k, relativos à Função de Demanda da Batatinha no Brasil, Modelo II.

| k | Fatores de Infl. da Var. das Est. dos Parâm. das variáveis | | | | | | | $S^*(k)$ |
|---|---|---|---|---|---|---|---|---|
| | tempo | renda | urban. | preço arroz | preço feijão | preço batat. | salário mínimo | |
| 0,0 | 485,77 | 47,35 | 546,11 | 1,57 | 3,36 | 5,73 | 3,63 | 0,125 |
| 0,1 | 0,55 | 0,93 | 0,47 | 0,94 | 1,18 | 1,71 | 1,16 | 0,160 |
| 0,2 | 0,28 | 0,33 | 0,23 | 0,75 | 0,78 | 0,91 | 0,86 | 0,170 |
| 0,3 | 0,19 | 0,19 | 0,16 | 0,62 | 0,58 | 0,59 | 0,67 | 0,179 |
| 0,4 | 0,15 | 0,14 | 0,12 | 0,52 | 0,47 | 0,42 | 0,54 | 0,187 |
| 0,5 | 0,12 | 0,11 | 0,10 | 0,44 | 0,39 | 0,32 | 0,44 | 0,194 |
| 0,6 | 0,10 | 0,09 | 0,09 | 0,38 | 0,33 | 0,26 | 0,37 | 0,202 |
| 0,7 | 0,09 | 0,08 | 0,08 | 0,34 | 0,29 | 0,22 | 0,32 | 0,209 |
| 0,8 | 0,08 | 0,07 | 0,07 | 0,30 | 0,25 | 0,18 | 0,28 | 0,216 |
| 0,9 | 0,07 | 0,06 | 0,06 | 0,26 | 0,22 | 0,16 | 0,24 | 0,224 |
| 1,0 | 0,07 | 0,06 | 0,06 | 0,24 | 0,20 | 0,14 | 0,21 | 0,231 |

Tabela 4. Estimativas dos "Fatores de Inflação da Variância das Estimativas dos Parâmetros" e Soma de Quadrados dos Resíduos, aos vários níveis de k, relativos à Função de Demanda da Carne no Brasil, Modelo II.

| k | Fatores de Infl. da Var. dos Est. dos Parâm. das variáveis | | | | | | $S^*(k)$ |
|---|---|---|---|---|---|---|---|
| | tempo | renda | urban. | preço carne | preço carne seca | salário mínimo | |
| 0,0 | 682,64 | 41,96 | 702,75 | 14,02 | 6,42 | 3,26 | 0,064 |
| 0,1 | 0,56 | 1,15 | 0,52 | 1,96 | 1,91 | 1,21 | 0,093 |
| 0,2 | 0,31 | 0,42 | 0,26 | 0,90 | 0,99 | 0,87 | 0,112 |
| 0,3 | 0,22 | 0,24 | 0,18 | 0,53 | 0,62 | 0,67 | 0,126 |
| 0,4 | 0,17 | 0,17 | 0,13 | 0,36 | 0,43 | 0,53 | 0,137 |
| 0,5 | 0,14 | 0,13 | 0,11 | 0,26 | 0,32 | 0,44 | 0,146 |
| 0,6 | 0,11 | 0,10 | 0,09 | 0,20 | 0,25 | 0,37 | 0,154 |
| 0,7 | 0,10 | 0,09 | 0,08 | 0,16 | 0,20 | 0,32 | 0,161 |
| 0,8 | 0,09 | 0,07 | 0,07 | 0,13 | 0,17 | 0,27 | 0,168 |
| 0,9 | 0,08 | 0,06 | 0,06 | 0,11 | 0,14 | 0,24 | 0,175 |
| 1,0 | 0,07 | 0,06 | 0,06 | 0,10 | 0,12 | 0,21 | 0,181 |

Tabela 5. Estimativas dos "Fatores de Inflação da Variância das Estimativas dos Parâmetros" e Soma de Quadrados dos Resíduos, aos vários níveis de k, relativos à Função de Demanda do Leite no Brasil, Modelo II.

| k | Fatores de Infl. da Var. das Est. dos Parâm. das variáveis | | | | | | $S^*(k)$ |
|---|---|---|---|---|---|---|---|
| | tempo | renda | urban. | preço leite | preço carne | salário mínimo | |
| 0,0 | 691,83 | 42,00 | 719,92 | 2,48 | 10,41 | 3,25 | 0,026 |
| 0,1 | 0,59 | 1,15 | 0,53 | 1,48 | 1,95 | 1,17 | 0,054 |
| 0,2 | 0,31 | 0,41 | 0,26 | 1,00 | 1,00 | 0,84 | 0,056 |
| 0,3 | 0,22 | 0,23 | 0,17 | 0,73 | 0,62 | 0,65 | 0,059 |
| 0,4 | 0,16 | 0,16 | 0,13 | 0,55 | 0,43 | 0,52 | 0,062 |
| 0,5 | 0,13 | 0,12 | 0,11 | 0,44 | 0,31 | 0,44 | 0,065 |
| 0,6 | 0,11 | 0,10 | 0,09 | 0,36 | 0,24 | 0,37 | 0,069 |
| 0,7 | 0,09 | 0,08 | 0,08 | 0,30 | 0,20 | 0,32 | 0,073 |
| 0,8 | 0,08 | 0,07 | 0,07 | 0,25 | 0,16 | 0,28 | 0,077 |
| 0,9 | 0,07 | 0,06 | 0,06 | 0,21 | 0,14 | 0,24 | 0,082 |
| 1,0 | 0,07 | 0,06 | 0,06 | 0,19 | 0,12 | 0,22 | 0,087 |

k = 0 (mínimos quadrados). São os coeficientes dessas variáveis portanto, os que estão mais afetados pela multicolinearidade. Logo abaixo dessas, a variável que apresentou o "fator de inflação da variância do parâmetro" mais elevado é a "preço da carne". No caso do leite,por exemplo, em cujo ajustamento ela participa, o seu fator é de 10,41,bem abaixo do fator apresentado pela urbanização, que é 719,92, e cerca de 25% do fator correspondente à renda (ver tabela 5).

Uma propriedade do método de regressão de cumeeira é que, com crescentes níveis de k, processam-se diminuições nas estimativas dos "fatores de inflação da variância das estimativas dos parâmetros". Essa diminuição é mais intensa no intervalo $0 \leqslant k \leqslant 0,1$. Por exemplo, no caso do arroz, o fator relativo à urbanização passou de 636,01 a 0,45, ou seja, houve a sensível diminuição de 99,93% quando k variou de 0 a 0,1.

### 4.1.2. Gráficos de Cumeeira

Obtidos os valores dos parâmetros, aos diferentes níveis de k, foram traçados os gráficos apresentados nas Figuras 2 a 6, que correspondem aos gráficos de cumeeira dos modelos II, para os vários produtos analisados. Eles foram construídos com base nas modificações nos valores dos coeficientes de regressão, estimados para as variáveis normalizadas, com a finalidade de compará-los, o que não seria possível se se utilizasse os coeficientes que correspondem às variáveis originais.

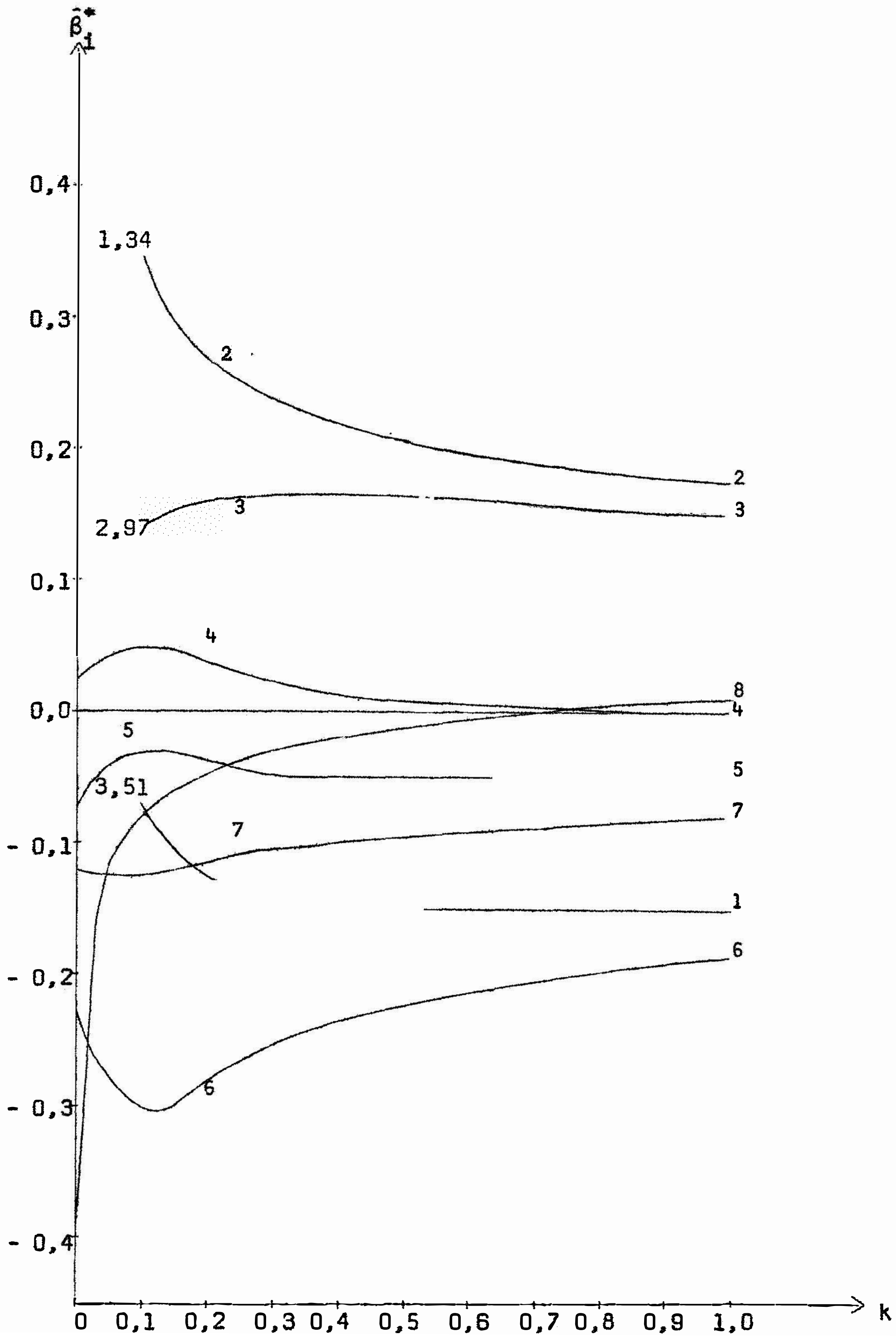


FIGURA 2 - Variação no valor dos coeficientes da regressão, com os níveis de k . Função de demanda de arroz, Modelo II, 8 variáveis independentes (Ver Tabela 6).

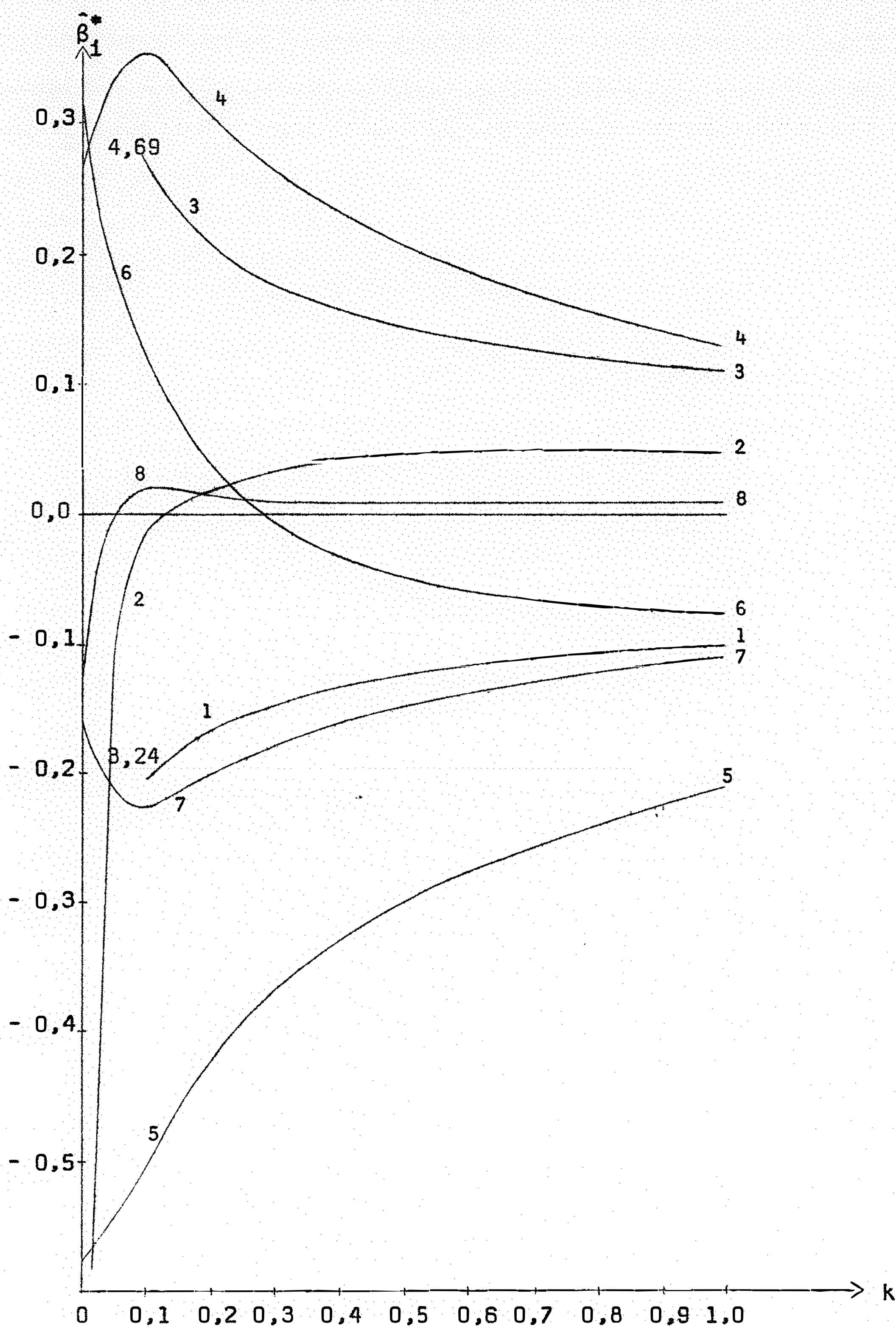


FIGURA 3 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função de demanda de feijão, Modelo II , 8 variáveis independentes (Ver Tabela 10).

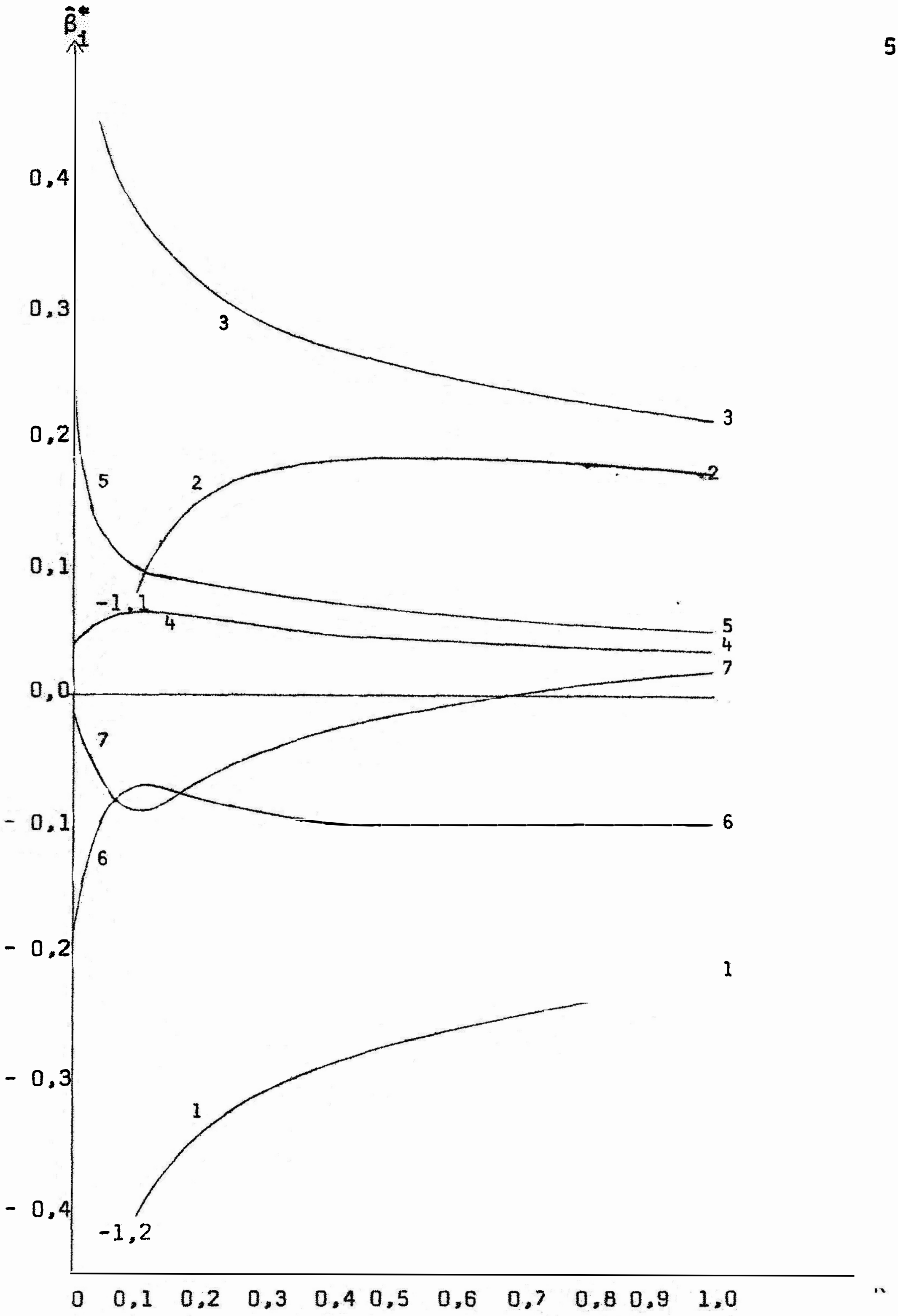


FIGURA 4 - Variação no valor dos coeficientes de regressão com os níveis de k . Função de demanda da batatinha. Modelo II , 7 variáveis independentes (Ver Tabela 14).

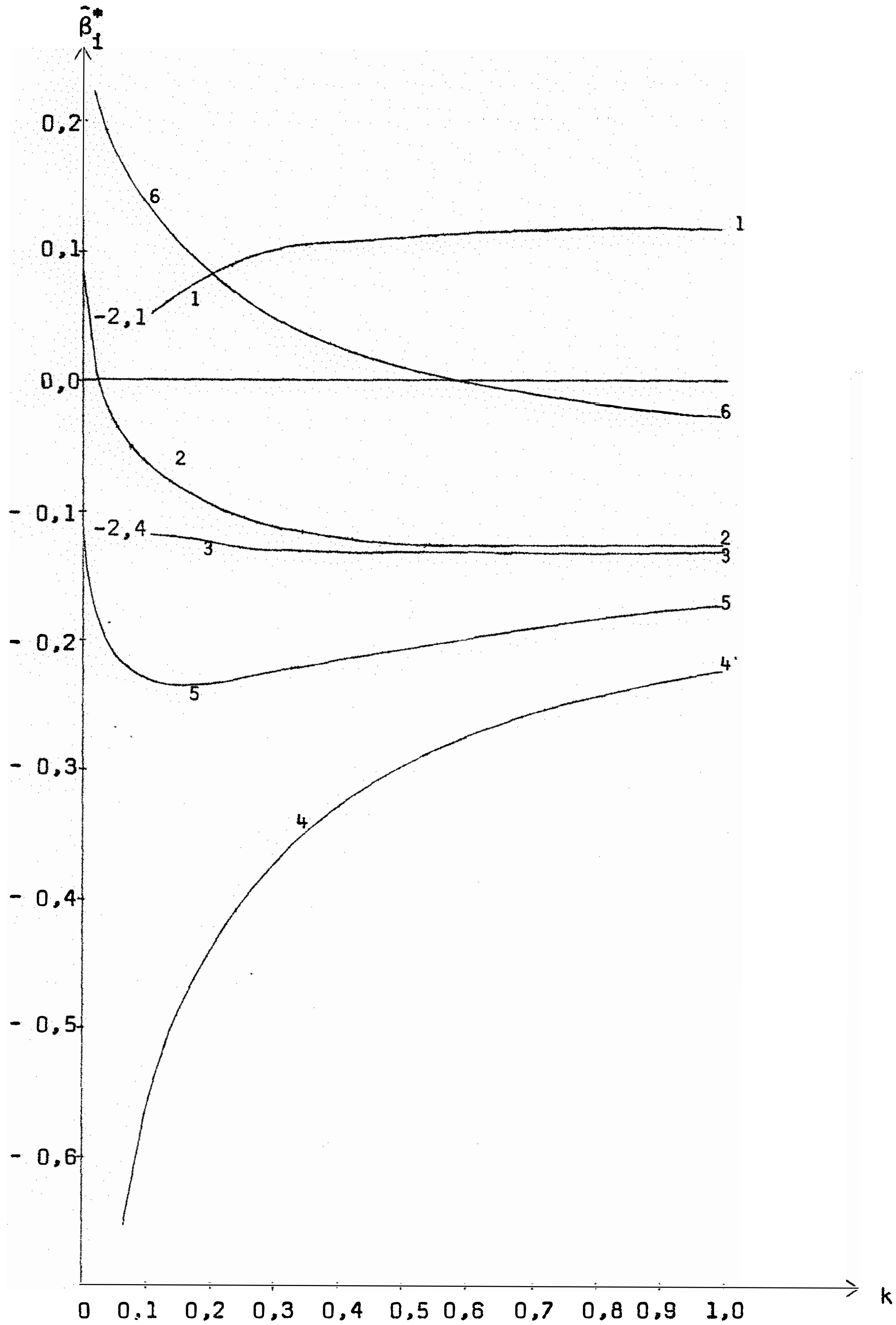


FIGURA 5 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função de demanda da carne, Modelo II , 6 variáveis independentes (Ver Tabela 18).

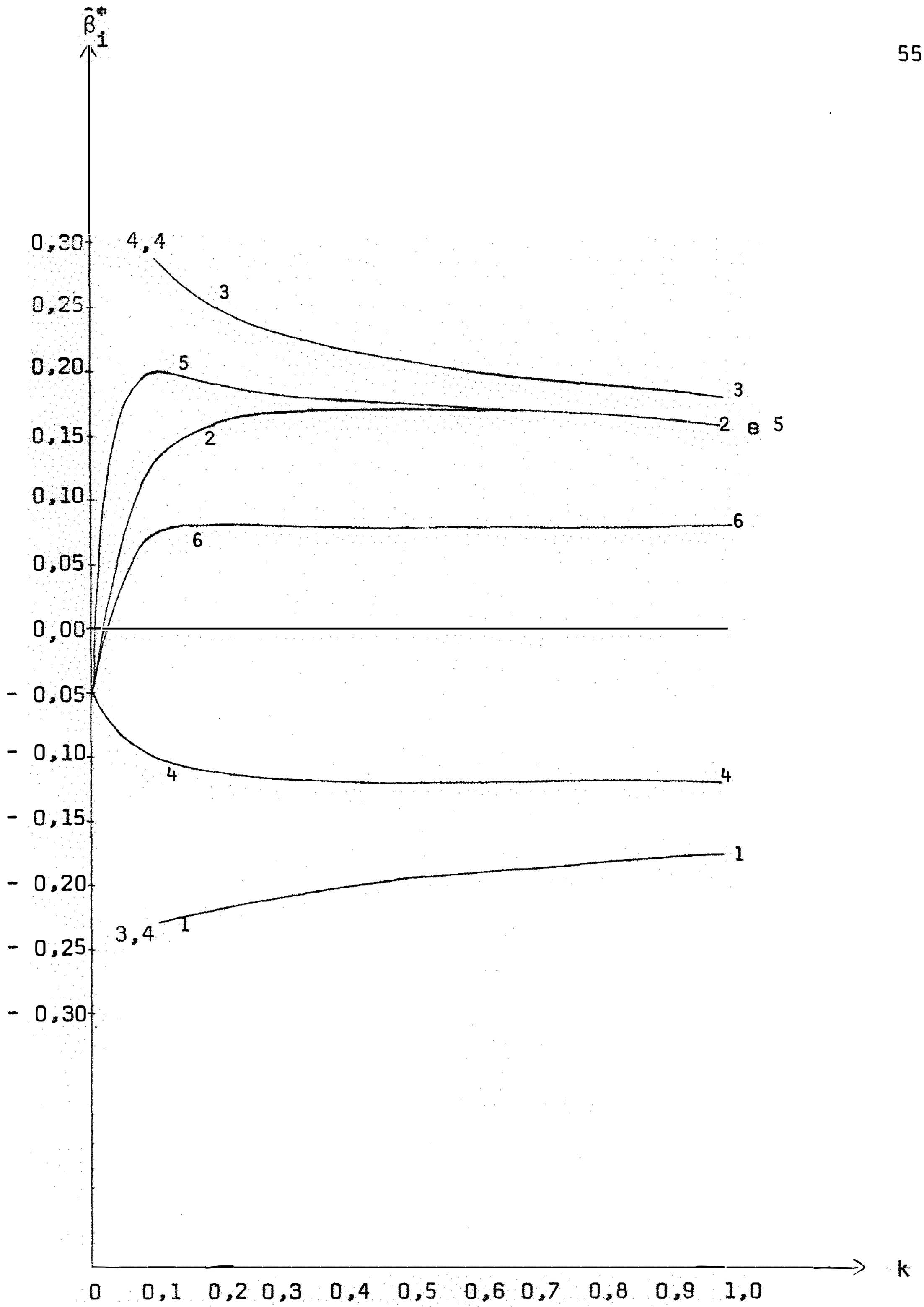


FIGURA 6 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função de demanda de leite, Modelo II , 6 variáveis independentes (Ver Tabela 22).

Constata-se, através da observação dos gráficos de cumeeira, que existe multicolinearidade envolvendo a grande maioria das variáveis incluídas. Os coeficientes de regressão apresentam um conjunto de alterações bastante sensíveis em seu valor absoluto, e mesmo alguns tiveram seu sinal mudado para o esperado, a um determinado nível de acréscimo (k). Pode-se perceber uma instabilidade muito grande nas estimativas dos parâmetros, para pequenos valores de k, em especial no intervalo $0 < k \leq 0,1$. Para um determinado nível de k, ocorre uma relativa estabilização dos parâmetros, e este valor foi utilizado na estimação da função de demanda. A estabilização dos vários sistemas no modelo II, de acordo com os critérios já mencionados, ocorre aos níveis: arroz, $k = 0,45$; feijão, $k = 0,20$; batatinha, $k = 0,35$; carne, $k = 0,35$; leite, $k = 0,20$.

Os parâmetros cujas alterações são maiores referem-se às variáveis tempo, renda e urbanização, que são os mais correlacionados entre si e com os outros fatores dos vários modelos. Ao nível $k = 0$ (mínimos quadrados), esses parâmetros mostraram-se com elevados valores absolutos, e em alguns casos com sinal contrário àquele que seria esperado pelo conhecimento anterior. As grandes modificações que se observa com pequenos aumentos em k estão retratando a existência de um nível alto de multicolinearidade.

SOBRAL (1973) escolheu como "melhores" regressões alguns ajustamentos a partir do modelo II (exceto para o leite, onde escolheu o modelo I), nos quais certas variáveis foram canceladas. Por exemplo, para o arroz, êle inclui apenas a renda, o preço do arroz e o preço da batatinha. Com estes melhores modelos foram construidos os gráficos de cumeeira que estão representados nas Figuras 7 a 11.

Exceto para o arroz, cujo sistema se mostra bastante estável, percebe-se que para os outros produtos há existência de multicolinearidade nos modelos ajustados por SOBRAL. Os coeficientes de regressão continuam instáveis após a eliminação de algumas variáveis do modelo. No caso do arroz, apesar dos parâmetros se apresentarem estáveis, tendo sido contornado o problema de multicolinearidade, apenas três variáveis foram incluidas no modelo. Não se utilizou a urbanização que, como se verá, pelo método de cumeeira demonstrou ser importante na explicação do consumo do arroz.

### 4.1.3. Estimativa dos Coeficientes de Elasticidade de Demanda

#### 4.1.3.1. Arroz

Na Tabela 6 se encontram as estimativas dos coeficientes de regressão e testes estatísticos correspondentes à melhor regressão ajustada ao modelo II, ao nível k = 0,45. As variáveis incluidas estão "explicando" apenas 63,77% na variação do consumo de arroz. As mais influentes nesse consumo são a renda, a urbanização e o preço da batatinha, cujos coeficientes de regressão são possivelmente significativos.

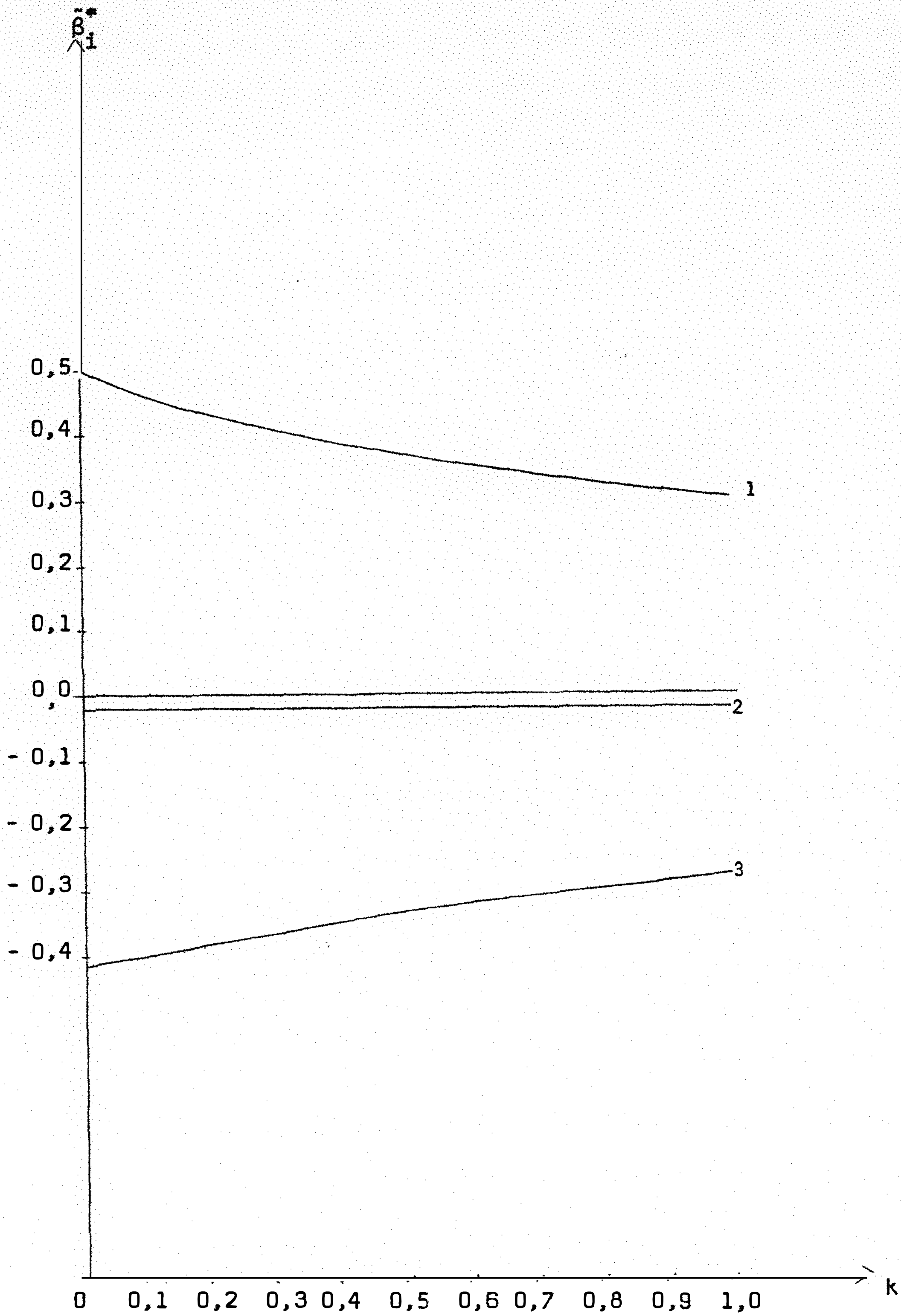


FIGURA 7 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função de demanda de arroz, Modelo II , 3 variáveis independentes: 1 - renda ; 2 - preço do arroz ; 3 - preço da batatinha.

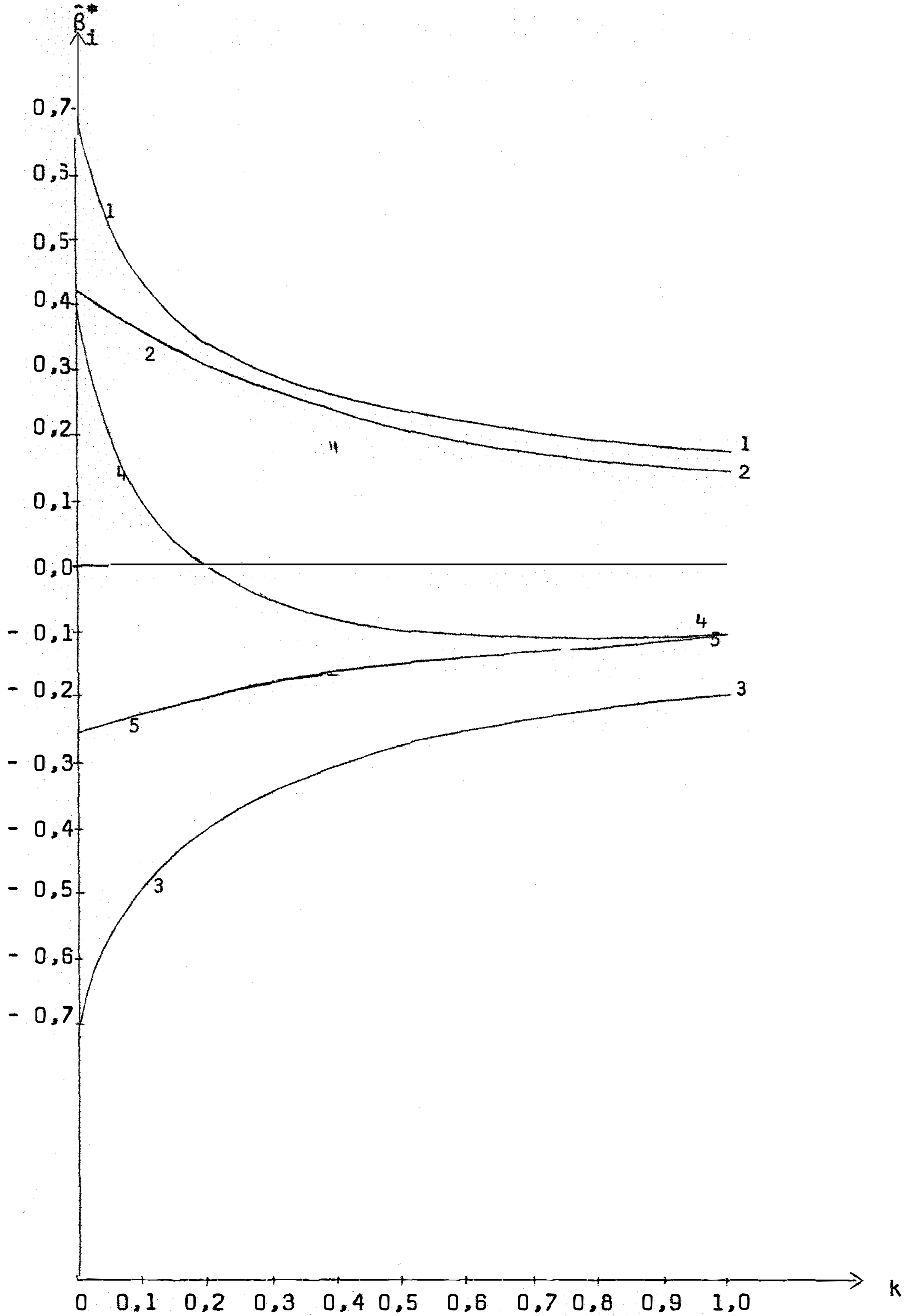


FIGURA 8 - Variação no valor dos coeficientes de regressão, com os níveis de $k$. Função de demanda do feijão. Mode II, 5 variáveis independentes: 1 - urbanização; 2 - preço do arroz ; 3 - preço do feijão ; 4 - preço da batatinha ; 5 - preço da farinha.

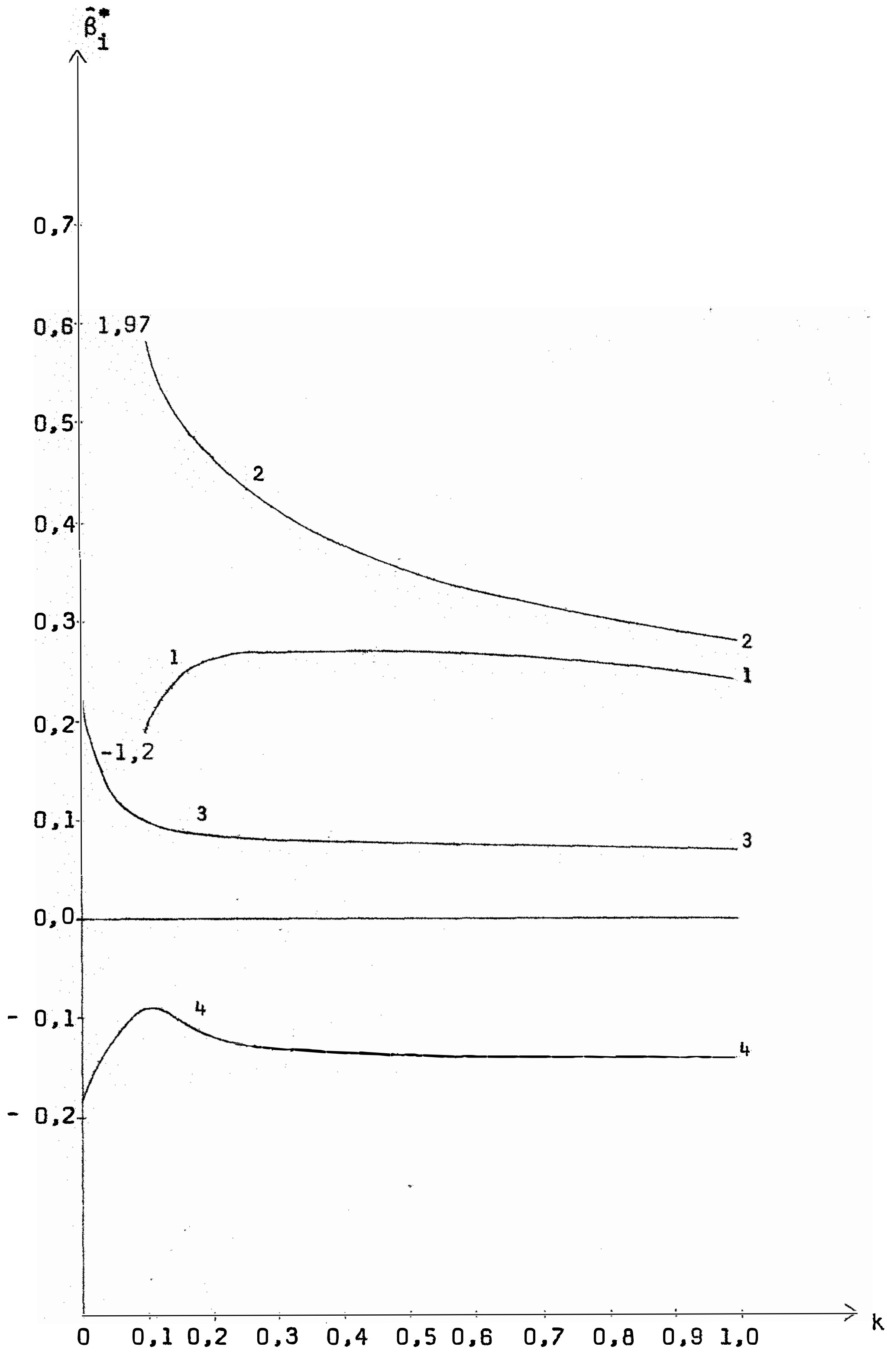


FIGURA 9 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função de demanda da batatinha. Modelo II , 4 variáveis independentes: 1 - renda ; 2 - urbanização ; 3 - preço do feijão ; 4 - preço da batatinha.

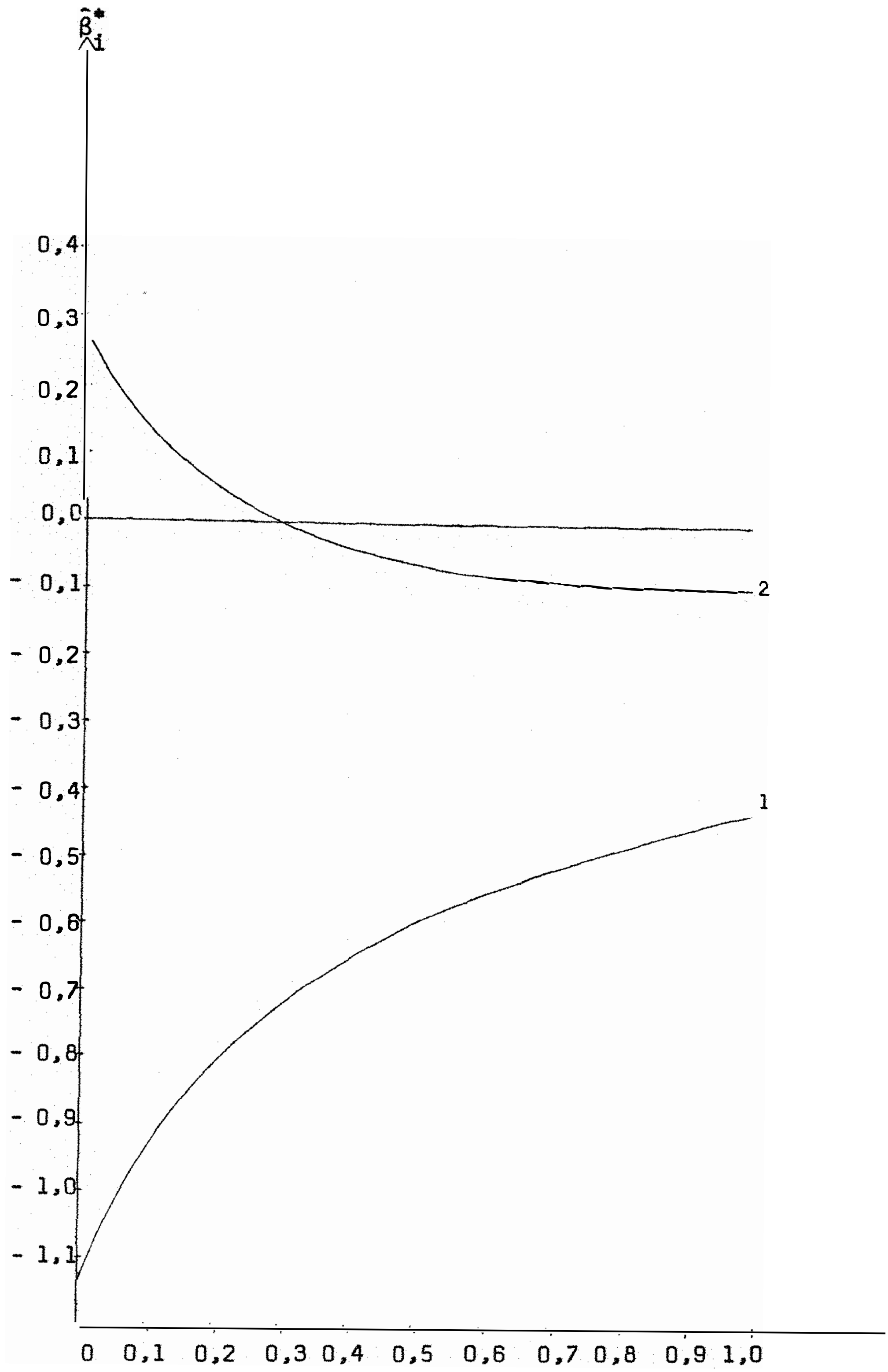


FIGURA 10 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função de demanda da carne, Modelo II , 2 variáveis independentes: 1 - preço da carne ; 2 - salário mínimo

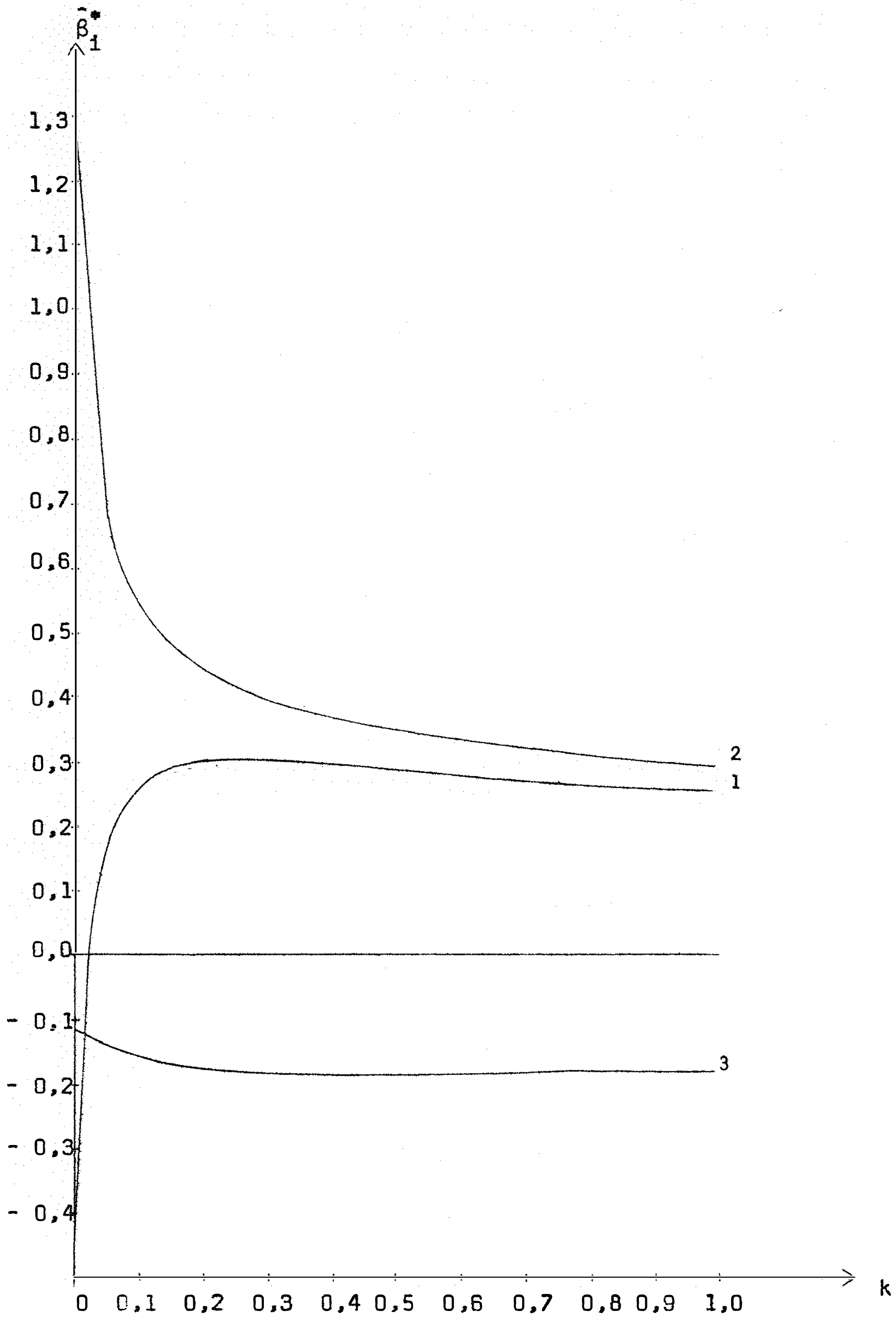


FIGURA 11 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função de demanda do leite, Modelo I , 3 variáveis independentes: 1 - renda ; 2 - urbanização ; 3 - preço do leite.

Já para o preço do arroz foi obtido um parâmetro cujo valor é bastante baixo ($\hat{\theta}_4^* = 0{,}008$) e não significativo. Com relação ao melhor modelo obtido por SOBRAL (1973, p. 67), apesar de estarem incluidas apenas tres variáveis independentes (urbanização, preço do arroz e preço da batatinha), elas explicam aproximadamente 71% da variação no consumo, mais do que aquelas incluidas na equação obtida pelo método de cumeeira. Encontrou-se um valor de Ø igual a 4,13, o qual está bem abaixo do valor (8,68 a 5%) da tabela de TORO-VIZCARRONDO e WALLACE (1968,p. 566). Houve um aumento de 31,76% da soma de quadrados dos resíduos da regressão, que pode ser considerado relativamente baixo.

De acordo com SOBRAL (1973), o modelo II é mais apropriado à análise da demanda de arroz, bem como aos alimentos calóricos,pois a quantidade demandada desse alimento atinge rapidamente um máximo quando a renda se eleva. Observando, na Tabela 8, os coeficientes de elasticidade calculados com base no Modelo II, para a média 1950/70, se deduzem as possíveis variações no consumo do arroz, "coeteris paribus".

Um acréscimo de 10% na renda conduz a um aumento de 1,71% no consumo.

Elevando-se de 10% o grau de urbanização, a demanda será aumentada de 1,96%.

10% de alta no preço da batatinha ocasionará uma diminuição de 1,04% no consumo.

Tabela 6. Estimativas dos parâmetros, respectivos testes "t" e testes estatísticos correspondentes à Função de Demanda do Arroz no Brasil, Modelo II.

| Variável | Coeficientes de Regressão ($\hat{\theta}_i^*$)a/ | Valor de t |
|---|---|---|
| 1 - Tempo | - 0,004 | -2,69 |
| 2 - Renda | 97,909 | 3,96 |
| 3 - Urbanização | 8,936 | 3,20 |
| 4 - Preço do Arroz | 0,008 | 0,14 |
| 5 - Preço do Feijão | -0,015 | -0,51 |
| 6 - Preço da Batatinha | -0,049 | -2,45 |
| 7 - Preço da Farinha | -0,024 | -0,91 |
| 8 - Salário Mínimo | -0,413 | -0,15 |

Termo Constante ($\hat{\alpha}^*$) = 3,482

Coeficiente de Determinação ($R^2$) = 0,638

Valor de F = 3,309

Valor do Teste de Durbin-Watson = 1,995

a/ Estimado ao nível k = 0,45.

Tabela 7. Estimativas dos parâmetros, respectivos testes "t" e testes estatísticos correspondentes à Função de Demanda do Arroz no Brasil, Modelo I.

| Variável | Coeficientes de Regressão $(\hat{\theta}_i^*)$ a/ | Valor de t |
|---|---|---|
| 1 - Tempo | 0,154 | 2,70 |
| 2 - Renda | 6,070 | 3,28 |
| 3 - Urbanização | 7,134 | 2,67 |
| 4 - Preço do Arroz | -0,016 | -0,003 |
| 5 - Preço do Feijão | -1,121 | -0,44 |
| 6 - Preço da Batatinha | -5,762 | -2,86 |
| 7 - Preço da Farinha | -3,007 | -0,75 |
| 8 - Salário Mínimo | 0,412 | 0,19 |

Termo constante $(\tilde{\alpha}^*)$ = -36,221

Coeficiente de Determinação $(R^2)$ = 0,623

Teste F = 3,0573

Teste de Durbin-Watson = 1,966

a/ Estimados ao nível k = 0,45.

Evidencia-se pois o arroz como um bem normal, cujo consumo se eleva quando a renda sobe. A batatinha é um produto complementar do arroz. Embora não estejam influenciando significativamente o consumo, os sinais encontrados dos preços do feijão e da farinha indicam serem alimentos complementares do arroz.

Contrariamente ao que se espera, o coeficiente do preço é positivo. Note-se, entretanto, que seu valor é muito pequeno, atingindo apenas a 14% da estimativa do próprio desvio padrão.

Tabela 8. Coeficientes de Elasticidade da Demanda do Arroz no Brasil, segundo os Modelos I e II.

| Variável | Modelo | 1950 | 1960 | 1970 | Média 1950/70 |
|---|---|---|---|---|---|
| renda | I | 0,167 | 0,143 | 0,127 | 0,149 |
| | II | 0,241 | 0,167 | 0,144 | 0,171 |
| urbanização | I | 0,196 | 0,168 | 0,149 | 0,176 |
| | II | 0,247 | 0,200 | 0,161 | 0,196 |
| preço da batatinha | I | -0,158 | -0,136 | -0,121 | -0,142 |
| | II | -0,081 | -0,082 | -0,145 | -0,104 |
| preço do arroz | I | 0 | 0 | 0 | 0 |
| | II | 0,017 | 0,017 | 0,019 | 0,016 |

Aqui também é interessante fazer-se uma comparação com os coeficientes de elasticidade deduzidos por SOBRAL (1973, p. 65), os quais são reproduzidos na Tabela 9.

Tabela 9. Coeficientes de Elasticidade da Demanda do Arroz no Brasil, obtidos pelo Método de Mínimos Quadrados.

| Variável | Modelo | 1950 | 1960 | 1970 | Média 1950/70 |
|---|---|---|---|---|---|
| renda | I | 0,391 | 0,342 | 0,270 | 0,320 |
| | II | 0,551 | 0,384 | 0,284 | 0,391 |
| preço da batatinha | I | -0,324 | -0,284 | -0,224 | -0,295 |
| | II | -0,147 | -0,148 | -0,263 | -0,189 |

Fonte: SOBRAL (1973, p. 65).

Pode-se comprovar através destes resultados alguns daqueles que aparecem na Tabela 8, ou seja, que o arroz é um bem normal,com coeficiente de demanda inelástico para a renda, e que a batatinha é complementar do mesmo.

#### 4.1.3.2. Feijão

Analisando, na Tabela 10, os coeficientes resultantes da melhor regressão ajustada ao modelo II, observa-se que o preço do feijão e a urbanização são as variáveis de maior efeito no consumo deste alimento. A regressão foi obtida ao nível k = 0,20. Os coeficientes desses dois fatores são os que apresentaram maiores valores de t entre as oito variáveis utilizadas para explicar a demanda do feijão. O coeficiente de determinação múltipla resultante é baixo, cerca de 45%. Já a melhor regressão ajustada por SOBRAL fornece um $R^2$ igual a 57,5%, onde foram incluidas as variáveis urbanização e preços do arroz, do feijão, da batatinha e da farinha.

Tabela 10 - Estimativas dos parâmetros, respectivos testes "t", e testes estatísticos correspondentes à Função de Demanda do Feijão no Brasil, Modelo II.

| Variável | Coeficientes de Regressão $(\hat{\theta}_i^*)$ a/ | Valor de t |
|---|---|---|
| 1 - Tempo | -0,002 | -1,59 |
| 2 - Renda | 4,511 | 0,16 |
| 3 - Urbanização | 5,933 | 2,19 |
| 4 - Preço do Arroz | 0,083 | 1,69 |
| 5 - Preço do Feijão | -0,066 | -2,38 |
| 6 - Preço da Batatinha | 0,004 | 0,20 |
| 7 - Preço da Farinha | -0,026 | -1,10 |
| 8 - Salário Mínimo | 0,189 | 0,08 |

Termo constante $(\hat{\alpha}^*)$ = 3,186
Coeficiente de Determinação $(R^2)$ = 0,447
Valor de F = 1,409
Teste de Durbin-Watson = 2,303

a/ Estimados ao nível k = 0,20.

Tabela 11 - Estimativas dos parâmetros, respectivos testes "t", e testes estatísticos correspondentes à Função de Demanda do Feijão no Brasil, modelo I.

| Variável | Coeficiente de Regressão ($\hat{\beta}_i^*$) a/ | Valor de t |
|---|---|---|
| 1 - Tempo | 0,048 | 1,84 |
| 2 - Renda | 0,106 | 0,12 |
| 3 - Urbanização | 2,503 | 2,03 |
| 4 - Preço do Arroz | 3,242 | 1,55 |
| 5 - Preço do Feijão | -2,357 | -2,05 |
| 6 - Preço da Batatinha | -0,316 | -0,32 |
| 7 - Preço da Farinha | -2,156 | -1,20 |
| 8 - Salário Mínimo | 0,102 | 0,11 |

Termo constante ($\hat{\alpha}^*$) = 10,439

Coeficiente de Determinação ($R^2$) = 0,399

Valor de F = 1,150

Valor do teste de Durbin-Watson = 2,152

a/ Estimados ao nível de k = 0,30.

O valor de Ø para o feijão é 2,64; em relação ao da tabela (8,68 a 5% de probabilidade), está bem abaixo, podendo-se concluir que a esperança do quadrado do desvio neste caso é menor para o estimador de cumeeira do que para o de mínimos quadrados. Houve um aumento de apenas 20,34% na soma de quadrados dos resíduos.

As conclusões sobre as elasticidades de demanda foram derivadas do modelo II, pelo mesmo motivo exposto no caso do arroz. Mantidas as outras variáveis constantes, deduz-se que:

Crescendo a urbanização em 10%, o consumo de feijão terá um incremento de 1,3%.

Aumentando o preço do feijão em 10%, seu consumo diminuirá em 1,28%.

Para aumento de 10% no preço da farinha, haverá um decréscimo correspondente a 0,95% no consumo de feijão.

Dessa forma, deduz-se que o feijão, como o arroz, também é um bem normal, cujo consumo deve aumentar com a renda,a urbanização, e com decréscimos no preço, sendo a farinha produto complementar do mesmo, e o arroz e a batatinha produtos substitutos.

Tabela 12. Coeficientes de Elasticidade da Demanda do Feijão no Brasil, segundo os Modelos I e II.

| Variável | Modelo | 1950 | 1960 | 1970 | Média 1950/70 |
|---|---|---|---|---|---|
| renda | I | 0,005 | 0,005 | 0,005 | 0,005 |
| | II | 0,011 | 0,008 | 0,006 | 0,008 |
| urbanização | I | 0,118 | 0,111 | 0,119 | 0,112 |
| | II | 0,164 | 0,133 | 0,107 | 0,130 |
| salário mínimo | I | 0,005 | 0,004 | 0,005 | 0,005 |
| | II | 0,005 | 0,002 | 0,002 | 0,002 |
| preço do feijão | I | -0,111 | -0,104 | -0,112 | -0,106 |
| | II | -0,179 | -0,078 | -0,098 | -0,128 |
| preço da farinha | I | -0,101 | -0,095 | -0,102 | -0,097 |
| | II | -0,096 | -0,113 | -0,096 | -0,095 |

Os valores que se encontram na Tabela 12 são bastante diferentes dos obtidos por SOBRAL (Tabela 13), principalmente a elasticidade da urbanização, que resultou aqui um valor menor.

Tabela 13 - Coeficientes de Elasticidade de Demanda do Feijão no Brasil, obtidos pelo Método dos Mínimos Quadrados.

| Variável | Modelo | 1950 | 1960 | 1970 | Média 1950/70 |
|---|---|---|---|---|---|
| renda | I | -0,714 | -0,774 | -0.706 | -0,691 |
| preço do feijão | I | -0,124 | -0,135 | -0,123 | -0,120 |
| | II | -0,297 | -0,129 | -0,163 | -0,213 |
| urbanização | I | 1,326 | 1,436 | 1,310 | 1,281 |
| | II | 0,548 | 0,443 | 0,356 | 0,435 |
| preço do arroz | I | 0,121 | 0,131 | 0,120 | 0,117 |
| | II | 0,248 | 0,244 | 0,271 | 0,228 |
| preço da batatinha | II | 0,073 | 0,073 | 0,130 | 0,093 |
| preço da farinha | II | -0,120 | -0,142 | -0,121 | -0,119 |

Fonte: SOBRAL (1973, p. 73).

### 4.1.3.3. Batatinha

A Tabela 14 inclui as estimativas dos coeficientes de regressão e testes estatísticos correspondentes à estimativa da regressão escolhida ao nível k = 0,35, para o modelo II. Renda e urbanização são as responsáveis pela maior influência no consumo da batatinha, tendo sido obtidos elevados valores de t para essas variáveis, sendo possivelmente significativas. As variáveis incluidas no modelo II estão "explicando" 73,42% da variação no consumo. O valor de Ø encontrado para a batatinha é igual a 5,49953, razoavelmente baixo quando comparado com o da tabela, que é igual a 8,83, a 5% de probabilidade. Houve aumento de 45,83% na soma de quadrados dos resíduos, que é aceitável.

Foram calculadas as elasticidades da demanda, que estão na Tabela 16. As variações que ocorrerão no consumo da batatinha, com base no modelo II, permanecendo as outras variáveis inalteradas, são as seguintes:

Um aumento de 10% na renda leva a um acréscimo de 1,19% no consumo.

Um aumento de 10% no grau de urbanização ocasionará elevação de 2,78% no consumo.

Elevando-se 10% o preço da batatinha, ter-se-á diminuição de 0,35% no consumo da mesma.

Tabela 14. Estimativas dos parâmetros, respectivos testes "t", e testes estatísticos correspondentes à Função de Demanda da Batatinha no Brasil, Modelo II.

| Variável | Coeficiente de Regressão $(\hat{\theta}_i^*)$ a/ | Valor de t |
|---|---|---|
| 1 - Tempo | -0,006 | -6,16 |
| 2 - Renda | 68,25 | 3,80 |
| 3 - Urbanização | 12,68 | 6,35 |
| 4 - Preço do Arroz | 0,022 | 0,58 |
| 5 - Preço do Feijão | 0,019 | 0,89 |
| 6 - Preço da Batatinha | -0,017 | -1,14 |
| 7 - Salário Mínimo | -0,748 | -0,38 |

Termo constante $(\hat{\alpha}^*)$ = 2,669

Coeficiente de Determinação $(R^2)$ = 0,741

Valor de F = 7,526

Valor do Teste de Durbin-Watson = 1,427

a/ Estimados ao nível k = 0,35.

Tabela 15. Estimativas dos parâmetros, respectivos testes "t", e testes estatísticos correspondentes à Função de Demanda da Batatinha no Brasil, Modelo I.

| Variável | Coeficientes de Regressão ($\hat{\theta}_i^*$)a/ | Teste t |
|---|---|---|
| 1 - Tempo | 0,059 | 6,43 |
| 2 - Renda | 1,149 | 3,76 |
| 3 - Urbanização | 2,734 | 6,25 |
| 4 - Preço do Arroz | 0,263 | 0,35 |
| 5 - Preço do Feijão | 0,369 | 0,91 |
| 6 - Preço da Batatinha | -0,337 | -1,00 |
| 7 - Salário Mínimo | -0,166 | -0,49 |

Termo constante ($\hat{\alpha}^*$) = -7,443

Coeficiente de Determinação ($R^2$) = 0,734

Valor de F = 7,205

Valor do Teste de Durbin-Watson = 1,354

a/ Estimados ao nível k = 0,35.

Tabela 16. Coeficientes de Elasticidade da Demanda da Batatinha no Brasil, Modelos I e II.

| Variável | Modelo | 1950 | 1960 | 1970 | Média 1950/70 |
|---|---|---|---|---|---|
| renda | I | 0,147 | 0,120 | 0,102 | 0,123 |
| | II | 0,168 | 0,117 | 0,086 | 0,119 |
| urbanização | I | 0,349 | 0,285 | 0,243 | 0,293 |
| | II | 0,351 | 0,283 | 0,228 | 0,278 |
| preço da batatinha | I | -0,043 | -0,035 | -0,030 | -0,036 |
| | II | -0,028 | -0,028 | -0,049 | -0,035 |
| salário mínimo | I | -0,021 | -0,017 | -0,015 | -0,018 |
| | II | -0,021 | -0,007 | -0,009 | -0,008 |

A batatinha é um bem normal, cujo consumo se eleva com a elevação na renda, da mesma forma que quando a urbanização se intensifica.

Os coeficientes de elasticidade obtidos por SOBRAL pelo método de mínimos quadrados são todos maiores em valor absoluto, e se encontram reproduzidos na Tabela 17, para fins de comparação.

Tabela 17. Coeficientes de Elasticidade da Demanda da Batatinha no Brasil, obtidos pelo Método de Mínimos Quadrados.

| Variável | Modelo | 1950 | 1960 | 1970 | Média 1950/70 |
|---|---|---|---|---|---|
| renda | I | -0,875 | -0,685 | -0,578 | -0,684 |
| | II | -0,944 | -0,657 | -0,487 | -0,670 |
| urbanização | I | 2,406 | 1,883 | 1,588 | 1,880 |
| | II | 2,363 | 1,908 | 1,536 | 1,876 |
| preço do feijão | I | 0,096 | 0,075 | 0,063 | 0,075 |
| | II | 0,093 | 0,040 | 0,051 | 0,066 |

Fonte: SOBRAL (1973, p.70).

Ainda que o coeficiente de elasticidade-renda tenha resultado negativo para SOBRAL, êle concluiu ser a batatinha um bem normal, através do ajustamento de uma regressão linear simples entre este produto e a renda o que, evidentemente, não é válido.

#### 4.1.3.4. Carne

Observa-se através da Tabela 18 (Modelo II, k = 0,35),que foram obtidas algumas estimativas de coeficientes de regressão com sinais incoerentes com o esperado,como é o caso dos parâmetros correspondentes

Tabela 18. Estimativas dos parâmetros, respectivos testes "t", e testes estatísticos correspondentes à Função de Demanda da Carne no Brasil, Modelo II.

| Variável | Coeficientes de Regressão ($\hat{\theta}_i^*$) a/ | Valor de t |
|---|---|---|
| 1 - Tempo | 0,0014 | 2,47 |
| 2 - Renda | -27,771 | -2,64 |
| 3 - Urbanização | - 3,741 | -3,44 |
| 4 - Preço da Carne | - 0,159 | -5,46 |
| 5 - Preço da Carne Seca | - 0,254 | -3,15 |
| 6 - Salário Mínimo | 0,5167 | 0,51 |

Termo constante ($\hat{\alpha}^*$) = 2,384
Coeficiente de Determinação ($R^2$) = 0,7945
Valor de F = 14,068
Valor do teste de Durbin-Watson = 0,65

a/ Estimados ao nível k = 0,35.

Tabela 19. Estimativas dos parâmetros, respectivos testes "t", e testes estatísticos correspondentes à Função de Demanda da Carne no Brasil, Modelo I.

| Variável | Coeficientes de Regressão $(\hat{\theta}_i^*)$ a/ | Valor de t |
|---|---|---|
| 1 - Tempo | -0,0238 | -3,04 |
| 2 - Renda | -0,7569 | -2,52 |
| 3 - Urbanização | -1,2725 | -3,41 |
| 4 - Preço da Carne | -2,1752 | -5,62 |
| 5 - Preço da Carne Seca | -2,1082 | -3,26 |
| 6 - Salário Mínimo | -0,0492 | -0,18 |

Termo constante $(\hat{\alpha}^*)$ = 27,816

Coeficiente de Determinação $(R^2)$ = 0,8144

Valor de F = 16,896

Valor do teste de Durbin-Watson = 0,8023

a/ Estimados ao nível k = 0,35.

às variáveis renda e urbanização. Quase todos os parâmetros são possivelmente significativos, exceto salário mínimo. O coeficiente de determinação encontrado é alto, cerca de 80%. Pelo método de mínimos quadrados, SOBRAL ajustou uma equação ao Modelo II, com as variáveis preço da carne e salário mínimo. Mesmo apenas com esses dois fatores obteve 92% de "explicação" na variação do consumo. Mas não pode analisar as influências da renda e urbanização, consideradas bastante importantes no caso da carne e outros alimentos proteicos.

Foi encontrado um valor de Ø igual a 16,052, que está bastante alto em comparação com o da tabela (8,41 ao nível de 5% de probabilidade). Foi também bastante elevado o aumento na soma de quadrados dos resíduos, cerca de 107%, bem acima dos outros produtos até agora considerados.

Baseando-se nas estimativas resultantes, foram obtidos os coeficientes de elasticidade de demanda da carne, que se encontram na Tabela 20. Com relação às variáveis preço da carne, preço da carne seca e salário mínimo que se apresentaram coerentes com a análise prévia, e mantidas as outras variáveis inalteradas, conclui-se que:

Um aumento de 10% no preço da carne leva à diminuição de 1,23% no consumo.

Esse mesmo aumento no preço da carne seca leva à diminuição de cerca de 1,3% no consumo da carne.

Elevando-se em 10% o salário mínimo, o consumo da carne irá aumentar em 0,06%.

Tabela 20. Estimativa dos Coeficientes de Elasticidade da Demanda de Carne no Brasil, Modelos I e II.

| Variável | Modelo | 1950 | 1960 | 1970 | Média 1950/70 |
|---|---|---|---|---|---|
| renda | I | -0,042 | -0,048 | -0,049 | -0,047 |
| | II | -0,068 | -0,047 | -0,035 | -0,048 |
| urbanização | I | -0,070 | -0,081 | -0,083 | -0,079 |
| | II | -0,103 | -0,084 | -0,067 | -0,082 |
| preço da carne | I | -0,119 | -0,139 | -0,142 | -0,136 |
| | II | -0,192 | -0,127 | -0,107 | -0,123 |
| preço da carne seca | I | -0,116 | -0,135 | -0,137 | -0,131 |
| | II | -0,172 | -0,193 | -0,118 | -1,297 |
| salário mínimo | I | -0,003 | -0,003 | -0,003 | -0,003 |
| | II | 0,015 | 0,005 | 0,006 | 0,006 |

Os melhores resultados obtidos por SOBRAL estão na Tabela 21.

Tabela 21. Coeficientes de Elasticidade de Demanda da Carne no Brasil, estimados pelo Método de Mínimos Quadrados.

| Variável | Modelo | 1950 | 1960 | 1970 | Média 1950/70 |
|---|---|---|---|---|---|
| urbanização | I | -0,096 | -0,109 | -0,119 | -0,110 |
| preço da carne | I | -0,291 | -0,330 | -0,361 | -0,333 |
| | II | -0,625 | -0,413 | -0,347 | -0,398 |
| salário mínimo | I | 0,021 | 0,024 | 0,027 | 0,024 |
| | II | 0,106 | 0,037 | 0,045 | 0,042 |

Fonte: SOBRAL (1973, p. 77).

Reconhece-se a falha do Método de Regressão de Cumeeira neste caso, não tendo-se conseguido melhores resultados que pelo Método de Mínimos Quadrados. Possivelmente a não inclusão de outra variável importante à "explicação" do consumo da carne, esteja afetando as estimativas aqui. Espera-se que nas camadas de renda mais baixa a elasticidade renda da carne possua valor elevado. A não disponibilidade de uma variável que represente a concentração da renda no período 1950/70, não permitiu a inclusão da mesma no modelo.

### 4.1.3.5. Leite

A equação de demanda para o modelo II ao nível k igual a 0,2 encontra-se na Tabela 22. As variáveis urbanização e renda são, como no caso da carne, as que possuem maior influência no consumo do leite, com valor de t bastante alto, principalmente para a urbanização. As variáveis incluidas estão "explicando" aproximadamente 92% do consumo do leite. SOBRAL ajustou uma regressão ao modelo I, incluindo a renda, a urbanização e o preço do leite, obtendo para $R^2$ o valor de 93% aproximadamente. Para a renda, houve incoerência no valor encontrado, cujo sinal é negativo. Houve também problema de autocorrelação nos resíduos significativa.

Resultou para o leite o maior valor do teste para a esperança do quadrado do desvio, ou seja, Ø é igual a 17,74. Resultou também um maior acréscimo na soma de quadrados dos resíduos, que é 118%, muito alto para ser aceitável. Porisso, conclui-se que neste caso é possível que a EQD para $\hat{\beta}^*$ seja maior que para $\hat{\beta}$, não sendo mais conveniente estimar a regressão pelo método de cumeeira.

O modelo II foi utilizado para análise das elasticidades de demanda. Com base na Tabela 24, e supondo-se "coeteris paribus", conclui-se que:

Tabela 22. Estimativas dos parâmetros, respectivos testes "t", e testes estatísticos correspondentes à Função de Demanda do Leite no Brasil, Modelo II.

| Variável | Coeficientes de Regressão $(\hat{\theta}_i^*)$ a/ | Valor de t |
|---|---|---|
| 1 - Tempo | - 0,011 | - 6,16 |
| 2 - Renda | 145,437 | 3,95 |
| 3 - Urbanização | 26,651 | 7,75 |
| 4 - Preço do Leite | -0,099 | - 1,73 |
| 5 - Preço da Carne | 0,323 | 3,00 |
| 6 - Salário Mínimo | 3,965 | 1,35 |

Termo constante $(\hat{\alpha}^*)$ = 4,578

Coeficiente de Determinação $(R^2)$ = 0,915

Valor de F = 38,738

Valor do teste de Durbin-Watson = 1,122

a/ Estimados ao nível k = 0,20.

Tabela 23. Estimativas dos parâmetros, respectivos testes "t", e testes estatísticos correspondentes à Função de Demanda do Leite no Brasil, Modelo I.

| Variável | Coeficientes de Regressão $(\theta_i^* )$ a/ | Valor de t |
|---|---|---|
| 1 - Tempo | 0,497 | 6,90 |
| 2 - Renda | 9,084 | 3,20 |
| 3 - Urbanização | 24,549 | 7,18 |
| 4 - Preço do Leite | -14,529 | -1,70 |
| 5 - Preço da Carne | 10,863 | 2,78 |
| 6 - Salário Mínimo | 0,164 | 0,07 |

Termo constante $(\hat{\alpha}^*)$ = -129,39

Coeficiente de Determinação $(R^2)$ = 0,869

Valor de F = 24,90

Valor do teste de Durbin-Watson = 0,882

a/ Estimados ao nível k = 0,25.

Uma elevação de 10% na renda será responsável por um aumento de 2,54% no consumo de leite.

10% de aumento no grau de urbanização irá provocar acréscimo de 5,85% no consumo do leite.

Se o preço do leite aumentar em 10%, haverá diminuição de 3,41% no consumo do mesmo.

Aumentando o preço da carne em 10%, aumentará em 2,5% o consumo do leite.

Os resultados acima permitem concluir que o leite é um bem normal, com as variáveis renda e urbanização influindo positivamente no seu consumo. Conclui-se também que: a carne se apresenta como um produto substituto do leite; a variável salário mínimo, apesar de não estar influenciando significativamente o consumo do leite, mostra efeito positivo no mesmo.

Tabela 24. Estimativa dos Coeficientes de Elasticidade da Demanda do Leite no Brasil, Modelos I e II.

| Variável | Modelo | 1950 | 1960 | 1970 | Média 1950/70 |
|---|---|---|---|---|---|
| renda | I | 0,427 | 0,207 | 0,167 | 0,211 |
| | II | 0,357 | 0,249 | 0,184 | 0,254 |
| urbanização | I | 1,155 | 0,560 | 0,452 | 0,570 |
| | II | 0,737 | 0,595 | 0,479 | 0,585 |
| preço do leite | I | -0,684 | -0,331 | -0,267 | -0,330 |
| | II | -0,292 | -0,346 | -0,392 | -0,341 |
| preço da carne | I | 0,511 | 0,245 | 0,200 | 0,250 |
| | II | 0,391 | 0,258 | 0,218 | 0,250 |
| salário mínimo | I | 0,008 | 0,004 | 0,003 | 0,004 |
| | II | 0,113 | 0,039 | 0,048 | 0,045 |

Os coeficientes de elasticidade calculados pelo método de regressão de cumeeira diferem bastante dos encontrados por SOBRAL, especialmente as elasticidades renda e urbanização (ver Tabela 25).

Tabela 25. Coeficientes de Elasticidade de Demanda do Leite no Brasil, estimados pelo Método de Mínimos Quadrados.

| Variável | Modelo | 1950 | 1960 | 1970 | Média 1950/70 |
|---|---|---|---|---|---|
| renda | I | -1,223 | -0,677 | -0,457 | -0,646 |
| preço do leite | I | -0,618 | -0,342 | -0,231 | -0,326 |
| urbanização | I | 5,352 | 2,962 | 1,998 | 2,826 |
| | II | 2,650 | 2,140 | 1,723 | 2,103 |
| salário mínimo | II | 0,191 | 0,066 | 0,080 | 0,075 |

Fonte: SOBRAL (1973, p. 80).

## 4.2. Resultados para as Funções de Produção

Os resultados referentes à função de produção do milho foram obtidos a partir do modelo na forma Cobb-Douglas. Não foi testado o modelo na forma de "produtividade", porque êle tem sempre menor poder de "explicação".

### 4.2.1. Multicolinearidade

Como era esperado, constatou-se tanto em Jardinópolis como em Guaíra, alta correlação entre a área cultivada e as outras variáveis, com exceção de "Educação Formal" e "Extensão Rural".

Segundo as Tabelas 26 e 27, onde estão representados os "fatores de inflação da variância dos parâmetros", encontram-se afetadas pela multicolinearidade as variáveis $X_1$ (área), $X_2$ (trabalho humano) e $X_5$ (máquinas e implementos), que são as que apresentaram os maiores fatores quando $k = 0$. Com os acréscimos em k, êles diminuem bastante. Para $\hat{\beta}_1^*$, a $k = 0,1$, o fator passa a ser apenas cerca de 10% daquele para $k = 0$, no caso de Jardinópolis.

Comparando com as funções de demanda, para as funções de produção os valores dos "fatores de inflação da variância dos parâmetros" não foram tão elevados. Não houve variação tão acentuada entre os fatores para todas as variáveis, como aconteceu na demanda. Por exemplo, depois de $X_5$, a variável cujo fator é maior é $X_6$ (despesas de custeio), com valor igual a 3,26 para Jardinópolis e 3,53 para Guaíra, a $k = 0$.

Tabela 26. Estimativas dos "Fatores de Inflação da Variância das Estimativas dos Parâmetros" e Soma de Quadrados dos Resíduos, aos vários níveis de k, relativos à Função de Produção de Milho, Município de Jardinópolis, São Paulo, 1969/70.

| k | Fatores de Infl. da Var. das Estim. dos Parâm. das variáveis | | | | | | | | $S^*(k)$ |
|---|---|---|---|---|---|---|---|---|---|
| | área cult. | trab. humano | fert. | semen tes | máq. impl. | desp. cust. | educ. form. | ext. rural | |
| 0,0 | 17,85 | 8,85 | 4,25 | 1,88 | 5,00 | 3,26 | 1,17 | 1,43 | 0,042 |
| 0,1 | 1,63 | 1,71 | 1,77 | 1,26 | 1,75 | 1,67 | 0,90 | 1,02 | 0,047 |
| 0,2 | 0,65 | 0,90 | 1,04 | 0,92 | 1,00 | 1,04 | 0,73 | 0,80 | 0,051 |
| 0,3 | 0,37 | 0,58 | 0,69 | 0,70 | 0,66 | 0,71 | 0,60 | 0,64 | 0,054 |
| 0,4 | 0,25 | 0,41 | 0,50 | 0,56 | 0,48 | 0,53 | 0,51 | 0,53 | 0,057 |
| 0,5 | 0,19 | 0,31 | 0,38 | 0,45 | 0,36 | 0,40 | 0,44 | 0,45 | 0,061 |
| 0,6 | 0,15 | 0,25 | 0,30 | 0,38 | 0,29 | 0,32 | 0,38 | 0,39 | 0,065 |
| 0,7 | 0,12 | 0,20 | 0,24 | 0,32 | 0,24 | 0,26 | 0,34 | 0,34 | 0,069 |
| 0,8 | 0,10 | 0,17 | 0,20 | 0,28 | 0,20 | 0,22 | 0,30 | 0,29 | 0,073 |
| 0,9 | 0,09 | 0,14 | 0,17 | 0,24 | 0,17 | 0,19 | 0,27 | 0,26 | 0,078 |
| 1,0 | 0,08 | 0,12 | 0,15 | 0,21 | 0,15 | 0,16 | 0,24 | 0,23 | 0,082 |

Tabela 27. Estimativas dos "Fatores de Inflação da Variância das Estimativas dos Parâmetros" e Soma de Quadrados dos Resíduos, aos vários níveis de k, relativos à Função de Produção de Milho, Município de Guaíra, São Paulo, 1969/70.

| k | Fatores de infl. da Var. das Estim. dos Parâm. das variáveis | | | | | | | S*(k) |
|---|---|---|---|---|---|---|---|---|
| | área cult. | trab. humano | fert. | sementes | máq. impl. | desp. cust. | educação formal | |
| 0,0 | 8,44 | 4,95 | 2,48 | 2,77 | 5,37 | 3,53 | 1,08 | 0,099 |
| 0,1 | 1,99 | 1,77 | 1,50 | 1,57 | 1,96 | 1,67 | 0,86 | 0,102 |
| 0,2 | 0,91 | 0,99 | 1,01 | 1,03 | 1,03 | 1,02 | 0,71 | 0,106 |
| 0,3 | 0,53 | 0,66 | 0,73 | 0,73 | 0,64 | 0,70 | 0,60 | 0,110 |
| 0,4 | 0,36 | 0,47 | 0,56 | 0,55 | 0,44 | 0,52 | 0,51 | 0,114 |
| 0,5 | 0,26 | 0,36 | 0,44 | 0,43 | 0,33 | 0,40 | 0,45 | 0,119 |
| 0,6 | 0,20 | 0,28 | 0,35 | 0,35 | 0,25 | 0,32 | 0,39 | 0,123 |
| 0,7 | 0,16 | 0,23 | 0,29 | 0,28 | 0,20 | 0,26 | 0,34 | 0,128 |
| 0,8 | 0,13 | 0,19 | 0,25 | 0,24 | 0,17 | 0,22 | 0,31 | 0,133 |
| 0,9 | 0,11 | 0,16 | 0,21 | 0,20 | 0,14 | 0,19 | 0,27 | 0,138 |
| 1,0 | 0,10 | 0,14 | 0,18 | 0,18 | 0,12 | 0,16 | 0,25 | 0,143 |

### 4.2.2. Gráficos de Cumeeira

Nas figuras 12 e 13 se encontram os gráficos de cumeeira correspondentes respectivamente às funções de produção para Jardinópolis e Guaíra. Êles estão representando as variações verificadas nas estimativas dos parâmetros no intervalo $0 < k \leq 1$. Para Jardinópolis, a maior oscilação foi notada em $\hat{\beta}_1^*$ (coeficiente de regressão do fator área cultivada), que teve seu valor diminuido de 0,62 quando k = 0 (mínimos quadrados) para 0,33 quando k atingiu 0,1. Para $k > 0,3$, os parâmetros se apresentaram bastante estáveis, e as estimativas da equação ajustada ao valor k = 0,3 foram utilizadas para obtenção dos produtos marginais. Em Guaíra, os parâmetros que se apresentavam com valores absolutos aumentados tiveram diminuição nessa magnitude, sendo que $\hat{\beta}_3^*$(coeficiente de regressão correspondente à variável fertilizantes), teve seu sinal trocado. Ao valor k = 0,3 o sistema mostrou-se estável, sendo este nível utilizado na obtenção da função de produção de milho em Guaíra.

Em seu trabalho BISERRA selecionou duas equações para Jardinópolis que foram denominadas Modelo VII e Modelo XII, e duas para Guaíra, que correspondem ao Modelo II e ao Modelo V. Foram construidos os gráficos de cumeeira para estas equações, apresentados, respectivamente, nas Figuras 14 a 17. De modo geral percebe-se que os sistemas são

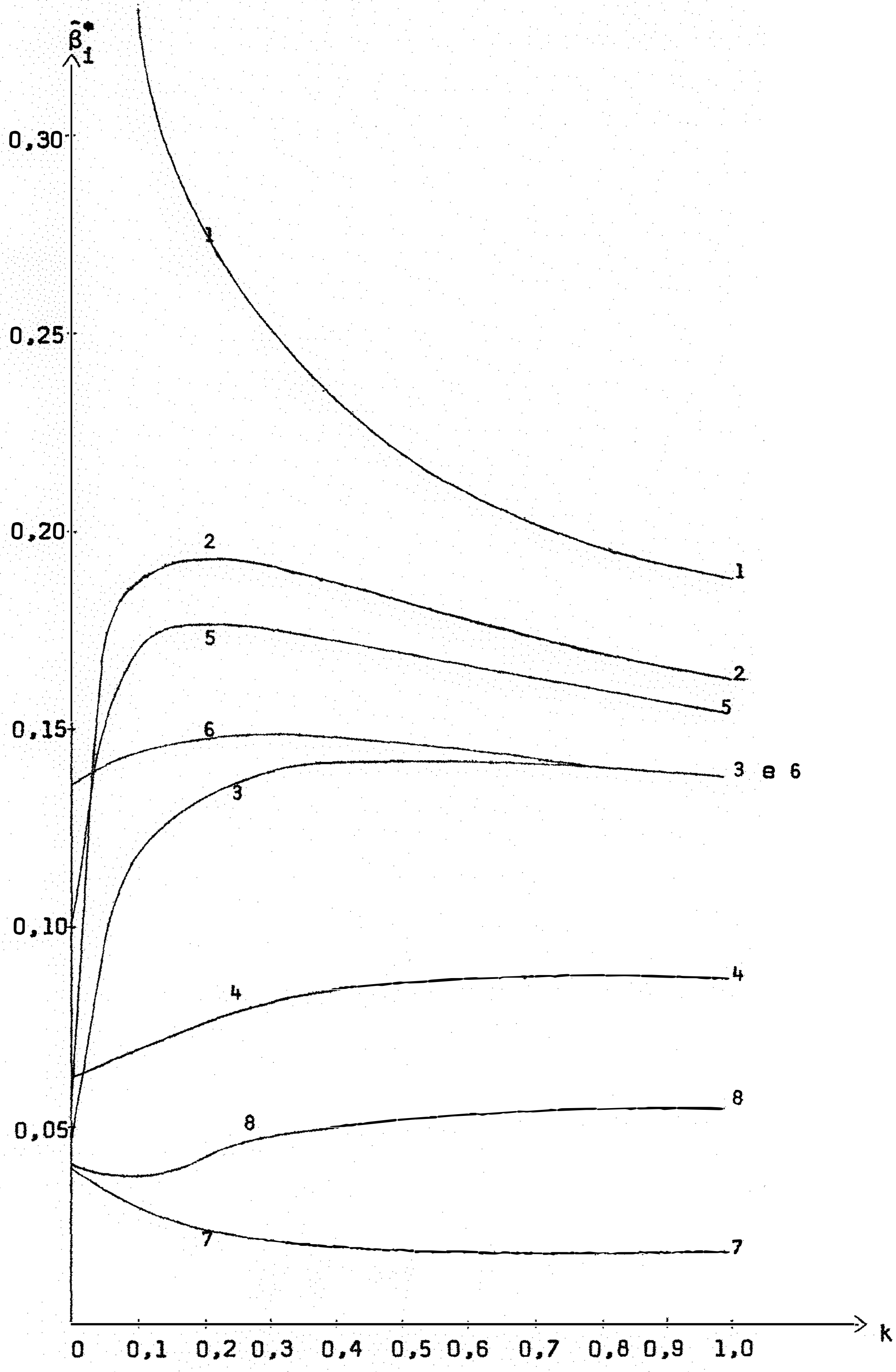


FIGURA 12 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função de produção de milho, Município de Jardinópolis, 8 variáveis independentes (Ver Tabela 28).

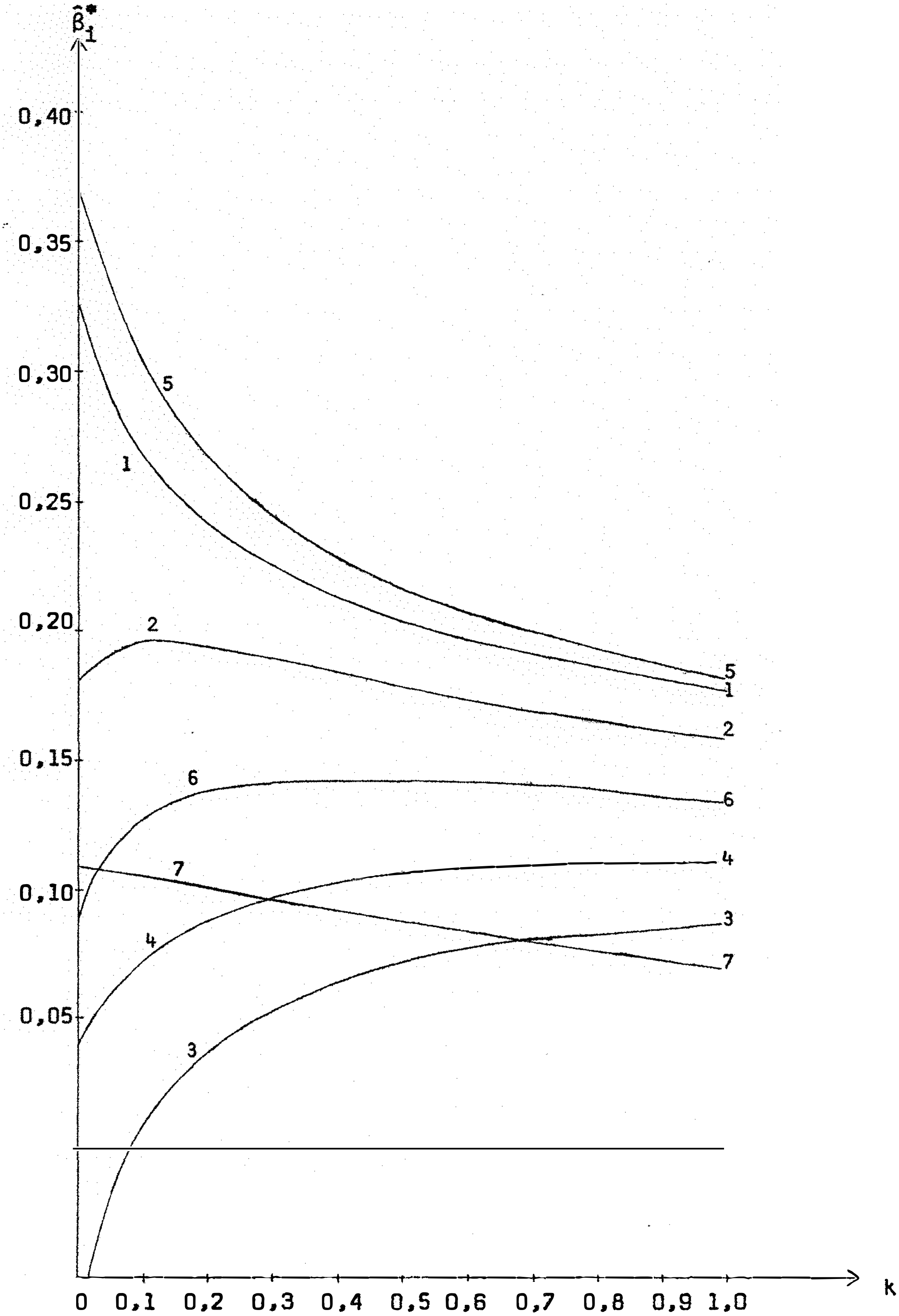


FIGURA 13 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função da produção de milho. Município de Guaíra, 7 variáveis independentes. (Ver Tabela 30).

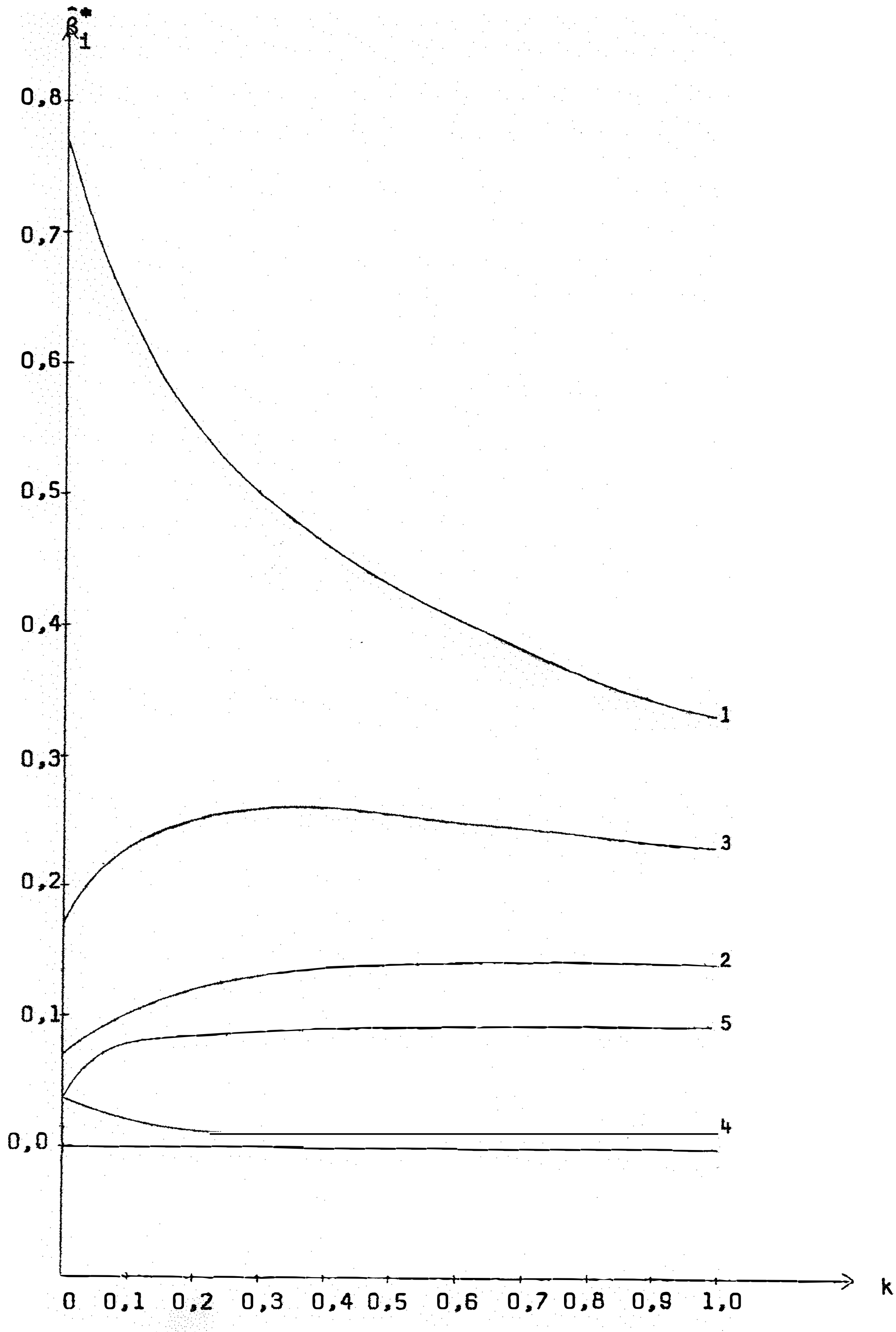


FIGURA 14 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função de produção de milho, Município de Jardinópolis, 5 variáveis independentes : $X_1$ - Área cultivada ; $X_2$ - Sementes ; $X_3$ - Despesas de Custeio; $X_4$ - Anos de Educação ; $X_5$ - Contatos com extensionista (Modelo VII)

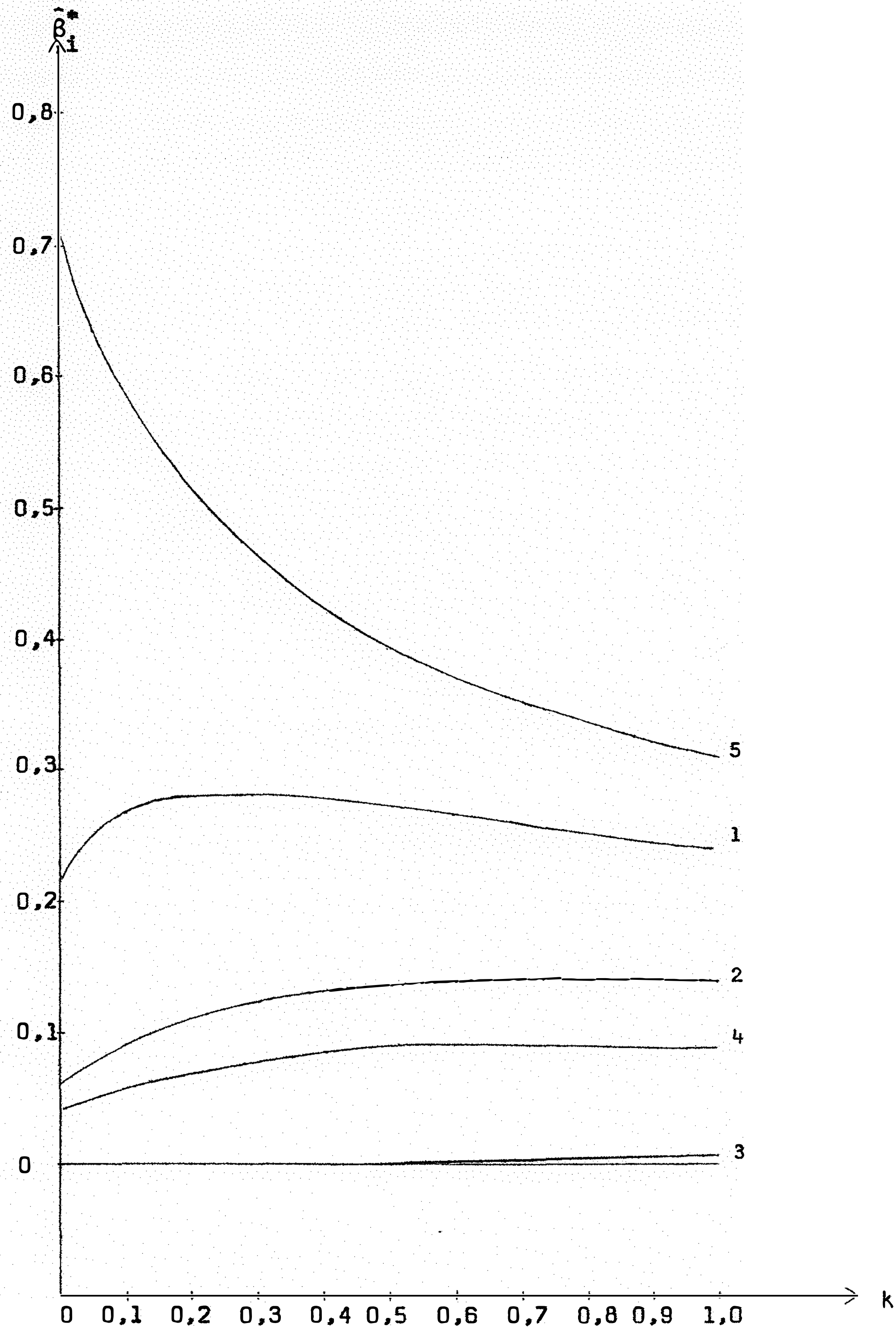


FIGURA 15 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função de produção de milho, Município de Jardinópolis, 5 variáveis independentes: $X_1$ - Fertilizantes ; $X_2$ - Sementes ; $X_3$ - Anos de educação ; $X_4$ - Contatos com extensionista ; $X_5$ - Despesa com mão de obra e despesas de custeio. (Modelo XII)

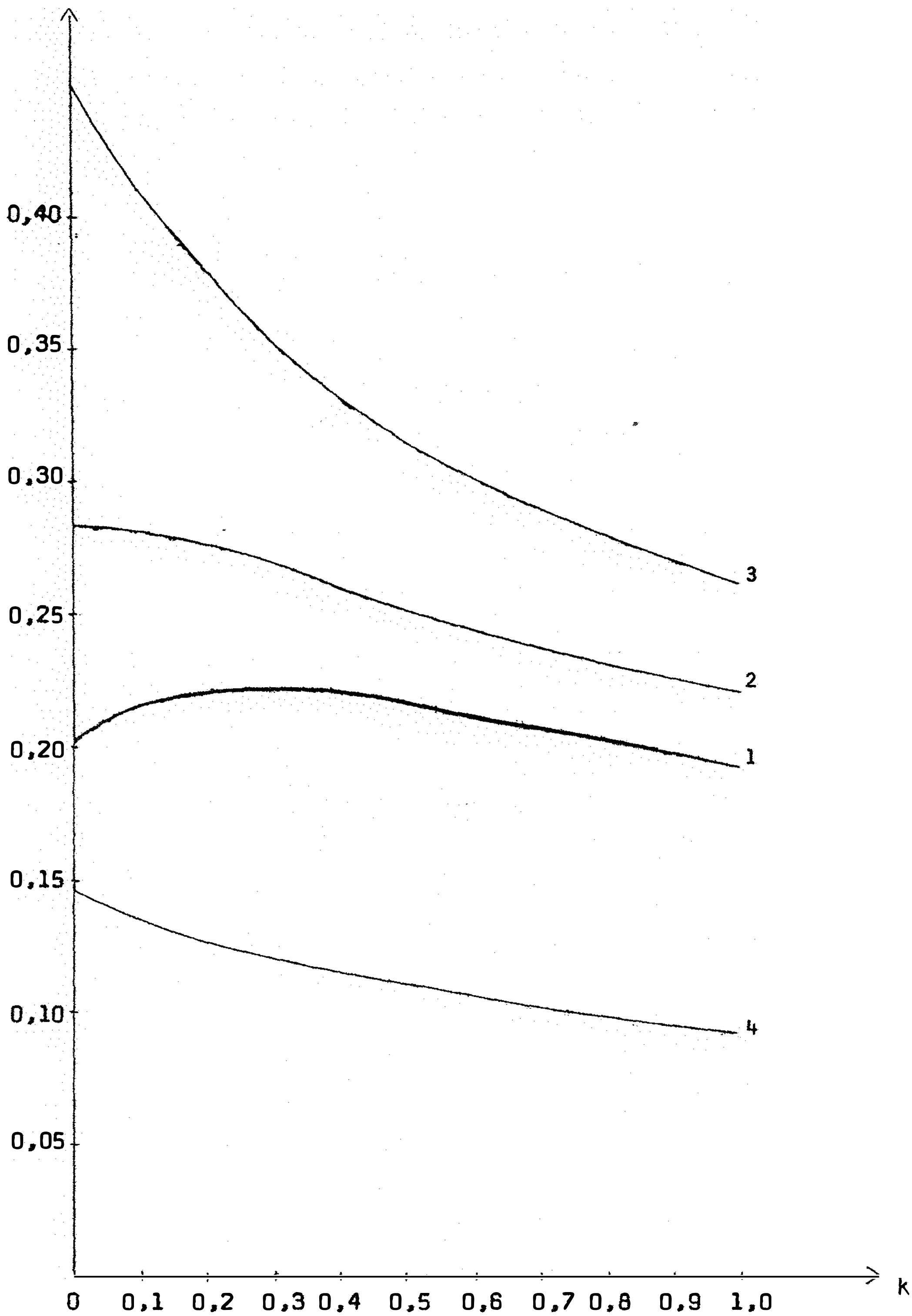


FIGURA 16 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função de produção de milho, Município de Guaíra, 4 variáveis independentes: 1 - Fertilizantes ; 2 - Sementes ; 3 - Despesas de custeio ; 4 - Educação. (Modelo II)

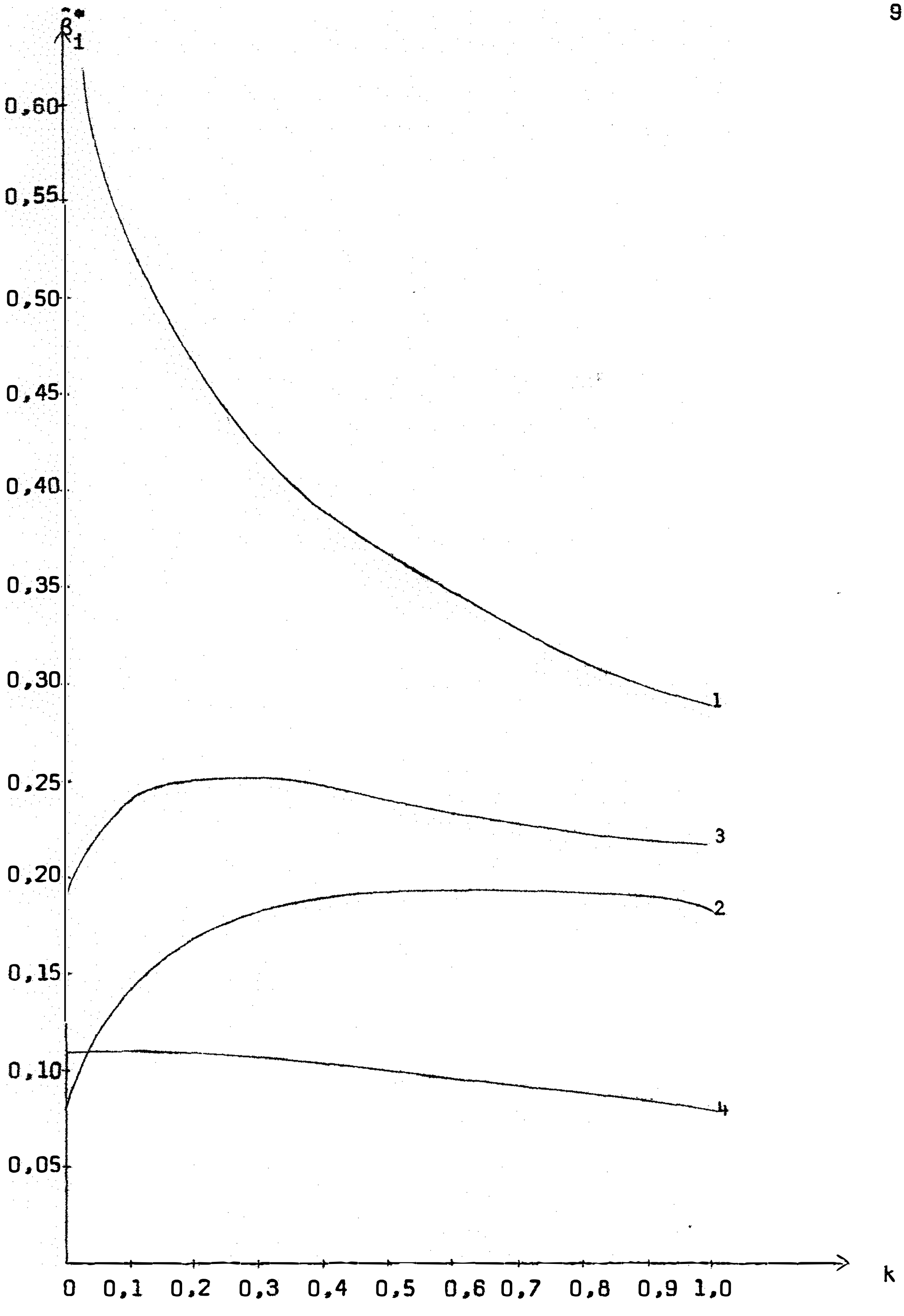


FIGURA 17 - Variação no valor dos coeficientes de regressão, com os níveis de k . Função da produção de milho, Município de Guaíra, 4 variáveis independentes: 1 - Área ; 2 - Sementes ; 3 - Despesas de custeio ; 4 - Educação. (Modelo V) instáveis, revelando que existem insumos correlacionados de forma a prejudicar as estimativas. Os parâmetros referentes à área cultivada, despesas de custeio e despesas com mão-de-obra e custeio, tiveram seus valores mais alterados com acréscimos em k. Então as funções de produção selecionadas, nas quais se retirou algumas variáveis, ainda se apresentam com problema de multicolinearidade afetando as estimativas.

### 4.2.3. Análise do Nível de Utilização dos Fatores

#### 4.2.3.1. Função de Produção de Milho em Jardinópolis

A melhor equação estimada ao nível k igual a 0,3, revela que todos os parâmetros são possivelmente significativos, com exceção daquele correspondente à variável "educação formal", como se pode observar através da Tabela 28. O valor de $R^2$ encontrado (0,89) evidencia que as variáveis que contribuem na produção do milho em Jardinópolis "explicam" aproximadamente 89% da variação nessa produção. Esses resultados podem ser considerados bons.

Comparando com os melhores resultados obtidos por BISERRA (1971, p. 41), verifica-se que êle obteve valores de $R^2$ em torno de 93%, tendo incluido cinco variáveis nas regressões ajustadas, onde quase todas essas variáveis mostraram ser significativas:

Tabela 28. Estimativas dos parâmetros, respectivos testes "t" e testes estatísticos correspondentes à Função de Produção da Cultura do Milho. Modelo Cobb-Douglas, estimado pelo Método de Regressão de Cumeeira com k = 0,30. Município de Jardinópolis, Estado de São Paulo, Ano Agrícola 1969/70.

| Variáveis | Coeficientes de Regressão ($\hat{\theta}_i^*$) | Valor de t |
|---|---|---|
| $X_1$ - Área Cultivada (ha) | 0,272 | 12,80 |
| $X_2$ - Trabalho Humano (d-h) | 0,243 | 7,70 |
| $X_3$ - Fertilizantes (Cr$) | 0,109 | 5,16 |
| $X_4$ - Sementes (Cr$) | 0,0576 | 2,96 |
| $X_5$ - Máq. e Impl. (Cr$) | 0,148 | 6,64 |
| $X_6$ - Despesas de Custeio (Cr$) | 0,0974 | 5,40 |
| $X_7$ - Educação Formal (Anos) | 0,0353 | 0,91 |
| $X_8$ - Extensão Rural (Nº contatos) | 0,0473 | 1,79 |

Termo Constante ($\hat{\alpha}^*$) = 3,99

Coeficiente de Determinação ($R^2$) = 0,891

Valor de F = 105,59

O valor que foi obtido para o teste correspondente à esperança do quadrado do desvio é $\emptyset$ = 13,9, o qual se encontra acima daquele da tabela, que é 7,33 ao nível de 5% de probabilidade. No entanto houve aumento de apenas 27% na soma de quadrados dos resíduos, que não é considerado elevado, sendo portanto aceitável.

As elasticidades parciais de produção correspondem, no modelo Cobb-Douglas, aos valores estimados dos coeficientes de regressão ($\hat{\beta}_i^*$) (ver Tabela 28). Elas permitem concluir que, se houver uma variação de 10% na utilização dos fatores isoladamente, mantendo-se constantes os níveis dos demais, seria incrementado o valor da produção de milho em:

a) 2,7% caso se eleve em 10% a área de plantio;

b) 2,4% se se aumentar em 10% a utilização de trabalho humano;

c) 1,1% se houver incremento de 10% na utilização de fertilizantes;

d) 0,6% se houver acréscimo de 10% na utilização de sementes de milho;

e) 1,48% se aumentada em 10% a utilização das máquinas e implementos;

f) 1% se incrementado em 10% as "despesas de custeio";

g) 0,5% se se elevasse os contatos com técnico extensionista em 10%.

A tabela 29 exibe os valores calculados dos produtos médios e marginais, os preços correntes dos insumos na época do levantamento, e as relações entre os valores dos produtos marginais para cada fator e seu respectivo preço. O valor do produto médio é o mesmo apresentado no trabalho de BISERRA (1971), onde foi calculado com base nas médias geométricas dos valores observados na amostra analisada. Os preços dos fatores também foram obtidos deste mesmo trabalho, com exceção do trabalho e da remuneração por cruzeiro gasto em máquinas e implementos cujos valores não são fornecidos por BISERRA, e foram calculados aqui[10/].

Os insumos estão sendo utilizados no estágio racional de produção (estágio II), já que todos os parâmetros obtidos são positivos e inferiores à unidade.

No caso da área plantada com milho, se houvesse aumento de um hectare, a renda bruta com este produto seria ampliada em Cr$ 119,00. A relação $VPMa_x/P_x$ é pouco superior à unidade (1,37), indicando ligeira sub-utilização deste fator. BISERRA (1971), encontrou valor superestimado (4,19) devido à multicolinearidade e erros de especificação.

---

10/ Ver Apêndice B.

Tabela 29 - Valores dos Produtos Médio e Marginal, Preços dos Insumos e Relação entre o Valor do Produto Marginal e o Preço de respectivo Insumo, referentes à Produção de Milho, Município de Jardinópolis, Estado de São Paulo, Ano Agrícola 1969/70.

| Item | Área ($X_1$) | Trab. ($X_2$) | Fert. ($X_3$) | Sem. ($X_4$) | Máq.e Impl. ($X_5$) | Desp.Cust. ($X_6$) | Educ. ($X_7$) | Ext. ($X_8$) |
|---|---|---|---|---|---|---|---|---|
| $VPMe_{x_i}$ a/ b/ | 437,09 | 18,08 | 7,47 | 68,32 | 8,51 | 24,57 | 2431,47 | 1629,89 |
| $VPMa_{x_i}$ a/ b/ | 118,89 | 4,40 | 0,82 | 3,93 | 1,26 | 2,39 | 85,83 | 77,09 |
| $P_{x_i}$ a/ | 87,00 | 5,40 | 1,07 | 1,17 | 1,17 | 1,17 | c/ | c/ |
| $VPMa_{x_i}/P_{x_i}$ | 1,37 | 0,81 | 0,77 | 3,36 | 1,09 | 2,04 | - | - |

a/ Referem-se a cruzeiros por unidade de insumo.

b/ Os valores dos produtos médios e marginais foram calculados utilizando-se as médias geométricas dos valores observados na amostra.

c/ O trabalho de BISERRA (1971) não apresenta informação sobre o preço desses insumos.

Analisando o VPMa do trabalho humano, verifica-se que um aumento de 10 dias-homens ocasionaria elevação de Cr$44,00 na renda bruta. O valor da relação $VPMa_x/P_x$, igual a 0,81, indica superutilização desse fator.

Aumentando em Cr$10,00 a despesa com fertilizantes, haveria aumento na renda bruta correspondente ao milho, de Cr$8,20. A utilização deste insumo estava sendo feita além do nível de ótimo econômico ($VPMa_x/P_x = 0,77$), ao contrário do estimado através do método de mínimos quadrados ($VPMa_x/P_x = 1,36$).

Quanto à variável sementes, haveria elevação de Cr$ 39,30 se ocorresse acréscimo de Cr$10,00 no dispêndio com esse insumo. Sua utilização estava se processando em nível inferior à de ótimo econômico ($VPMa_x/P_x = 3,36$), confirmando os resultados obtidos por BISERRA (1971) (relação $VPMa_x/P_x$ igual a 2,97 ou 4,03 conforme o modelo considerado).

Com relação à variável máquinas e implementos agrícolas, conclui-se que um aumento de Cr$10,00 na despesa com este fator seria responsável em elevar em Cr$12,60 a renda bruta do milho. A utilização estava sendo feita no nível de ótimo econômico ($VPMa_x/P_x$=1,09).

O valor do produto marginal calculado à "despesas de custeio", indica que haveria acréscimo de Cr$23,90 no valor da produção caso os agricultores elevassem as despesas de custeio em Cr$10,00. Este fator estava sendo sub-utilizado em Jardinópolis, como indica o resultado $VPMa_x/P_x$ = 2,04. Isto confirma as conclusões de BISERRA (1971) que obteve o valor 2,26 para esta relação.

### 4.2.3.2. Função de Produção de Milho em Guaíra

Observa-se, na Tabela 30, que as variáveis incluidas estão "explicando" 83% da variação no valor da produção do milho. Todas elas se mostram possivelmente significativas, e as de maior efeito são trabalho humano, área cultivada e máquinas e implementos. A de menor efeito é a variável fertilizantes.

As duas melhores equações selecionadas por BISERRA(1971, p.55) estão explicando cerca de 76% e 91% das variações no valor da produção de milho. Elas incluem quatro variáveis independentes e quase todas são significativas. Não incluem máquinas e implementos e mão-de-obra, duas variáveis importantes à especificação do modelo.

Em Guaíra o valor de Ø para a regressão de Cumeeira selecionada é igual a 6,33. Comparando com o da tabela (7,33 ao nível de 5% de probabilidade), este mostra-se inferior. Por sua vez, a soma de quadrados dos resíduos teve o pequeno aumento de 11,11%, bastante baixo.

Verificando, ainda na Tabela 30, os valores dos coeficientes de regressão resultantes para Guaíra, conclui-se que todos os fatores estavam sendo utilizados no estágio racional de produção (estágio II). Estes mesmos coeficientes representam aqui as elasticidades

Tabela 30. Estimativa dos parâmetros, respectivos testes "t" e testes estatísticos correspondentes à Função de Produção da Cultura do Milho. Modelo Cobb-Douglas, estimado pelo Método de Regressão de Cumeeira ao nível k = 0,30. Município de Guaíra, Estado de São Paulo, Ano Agrícola 1969/70.

| Variáveis | Coeficientes de Regressão ($\hat{\theta}_i^*$) | Valor de t |
|---|---|---|
| $X_1$ - Área Cultivada (ha) | 0,2455 | 6,91 |
| $X_2$ - Trabalho Humaho (d-h) | 0,2537 | 5,22 |
| $X_3$ - Fertilizantes (Cr$) | 0,0308 | 1,38 |
| $X_4$ - Sementes (Cr$) | 0,0844 | 2,54 |
| $X_5$ - Máq. e Impl. (Cr$) | 0,198 | 6,84 |
| $X_6$ - Desp. de Custeio (Cr$) | 0,086 | 3,77 |
| $X_7$ - Educação Formal (anos) | 0,176 | 2,75 |

Termo Constante ($\hat{\alpha}^*$) = 3,378

Coeficiente de Determinação ($R^2$) = 0,834

Valor de F = 60,626

Tabela 31. Valor do Produto Médio, Valor do Produto Marginal, Preços dos Insumos e Relação entre o Valor do Produto Marginal e o Preço do respectivo Insumo, referentes à Produção de Milho, Município de Guaíra, Estado de São Paulo. Ano Agrícola 1969/70.

| Item | Área | Trab. | Fert. | Sem. | Máq. e Impl. | Desp.Cust. | Educ. |
|---|---|---|---|---|---|---|---|
| $VPMe_{x_i}$ a/b/ | 383,68 | 18,95 | 8,97 | 44,61 | 7,26 | 20,65 | 3136,16 |
| $VPMa_{x_i}$ a/b/ | 94,19 | 4,81 | 0,28 | 3,76 | 1,44 | 1,78 | 551,96 |
| $P_{x_i}$ a/ | 107,00 | 5,40 | 1,07 | 1,17 | 1,17 | 1,17 | c/ |
| $VPMa_{x_i}/P_{x_i}$ | 0,88 | 0,89 | 0,26 | 3,21 | 1,23 | 1,52 | - |

a/ Referem-se a cruzeiros por unidade de insumo.

b/ Os valores correspondentes aos produtos médios e marginais foram calculados utilizando-se as médias geométricas dos valores observados na amostra.

c/ O trabalho de BISERRA (1971) não apresenta informação sobre o preço desse insumo.

parciais de produção, que permitem inferir que, permanecendo outros fatores constantes:

a) um acréscimo de 10% na área iria produzir aumento de 2,5% no valor da produção;

b) elevando-se 10% o emprego do fator trabalho ter-se-ia aumento de 2,5% no valor da produção;

c) incrementando-se o uso de fertilizantes em 10%, a renda do milho se elevará em 0,3%;

d) se houvesse um aumento de 10% na utilização de sementes, a área aumentaria de 0,8%;

e) uma elevação de 10% na utilização das máquinas e implementos seria responsável pelo aumento aproximado de 2% na renda bruta correspondente ao milho;

f) para um aumento de 10% nas despesas com custeio ter-se-ia acréscimo de 0,9% no valor total da produção.

De acordo com a Tabela 31, conclui-se que aumentando-se em um hectare a área cultivada com milho, se terá renda adicional de Cr$94,19. Há utilização em excesso deste fator, como mostra a relação $VPMa_x/P_x$ (igual a 0,88), e ao contrário da sub-utilização constatada por BISERRA ($VPMa_x/P_x = 3,69$).

Com relação à variável trabalho humano, o aumento de 10 dias-homens na sua utilização iria ocasionar elevação de Cr$48,00 na renda dos agricultores. O valor da relação $VPMa_x/P_x$, igual a 0,89, indica ligeira super-utilização da mão-de-obra na produção de milho.

O valor do produto marginal de fertilizantes, está indicando que um aumento de Cr$10,00 na despesa com esse insumo, produzirá acréscimo de Cr$2,80 na renda. Estava havendo aqui também super-utilização do fator ($VPMa_x/P_x$ = 0,26). Segundo a regressão estimada por BISERRA a relação $VPMa_x/P_x$ é igual a 0,99.

Uma elevação de Cr$10,00 nas despesas com sementes iria ocasionar aumento de Cr$37,60 no valor do produto. Este insumo estava sendo sub-utilizado ($VPMa_x/P_x$ = 3,21), confirmando, embora em nível menor, a estimativa por BISERRA ($VPMa_x/P_x$ = 9,89), de sub-utilização do fator.

O valor do produto marginal calculado para máquinas e implementos indica que uma elevação de Cr$10,00 nas despesas com sua utilização ocasionaria aumento de Cr$14,40 no valor da produção. Houve pequena sub-utilização de maquinaria, como indica a relação $VPMa_x/P_x$ = 1,23.

No caso das "despesas de custeio", se se elevasse em Cr$10,00 sua utilização, ocorreria aumento de Cr$17,80 na renda. Este insumo também estava sendo sub-utilizado ($VPMa_x/P_x$ = 1,25), confirmando, embora em nível muito menor, o resultado de BISERRA ($VPMa_x/P_x$=5,04).

## 4.3. Sugestões para Pesquisas Futuras

Para exame da esperança do quadrado do desvio, seria interessante o desenvolvimento de um teste estatístico que verificasse, num modelo com restrição não linear como é o caso do método de regressão de cumeeira, se êle é mais conveniente que o modelo sem restrição e com problemas de multicolinearidade, como acontece com método de mínimos quadrados ordinários nos exemplos aqui estudados.

# 5. CONCLUSÕES

1. Constatou-se não haver uma regra simples para a verificação da possível vantagem do método de regressão de cumeeira. Não há um teste estatístico que se possa utilizar para verificar se a esperança do quadrado do desvio é menor para $\hat{\beta}^*$ (estimado pelo método de regressão de cumeeira) do que para $\hat{\beta}$ (estimado pelo método de mínimos quadrados). Porisso é difícil decidir se as estimativas obtidas pelo método de regressão de cumeeira são melhores do que as de mínimos quadrados.

2. Comprovou-se que no método de regressão de cumeeira há sempre aumento da precisão das estimativas dos parâmetros. Ocorre também diminuição nos valores absolutos de todos os coeficientes de regressão, em proporção maior para aqueles com multicolinearidade alta, que possivelmente estavam superestimados devido aos erros prove-

nientes desta. Dessa forma, para as regressões estimadas a um dado valor de k > 0, resultaram variâncias menores e diminuição dos valores absolutos dos coeficientes de regressão, em relação à regressão para k = 0.

3. O valor da soma de quadrados do resíduo da regressão aumenta sempre com o aumento de k. Verificou-se que para as funções de produção o acréscimo relativo na soma de quadrados do resíduo foi menor que para as funções de demanda. Note-se, entretanto, que a importância de um aumento relativo no valor da soma de quadrados do resíduo varia em proporção inversa com o número de graus de liberdade correspondente, e que no caso das funções de produção o número de graus de liberdade do resíduo é bem maior do que no caso das funções de demanda analisadas.

4. Os resultados obtidos através do método de regressão de cumeeira para as funções de produção são melhores que aqueles obtidos para as funções de demanda:

a. No caso das funções de produção, obteve-se sempre resultados coerentes com o esperado com base em análise prévia. Na demanda, no caso da carne não se conseguiu bons resultados, com alguns coeficientes apresentando sinais incorretos.

b. Quanto ao aumento na soma de quadrados do resíduo, observa-se que para as estimativas do leite, onde se conseguiu parâmetros consistentes com o esperado, houve um aumento relativo muito gran de na soma de quadrados do resíduo , o que é inaceitável.

5. Os testes estatísticos normalmente usados em análise de regressão (teste F, teste t e teste de Durbin-Watson), não se aplicam ao método de regressão de cumeeira, pois os estimadores são tendenciosos.

6. A aplicação do coeficiente de determinação múltipla como medida descritiva no método de regressão de cumeeira é válida. Ocorre sempre diminuição do $R^2$ com o aumento de k. Então existe um maior resíduo ou variação "não explicada" dos valores da variável dependente, no caso do método de regressão de cumeeira.

7. O critério utilizado para a escolha das "melhores" regressões[11/], ou seja, o critério para a escolha do valor de k através do gráfico de cumeeira, possui a desvantagem de ser subjetivo.

8. Os gráficos de cumeeira mostraram ser úteis na localização das variáveis afetadas pela multicolinearidade, e na evidência da instabilidade dos coeficientes de regressão. A oscilação nos valores dos parâmetros representada nos gráficos de cumeeira foi maior no caso das funções de demanda analisadas do que para as funções de produção de milho.

---

11/ Ver p. 43 deste trabalho.

9. A análise dos "fatores de inflação da variância das estimativas dos parâmetros" é um complemento também útil à detecção da multicolinearidade nos dados da amostra. As maiores diminuições nestes "fatores" no intervalo $0 \leq k \leq 0,1$ situam-se entre 85 a 95% para as funções de produção e 98 a 99,9% para as funções de demanda.

10. A eliminação de variáveis importantes nos modelos origina estimativas viesadas. É o caso da estimação da função de produção de milho no município de Guaíra, pelo método de mínimos quadrados, em que não se incluiu mão-de-obra e maquinaria. Observou-se através do gráfico de cumeeira que, em alguns casos o procedimento de se cancelar variáveis não elimina o problema da multicolinearidade, pois os parâmetros continuam "instáveis".

11. Verificou-se que os coeficientes de demanda obtidos pelo método de regressão de cumeeira são sempre mais inelásticos do que os calculados pelo método de mínimos quadrados ordinários. Nas funções de produção também obteve-se estimativas menores pelo método de cumeeira para os valores da relação entre o valor do produto marginal e o preço do insumo.

12. Todos os coeficientes de elasticidade resultantes diferem bastante daqueles encontrados por SOBRAL (1973), principalmente os de elasticidade em relação à renda e à urbanização. Encontrou-se coeficientes de demanda inelásticos para a renda em todos os produtos. No

entanto, se fossem consideradas diferentes classes de renda, as estimativas das elasticidades de demanda iriam variar entre as classes.

13. Com exceção da carne, as variáveis que possuem maior influência no consumo de alimentos, são, pela ordem, a renda, a urbanização e o preço do produto. Todos os produtos revelaram-se bens normais.

14. Com relação à carne, não foi possível a obtenção de resultados satisfatórios pelo método empregado neste trabalho. Provavelmente o comportamento dos consumidores para com este alimento é bastante influenciado pela classe de renda; não se dispunha, no entanto, de índices anuais para o grau de concentração da renda no país para o período estudado.

15. Na estimativa da função de produção de milho pelo método de regressão de cumeeira as variáveis que se revelaram mais influentes na explicação da produção são, pela ordem, a área cultivada, o trabalho humano, as máquinas e os implementos e as despesas de custeio. Todos os insumos estavam sendo utilizados no estágio racional de produção, havendo porém sub-utilização dos fatores sementes e despesas de custeio. Os coeficientes obtidos ao nível k = 0,3 revelaram-se melhores que aqueles a k = 0, tendo-se encontrado coeficientes de determinação altos, entre 0,83 e 0,89.

## SUMMARY

The main purpose of this research was to apply the ridge regression method to demand and production functions, and to compare the results with those obteined by the least squares method. The possibility of achieving "better" results in economic studies when explanatory variables present multicollinearity was verified.

Causes ,effects and measures to detect multicollinearity were examined, as well as problems that arise with use of methods generally utilized to mitigate the problem. Researches characterized by this type of problem are likely to result in imprecise and, thus, unreliable estimates.

Studies by SOBRAL (1973), on demand for food products (rice, beans, potatoes, meat and milk) and by BISERRA (1971), of corn production functions in Jardinópolis and Guaíra, municipalities of Ribeirão Preto's DIRA,State of São Paulo, were used to compare results.

Their data was reanalysed in the present study. Data corresponding to the demand function are from different sources, are yearly and for the period 1950-1970. Data corresponding to the production function analysis are from the agricultural year 1969-1970, and were obtained through direct interviews with farmers.

In adjusting the equations the ridge regression method was used. It provides smaller estimates of the parameters vector than those obteined by least squares. The technique was originally proposed by HOERL and KENNARD (1970 a), and seeks to achieve better regression analysis results where the problem of multicollinearity seriously affects estimates obtained by the usual least squares method. It is based on the increment of small positive quantities (0 to 1 interval), to elements of the main diagonal of the X'X correlation matrix. There are increment values for which the ridge estimates vector ($\hat{\beta}^*$) is closer on the average to $\beta$ (the parameters vector) than to $\hat{\beta}$ (the least squares estimates vector), i. e., the mean square error is smaller for $\hat{\beta}^*$ than for $\hat{\beta}$. An introduction to this method was elaborated in the thesis. It seeks to clarify the theoretical bases for the method.

The existence of importante cases of multicollinearity in the sample data was verified through a ridge trace and the analysis of "parameters variance inflation factors".

Income and price elasticity coefficients were obtained for demand functions of food products. They showed demand for them to be inelastic.

Results corresponding to the corn production function demonstrated that all factors were being utilized in the rational stage of production, and that the variables "seeds" and "defrayed costs" were being under-utilized.

It was concluded that the ridge regression method is a valid alternative for analyses of data with serious problems of multicollinearity. More satisfactory results were obtained in applying this method to production functions than to demand functions.

## BIBLIOGRAFIA CONSULTADA


ARAUJO, P.F.C. et allii, 1974. Crescimento e Desenvolvimento da Agricultura Paulista. Agricultura em São Paulo. Ano XXI, Tomo III, p. 169-199.

BISERRA, J.V., 1971. Análise de Relações Fator-Produto na Cultura do Milho em Jardinópolis e Guaíra, Estado de São Paulo, Ano Agrícola 1969/70. Piracicaba, ESALQ/USP, 119 p.. (Tese de Mestrado).

BROWN, W.G. e B.R. BEATTIE, 1975. Improving Estimates of Economic Parameters by Use of Ridge Regression with Production Function Applications. American Journal of Agricultural Economics. 57:21-32.

CAMARGO, J.R.V., 1974. Análise da Produtividade nas Culturas de Algodão e Soja com a Aplicação do Modelo Ulveling-Fletcher. Piracicaba, ESALQ/USP, 131 p.. (Tese de Mestrado).

DOLL, J.P., 1974. On Exact Multicollinearity and the Estimation of the Cobb-Douglas Production Function. American Journal of Agricultural Economics. 56:556-563.

DURBIN, J. e G.S. WATSON, 1951. Testing for Serial Correlation in Least Squares Regression II. Biometrika. 38:159-178.

ENGLER, J.J.C., 1968. Análise da Produtividade de Recursos na Agricultura. Piracicaba, ESALQ/USP, 102 p.. (Tese de Doutoramento).

FARRAR, D.E. e R.R. GLAUBER, 1967. Multicollinearity in Regression Analysis: The Problem Revisited. The Review of Economics and Statistics. 49:92-107.

GIRÃO, J.A., 1965. A Função de Produção de Cobb-Douglas e a Análise Inter-Regional da Produção Agrícola. Lisboa, Fundação Calouste Gulbenkian, Centro de Estudos de Economia Agrária, 117p..

HEADY, E.O. e J.L. DILLON, 1969. Agricultural Production Functions. Ames, The Iowa State University Press, 667 p..

HOERL, A.E., 1962. Aplication of Ridge Analysis to Regression Problems. Chemical Engineering Progress. 58:54-59.

HOERL, A.E., 1964. Ridge Analysis. Chemical Engineering Progress Symposium. 60:67-77.

HOERL, A.E. e R.W. KENNARD, 1970 (a). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. 12:55-67.

HOERL, A.E. e R.W. KENNARD, 1970 (b). Ridge Regression: Applications to Nonorthogonal Problems. Tecnometrics. 12:69-82.

HOFFMANN, R., 1973. Análise de Regressão: Uma Introdução à Econometria. Piracicaba, ESALQ/USP, Departamento de Ciências Sociais Aplicadas.

JOHNSTON, J., 1972. Econometric Methods. Segunda Edição, New York, Mc Graw-Hill Book Company, 437 p..

JUNQUEIRA, P.C., 1964. Demand Analysis for Selected Agricultural Products in the State of São Paulo. Ohio, U.S.A., The Ohio State University, 174 p.. (Tese de Master of Science).

KALIL, M.N., 1977. Programa "Análise de Regressão pelo Método de Cumeeira ('Ridge Regression')". Piracicaba, ESALQ/USP, 29 p..

KLEIN, L.R., 1962. An Introduction to Econometrics. Engle wood Cliffs, Prentice-Hall Inc., 280 p..

KMENTA, J., 1971. Elements of Econometrics. New York, The Mac Millan Company, 655 p..

McDONALD, G.C. e D.I. GALARNEAU, 1975. A Monte Carlo Evaluation of Some Ridge-Tipe Estimators. Journal of the American Statistical Association. 70:407-417.

MARQUARDT, D.W., 1970. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. Tecnometrics. 12:591--612.

NEELEMAN, D., 1973. Multicollinearity in Linear Economic Models. The Netherlands, Tilburg University Press, 103 p..

PEREZ, M.C.R.C., 1973. Contribuição ao Estudo da Elasticidade-Renda do Consumo de Alimentos. Piracicaba, ESALQ/USP, 94 p.. (Tese de Mestrado).

SOBRAL, G., 1973. Demanda de Alimentos no Brasil: Arroz, Batatinha, Feijão, Carne e Leite, Período 1950-70. Piracicaba, ESALQ/USP, 133 p.. (Tese de Mestrado).

THEIL, H., 1971. Principles of Econometrics. New York, John Wiley & Sons, Inc., 736 p..

TORO-VIZCARRONDO, C. e T.D. WALLACE, 1968. A Test of the Mean Square Error Criterion for Restrictions in Linear Regression. Journal of the American Statistical Association. 63:558-572.

VAUGH, F.V., 1973. Análise de Demanda e Preços na Agricultura. Piracicaba, ESALQ/USP, Departamento de Ciências Sociais Aplicadas, 192p..

WONNACOTT, R.J. e T.H. WONNACOTT, 1970. Econometrics. New York, John Willey & Sons, Inc., 445 p..

WRIGHT, C.L., 1973. Análise Econômica de Adubação em Culturas Anuais na Região de Ribeirão Preto, Ano Agrícola 1971/72. Piracicaba, ESALQ//USP, 162 p.. (Tese de Mestrado).

YAMANE, T., 1974. Matemática para economistas. Segunda Edição, São Paulo, Editora Atlas S.A., 656 p..

# A P Ê N D I C E     A

## MÉTODO DE REGRESSÃO DE CUMEEIRA

# Método de Regressão de Cumeeira

## 1. Estimador de Cumeeira

Considerando o modelo linear de regressão múltipla

$$y = X\beta + u \tag{1}$$

com as variáveis normalizadas de forma que $X'X$ e $X'y$ se apresentem como matrizes de correlações simples, o estimador de cumeeira $(\hat{\beta}^*)$ é definido como sendo:

$$\hat{\beta}^* = (X'X + kI_p)^{-1}X'y \tag{2}$$

$$0 \leqslant k \leqslant 1.$$

Sendo z o vetor com as variáveis dependentes centradas e V a matriz das variáveis independentes também centradas, tem-se que

$$y = z(z'z)^{-\frac{1}{2}} \tag{3}$$

$$X = VH^{-1} \tag{4}$$

onde H é a matriz

$$H = \begin{bmatrix} \sqrt{\Sigma v_{1j}^2} & 0 & \cdots & 0 \\ 0 & \sqrt{\Sigma v_{2j}^2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sqrt{\Sigma v_{pj}^2} \end{bmatrix}$$

Substituindo, em (1), o vetor y por (3) e a matriz X por (4), vem:

$$z(z'z)^{-\frac{1}{2}} = VH^{-1}\beta + u$$

Logo

$$z = V(z'z)^{\frac{1}{2}} H^{-1}\beta + u(z'z)^{\frac{1}{2}}$$

Denominando θ o coeficiente de regressão com as variáveis originais, e s os erros originais, fica

$$z = V\theta + s$$

onde

$$\theta = (z'z)^{\frac{1}{2}} H^{-1}\beta$$

$$s = u(z'z)^{\frac{1}{2}}$$

Então o estimador de cumeeira com as variáveis na forma original, é

$$\hat{\theta}^* = (z'z)^{\frac{1}{2}} H^{-1} \hat{\beta}^* \qquad (5)$$

Por outro lado, substituindo em (2) o vetor y por (3) e a matriz X por (4), tem-se

$$\hat{\beta}^* = (H^{-1}V'VH^{-1} + kI_p)^{-1} H^{-1}V'z(z'z)^{-\frac{1}{2}}$$

Substituindo, em (5), $\hat{\beta}^*$ pela expressão acima, resulta

$$\hat{\theta}^* = (z'z)^{\frac{1}{2}} H^{-1}(H^{-1}V'VH^{-1} + kI_p)^{-1} H^{-1}V'z(z'z)^{-\frac{1}{2}}$$

$$= \left[H(H^{-1}V'VH^{-1} + kI_p)H\right]^{-1}V'z$$

$$= (V'V + kHH)^{-1}V'z$$

Denominando

$$\Lambda = kHH$$

tem-se

$$\Lambda = k \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & \lambda_p \end{bmatrix}$$

onde $\lambda_i = \sum_{j=1}^{n} v_{ij}^2$, $i=1,2,\ldots,p$

Então o estimador de cumeeira com as variáveis na forma original é

$$\hat{\theta}^* = (V'V + \Lambda)^{-1}V'z$$

## 2. Esperança do Estimador de Cumeeira

Em (2), pode-se denominar

$$W = (X'X + kI_p)^{-1} \tag{6}$$

Então tem-se

$$\hat{\beta}^* = WX'y \tag{7}$$

Substituindo, em (7), y por (1), fica

$$\begin{aligned} \hat{\beta}^* &= WX'(X\beta + u) \\ &= WX'X\beta + WX'u \\ &= \beta - (I - WX'X)\beta + WX'u \end{aligned}$$

Mas, de (6), tem-se que

$$X'X = W^{-1} - kI \tag{8}$$

Então

$$\hat{\beta}^* = \beta - (I - I + kW)\beta + WX'u$$

Portanto

$$\hat{\beta}^* = \beta - kW\beta + WX'u \tag{9}$$

Sendo $E(u) = 0$, segue-se que

$$E(\hat{\beta}^*) = \beta - kW\beta \qquad (10)$$

$$E(\hat{\beta}^*) - \beta = -kW\beta \qquad (11)$$

$$= -k(X'X + kI)^{-1}\beta$$

A expressão (11) evidencia que a estimativa de cumeeira é viesada.

3. Variâncias das Estimativas dos Parâmetros

A matriz de variâncias e covariâncias de $\hat{\beta}^*$ é

$$\text{var-cov}(\hat{\beta}^*) = E\left[\hat{\beta}^* - E(\hat{\beta}^*)\right]\left[\hat{\beta}^* - E(\hat{\beta}^*)\right]'$$

De (9) e (10), vem

$$\hat{\beta}^* - E(\hat{\beta}^*) = WX'u$$

Portanto

$$\text{var-cov}(\hat{\beta}^*) = E\left[(WX'u)(WX'u)'\right]$$

$$= E(WX'uu'XW')$$

$$= \sigma_u^2 WX'XW \qquad (12)$$

pois W sendo simétrica, $W' = W$, e $E(uu') = I\sigma_u^2$ .

Então

$$\text{var-cov}(\hat{\beta}^*) = \sigma_u^2(X'X + kI)^{-1}(X'X)(X'X + kI)^{-1}$$

4. Esperança do Quadrado do Desvio

Para o vetor de parâmetros $\hat{\beta}^*$, a esperança do quadrado do desvio é:

$$EQD(\hat{\beta}^*) = E\left[(\hat{\beta}^* - \beta)'(\hat{\beta}^* - \beta)\right]$$

De (9), obtém-se:

$$\hat{\beta}^* - \beta = -kW\beta + WX'u$$

Portanto

$$EQD(\hat{\beta}^*) = E\left[(-\beta'W'k + u'XW')(-kW\beta + WX'u)\right]$$

$$= E(k^2\beta'W'W\beta - k\beta'W'WX'u - ku'XW'W\beta + u'XW'WX'u)$$

$$= E\left[k^2\beta'W'W\beta + tr(u'XW'WX'u)\right]$$

pois $E(u) = 0$ e $u'XW'WX'u$ é de ordem 1x1.

Sendo

$$E(uu') = I\sigma_u^2$$

e também

$$tr(u'XW'WX'u) = tr(WX'uu'XW')$$

segue-se que

$$EQD(\hat{\beta}^*) = k^2\beta'W'W + \sigma_u^2 tr(WX'XW')$$

Mas $W' = W$. Então, lembrando (11) e (12), vem

$$EQD(\hat{\beta}^*) = \Sigma\left[viés(\hat{\beta}_i^*)\right]^2 + \Sigma var(\hat{\beta}_i^*)$$

5. Soma de Quadrados dos Resíduos para o Estimador de Cumeeira

Considerando vários níveis de k, a soma de quadrados dos resíduos da equação estimada pelo método de regressão de cumeeira é

$$S^*(k) = (y - X\hat{\beta}^*)'(y - X\hat{\beta}^*)$$

Efetuando a multiplicação das expressões entre parêntesis, fica

$$S^*(k) = y'y - (\hat{\beta}^*)'X'y - y'X\hat{\beta}^* + (\hat{\beta}^*)'X'X\hat{\beta}^*$$

De (8), obtém-se

$$S^*(k) = y'y - (\hat{\beta}^*)'X'y - y'X\hat{\beta}^* + (\hat{\beta}^*)'(W^{-1} - kI_p)\hat{\beta}^*$$

$$= y'y - (\hat{\beta}^*)'X'y - y'X\hat{\beta}^* + (\hat{\beta}^*)'W^{-1}\hat{\beta}^* - k(\hat{\beta}^*)'(\hat{\beta}^*)$$

Considerando (7), vem

$$S^*(k) = y'y - (\hat{\beta}^*)'X'y - y'X\hat{\beta}^* + y'XWW^{-1}\hat{\beta}^* - k(\hat{\beta}^*)'(\hat{\beta}^*)$$

Portanto:

$$S^*(k) = y'y - (\hat{\beta}^*)'X'y - k(\hat{\beta}^*)'(\hat{\beta}^*) \qquad (13)$$

Como se está utilizando as variáveis normalizadas, a soma de quadrados dos resíduos pode ser expressa da seguinte forma

$$S*(k) = 1 - \hat{\beta}^{*\prime}X'y - k\left[\Sigma(\hat{\beta}_i^*)^2\right]$$

Esta fórmula é útil para fins de cálculo.

A equação correspondente às variáveis originais é

$$S^*(\Lambda) = (z - V\hat{\theta}^*)'(z - V\hat{\theta}^*)$$

De acordo com (3), (4) e (5), vem:

$$S^*(\Lambda) = \left[y(z'z)^{\frac{1}{2}} - XH(z'z)^{\frac{1}{2}} H^{-1}\hat{\beta}^*\right]' \left[y(z'z)^{\frac{1}{2}} - XH(z'z)^{\frac{1}{2}} H^{-1}\hat{\beta}^*\right]$$

$$= (z'z)^{\frac{1}{2}} \left[y - X\hat{\beta}^*\right]' (z'z)^{\frac{1}{2}} \left[y - X\hat{\beta}^*\right]$$

$$= (z'z) \left[y - X\hat{\beta}^*\right]' \left[y - X\hat{\beta}^*\right]$$

$$= (z'z)S^*(k)$$

Prova-se que a equação acima corresponde a

$$S^*(\Lambda) = z'z - (\hat{\theta}^*)'V'z - (\hat{\theta}^*)' \Lambda(\hat{\theta}^*)$$

## 6. A Definição do Coeficiente de Regressão Múltipla no Caso da Regressão de Cumeeira

Considerando que a soma de quadrados da regressão de cumeeira é:

$$\Sigma\hat{y}_j^{*2} = (X\hat{\beta}^*)'(X\hat{\beta}^*)$$

$$= (\hat{\beta}^*)'X'X\hat{\beta}^*$$

De acordo com (8), vem

$$\Sigma\hat{y}_j^{*2} = (\hat{\beta}^*)'(W^{-1} - kI)\hat{\beta}^*$$

$$= (\hat{\beta}^*)'W^{-1}\hat{\beta}^* - k(\hat{\beta}^*)'\hat{\beta}^*$$

Mas, de (7), tem-se que:

$$\beta^* = WX'y$$

Então

$$\Sigma \hat{y}_j^{*2} = (\hat{\beta}^*)'X'y - k(\hat{\beta}^*)'\hat{\beta}^* \tag{14}$$

Definindo-se o coeficiente de determinação múltipla como sendo

$$R^2 = 1 - \frac{S^*(k)}{y'y}$$

tem-se, de acordo com (13)

$$R^2 = \frac{(\hat{\beta}^*)'X'y + k(\hat{\beta}^*)'\hat{\beta}^*}{y'y} \tag{15}$$

Por outro lado, se se definir

$$R^2 = \frac{SQRegr}{y'y}$$

tem-se, de acordo com (14):

$$R^2 = \frac{(\hat{\beta}^*)'X'y - k(\hat{\beta}^*)'\hat{\beta}^*}{y'y} \tag{16}$$

Outra alternativa seria definir

$$R^2 = \frac{(\hat{\beta}^*)'X'y}{y'y} \tag{17}$$

por analogia com o caso de uma regressão múltipla ajustada pelo método de mínimos quadrados. Os coeficientes de determinação múltipla apresentados neste trabalho foram calculados através de (17).

## 7. Comprimento de $\hat{\beta}^*$

Para determinado valor da soma de quadrados dos resíduos, pode-se demonstrar que $\hat{\beta}^*$ é a estimativa de $\beta$ que possui o menor comprimento.

Considerando b como sendo uma estimativa qualquer de $\beta$, a soma de quadrados do resíduo correspondente é

$$S = (y - Xb)'(y - Xb)$$

Efetuando as multiplicações, e somando e subtraindo $\hat{\beta}'X'y$ à equação acima, vem:

$$S = y'y - \hat{\beta}'X'y + b'X'Xb - b'X'X(X'X)^{-1}X'y - y'X(X'X)^{-1}X'Xb + \hat{\beta}'X'X(X'X)^{-1}X'y$$

Substituindo $(X'X)^{-1}X'y$ por $\hat{\beta}$, e fatorando, vem

$$S = (y - X\hat{\beta})'(y - X\hat{\beta}) + (b - \hat{\beta})'X'X(b - \hat{\beta})$$

Para determinar o valor de b que irá minimizar $b'b$ sujeito a

$$(b - \hat{\beta})'X'X(b - \hat{\beta}) = \gamma$$

forma-se a função

$$F = b'b + (1/k)\left[(b - \hat{\beta})'X'X(b - \hat{\beta}) - \gamma\right]$$

onde $(1/k)$ é o multiplicador de Lagrange.

Então

$$\frac{\partial F}{\partial b} = 2b + (1/k)\left[2(X'X)b - 2(X'X)\hat{\beta}\right] = 0$$

Lembrando novamente que

$$\hat{\beta} = (X'X)^{-1}X'y$$

e substituindo na expressão anterior, resulta:

$$b = (X'X + kI_p)^{-1}X'y$$

Mas este é o estimador de cumeeira. Então conclui-se que $\hat{\beta}^*$ é o valor de $\beta$ que minimiza o quadrado do comprimento do vetor das estimativas de parâmetros, para um dado valor da soma de quadrados do resíduo.

# A P Ê N D I C E B

*PREÇOS DOS FATORES*
*MÃO-DE-OBRA E*
*MÁQUINAS-IMPLEMENTOS*

1. Mão-de-Obra

Segundo publicação do Instituto de Economia Agrícola, o salário de diarista residente em 1970 era Cr$154,05 e em 1969 era Cr$116,25 (ARAUJO, 1974, p. 192). Então tomando-se para 1969/70 a média desses valores, obtém-se Cr$135,10. Considerando 25 dias úteis por mês, a diária era de Cr$5,40, aproximadamente.

2. Máquinas e Implementos

Considera-se que o rendimento de cada unidade monetária investida em máquinas e implementos proporcionaria juros a uma taxa de 17% ao ano. Então o preço por cruzeiro investido nesse fator é Cr$1,17.