REVISTA BRASILEIRA DE ESTATÍSTICA

Órgão oficial do IBGE
e da Sociedade Brasileira de Estatística

A Revista não se responsabiliza
pelos conceitos emitidos
em artigos assinados

PUBLICAÇÃO TRIMESTRAL

ISSN 0034-7175

Pedidos de assinatura e número avulso ou atrasado para:

Diretoria de Administração — Sergraf: Av. Brasil, 15.671 — Lucas — Rio de Janeiro — Brasil
CEP — 21.241
Tel.: 391-7788

Livraria do IBGE: Av. Franklin Roosevelt, 146/loja — Centro — RJ — Brasil
CEP — 20.021
Tel.: 220-9147 / 220-8163
DDD: 021

SUMÁRIO

Artigos

As variáveis binárias em análise de regressão: teoria e aplicação
José Welisson Rossi 231

Ajuste sazonal para os indicadores de produção física
Departamento de Indicadores Conjunturais (DEICO) 319

Bibliografia

Publicações de interesse para a Estatística editadas pelo IBGE no período de abril a junho de 1983 377

ISSN 0034-7175

R. bras. Estat.	Rio de Janeiro	v. 44	n.º 175	p. 229 a 378	jul./set. 1983

FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA — IBGE
Av. Franklin Roosevelt, 166 — Centro
20 021 — Rio de Janeiro, RJ — Brasil

ISSN 0034-7175

Revista brasileira de estatística / Fundação Instituto Brasileiro de Geografia e Estatística · — ano 1, n. 1(1940, jan./mar.)- . —

Rio de Janeiro : IBGE, 1940-

Trimestral.

Órgão oficial do IBGE e da Sociedade Brasileira de Estatística.

Continuação de : Revista de economia e estatística.

Índices : autor-assunto, v. 22-24(1961-1963) no v. 25, n. 1 ; v. 25-26 (1964-1965) no v. 27, n. 1 ; v. 27-28(1966-1967) no v. 29, n. 1 ; . Índices anuais de autor-assunto, publicados no n. 1 (jan./mar.) do ano seguinte (período 1968-1977, v. 29-38).

ISSN 0034-7175 = Revista brasileira de estatística.

1. Estatística — Periódicos. I. IBGE.

IBGE. Biblioteca Central
RJ-IBGE/81-41

CDU 31(05)

Impresso no Brasil/Printed in Brazil

ARTIGOS

AS VARIÁVEIS BINÁRIAS EM ANÁLISE DE REGRESSÃO: TEORIA E APLICAÇÃO*

José Welisson Rossi **

SUMÁRIO

1 — Introdução
2 — Mudanças nos parâmetros do modelo linear
3 — Equivalência entre análise de variância e análise de regressão com variáveis binárias
4 — Dessazonalização de séries
5 — As variáveis binárias na previsão da regressão
6 — A regressão com variável dependente qualitativa
7 — Apêndices
8 — Bibliografia

1 — INTRODUÇÃO

As variáveis binárias ou *dummy* são assim denominadas por assumirem os valores de zero e um (ou -1 e $+1$) quando utilizadas em análise de regressão. No seu papel de variáveis explicativas da regressão,

* O professor Rodolfo Hoffmann leu os manuscritos do trabalho e além de apontar alguns erros, fez várias sugestões úteis. Obviamente se erros persistem são de inteira responsabilidade do autor.

** Professor da Universidade Federal do Rio de Janeiro — UFRJ. O autor agradece ao professor Cesar das Neves pela orientação dada à parte computacional dos itens 5.2.1 e 5.3.1.

elas procuram captar a influência de fatores qualitativos (por exemplo, a influência do sexo do indivíduo na demanda por certo produto). Mas as variáveis binárias são também empregadas na regressão como um instrumento na eliminação de influências sazonais porventura presentes numa série histórica de dados, além de poderem ainda ser utilizadas numa construção artificial, cujo objetivo principal é a obtenção prática de previsões na regressão. Em todos esses três casos as variáveis entram na regressão como variáveis independentes. Como variável dependente na regressão, por outro lado, as variáveis binárias podem ser úteis em modelos que pretendem explicar a posse de um dado bem do tipo casa própria, por exemplo.

Todos esses aspectos, além de alguns outros, serão exaustivamente discutidos nas seções seguintes deste estudo, que representa um esforço no sentido de fornecer um tratamento unificado dos usos mais importantes das variáveis binárias em análise de regressão. Na verdade, os tópicos aqui discutidos encontram-se dispersos na literatura acadêmica, alguns sendo abordados nos livros textos de Econometria geralmente em uso, outros porém, de desenvolvimento mais recente, estão disponíveis apenas em revistas especializadas. Na primeira dessas categorias encontramos, por exemplo, a utilização das variáveis binárias para captar mudanças nos parâmetros da regressão, que é discutido na seção 2. Este é também o contexto mais conhecido do uso das variáveis binárias em análise de regressão. Na segunda, temos o uso das variáveis binárias em estudo de previsão. Tal assunto (tratado na Seção 5), por ser de alto interesse prático e não ter sido ainda incorporado em nenhum livro texto, mereceu uma boa parte do espaço da nossa exposição e é, provavelmente, o tópico tecnicamente mais difícil; na tentativa de tornar a leitura do texto mais amena, entretanto, restringimos os aspectos mais técnicos da discussão ao apêndice técnico apropriado.

Em resumo, além da preocupação em oferecer um tratamento unificado do uso das variáveis binárias em análise de regressão, procuramos também executar tal tarefa de maneira tanto quanto possível didática.

O público a que nos dirigimos é, sobretudo, aquele constituído pelos estudantes de Economia (graduação e pós-graduação), já que têm Econometria como disciplina regular nos seus currículos. Entretanto, notas de aula preliminares cobrindo alguns dos tópicos aqui tratados foram recentemente utilizados pelo autor, com bons resultados, em cursos de Estatística Aplicada e Econometria, na Universidade Federal do Rio de Janeiro, a nível tanto de graduação como de pós-graduação.

2 — MUDANÇAS NOS PARÂMETROS DO MODELO LINEAR

Um exemplo freqüentemente apresentado para a discussão da mudança de parâmetros de um modelo, entre dois períodos distintos, é o

da função de consumo agregado em anos de paz e de guerra. O modelo é, na sua forma mais simples, dado por:

$$y_i = a + bx_i + \mu_i \,, \tag{2-1}$$

onde y_i representa o consumo; x_i a renda e μ_i o erro estocástico (para $i = 1, 2, \ldots, n$). Para simplificar a notação, daqui em diante omitiremos, nesta seção, o subíndice i na regressão.

À luz do modelo em (2-1), vamos supor que tenha havido modificações no comportamento do consumo entre os períodos de paz e guerra. Tais modificações poderiam assumir as seguintes características:

1 — deslocamento no nível (interseção) da função, sem alteração na sua inclinação;

2 — mudança na inclinação, com mesma interseção;

3 — alteração em ambos os parâmetros: a — com descontinuidade na função e b — sem descontinuidade na função.

Vejamos cada uma dessas situações por seu turno.

2.1 — Deslocamento do nível da função sem alteração na sua inclinação

Essa situação poderia ser convenientemente representada como na figura 1, a seguir:

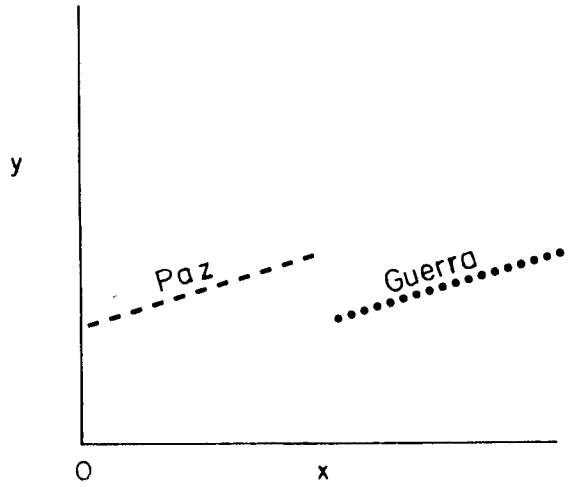

FIGURA 1

A função em (2-1) poderia ser modificada, para representar esta mudança de comportamento, como segue:

$$y = a + bx + cD + \mu \tag{2-2}$$

onde $D = \begin{cases} 0 \text{ em anos de paz} \\ 1 \text{ em anos de guerra} \end{cases}$

(Nota: não faz diferença que período recebe o valor zero).

Assim, a inclusão da variável D permite, com um modelo apenas, representar duas funções distintas:

$$y = a + bx + \mu \tag{2-3}$$

para os anos de paz, e

$$y = (a + c) + bx + \mu \tag{2-4}$$

para os anos de guerra.

Deve ser observado que, caso os fatos demonstrem realmente um deslocamento da função, então os estimadores de Mínimos Quadrados Ordinários (MQO) são mais eficientes em (2-2) que aqueles obtidos com dois ajustamentos separados: um para (2-3) e outro para (2-4); por uma simples razão: se a inclinação não se modificou nos dois períodos, então dispõe-se de maior número de observações para estimar a inclinação em (2-2) que em (2-3) ou em (2-4).

Para submeter a hipótese do deslocamento da função ao teste estatístico, basta testar a hipótese de que o parâmetro c em (2-2) seja zero. Tal teste é, naturalmente, realizado através do uso da distribuição de Student, que tem aplicação no exemplo ilustrativo, discutido no final desta seção.

2.2 — Mudança de inclinação com mesma interseção

A representação gráfica dessa situação será omitida por parecer suficientemente óbvia. Matematicamente, a função assumiria a seguinte representação:

$$y = a + bx + cDx + \mu \ , \tag{2-5}$$

onde D tem os mesmos valores que em (2-2). Desta forma, a função dos anos de paz permaneceria como em (2-3) e a dos anos de guerra seria:

$$y = a + (b + c)x + \mu \ . \tag{2-6}$$

A hipótese de mudança na inclinação poderá ser efetuada através do teste do parâmetro c na função (2-5), ainda com o uso da distribuição de Student. As mesmas considerações, sobre a eficiência da estimação, apresentadas acima, seriam também válidas aqui.

2.3 — Alteração em ambos os parâmetros

2.3.1 — Com descontinuidade da função

Um modelo capaz de conter tais mudanças seria:

$$y = a + bx + cD + dDx + \mu \ , \qquad (2\text{-}7)$$

onde D é definido como em (2-2). Note-se que a função dos anos de paz permaneceria como em (2-3) e a dos anos de guerra seria:

$$y = (a + c) + (b + d)x + \mu \ . \qquad (2\text{-}8)$$

Uma vez mais, poderíamos testar as hipóteses de que c e/ou d, em (2-7), são zero. Tais hipóteses seriam testadas através da distribuição de Student e de Fischer (F), conforme sejam os parâmetros testados individualmente ou simultaneamente {Kmenta (1971: 366-71)}.

Um aspecto interessante deste último modelo é que os mesmos estimadores de MQO obtidos para as funções de paz e guerra através de (2-7), seriam também obtidos com ajustamentos separados de (2-3) e (2-8) aos respectivos períodos de paz e guerra. A confirmação deste resultado é obtida no exemplo ilustrativo e uma prova matemática é fornecida no apêndice técnico.

Cumpre salientar que, apesar dessa igualdade nos valores dos estimadores que acabamos de mencionar, o mesmo não poderia ser dito sobre as estimativas das suas variâncias. Estas provavelmente seriam diferentes num caso e noutro, tendo, portanto, implicações nos testes de hipóteses sobre os parâmetros. Aliás, mesmo que as variâncias fossem idênticas, ainda assim as decisões nos testes de hipóteses poderiam diferir, dado serem as amostras baseadas em tamanhos distintos nos dois casos. O exemplo ilustrativo e o apêndice técnico, exploram estes pontos mais detalhadamente.

2.3.2 — Sem descontinuidade na função [1]

Caso deseje-se que a função tenha a sua inclinação alterada entre dois períodos, mas sem sofrer qualquer descontinuidade entre esses

1 Para uma discussão mais detalhada sobre esta questão ver, por exemplo, Hoffmann & Vieira (1983:208-15).

períodos (separados na Figura 2 pelo ponto representado por x^*) então a especificação apropriada seria:

$$y = a + bx + c(x - x^*)D + \mu \tag{2-9}$$

para $D = \begin{cases} 1 \text{ para observações com } x \geqslant x^* \\ 0 \text{ de outro modo} \end{cases}$

A representação gráfica da especificação em (2-9) é mostrada em seguida (Figura 2).

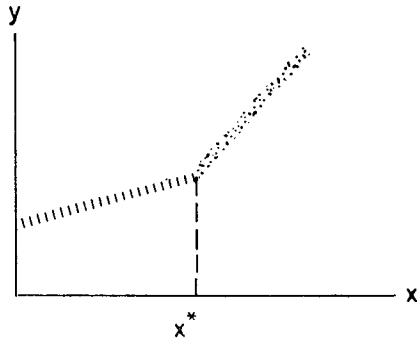

FIGURA 2

Note-se que, para $x \geqslant x^*$, temos

$$E(y) = (a - cx^*) + (b + c)x \tag{2-10}$$

que para $x = x^*$ produz

$$E(y^*) = a + bx^* \tag{2-11}$$

isto é, parâmetros idênticos aos prevalescentes para o primeiro segmento — vale dizer, as retas dos dois segmentos efetivamente se interceptam no ponto $x = x^*$.

Finalmente, observe-se que, como as inclinações nos dois segmentos são dadas respectivamente por b e $b + c$, então é claro que, pela hipótese nula $H_0: c = o$, testa-se a igualdade dessas inclinações na regressão.

2.4 — Generalizações dos modelos

Os modelos discutidos acima poderão ser estendidos para situações envolvendo mais de dois períodos distintos e com mais de um regressor, sem dificuldade. Por exemplo, adaptando-se (2-7) para o caso com dois regressores (x_1 e x_2) e três períodos distintos (antes, durante e após a guerra), teríamos:

$$\begin{aligned} y = a + bx_1 + cx_2 + dD_1 + eD_2 + fD_1x_1 + gD_2x_1 + \\ + hD_1x_2 + jD_2x_2 + \mu \end{aligned} \tag{2-12}$$

onde $D_1 = \begin{cases} 1 \text{ durante a guerra} \\ 0 \text{ outros anos} \end{cases}$ e $D_2 = \begin{cases} 1 \text{ após a guerra} \\ 0 \text{ outros anos} \end{cases}$

Temos portanto:

$$y = a + bx_1 + cx_2 + \mu \tag{2-13}$$

para antes da guerra,

$$y = (a + d) + (b + f)x_1 + (c + h)x_2 + \mu \tag{2-14}$$

para durante a guerra e

$$y = (a + e) + (b + g)x_1 + (c + j)x_2 + \mu \tag{2-15}$$

para após a guerra. (Observe-se que usamos sempre uma variável D a menos que o número de categorias representadas. Isso para evitar o problema da perfeita multicolinearidade, que impossibilita a estimação dos parâmetros de MQO. Retornaremos a este ponto na Seção 3).

A extensão desse modelo, para casos mais gerais, envolvendo maior número de regressores e de períodos, parece suficientemente óbvia para merecer maiores considerações aqui. Também os testes estatísticos sobre os parâmetros do modelo geral seguem a mesma orientação dos casos já abordados para o modelo mais simples.

Quanto à adaptação do modelo em (2-9) para o caso onde temos três segmentos de reta (isto é, onde há a possibilidade de obter-se inclinações distintas nesses segmentos mas sem que haja qualquer descontinuidade na função) a especificação apropriada seria:

$$y = a + bx + c_1D_1(x - x^*) + c_2D_2(x - x^{**}) + \mu \tag{2-16}$$

onde $D_1 = \begin{cases} 1 \text{ para observações com } x \geqslant x^* \\ 0 \text{ de outro modo} \end{cases}$

e $D_2 = \begin{cases} 1 \text{ para observações com } x \geqslant x^{**} \\ 0 \text{ de outro modo.} \end{cases}$

A representação gráfica desse modelo é apresentada na figura 3, a seguir.

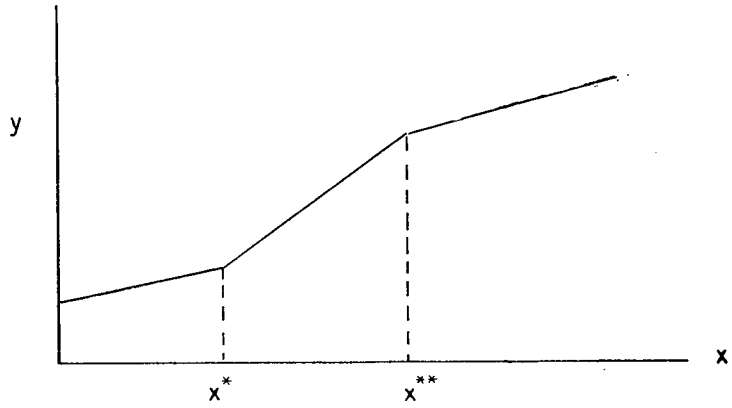

FIGURA 3

Como as inclinações nos três segmentos, no modelo em (2-16), são agora dadas respectivamente por b, $b + c_1$ e $b + c_1 + c_2$, então podemos testar se: 1 — as inclinações são idênticas entre o primeiro e o segundo segmentos (isto é, H_0: $c_1 = 0$); 2 — se as inclinações são iguais entre o segundo e terceiro segmentos (isto é, H_0: $c_2 = 0$); e 3 — se são as inclinações iguais no primeiro e terceiro segmentos (isto é, H_0: $c_1 + c_2 = 0$).

A especificação (e correspondentes testes de hipóteses) para o caso mais geral com mais de três segmentos parece óbvia e será, portanto, omitida aqui.

2.5 — Exemplos

Com o objetivo de ilustrar muitos dos pontos discutidos acima, consideremos agora algumas aplicações:

a — inicialmente, para ilustrar as seções (2-1), (2-2) e (2-3), sejam os dados {Wesolowsky (1976:106)}:

y	10	10	11	13	14	14	16	16	15	16	16	17	17	17	18
x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Uma representação gráfica desses dados nas coordenadas (x, y) mostraria uma mudança de comportamento da relação entre as variáveis a partir de $x = 9$. O ajustamento de MQO do modelo (2-1) aos dados produziu:

$$\hat{y} = 10{,}30 + 0{,}546\ x\ ;\ R^2 = 0{,}877\ . \tag{2-17}$$

Entretanto, dado o deslocamento da função, o ajustamento do modelo (2-2) parece mais apropriado:

$$\hat{y} = \underset{(0{,}512)}{9{,}59} + \underset{(0{,}095)}{0{,}757}\ x - \underset{(0{,}823)}{2{,}110}\ D\ ;\ R^2 = 0{,}920 \tag{2-18}$$

onde $D = \begin{cases} 0 \text{ para } x < 9 \\ 1 \text{ para } x \geqslant 9 \end{cases}$

e onde os dados entre parênteses representam estimativas dos desvios padrão dos estimadores.

Portanto, de (2-14), temos que:

$$\hat{y} = 9{,}59 + 0{,}757\ x\ ,\ \text{para } x < 9 \tag{2-19}$$

e

$$\hat{y} = 7{,}49 + 0{,}757\ x\ ,\ \text{para } x \geqslant 9\ . \tag{2-20}$$

Note-se que a inclusão da variável D em (2-18) ocasionou um aumento sensível no valor do coeficiente de determinação (R^2) em relação aquele em (2-13). Tal resultado não deve surpreender aliás, pois a inclusão daquela variável permite uma maior flexibilidade para o ajustamento da reta aos pontos, dando-lhe uma clara opção de deslocamento nos dois períodos. Neste sentido, observe-se que o estimador do parâmetro da variável D é significativo ao nível de 5% — isto é, sob a hipótese nula de que $c = 0$, temos:

$t = -\frac{2{,}11}{0{,}82} = -2{,}57$, que equivale, aproximadamente, ao valor crítico de 2,5% da distribuição de Student com 12 graus de liberdade. Assim, podemos afirmar que o salto da função é estatisticamente significativo.

O ajustamento do modelo (2-7), por outro lado, produziu:

$$\hat{y} = \underset{(0{,}375)}{8{,}61} + \underset{(0{,}074)}{0{,}976}\, x + \underset{(1{,}168)}{2{,}820}\, D - \underset{(0{,}117)}{0{,}548}\, Dx \; ; \; R^2 = 0{,}973 \qquad (2\text{-}21)$$

resultando então em:

$$\hat{y} = 8{,}61 + 0{,}976\, x \text{ para } x < 9 \qquad (2\text{-}22)$$

e

$$\hat{y} = 11{,}43 + 0{,}429\, x \text{ para } x \geqslant 9\,. \qquad (2\text{-}23)$$

Vale a pena observar a melhoria do R^2 em (2-21) sobre aquele em (2-18). Este resultado já era esperado, pois a flexibilidade do modelo (2-21) é ainda maior que aquele em (2-18) pois, agora, tanto a interseção quanto a inclinação têm opção para variar entre os dois períodos, permitindo então que a reta se ajuste da melhor maneira possível dentro de cada um dos grupos. Neste contexto aliás (como é demonstrado no apêndice), as mesmas equações (2-22) e (2-23) seriam obtidas caso ajustássemos 2 funções, uma do tipo (2-3) para $x < 9$ e outra do tipo (2-8) para $x \geqslant 9$. Tais ajustamentos forneceram:

$$\hat{y} = \underset{(0{,}41)}{8{,}61} + \underset{(0{,}089)}{0{,}976}\, x \; ; \; R^2 = 0{,}953 \text{ para } x < 9 \qquad (2\text{-}24)$$

e

$$\hat{y} = 11{,}43 + 0{,}429\, x \; ; \; R^2 = 0{,}900 \text{ para } x \geqslant 9\,. \qquad (2\text{-}25)$$

Note-se que os estimadores em (2-24) e (2-22) são idênticos, e, também o são, aqueles em (2-25) e (2-23). Mas se comparássemos seus desvios padrão, então ficariam evidenciadas as suas diferenças, com implicações, naturalmente, nas decisões sobre os testes de hipóteses. Por exemplo, os desvios padrão em (2-24) são maiores que aqueles em (2-21). Se adicionarmos a isto o fato de que aqueles são baseados em

menor número de graus de liberdade (6 contra 11), então fica clara a existência de duas forças favorecendo os estimadores em (2-21) como mais significativos. Primeiramente, os valores calculados de t, sob a hipótese nula de que o parâmetro seja zero, serão maiores. Em segundo lugar, os valores críticos de t serão menores, já que são baseados numa distribuição menos espalhada. E como sabemos, as decisões nos testes de hipóteses decorrem da comparação entre os valores calculados e valores críticos da distribuição. Tomemos como exemplo o caso específico do parâmetro b. Das equações (2-21) e (2-24) temos, sob a hipótese nula de que $b = 0$ e com nível de significância de 5%, que:

$$t = \frac{0{,}976}{0{,}074} = 13, \text{ contra } t^{(11)} \text{ crítico} = 3{,}10$$

e

$$t = \frac{0{,}976}{0{,}089} = 10.8, \text{ contra } t^{(6)} \text{ crítico} = 3{,}70$$

(os expoentes de t denotam graus de liberdade).

É pois óbvio, neste caso, que a estimação é mais eficiente em (2-21). Mas mesmo que as variâncias fossem iguais, ainda assim esses estimadores seriam mais eficientes, como pode-se depreender da discussão acima, relativa à diferença de graus de liberdade das duas estimações.

b — Para ilustrar o caso de alterações, tanto nas interseções quanto nas inclinações da regressão, sem que haja qualquer salto na função, utilizaremos uma aplicação efetuada por Hoffmann (1983), relativa ao cálculo das elasticidades de Engel, para os dispêndios familiares na Cidade do Rio de Janeiro, entre 1974 e 1975 (dados do Estudo Nacional da Despesa Familiar, IBGE). Com efeito, seja a especificação para os dispêndios em alimentação, por exemplo, dada por:

$$Log\ y = a + b\ Logx + c_1D_1(Logx - Logx^*) + \\ + c_2D_2(Logx - Logx^{**}) + \mu$$

onde: y = dispêndio médio familiar no item alimentação em dada classe de renda;

x = dispêndio médio familiar com todos os itens do orçamento em dada classe;

x^* = 15.800 cruzeiros;

x^{**} = 31.600 cruzeiros;

$$D_1 = \begin{cases} 1 \text{ se } x \geqslant 15.800 \\ 0 \text{ de outro modo} \end{cases}$$

$$D_2 = \begin{cases} 1 \text{ se } x \geqslant 31.600 \\ 0 \text{ de outro modo} \end{cases}$$

μ = erro estocástico com as propriedades clássicas.

Os dados utilizados no ajustamento deste modelo têm as seguintes características básicas: os dispêndios são distribuídos entre 9 (nove) faixas de dispêndio familiar, sendo que a faixa mais elevada é aberta à direita, tendo como limite inferior o valor de 134.799 cruzeiros; há ainda 3 (três) faixas de dispêndio abaixo do valor de 15.800 cruzeiros, 2 (duas) classes entre os valores de 15.800 e 31.600 e 4 (quatro) classes com valores maiores que 31.600 cruzeiros.

Os resultados obtidos por Hoffmann (1983), no ajustamento do modelo, foram:

$$Log\ y = a + 0{,}98\ Logx - \underset{(-4{,}3)}{0{,}30}\ (Logx - Log\ 15.800) -$$

$$- \underset{(-5{,}4)}{0{,}33}\ (Logx - Log\ 31.600),$$

com $R^2 = 0{,}9985$ (dados entre parênteses são valores de t), permitindo obter as seguintes elasticidades nos três segmentos em que foram divididos os dispêndios:

$$\xi_1 = 0{,}98\ ;\ \xi_2 = 0{,}68 \text{ e } \xi_3 = 0{,}35\ .$$

Note-se que, como essas elasticidades equivalem no modelo aos parâmetros b, $b + c_1$ e $b + c_1 + c_2$, respectivamente, podemos testar então se tais elasticidades são estatisticamente diferentes entre o primeiro e o segundo segmento, e também entre o segundo e terceiro segmentos, bastando, para isso, testar as correspondentes hipóteses nulas,

$$H_0 : c_1 = 0 \text{ e } H_0 : c_2 = 0.$$

Assim, como os valores de t, para c_1 e c_2, são respectivamente de $-4{,}3$ e $-5{,}4$, são rejeitadas, ao nível de significância de 5%, as hipóteses de igualdade das elasticidades entre esses dois pares de segmentos da renda. Observe-se, ainda, que, para testarmos a significância estatística das elasticidades individuais, nos três segmentos, faz-se necessário o conhecimento dos desvios padrão de b, $b + c_1$ e $b + c_1 + c_2$ que seriam obtidos a partir da matriz de covariâncias dos estimadores da regressão.

c — Como ilustração final, sobre o uso das variáveis binárias para captar deslocamentos na linha de regressão, utilizaremos alguns dos resultados de uma aplicação deste autor (Rossi, 1981), relativa à questão da distribuição de renda no Brasil, no período 1970-1974. Com efeito, a partir dos dados regionais (10 regiões) das declarações de renda das pessoas físicas, obtivemos a seguinte relação gráfica (Figura 4) entre o

grau de concentração de renda {dado pelo índice de redundância de Theil (1967)} e o nível da renda média dos declarantes, em cada uma das regiões.

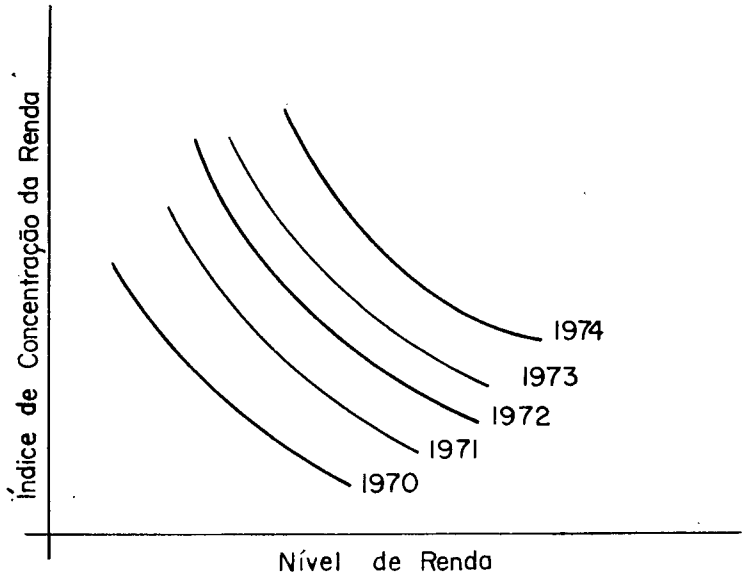

FIGURA 4

Isto é, parece existir uma relação parabólica entre as variáveis grau de concentração da renda e nível da renda, quando essas observações são consideradas num dado ano (isto é, num corte transversal). Além disso, parece clara também a existência de um deslocamento aproximadamente paralelo (para cima e para a direita) dessa relação com o passar dos anos. Assim, uma especificação apropriada poderia ser do tipo

$$y = a + b\,x + c\,x^2 + dD_1 + eD_2 + fD_3 + gD_4 + \mu$$

onde $D_i = \begin{cases} 1 \text{ se observações pertencem ao ano } i \\ 0 \text{ de outro modo} \end{cases}$

para $i = 1, 2, 3$ e 4 correspondendo aos anos 1970, 1971, 1972 e 1973, respectivamente; segue-se então que, se todos os $D_i = 0$, as observações referem-se ao ano de 1974.

Desta forma, essa especificação permite combinar dados do corte transversal com dados temporais. A vantagem deste procedimento é permitir uma estimação estatisticamente mais eficiente da relação parabólica, já que utilizamos todas as 50 observações na estimação da

relação em vez das 10 observações que seriam utilizadas caso adotássemos um corte transversal apenas. Quanto aos resultados, obtivemos

$$y = 2{,}15 - \underset{(-3{,}3)}{0{,}202x} + \underset{(2{,}4)}{0{,}009x^2} - \underset{(-11{,}9)}{0{,}452D_1} - \underset{(-7{,}9)}{0{,}270D_2} - \underset{(-7{,}3)}{0{,}247D_3} - \underset{(-0{,}4)}{0{,}042D_4} \; ; \; R^2 = 0{,}91$$

(Dados entre parênteses são valores de t).

Note-se que não só é, do ponto de vista estatístico, significativa a relação parabólica como também são significativos, em geral (a exceção aqui é a variável D_4), os deslocamentos da função ao longo dos anos.

3 — EQUIVALÊNCIA ENTRE A ANÁLISE DE VARIÂNCIA E A ANÁLISE DE REGRESSÃO COM VARIÁVEIS BINÁRIAS

3.1 — A análise de variância e a análise de regressão

Mostraremos nesta seção que a questão da análise de variância poderá ser tratada como um problema de regressão linear com variáveis binárias como regressores. A matéria será aqui exposta apenas no contexto da análise de variância com um fator interveniente (*one-way* ANOVA), mas a equivalência entre a análise de variância e a análise de regressão se estende também para os casos mais complexos {Kmenta (1971: 414-9)}.

Para os nossos propósitos, podemos afirmar que a análise de variância com um fator se constitui num procedimento estatístico que visa testar se dois ou mais grupos de dados apresentam a mesma média. Para tanto, a análise adota alguns pressupostos básicos {Costa Neto (1977: 153)} sobre os dados (por exemplo, que tem distribuição normal), que não cabe aqui analisar.

Com o objetivo de estabelecer a equivalência entre a análise de variância e a análise de regressão linear, utilizaremos uma especificação para dados classificados em três grupos. Mais precisamente:

$$y_{ij} = b_0 + b_1 D_{1_{ij}} + b_2 D_{2_{ij}} + \mu_{ij} \qquad (3\text{-}1)$$

para $i = 1, 2, \ldots, n_j$ e $j = 1, 2,$ e $3,$ e

onde $D_1 = \begin{cases} 1 \text{ se } y \text{ estiver no grupo } 1 \\ 0 \text{ se estiver nos outros grupos} \end{cases}$

e $D_2 = \begin{cases} 1 \text{ se } y \text{ estiver no grupo } 2 \\ 0 \text{ se estiver nos outros grupos.} \end{cases}$

Note-se que quando D_1 e D_2 são ambos zero, então y pertence ao grupo três, fazendo pois desnecessário o uso de uma terceira variável binária, D_3. Aliás, o uso de D_3 no modelo (3-1) ocasionaria um problema de perfeita multicolinearidade (isto é, as colunas da matriz dos dados dos

regressores não seriam linearmente independentes entre si) tornando impossível o cálculo dos estimadores de MQO.

Como é sabido, os estimadores de *MQO* são dados por {Johnston (1972: 124)}:

$$\hat{B} = (X'X)^{-1}X'Y \tag{3-2}$$

onde X e Y são a matriz e o vetor dos dados de x e y, respectivamente (sendo todos os elementos da primeira coluna de X iguais a um, a segunda coluna é formada pelos dados da variável x_1 e a terceira formada por x_2, etc.); X' é o transposto de X e, o expoente -1 significa o inverso da matriz. Aplicando-se então (3-2) ao modelo (3-1), temos:

$$\hat{B} = (D'D)^{-1}D'Y \tag{3-3}$$

onde:

$$D = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ . & . & . \\ . & 1 & 0 \\ . & 0 & 1 \\ . & . & . \\ . & . & 1 \\ . & 0 & 0 \\ . & . & . \\ 1 & 0 & 0 \end{bmatrix} \tag{3-4}$$

Naturalmente, este resultado, para a matriz D, é obtido após efetuada a ordenação dos dados, de tal maneira que os y_s do primeiro grupo sejam seguidos daqueles do segundo e estes por aqueles do terceiro. Para simplificar os cálculos, suporemos inicialmente que o número de observações em cada grupo seja o mesmo, digamos n; este pressuposto não se constitui em condição *sine qua non* para a obtenção do resultado abaixo, conforme ficará evidente no final desta seção, mas facilitará a exposição. Com essas considerações, temos então:

$$D'D = \begin{bmatrix} 3n & n & n \\ n & n & 0 \\ n & 0 & n \end{bmatrix}; \quad (D'D)^{-1} = \frac{1}{n}\begin{bmatrix} 1 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix} \quad \text{e}$$

$$D'Y = \begin{bmatrix} \sum_j \sum_i y_{ij} \\ \sum_i y_{i1} \\ \sum_i y_{i2} \end{bmatrix}$$

Assim, se chamarmos $\overline{Y}$ a média geral dos y_i e $\overline{Y}_j$ a média de cada grupo, e dado ainda que $3\overline{Y} = \sum_{j=1}^{3} \overline{Y}_j$, vem:

$$\hat{B} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} 3\overline{Y} - \overline{Y}_1 - \overline{Y}_2 \\ -3\overline{Y} + 2\overline{Y}_1 + \overline{Y}_2 \\ -3\overline{Y} + \overline{Y}_1 + 2\overline{Y}_2 \end{bmatrix} = \begin{bmatrix} \overline{Y}_3 \\ \overline{Y}_1 - \overline{Y}_3 \\ \overline{Y}_2 - \overline{Y}_3 \end{bmatrix} \tag{3-5}$$

Desse modo, a hipótese normalmente testada em análise de variância de que as médias dos grupos sejam iguais (isto é, no caso aqui: $\overline{Y}_1 = \overline{Y}_2 = \overline{Y}_3$ ou $\overline{Y}_1 - \overline{Y}_3 = \overline{Y}_2 - \overline{Y}_3 = 0$) é equivalente ao teste $b_1 = b_2 = 0$, no modelo de regressão, que é naturalmente efetuado através da seguinte distribuição F {Kmenta (1971: 367)}:

$$\frac{R^2}{2} \div \frac{1 - R^2}{3n - 3} = F_{(2;\, 3n-3)}, \tag{3-6}$$

onde R^2 é o coeficiente de determinação da regressão e os subíndices de F indicam os respectivos graus de liberdade da distribuição.

Para estabelecer a equivalência entre as duas técnicas, ainda sob um outro ângulo, mostraremos que a relação em (3-6) nada mais representa que a fórmula clássica para o teste das médias na análise de variância — isto é, a razão das variâncias entre os grupos e dentro dos grupos, após considerados os respectivos graus de liberdade. Senão, vejamos. Por definição

$$R^2 = 1 - \frac{\Sigma e_i^2}{\Sigma (y_i - \overline{Y})^2} \tag{3-7}$$

onde e_i é o resíduo dos *MQO*, definido como

$$e_i = y_i - X\hat{B} \ . \tag{3-8}$$

A adaptação para o modelo sob consideração produziria:

$$e_{ij} = y_{ij} - D\hat{B} = y_{ij} - \overline{Y}_j \tag{3-9}$$

Substituindo agora esta expressão em (3-7) obtemos, após a adaptação de (3-7) para (3-1):

$$R^2 = \frac{\Sigma\Sigma (y_{ij} - \overline{Y})^2 - \Sigma\Sigma (y_{ij} - \overline{Y}_j)^2}{\Sigma\Sigma (y_{ij} - \overline{Y})^2} = \frac{n\Sigma (\overline{Y}_j - \overline{Y})^2}{\Sigma\Sigma (y_{ij} - \overline{Y})^2}, \tag{3-10}$$

onde utilizamos a identidade básica da análise de variância: variação total = variação dentro dos grupos + variação entre os grupos, ou $\Sigma\Sigma (y_{ij} - \overline{Y})^2 = \Sigma\Sigma (y_{ij} - \overline{Y}_j)^2 + n\Sigma (\overline{Y}_j - \overline{Y})^2$ {Costa Neto (1977: 156)}.

Substituindo finalmente (3-10) em (3-6), vem:

$$\frac{n\Sigma(\overline{Y}_j - \overline{Y})^2}{2} \div \frac{\Sigma\Sigma(y_{ij} - \overline{Y}_j)^2}{3n - 3} \tag{3-11}$$

que é a relação utilizada para testes de médias na análise de variância, mostrando, portanto, que as duas técnicas se equivalem.

Observe-se que o modelo em (3-1) e os resultados em (3-5) e (3-6) podem ser respectivamente adaptados para o caso geral, com K grupos, como segue {Kooyman (1976: 99-104)}:

$$y_{ij} = b_0 + b_1 D_{1ij} + \ldots + b_{k-1} D_{k-1,ij} + \mu_{ij} \tag{3-12}$$

$$\hat{B} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \hat{b}_{k-1} \end{bmatrix} = \begin{bmatrix} \overline{Y}_k \\ \overline{Y}_1 - \overline{Y}_k \\ \overline{Y}_2 - \overline{Y}_k \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \overline{Y}_{k-1} - \overline{Y}_k \end{bmatrix} \tag{3-13}$$

e

$$\frac{R^2}{k - 1} \div \frac{1 - R^2}{Kn - K} = F_{(k-1;\, kn-k)}. \tag{3-14}$$

Cumpre ainda notar que, caso adotássemos um modelo com tantas variáveis binárias quanto o número de categorias, mas eliminando a interseção linear, então o problema da multicolinearidade seria contornado e tal modelo guardaria relação com o modelo com interseção. Dessa relação ver-se-á que os resultados em (3-13) independem de serem iguais o número de observações nos vários grupos. Com efeito, seja o modelo geral, sem interseção, dado por [Aigner (1971: 176)].

$$y_{ij} = b'_1 D_{1ij} + b'_2 D_{2ij} + \ldots + b'_k D_{kij} + \mu_{ij} \tag{3-15}$$

Subtraindo-se $b'_k \sum_{i=1}^{k} D_i$ de ambos os lados desta relação, e dado que $\sum_{i=1}^{k} D_i = 1$, temos:

$$y_{ij} = b'_k + (b'_1 - b'_k) D_{1ij} + \ldots + (b'_{k-1} - b'_k) D_{k-1,ij} + \mu_{ij} \tag{3-16}$$

Comparando-se agora (3-16) com (3-12), vem:

$$b_0 = b'_k;\ b_1 = b'_1 - b'_k;\ \ldots;\ b_{k-1} = b'_{k-1} - b'_k \tag{3-17}$$

Uma vez determinado, portanto, o vetor dos estimadores em (3-15) podemos, através de (3-17), obter os estimadores em (3-12). Assim sendo, a aplicação de *MQO* em (3-15) produz:

$$\hat{B} = (\tilde{D}'\tilde{D})^{-1}\tilde{D}'Y \tag{3-18}$$

$$\text{onde} \quad \tilde{D} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & . & . \\ . & 0 & . \\ 0 & 1 & . \\ . & . & . \\ . & 1 & 0 \\ . & 0 & 1 \\ . & . & . \\ . & . & . \\ 0 & 0 & 1 \end{bmatrix} \tag{3-19}$$

isto é, com $\tilde{D}$ diferindo de D em (3-4). Segue-se então que:

$$\tilde{D}'\tilde{D} = \begin{bmatrix} n_1 & & & 0 \\ & n_2 & & \\ & & \ddots & \\ 0 & & & n_k \end{bmatrix};$$

$$(\tilde{D}'\tilde{D})^{-1} = \begin{bmatrix} \frac{1}{n_1} & & & 0 \\ & \frac{1}{n_2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{n_k} \end{bmatrix} \quad \text{e}$$

$$\tilde{D}'Y = \begin{bmatrix} n_1\overline{Y}_1 \\ n_2\overline{Y}_2 \\ \\ \\ n_k\overline{Y}_k \end{bmatrix}$$

Desse modo temos:

$$\hat{B}' = \begin{bmatrix} \hat{b}'_1 \\ \hat{b}'_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{b}'_k \end{bmatrix} = \begin{bmatrix} \overline{Y}_1 \\ \overline{Y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \overline{Y}_k \end{bmatrix} \tag{3-20}$$

Observe-se que o vetor em (3-20) independe do número de observações dentro de cada grupo e, tendo em vista as relações em (3-17), conclui-se então que o vetor em (3-13) é independente também do número de observações dos grupos, apesar de termos obtido (3-5), ou (3-13), considerando número idêntico de observações para cada grupo.

3.2 — Exemplo

Sejam os três grupos de dados abaixo {Wesolowsky (1976: 152)}

GRUPO *1*	*52*	*36*	*24*	*48*	*52*	*27*	*48*	*41*
GRUPO *2*	*23*	*36*	*29*	*16*	*27*			
GRUPO *3*	*21*	*32*	*17*	*24*	*15*			

Para testar se há diferença significativa entre as médias dos três grupos, necessitamos inicialmente do valor da relação em (3-6). Assim, como o ajustamento do modelo em (3-1) forneceu

$$\hat{y} = 21{,}8 + 19{,}2D_1 + 4{,}4D_2;\ R^2 = 0{,}5165,$$

da expressão em (3-6) obteve-se $F = 8{,}012$, que é significativo ao nível de 0,4%. Naturalmente, estimadores idênticos seriam obtidos caso tivéssemos ajustado um modelo sem interseção e depois utilizado as relações em (3-17), conforme é mostrado em seguida:

$$\hat{B}' = \begin{bmatrix} \overline{Y}_1 \\ \overline{Y}_2 \\ \overline{Y}_3 \end{bmatrix} = \begin{bmatrix} 41{,}0 \\ 26{,}2 \\ 21{,}8 \end{bmatrix} \Longrightarrow \hat{B} = \begin{bmatrix} \overline{Y}_3 \\ \overline{Y}_1 - \overline{Y}_3 \\ \overline{Y}_2 - \overline{Y}_3 \end{bmatrix} = \begin{bmatrix} 21{,}8 \\ 19{,}2 \\ 4{,}4 \end{bmatrix}$$

Esses resultados confirmam, portanto, os pontos discutidos na seção anterior.

4 — DESSAZONALIZAÇÃO DE SÉRIES

4.1 — Métodos de ajustamento sazonal na regressão

Seguindo-se o tratamento apresentado em Johnston (1972), suponha-se que dispomos de $4n$ observações trimestrais para uma variável y. Poderíamos, então, imaginar o seguinte modelo para representar os dados:

$$y = b_1 D_1 + b_2 D_2 + b_3 D_3 + b_4 D_4 + \mu \tag{4-1}$$

onde $D_j = \begin{cases} 1 \text{ para o trimestre } j (j = 1, 2, 3 \text{ e } 4) \\ 0 \text{ para os outros trimestres.} \end{cases}$

Este modelo em notação matricial seria $Y = DB + U$, que com a aplicação de MQO produziria o seguinte estimador para o vetor de parâmetros B:

$$\hat{B} = (D'D)^{-1} D'Y, \tag{4-2}$$

$$\text{onde } D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4-3}$$

que é uma matriz de ordem $(4n \times 4)$. Assim, os resíduos de MQO do modelo em (4-1) seriam dados por:

$$e = Y^d = Y - D\hat{B} = Y - D(D'D)^{-1}D'Y = A\ Y \tag{4-4}$$

onde $A = I - D(D'D)^{-1} D'$.

À primeira vista poder-se-ia pensar na relação em (4-4) para representar a série dessazonalizada, já que $D\hat{B}$ seria uma espécie de componente sazonal da série. Entretanto, uma primeira objeção a tal uso

é que os valores de Y^d não têm soma igual à série original, pois sendo resíduos de *MQO*, sua soma é igual a zero. Esta dificuldade, no entanto, poderá ser facilmente contornada somando-se $\overline{Y}$ (a média dos y originais) a tais resíduos.

Uma objeção mais importante ao uso de Y^d para representar a série dessazonalizada é baseada no fato de que seus valores são as diferenças entre os y originais e os valores médios de cada trimestre (esta constatação fica evidente considerando-se que, conforme mostrado em (3-20) da seção anterior, $\hat{B} = (\overline{Y}_1, \overline{Y}_2, \overline{Y}_3, \overline{Y}_4)$). Então, se a série original contiver tendência e ciclos, essas médias de cada série de trimestres seriam afetadas e, portanto, afetariam a série dessazonalizada. Nestas circunstâncias, um modelo mais apropriado diante da presença de tendência, ciclos e influências sazonais, seria {Johnston (1972: 187)}:

$$y = a_1 t + a_2 t^2 + \ldots + a_p t^p + b_1 D_1 + \ldots + b_4 D_4 + \mu \qquad (4\text{-}5)$$

onde t representa a variável tempo. Em notação matricial, este modelo poderia ser escrito como:

$$Y = TA + DB + U, \qquad (4\text{-}6)$$

onde D é definido como em (4-3), A é a matriz dos a_i, não devendo ser confundida com aquela em (4-4), e T seria:

$$T = \begin{bmatrix} 1 & 1^2 & \ldots\ldots\ldots\ldots & 1^p \\ 2 & 2^2 & \ldots\ldots\ldots\ldots & 2^p \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 4n & (4n)^2 & \ldots\ldots\ldots\ldots & (4n)^p \end{bmatrix} \qquad (4\text{-}7)$$

A aplicação de *MQO* em (4-6), para a estimação do vetor B, já que estamos interessados em determinar o componente sazonal apenas, forneceria {Johnston (1972: 188)}:

$$\hat{B}^* = (D'MD)^{-1}D'MY \qquad (4\text{-}8)$$

onde $M = I - T(T'T)^{-1}T'$. Assim, a série dessazonalizada seria, agora:

$$Y^{ds} = Y - D\hat{B}^* = A^*Y, \qquad (4\text{-}9)$$

onde $A^* = I - D(D'MD)^{-1}D'M$.

4.1.1 — Exemplos

Apresentamos, aqui, duas aplicações para comparar o desempenho dos métodos de dessazonalização de séries que acabamos de discutir. Primeiramente, consideramos os resultados obtidos por Johnston (1972), utilizando dados trimestrais do índice de Produção Industrial no Reino Unido, no período 1948-1957. Inicialmente Johnston obteve a estimativa da componente sazonal com base no método aditivo das médias móveis com quatro períodos, a qual, em vista de seus reconhecidos méritos, fora utilizada como padrão para julgar-se o desempenho dos outros métodos. Os resultados obtidos para a componente sazonal, nos vários métodos, são então reproduzidos no quadro 1 que se segue.

QUADRO 1

COMPARAÇÃO ENTRE OS VÁRIOS MÉTODOS DE AJUSTE SAZONAL

MÉTODOS	COMPONENTE SAZONAL POR TRIMESTRE				ERRO QUADRÁTICO MÉDIO
	b_1	b_2	b_3	b_4	
Média móvel (aditiva).....	3,28	0,77	—7,13	3,08	0
Regressão com D.....	1,85	0,35	—7,15	4,95	1,430
Regressão com T e D					
a) para p = 4.....	4,87	0,36	—7,17	2,95	0,679
b) para p = 6.....	3,35	0,95	—7,54	3,25	0,059

FONTE — Adaptado de Johnston (1972: 189).

Desses resultados, verifica-se que, em geral, o método que utiliza apenas *D* na regressão (isto é, apenas as variáveis binárias) fornece resultados menos satisfatórios que aqueles obtidos com a regressão em *T* e *D*. Este fato poderá ser facilmente constatado quando utilizamos, conforme proposto por Sapsford (1980), o erro quadrático médio, definido em relação aos resultados da média móvel (isto é, erro quadrático $= \sum_{i=1}^{4} (b_i - b_i^{m.v})^2/4)$, como critério estatístico para avaliar o desempenho relativo dos vários métodos; os resultados obtidos estão na última coluna do quadro. Assim, por este critério, o melhor desempenho é obtido pelo método que utiliza *T* e *D* na regressão e quando o polinômio, na variável tempo, for de grau 6 em vez de 4.

Como segunda aplicação, utilizamos os resultados obtidos por Sapsford (1980), com dados mensais sobre o número de greves no Reino Unido, no período 1893-1971 (isto é, num total de 948 observações). Uma vez mais, a componente sazonal obtida pelo método das médias móveis é apresentado como padrão de julgamento dos outros métodos, sendo o erro quadrático médio novamente utilizado como critério estatístico para a avaliação desses métodos. Os resultados obtidos são mos-

QUADRO 2

COMPONENTE SAZONAL NOS MESES, SEGUNDO OS VÁRIOS MÉTODOS

MÉTODOS	COMPONENTE SAZONAL NOS VÁRIOS MESES						
	Janeiro	Fevereiro	Março	Abril	Maio	Junho	Julho
Média móvel (aditivo).....	−2,68	3,45	13,08	10,18	17,82	4,09	−18,99
Regressão com T e D:							
p = 3	−2,69	3,41	13,06	10,53	18,41	4,37	−18,94
p = 5	−2,77	3,34	13,00	10,49	18,38	4,35	−18,93

MÉTODOS	COMPONENTE SAZONAL NOS VÁRIOS MESES					ERRO QUADRÁTICO MÉDIO
	Agosto	Setembro	Outubro	Novembro	Dezembro	
Média móvel (aditivo).....	−3,76	4,98	12,32	−1,73	−38,63	0
Regressão com T e D:						
p = 3	−3,88	5,12	12,21	−1,97	−39,78	0,17
p = 5	−3,85	5,17	12,88	−1,87	−39,67	0,14

FONTE — Adaptado de Sapsford (1980).

trados no quadro 2, onde verificamos serem melhores os resultados produzidos pela regressão em T e D, quando $p = 5$, em vez de $p = 3$. Infelizmente não foram estimadas as componentes sazonais da regressão com D apenas. Aliás, deve ser ressaltado que em seu trabalho Sapsford (1980) tinha como objetivo principal a comparação desses vários métodos com aquele obtido pela chamada *spline regression,* que no caso produziu os melhores resultados, mas os quais não são mostrados aqui, por fugirem ao escopo deste estudo.

4.2 — Estimação de relações econométricas diante de influências sazonais

Quando pretendemos estimar uma relação econométrica com séries trimestrais ou mensais, uma questão que logo surge é saber se as variáveis devem figurar na especificação em sua forma original ou na forma dessazonalizada. Por exemplo, consideremos as seguintes situações: suponha-se que sabemos ser o consumo pessoal explicado pelo comportamento da renda pessoal. Entretanto, enquanto o consumo apresenta uma trajetória sazonal, a renda não possui qualquer elemento sazonal no seu comportamento. Neste caso, o consumo dessazonalizado deverá ser tomado como uma função da renda, e como o consumo observado seria igual ao consumo dessazonalizado mais uma componente sazonal, então segue-se que a regressão apropriada deveria ser entre o consumo observado e a renda mais um conjunto de variáveis sazonais apropriadas. Se, no entanto, também a renda apresentar uma trajetória sazonal, e uma que difira daquela do consumo (por exemplo, a série da

renda atinge o seu pico no primeiro trimestre e o consumo tem o seu pico no último trimestre do ano) então a regressão deveria ser entre o consumo observado e a renda dessazonalizada mais um conjunto de variáveis sazonais apropriadas.

Muitas vezes, entretanto, não é, do ponto de vista teórico, muito claro se as variáveis devem entrar na regressão em suas formas dessazonalizadas ou não. Além disso, o fato dessas séries estarem freqüentemente disponíveis em ambas as formas parece só contribuir para tornar mais difícil a decisão sobre a escolha da forma da variável. Felizmente, entretanto, a questão da especificação da forma em que a variável deve entrar na regressão parece não ser crucial, como o conjunto de resultados obtidos a seguir sugere claramente.

Com efeito, consideremos inicialmente os modelos:

$$Y = XB_1 + DB + U \tag{4-10}$$

e

$$Y^d = X^d B_2 + U \tag{4-11}$$

onde $Y^d = AY$, $X^d = AX$ e A é definido como em (4-4). Mostraremos que, se D assumir a forma dada em (4-3), então a aplicação de *MQO* em (4-10) e (4-11) resultaria em $\hat{B}_1 = \hat{B}_2$. Senão, vejamos. A aplicação de *MQO* em (4-10) fornece as seguintes equações normais da regressão {Johnston (1972: 196)}:

$$(X'X)\hat{B}_1 + (X'D)\hat{B} = X'Y \tag{4-12}$$

e

$$(D'X)\hat{B}_1 + (D'D)\hat{B} = D'Y \tag{4-13}$$

Desta última relação, temos:

$$\hat{B} = (D'D)^{-1}\,[D'Y - (D'X)\hat{B}_1] \tag{4-14}$$

que quando substituído em (4-12), fornece:

$$\hat{B}_1 = (X'AX)^{-1}\,X'AY \tag{4-15}$$

Por outro lado, aplicando *MQO* em (4-11), vem:

$$\hat{B}_2 = (X^{d\prime}X^d)^{-1}X^{d\prime}Y^d = (X'AX)^{-1}X'AY \tag{4-16}$$

já que, como poderá ser facilmente verificado, a matriz A é simétrica (isto é, $A' = A$) e idempotente (isto é, $A.A. = A$). Assim, de (4-15) e (4-16), temos que $\hat{B}_1 = \hat{B}_2$. Isto é, não há necessidade em dessazonalizarmos as séries via a matriz D para usá-las na regressão.

Para um resultado mais geral, consideremos os modelos:

$$Y = X^d B_3 + U \tag{4-17}$$

e

$$Y = X^d B_4 + DB + U, \quad \text{onde } X^d = AX \tag{4-18}$$

Então temos que {Lovell (1963: 1001)}:

$$\hat{B}_1 = \hat{B}_2 = \hat{B}_3 = \hat{B}_4 \tag{4-19}$$

Isto é, a regressão de Y com: 1) X e D; 2) X^d ou 3) X^d e D produziria o mesmo estimador para o coeficiente de X (ou X^d) que aquele obtido para a regressão de Y^d com X^d.

Na verdade, a igualdade $B_1 = B_2 = B_3$, poderá ser alternativamente obtida recorrendo-se a alguns resultados demonstrados por Goldberger (1968). Para facilitar a exposição, adotaremos aqui a mesma notação utilizada por Goldberger. Assim, seja a regressão múltipla, cujas variáveis têm como valores os desvios das respectivas médias:

$$y = b_{y1.2}\, x_1 + b_{y2.1}\, x_2 + e_{y.12}\ , \tag{4-20}$$

onde $e_{y.12}$ é o resíduo de *MQO* e onde $b_{y.2}$ denota, naturalmente, a inclinação de y com relação a x_1 quando x_2 for o outro regressor. Consideremos ainda a regressão simples:

$$x_2 = b_{21}\, x_1 + x_{2.1} \tag{4-21}$$

onde $x_{2.1}$ é o resíduo de *MQO*. Goldberger mostra então que, se efetuarmos a regressão

$$y = b_{y(2.1)} x_{2.1} + e_{y.(2.1)} \tag{4-22}$$

onde $e_{y.(2.1)}$ é o resíduo de *MQO*, segue-se que

$$b_{y(2.1)} = b_{y2.1} \tag{4-23}$$

Isto é, $b_{y2.1}$ na regressão múltipla mostra o efeito de x_2 sobre y depois do expurgo do efeito de x_1 sobre x_2.

Consideremos ainda a regressão simples:

$$y = b_{y1} x_1 + e_{y.1} \tag{4-24}$$

onde $e_{y.1}$ é o resíduo de *MQO*. Efetuemos em seguida a regressão entre estes resíduos e os resíduos $x_{2.1}$ da regressão em (4-21), isto é:

$$e_{y.1} = b_{(y.1)(2.1)} x_{2.1} + v. \tag{4-25}$$

Goldberger (1968) mostra então, que:

$$b_{(y.1)(2.1)} = b_{y(2.1)}\ , \tag{4-26}$$

ou seja, a inclinação da regressão entre os resíduos da regressão de y em x_1 com os resíduos da regressão de x_2 em x_1 é idêntica à inclinação da regressão de y com os resíduos da regressão de x_2 em x_1.

Com essas considerações, verificamos que, se X e D em (4-10) assumirem os lugares de x_1 e x_2 em (4-20), e já que Y^d em (4-11) é o resíduo de *MQO* da regressão de Y em D e X^d o resíduo da regressão de X em D, então é claro que consideradas as regressões $Y = XB_1 + DB + U$ e $Y^d = X^dB_2 + U$, teremos que $B_1 = B_2$. Além disso, se efetuarmos a regressão $Y = X^dB_3 + U$, então, de acordo com o resultado em (4-26), teremos $B_1 = B_2 = B_3$.

Se esses resultados são, por um lado, interessantes, eles parecem, por outro lado, suscitar algumas importantes questões. Por exemplo, se D é meramente uma matriz de variáveis binárias sazonais, então, como vimos, Y^d e X^d em (4-11) dificilmente poderiam ser aceitos como séries dessazonalizadas. Nestas circunstâncias, poder-se-ia então imaginar que o resultado em (4-19) seria apenas uma curiosidade econométrica, isso não fossem as suas reais possibilidades práticas de aplicação, conforme veremos nos exemplos apresentados no final desta seção. Ainda dentro dessa mesma perspectiva, temos, em contrapartida, que embora séries adequadamente dessazonalizadas possam ser obtidas com a adoção da matriz A^* em (4-9), a utilização na regressão em (4-11) de tais séries não permitiria provar a igualdade entre B_1 e B_2 nas relações em (4-10) e (4-11) {ver Johnston (1972: 191)}. Assim, duas questões logo surgem. Primeiramente, de quanto difeririam B_1 e B_2 quando, este último embora estimado na regressão em (4-11), a dessazonalização de Y e X fosse efetuada via matriz A^*, em vez da matriz A em (4-4). Em segundo lugar, há que se atentar para o fato de que as séries são na prática freqüentemente dessazonalizadas pelo método das médias móveis, em vez do método de *MQO* aqui utilizado. Assim sendo, caberia indagar então se não seriam muito distintos os coeficientes B_2 em (4-11), caso as séries fossem dessazonalizadas via médias móveis em vez de via matriz D. Apesar de não se dispor de resposta definitiva a este respeito, parece poder-se afirmar que esses diferentes métodos de dessazonalização geralmente produzem valores para B_2 em (4-11) bastante próximos entre si; aliás, os exemplos apresentados no final desta seção parecem corroborar este fato.

4.2.1 — A questão do erro de especificação pela inclusão ou exclusão indevida das variáveis sazonais

Dois outros pontos que merecem algumas considerações, estão relacionados ao problema do erro de especificação causado pela exclusão ou inclusão indevida de variáveis binárias no modelo de regressão.

Considere-se, primeiramente, o caso onde as variações sazonais de Y são conseqüência apenas das influências sazonais de X agindo sobre Y.

Nestas circunstâncias, a especificação apropriada do modelo seria simplesmente:

$$Y = Xb + U \tag{4-27}$$

Se usarmos, entretanto, o modelo em (4-10), incorremos num problema de erro de especificação, com a conseqüência de que, apesar de $\hat{B}_1$ ser estimador não tendencioso de b, temos que $\hat{B}_1 \neq \hat{b}$. Além disso, em geral a eficiência de $\hat{B}_1$ é menor que aquela de $\hat{b}$, só não sendo menor se as variáveis sazonais D, indevidamente incluídas, forem não correlacionadas com as variáveis explicativas X, o que seria uma situação pouco provável no presente contexto, pois o modelo em (4-27) assume que são as variações sazonais nas variáveis X que explicam as variações sazonais da variável dependente {conforme Thomas & Wallis (1971: 61)}; assim, X e D devem ser correlacionadas. Dessas considerações depreende-se ainda que se as variações sazonais em Y são causadas por variações sazonais em X, então a especificação em (4-11) seria também inadequada.

Imagine-se agora a situação onde as variações sazonais em Y não são explicadas pelas variáveis X. Neste caso, a especificação apropriada é aquela dada em (4-10). Suponha-se, porém, que o modelo usado seja aquele em (4-27), isto é, com a exclusão indevida das variáveis sazonais D. Conforme é sabido, a exclusão indevida de variáveis poderá causar tendenciosidade nos estimadores dos coeficientes daquelas variáveis incluídas no modelo. Na verdade, o viés só seria nulo se as variáveis excluídas (no caso, as variáveis binárias) forem não correlacionadas com as variáveis incluídas (no caso, os X). Como na situação analisada agora as variações sazonais em Y seriam explicadas pelas variáveis sazonais D, segue-se que D e X devem ser pouco correlacionadas. Portanto, a questão da tendenciosidade dos estimadores não deverá se constituir num problema. De qualquer maneira, é costume recomendar-se que "... se há de se errar, que o erro seja cometido incluindo variáveis na regressão em vez de excluir variáveis" {Johnston (1972: 169)}. Esta é naturalmente uma escolha entre eficiência e tendenciosidade do estimador e que, portanto, deve merecer muita cautela quando da decisão do pesquisador.

4.2.2 — Exemplos

Investigações empíricas efetuadas por Johnston (1972: 191-2) na estimação de relações dos tipos (4-10) e (4-11), onde D assumiu a forma em (4-3) e a dessazonalização fora processada tanto pelo método das médias móveis como pelas relações em (4-4) e (4-9), revelaram concordância de valores, em várias casas decimais, para os estimadores $\hat{B}_1$ e $\hat{B}_2$. Mas aquele autor chama a atenção para o fato de que as suas séries apresentavam crescimento ininterrupto e oscilações cíclicas desprezíveis,

e que, portanto, resultados menos favoráveis deveriam ser esperados quando do uso de séries com efeitos cíclicos mais fortes. De qualquer modo, podemos ainda utilizar os resultados de uma estimação apresentada em Thomas (1978: 222-31) como subsídio adicional ao problema. Após utilizar séries históricas, de 1948 à 1964, para a produção industrial de alimento (P) e consumo de alimento (C), na Inglaterra, Thomas estimou funções do tipo (4-10) e (4-11), mas com a dessazonalização em (4-11) sendo efetuada pelo método das médias móveis para quatro períodos, embora D em (4-10) assumisse a forma em (4-3). Os resultados obtidos foram, respectivamente (valores de t entre parênteses):

$$P_t = \underset{(0,045)}{1,434C_{t-1}} - 55,05D_1 - 26,68D_2 - 43,90D_3 - 42,13D_4;\ R^2 = 0,946$$

e

$$P_t^d = \underset{(0,044)}{1,435C_{t-1}^d} - \underset{(5,10)}{42,03};\ R^2 = 0,941$$

Conforme podemos observar, destes resultados, os coeficientes da inclinação são bastante semelhantes nas duas situações, e isto só vem reforçar a validade dos resultados em (4-15) e (4-16).

5 — AS VARIÁVEIS BINÁRIAS NA PREVISÃO DA REGRESSÃO

Mostraremos, a seguir, que a adição de variáveis binárias no modelo de regressão poderá oferecer grandes vantagens computacionais nos estudos de previsão, quer seja o método de estimação o dos *MQO*, Mínimos Quadrados Generalizados (MQG) ou ainda métodos não lineares de estimação, ou mesmo a estimação de sistema de equações simultâneas. Estas questões serão discutidas nesta seção.

5.1 — Previsão na regressão linear

5.1.1 — Considerações gerais

Como é sabido, no modelo de regressão linear o erro de previsão para o período $n + i$ (para $i = 1, 2, \ldots$) e a sua variância, são dados, respectivamente, por {Johnston (1972: 154)}:

$$ep_{n+i} = y_{n+i} - \hat{y}_{n+i} = y_{n+i} - a\hat{b} \tag{5-1}$$

$$V(ep_{n+i}) = \sigma^2\{1 + a(x'x)^{-1}\, a'\}, \tag{5-2}$$

onde y_{n+i} e $\hat{y}_{n+i}$ são os valores observado e previsto de y no período da previsão $n + i$; a é o vetor linha dos dados dos regressores utilizados

na previsão (inclui o elemento *1* na primeira posição); a' é o transposto de a; $\hat{b}$ é o vetor coluna dos estimadores de *MQO*; σ^2 é a variância dos erros (estimada através de $\hat{\sigma}^2 = \Sigma e^2_{i/(n-k)}$, onde e_i é um resíduo dos *MQO*; n é o número de observações amostrais; k é o número de parâmetros estimados no modelo); x é a matriz de dados amostrais já antes definida, com todos os elementos da primeira coluna sendo iguais a *1*; o expoente -1 significa o inverso da matriz.

Na abordagem tradicional, as expressões em (5-1) e (5-2), importantes no estabelecimento de intervalo de confiança para y_{n+i} e nos testes de hipóteses sobre os erros de previsão, não têm avaliação direta nos pacotes de computação para regressão geralmente em uso. Entretanto, com uma importante contribuição de Salkever (1976), tornou-se possível a avaliação dessas expressões com os pacotes em uso corrente, o que por si só já indicaria a sua importância. Há outro aspecto de relevo, entretanto: permitir diversas estimações dos tipos em (5-1) e (5-2), simultaneamente, conforme mostraremos em seguida.

Em linhas gerais, a técnica de Salkever poderia ser assim resumida. Aos dados amostrais dos x_i e de y, juntamos os dados destas variáveis correspondentes aos períodos de previsão. Aos regressores dos períodos de previsão, adicionamos variáveis binárias: uma para cada período (isto é, se uma variável binária assume o valor *1* para um dado período, as demais assumem naquele período o valor zero). Ajustando tal modelo global (isto é, composto dos dados amostrais mais os dados de previsão, incluindo as variáveis binárias) obteríamos, então, dos estimadores dos coeficientes das variáveis binárias, os respectivos valores previstos da regressão e as variâncias daqueles estimadores nada mais seriam que as respectivas variâncias dos erros de previsão.

Cabe acrescentar que os estimadores do modelo amostral, com respectivas variâncias, seriam diretamente fornecidos da estimação do modelo global. O coeficiente de determinação (R^2) amostral, por sua vez, poderia ser o mesmo do modelo global, bastando para isso fazer os y das previsões iguais ao y médio ($\overline{Y}$) da amostra. Passamos pois a uma demonstração formal destes resultados.

5.1.2 — Os coeficientes de X e D no modelo global

Seja o modelo global desmembrado como:

$$
\begin{aligned}
y_1 &= b_1 + b_2x_{21} + \dots + b_kx_{k1} + 0 + \dots + 0 + \mu_1 \\
&\vdots \\
y_n &= b_1 + b_2x_{2n} + \dots + b_kx_{kn} + 0 + \dots + 0 + \mu_n \\
y_{n+1} &= b_1 + b_2x_{2,n+1} + \dots + b_kx_{k,n+1} + c_1D_1 + 0 + \dots + 0 + \mu_{n+1} \\
&\vdots \\
y_{n+m} &= b_1 + b_2x_{2,n+m} + \dots + b_kx_{k,n+m} + 0 + \dots + 0 + c^mD^m + \mu_{n+}{}^m
\end{aligned}
\tag{5-3}
$$

onde as observações de 1 a n se referem ao período amostral, e as demais ao período de previsão, e, onde $D_\alpha = 1$, para $\alpha = 1, 2, \ldots, m$. Simplificando a representação de (5-3) e observando as suas partições, teríamos, em forma matricial:

$$\begin{pmatrix} y \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} x & 0 \\ \tilde{x} & I \end{pmatrix} \begin{pmatrix} b \\ c \end{pmatrix} + \begin{pmatrix} u \\ \tilde{u} \end{pmatrix}, \tag{5-4}$$

onde I é uma matriz unitária; as razões para tal formulação estão expostas no apêndice técnico.

Conforme já fora salientado anteriormente, os estimadores de MQO são obtidos através da fórmula

$$\hat{B} = (X'X)^{-1} X'Y, \tag{5-5}$$

onde X e Y são a matriz e vetor de dados. Aplicando pois esta fórmula no modelo em (5-4), temos:

$$\hat{B} = \begin{pmatrix} \hat{b} \\ \hat{c} \end{pmatrix} = \left\{ \begin{pmatrix} x' & \tilde{x}' \\ 0' & I \end{pmatrix} \begin{pmatrix} x & 0 \\ \tilde{x} & I \end{pmatrix} \right\}^{-1} \begin{pmatrix} x' & \tilde{x}' \\ 0' & I \end{pmatrix} \begin{pmatrix} y \\ \tilde{y} \end{pmatrix}; \tag{5-6}$$

ou

$$\hat{B} = \begin{pmatrix} x'x + \tilde{x}'\tilde{x} & \tilde{x}' \\ \tilde{x} & I \end{pmatrix}^{-1} \begin{pmatrix} x' & y + \tilde{x}' & \tilde{y} \\ & \tilde{y} \end{pmatrix}. \tag{5-7}$$

Se reescrevermos esta última fórmula como $\hat{B} = A^{-1}Z$, onde:

$$A = \begin{Bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{Bmatrix} \quad \text{e} \quad Z = \begin{Bmatrix} Z_1 \\ Z_2 \end{Bmatrix}$$

temos então que {Johnston (1972: 93)}:

$$A^{-1} = \begin{Bmatrix} E & -E\, A_{12}\, A_{22}^{-1} \\ -A_{22}^{-1}\, A_{21}\, E & A_{22}^{-1} + A_{22}^{-1}\, A_{21}\, E\, A_{12}\, A_{22}^{-1} \end{Bmatrix} \tag{5-8}$$

onde $E = \{A_{11} - A_{12}\, A_{22}^{-1}\, A_{21}\}^{-1}$.

Aplicando agora (5-8) em (5-7), temos:

$$\hat{B} = \begin{pmatrix} \hat{b} \\ \hat{c} \end{pmatrix} = \begin{Bmatrix} (x'\ x)^{-1} & -(x'\ x)^{-1}\ \tilde{x}' \\ -\tilde{x}\,(x'\ x)^{-1} & I + \tilde{x}\,(x'\ x)^{-1}\ \tilde{x}' \end{Bmatrix} \begin{Bmatrix} x'\ y + \tilde{x}'\ \tilde{y} \\ \tilde{y} \end{Bmatrix}, \tag{5-9}$$

ou

$$\hat{B} = \begin{Bmatrix} (x'\ x)^{-1}\ x'\ y \\ \tilde{y} - \tilde{x}\ (x'\ x)^{-1}\ x'\ y \end{Bmatrix}, \tag{5-10}$$

de onde se deduz que:

$$\hat{c} = \hat{y} - \tilde{x}\hat{b}. \tag{5-11}$$

Desses resultados, percebe-se que para $\hat{B}_i = \hat{b}_i$ (onde $i = 1, 2, \ldots, k$) temos os estimadores dos parâmetros do modelo amostral e, para $\hat{B}_j = \hat{c}_\alpha$ (onde $j = k + \alpha$ e $\alpha = 1, 2, \ldots, m$), temos os estimadores dos coeficientes das variáveis binárias que são, de acordo com (5-11), facilmente convertidos em erros de previsão, como definido em (5-1). Assim, se a previsão é efetuada para um período apenas, então (5-11) transforma-se virtualmente em (5-1) — bastando apenas a substituição apropriada do elemento $\tilde{y}$ pelo y_{n+1} observado — já que a matriz $\tilde{x}$ se converteria no vetor a definido em (5-1). No caso mais geral, de várias previsões, se considerarmos $\tilde{y}$ como sendo o vetor dos valores observados de y para os respectivos períodos, então (5-11) produz os correspondentes erros de previsão; note-se que o valor previsto poderia ser obtido diretamente no modelo em (5-4) fazendo $\tilde{y} = 0$ e substituindo I por $-I$; ver também (5-10).

5.1.3 — As variâncias dos coeficientes de X e D no modelo global

Uma preocupação freqüente nos estudos de previsão, é o de estabelecer intervalos de confiança para o valor a prever, bem como testar hipótese sobre o erro de previsão. Para desempenhar tais tarefas é necessário, conforme já foi mencionado, avaliar expressões dos tipos (5-1) e (5-2), que no modelo global corresponde a estimação de $\hat{c}$, como já foi demonstrado, e de sua variância, a qual passamos a discutir.

Como é sabido, no modelo linear geral as variâncias dos estimadores de *MQO* são dadas pelos elementos da diagonal principal (as demais posições representam covariâncias) da seguinte matriz {Johnston (1972: 126)}:

$$V(\hat{B}) = \sigma^2 (x'x)^{-1} \tag{5-12}$$

onde σ^2 e x são definidos como em (5-2). Aplicando então esta fórmula para o modelo em (5-3), temos:

$$V(\hat{B}) = \begin{Bmatrix} V(\hat{b}) & COV(\hat{b}, \hat{c}) \\ COV(\hat{c}, \hat{b}) & V(\hat{c}) \end{Bmatrix} = \sigma^2 \begin{Bmatrix} (x'x)^{-1} & -(x'x)^{-1}\tilde{x}' \\ -\tilde{x}(x'x)^{-1} & I + \tilde{x}(x'x)^{-1}\tilde{x}' \end{Bmatrix} \tag{5-13}$$

onde σ^2 é a variância dos erros do modelo global que é estimada através de:

$$\hat{\sigma}^2 = \sum_{i=1}^{n+m} \frac{e_i^2}{N - K}$$

onde N é o número de observações no modelo e K é o número de parâmetros estimados (note-se que $N-K = n+m-k-m = n-k$, ou seja, o número de graus de liberdade é igual ao do modelo amostral). Portanto, de (5-13) segue que:

$$V(\hat{c}) = \sigma^2 \{I + \tilde{x}(x'x)^{-1}\tilde{x}'\}, \tag{5-14}$$

onde as variâncias são apresentadas na diagonal principal, que são, conforme dado em (5-2), as próprias variâncias dos erros de previsão.

Ressalte-se, ainda, que a variância dos erros estimada no modelo global é igual aquela estimada no modelo amostral, sendo as $V(\hat{b})$ conseqüentemente igual em ambos os modelos. Senão, vejamos. Chamemos a variância dos erros para o período amostral de σ'^2, que sabemos ser estimada através de:

$$\hat{\sigma}'^2 = \sum_{i=1}^{n} e_i^2/(n-k) .$$

Conforme já foi mostrado acima, as fórmulas utilizadas para estimar σ^2 e σ'^2 apresentam o mesmo denominador. Cabe apenas demonstrar que:

$$\sum_{i=1}^{n+m} e_i^2 = \sum_{j=1}^{n} e_j^2 \tag{5-15}$$

onde e_i e e_j são os resíduos de *MQO* do modelo global e amostral, respectivamente. Como é sabido, no método dos *MQO* a seguinte propriedade é válida {Johnston (1972: 15)}:

$$e'X = 0, \tag{5-16}$$

onde e é o vetor dos resíduos e X é a matriz dos dados dos regressores; ou seja, o vetor dos resíduos é ortogonal a cada regressor. Assim sendo, no modelo (5-3) temos que:

$$e'D_\alpha = 0, \quad \text{para } \alpha = 1, 2, \ldots, m . \tag{5-17}$$

Com efeito, como os D_α assumem o valor de *1* apenas uma vez para cada α, e tão-somente para as observações $i = n + \alpha$, então fica claro de (5-17), que:

$$e_{n+1} D_1 = e_{n+2} D_2 = \ldots = e_{n+m} D_m = 0, \tag{5-18}$$

ou

$$e_{n+1} = e_{n+2} = \ldots = e_{n+m} = 0; \tag{5-19}$$

uma prova alternativa deste resultado é oferecida no apêndice técnico.

Considerando o resultado em (5-19), e ainda aquele em (5-10) que mostra ser $\hat{B} = \hat{b}_i$ para $i = 1, 2, \ldots, k$, tem-se a conclusão óbvia de que (ver apêndice técnico):

$$\sum_{i=1}^{n+m} e_i^2 = \sum_{j=1}^{n} e_i^2$$

e, conseqüentemente, que a estimativa de σ'^2 (amostral) é igual aquela de σ^2 (global).

5.1.4 — Efeitos sobre o R^2 e os testes t, F e Durbin-Watson (DW)

Dada a igualdade em (5-15) e visto ser $\tilde{y}$ um vetor de dados não especificados, é simples observar que, caso os elementos de $\tilde{y}$ sejam todos iguais à média dos y amostrais, segue-se que o coeficiente de determinação (R^2) da regressão global será igual aquele da regressão amostral, já que:

$$R^2 = 1 - \frac{\Sigma e_i^2}{\Sigma (Y_i - \bar{Y})^2}.$$

Por outro lado, se os $\hat{b}_i$ e suas variâncias estimadas não se alteram no ajustamento global relativamente ao ajustamento amostral, e se os graus de liberdade são também idênticos nos dois ajustamentos, segue-se então que a significância estatística dos parâmetros individuais (dada pelos testes t ou F) não se alterará. Entretanto, a significância do ajustamento, como um todo, poderia ficar alterada. Conforme é sabido, a significância estatística do ajustamento é dada pela relação {Kmenta (1971: 367)}

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

que tem distribuição F com $k-1$ e $n-k$ graus de liberdade, onde k é o número de parâmetros no modelo. Portanto, mesmo fazendo-se R^2 idêntico nos dois ajustamentos (através da escolha apropriada de $\tilde{y}$) e, sabendo-se serem os graus de liberdade $n-k$ idênticos nas duas situações, ainda assim, a expressão de F produziria valores diferentes, pois $(k-1)$ tem valores distintos nos dois casos. Naturalmente, o F calculado deverá ser comparado com o F crítico baseado em duas distribuições distintas. Assim, se houver mudança na significância estatística do ajustamento total, não parece claro qual seria a sua direção.

Ressalte-se ainda que também os valores do teste Durbin-Watson, para autocorrelação nos resíduos, seriam afetados nos dois ajustamentos. Esta constatação ficaria evidenciada após um simples exame da fórmula em que se baseia o teste, isto é:

$$D.W. = \sum_{t=2}^{n} (e_t - e_{t-1})^2 / \sum_{t=1}^{n} e_t^2 .$$

5.1.5 — Teste para mudança de estrutura do modelo

Finalmente, cabe salientar que informações extraídas de (5-14) seriam ainda úteis para testar a hipótese de alteração na estrutura do modelo de regressão entre os períodos da amostra e da previsão, pois, conforme é sabido, se não houve alteração naquela estrutura, temos a seguinte distribuição F {Christ (1966: 557-9)}:

$$F_{(m,\, n-k)} = \frac{vG^{-1}v'}{m\sigma'^2} \qquad (5\text{-}20)$$

onde v é um vetor linha dos erros de previsão (isto é, $v = ep_1, \ldots, ep_m$); v' é o transposto de v; G^{-1} é o inverso da matriz dada por $I + \tilde{x}(x'x)^{-1}\tilde{x}'$; m e σ^2 são como definidos anteriormente.

5.1.6 — *Resumo das conclusões*

As conclusões são, pois, que além da estimação do modelo em (5-3) oferecer vantagens computacionais nas previsões, ela também estima o modelo amostral como subproduto, permitindo os testes estatísticos usuais sobre os parâmetros da regressão. Além disso, conforme acabamos de expor, com simples adaptações podemos testar a hipótese de alteração na estrutura do modelo entre os períodos amostrais e de previsão.

5.1.7 — *Exemplo*

Sejam os seguintes dados amostrais quaisquer:

y_i	x_{2i}	x_{3i}
2	3	1
1	2	1
2	2	2
1	1	1
4	4	3

Analisemos as duas previsões para y quando:

$$x_{2,n+1} = 3 \text{ e } x_{3,n+1} = 2, \text{ e } x_{2,n+2} = 4 \text{ e } x_{3,n+2} = 3$$

$\left\{\text{isto é, } \tilde{x} = \begin{pmatrix} 1 & 3 & 2 \\ 1 & 4 & 3 \end{pmatrix}\right\}$. De acordo com o exposto nas seções anteriores, devemos ajustar o modelo global com base nos seguintes dados:

y_i	x_{2i}	x_{3i}	x_{4i}	x_{5i}
2	3	1	0	0
1	2	1	0	0
2	2	2	0	0
1	1	1	0	0
4	4	2	0	0
⋮	⋮	⋮	⋮	⋮
2	3	2	1	0
2	4	3	0	1

(Note-se que os y dos períodos de previsão foram feitos iguais ao y médio amostral $\left(\text{isto é, } \tilde{y} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)$ para reproduzir o R^2 amostral, conforme já discutido).

Os ajustamentos de *MQO* aos dados amostrais e globais forneceram, respectivamente:

$$\hat{y}_i = -1{,}08333 + \underset{(0{,}22822)}{0{,}75000x_{2i}} + \underset{(0{,}47507)}{0{,}91667x_{3i}}; \quad R^2 = 0{,}93056$$

e

$$\hat{y}_i = -1{,}0833 + \underset{(0{,}22822)}{0{,}7500x_{2i}} + \underset{(0{,}47507)}{0{,}91667x_{3i}} - \underset{(0{,}55902)}{1{,}00000x_{4i}} - \underset{(0{,}83333)}{2{,}66667_{5i}};$$

$$R^2 = 0{,}93056$$

(os dados entre parênteses significam os respectivos desvios padrão dos estimadores).

As equações acima vêm confirmar que os resultados da estimação amostral são obtidos no bojo da estimação global. Observe-se, ainda, que as variâncias dos $\hat{b}_i$ são as mesmas nas duas equações, levando a concluir que os níveis de significância dos testes são também iguais, já que os graus de liberdade dos dois modelos são idênticos, como já foi visto.

Retornando à análise das previsões temos, da função global estimada, que se $x_{2n+1} = 3$ e $x_{3n+1} = 2$, então a variância do erro de previsão será dada pela variância de $\hat{c}_1 = \hat{B}_4$, ou seja, $(0{,}55902)^2$; e se $x_{2,\,n+2} = 4$ e $x_{3,\,n+2} = 3$, então a variância do erro de previsão será $(0{,}83333)^2$.

O valor previsto, por sua vez, poderá ser obtido facilmente de $\hat{c}_\alpha$, para $\alpha = 1, 2$. Por exemplo, suponha-se que desejamos determinar o valor previsto quando $x_{2,n+3} = 3$ e $x_{3,n+1} = 2$. Como $\hat{c}_1 = \bar{y} - \hat{y}_{n+1}$, temos então que $\hat{y}_{n+1} = 3{,}00$.

Finalmente, para testar se a estrutura do período de previsão difere daquela do período amostral, basta procedermos como discutido em (5-20).

5.2 — As variáveis binárias na previsão de modelos não-lineares

O uso das variáveis binárias para previsão em modelos não-lineares, segue orientação semelhante àquela adotada na seção anterior com respeito ao modelo linear, como demonstrado por Fuller (1980). Antes de abordarmos, entretanto, o problema das variáveis binárias na previsão em regressão não-linear, apresentaremos uma breve exposição sobre a técnica de estimação utilizada em modelos não-lineares.

Seja o modelo não-linear dado por:

$$y = h(x, b) + u, \tag{5-21}$$

onde y e x são o vetor e matriz dos dados, respectivamente; b é o vetor dos parâmetros; u é o vetor dos erros estocásticos e h representa a forma funcional não-linear.

O método de estimação dos Mínimos Quadrados consiste em minimizar $\sum_{i=1}^{n} \{y_i - h(x,b)\}^2$, com relação ao vetor dos parâmetros b. Há dois processos normalmente utilizados para minimizar tal somatório quadrático. O primeiro envolve a derivação das equações normais da regressão que, geralmente, fornece complicadas equações não-lineares. O segundo consiste na adoção da Expansão de Taylor da forma não-linear, utilizando na estimação apenas a parte linear da expressão, sendo o estimador obtido de maneira iterativa. Este último procedimento é de uso freqüente em trabalhos aplicados e será, portanto, o método a ser adotado na exposição que se segue.

A Expansão de Taylor da função em (5-21) é, considerando apenas a parte linear da expansão, dada por:

$$h(x,b) = h\ (x,b_{(0)}) + \sum_{i=1}^{k} \left(\frac{\partial h}{\partial b_i}\right)_{b_i = b_{i(0)}} (b_i - b_{i(0)}) \qquad (5\text{-}22)$$

onde $b_{(0)} = (b_{1(0)}, b_{2(0)}, \ldots . b_{k(0)})$ é o vetor dos valores iniciais dos parâmetros, que são escolhidos arbitrariamente; e $\left(\frac{\partial h}{\partial b_i}\right) b_i = b_{i(0)}$ é a derivada parcial da função avaliada em $b_i = b_{i(0)}$.

Após a substituição dos valores iniciais dos parâmetros em (5-22), podemos utilizar tal informação para estimar

$$\sum_{i=1}^{k} \left(\frac{\partial h}{\partial b_i}\right)_{b_i = b_{i(0)}} (b_i - b_{i(0)}) = y - h(x,b_{(0)}) \qquad (5\text{-}23)$$

A estimação desta função é efetuada pelo método dos *MQO* e permite obter os estimadores para b_i, digamos $\hat{B}_i$. Tais estimadores entram então em (5-23) como $b_{i(1)}$ e nova estimação da função é obtida. O procedimento é repetido até que as diferenças entre os estimadores sucessivos sejam desprezíveis.

Denote-se agora: $q_i = b_i - b_{i(j)}$, onde (j) representa o número das iterações necessárias para a convergência;

$$\left[\left(\frac{\partial h}{\partial b_1}\right)_{b_1 = b_{1(j)}}, \left(\frac{\partial h}{\partial b_2}\right)_{b_2 = b_{2(j)}}, \ldots\right] = (a_{1t}, a_{2t}, \ldots) = a_t ,$$

para $t = 1, 2, \ldots, n$; e $y^* = y - h(x,\hat{b})$, onde $\hat{b}$ é o vetor dos estimadores finais (note-se que y^* mostra os resíduos de *MQO*). Então temos o seguinte vetor de estimadores, conforme indicado pela fórmula geral de *MQO*:

$$\hat{q} = (z'z)^{-1}\ z'\ \{y - h(x,\hat{b})\} = 0, \qquad (5\text{-}24)$$

onde z é a matriz formada pelos vetores a_t, sendo a igualdade a zero decorrência do conceito de convergência na estimação.

Quanto à previsão para um período, esta seria dada por:

$$\hat{y}_{n+1} = h(x_{n+1}, \hat{b}), \tag{5-25}$$

ou

$$\hat{y}^*_{n+1} = a_{n+1}\,\hat{q} \tag{5-26}$$

se o modelo em (5-23) for usado. Ainda utilizando o modelo em (5-23), temos o erro de previsão:

$$ep = y_{n+1} - \hat{y}^*_{n+1} = a_{n+1}q - a_{n+1}\hat{q} + u_{n+1} = u_{n+1} - a_{n+1}\,(\hat{q}-q), \tag{5-27}$$

cuja variância seria:

$$V(ep) = \sigma^2\,\{1 + a_{n+1}(z'z)^{-1}\,a'_{n+1}\}, \tag{5-28}$$

já que a variância de $\hat{q}$ é, no modelo geral de regressão linear, dada por $\sigma^2(z'z)^{-1}$. Note-se ainda que a variância de $\hat{q}$ é igual aquela de $\hat{b}$, pois, conforme foi mostrado acima, na estimação pelo processo iterativo temos que na diferença $q_i = b_i - b_{i(j)}$ o valor $b_{i(j)}$ é pré-fixado; para detalhes neste particular ver Aigner (1971: 180). Observe-se ainda que a variância do erro de previsão, em termos da variável original, seria idêntica àquela dada em (5-28), pois y^* é simplesmente y menos um conjunto de valores pré-fixados.

Incorporando agora a variável binária no modelo para a previsão de um período, temos:

$$Y_i = h(x_i,\hat{b}) + \hat{c}D + e_i \text{ ou } Y_i = H(X_i,\hat{B}) + e_i, \tag{5-29}$$

onde
$$Y_i = \begin{cases} y_i \text{ para } i = 1, 2, \ldots, n \\ 0 \text{ para } i = n+1 \end{cases}$$

$$D = \begin{cases} 0 \text{ para } i = 1, 2, \ldots, n \\ -1 \text{ para } i = n+1 \end{cases}$$

e e_i são os resíduos de *MQO* no modelo global.

Se adotarmos o mesmo procedimento utilizado para a obtenção em (5-24), temos:

$$\hat{Q} = (Z'Z)^{-1}\,Z'\,\{Y - H(X,\hat{B})\} = 0, \tag{5-30}$$

onde $Y - H(X,\hat{B}) = e$. Note-se que a última coluna de Z representa a derivada parcial $\frac{\partial H}{\partial c}$ que, de acordo com a definição da variável D em (5-29), fornece o elemento -1 na última posição e zero nas demais. Isto é, a última linha de Z' seria o vetor $(0, 0, \ldots, 0, -1)$ que, quando

substituído em (5-30), implica que $e_{n+1} = 0$. A substituição deste valor em (5-29) produz então:

$$\hat{c} = h(x_{n+1}, \hat{b}), \tag{5-31}$$

que é o valor previsto $\hat{y}_{n+1}$.

Se observada a partição da matriz em dados amostrais e do período da previsão, como efetuada no caso da regressão linear da seção anterior, sabemos então que a matriz das variâncias e covariâncias para $\hat{Q}$ (e $\hat{B}$, já que ambas são idênticas) seria dada por {ver a expressão em (5-13) no Item anterior}:

$$V(\hat{Q}) = V(\hat{B}) = \sigma^2(Z'Z)^{-1} = \sigma^2 \begin{bmatrix} (z'z)^{-1} & Cov(\hat{b},\hat{c}) \\ Cov(\hat{b}\ \hat{c}) & \{1 + a_{n+1}(z'z)^{-1}a'_{n+1}\} \end{bmatrix} \tag{5-32}$$

Isto é, de maneira análoga ao caso do item anterior, temos que a diagonal principal fornece, respectivamente, as variâncias dos estimadores amostrais e variância do erro de previsão. Note-se que esta última variância é, na verdade, idêntica à variância do estimador $\hat{c}$, sendo ainda igual a (5-28).

Devemos acrescentar que a técnica descrita se estende para casos envolvendo mais de uma previsão, bastando apenas acrescentar uma variável binária para cada previsão adicional; isto é, de maneira idêntica ao caso de regressão linear. As ilustrações apresentadas, a seguir, tratam de casos com duas previsões simultâneas e servem para esclarecer esta questão.

5.2.1 — Exemplos

a — inicialmente, consideremos as previsões para dois períodos quando a forma funcional do modelo é dada por {Aigner (1971: 179)}:

$$y = e^{-bx} + u.$$

Os dados apresentados a seguir foram devidamente subdivididos (linha pontilhada) entre a amostra e previsão:

Y	X	D_1	D_2
0,80	1	0	0
0,45	4	0	0
0,04	16	0	0
.....	.....	.....	.....
0	2	1	0
0	6	0	1

O ajustamento não-linear aos dados amostrais forneceu:

$$\hat{y} = e^{-0,2034(0,0060)x}$$

enquanto que o ajustamento do modelo global (isto é, $Y = e^{-bX} + c_1D_1 + c_2D_2 + U$) produziu

$$\hat{Y} = e^{-0,2034(0,0060)X} - \underset{(0,01465)}{0,6657D_1} - \underset{(0,01624)}{0,295D_2},$$

onde os dados entre parênteses significam desvios padrão dos estimadores. Note-se que o estimador amostral é obtido no bojo da estimação global e que os coeficientes de D_1 e D_2 são (afora o sinal trocado) os valores previstos, respectivamente, quando $x = 2$ e $x = 6$. Adicionalmente, os desvios padrão dos coeficientes de D_1 e D_2 são os próprios desvios padrão dos erros das previsões nas duas situações. Conforme era de se esperar ainda, do que foi exposto acima, constatou-se que os resíduos correspondentes aos períodos das previsões foram efetivamente iguais a zero, e os resíduos correspondentes aos períodos da amostra foram idênticos em ambos os ajustamentos (isto é, global e amostral). Portanto, todos os resultados derivados no texto acima, são comprovados com as estimações obtidas aqui.

b — Como segunda ilustração apresentamos o caso do ajustamento de uma função não-linear com dois parâmetros. Em vista de tal função estar ligada a um problema de distribuição de renda, cabem, inicialmente, as seguintes considerações. Conforme é sabido, se chamarmos x as freqüências acumuladas do número de famílias, quando estas são classificadas das mais pobres para as mais ricas e, se denominarmos, por outro lado, de y as correspondentes freqüências acumuladas das rendas recebidas por tais famílias, então, da conexão dos pontos representados pelas coordenadas x e y, temos a chamada curva de Lorenz.

Assim, a função para a curva de Lorenz, proposta por Rasche, Gaffney, Koo & Obst (1980), e aplicada anteriormente por este autor {Rossi (1982a e 1982b)} aos mesmos dados de renda aqui utilizados, é dada por:

$$y = \{1 - (1 - x)^{\alpha}\}^{1/B},$$

e será também aqui usada. Os dados de renda dos Censos de 1970 e 1980, que serviram de base para a ilustração, são reproduzidos no quadro 3.

Suponha-se agora que o objetivo principal da estimação da função acima seja o de efetuar duas previsões para cada um dos ajustamentos: uma para quando $x = 0,4$ e a outra para $x = 0,5$. De acordo com o

QUADRO 3

ESTIMATIVA NA DISTRIBUIÇÃO DE RENDA NO BRASIL — 1970-1980

CLASSES DE RENDA	ESTIMATIVA DE RENDA			
	X = % acumulada das famílias		Y = % acumulada das rendas	
	1970	1980	1970	1980
1	0,284	0,125	0,057	0,014
2	0,606	0,333	0,216	0,071
3	0,823	0,644	0,412	0,223
4	0,950	0,880	0,659	0,480
5	0,983	0,952	0,805	0,657
6	0,996	0,984	0,923	0,809
7	1,000	1,000	1,000	1,000

FONTE — ver o texto.

que já foi exposto, necessitamos realizar apenas o ajustamento do modelo global (para cada um dos anos), com a seguinte especificação:

$$Y = [1 - (1 - X)^{\alpha}]^{1/\beta} + c_1 D_1 + c_2 D_2 .$$

Desse modo, os dados relativos, por exemplo, ao ano de 1970, para o ajustamento desta função, teriam a seguinte forma:

Y	X	D_1	D_2
0,057	*0,284*	*0*	*0*
0,216	*0,606*	*0*	*0*
0,412	*0,823*	*0*	*0*
0,659	*0,950*	*0*	*0*
0,805	*0,983*	*0*	*0*
0,923	*0,996*	*0*	*0*
.....	.....	.....	.....
0	*0,4*	*1*	*0*
0	*0,5*	*0*	*1*

Note-se que omitimos deliberadamente os valores de x e y correspondentes à sétima classe de renda, já que $(1 - x)$ será ali zero e, portanto, traria alguma dificuldade quando da avaliação da função adotada.

Os resultados obtidos no ajustamento da função para os 2 anos foram, respectivamente:

$$\hat{Y}_{70} = \left[1-(1-X)^{0,5055(0,0213)} \right]^{\frac{1}{0,6219(0,0356)}} \qquad \begin{matrix} -0,09253D_1 & -0,1409D_2 \\ (0,0156) & (0,0162) \end{matrix}$$

e

$$\hat{Y}_{80} = \left[1-(1-X)^{0,4936(0,0182)} \right]^{\frac{1}{0,6089(0,0301)}} \qquad \begin{matrix} -0,08497D_1 & -0,1307D_2 \\ (0,0122) & (0,0127) \end{matrix}$$

onde os números entre parênteses representam os desvios padrão dos estimadores.

Uma vez mais, constata-se desses resultados que os coeficientes de D_1 e D_2 representam efetivamente, nas duas situações, os valores previstos para a função quando $x = 0,4$ e $x = 0,5$, respectivamente. Além disso, os desvios padrão dos coeficientes de D_1 e D_2 representam os desvios padrão dos respectivos erros de previsão, enquanto que os coeficientes (bem como os seus desvios padrão) da função de Lorenz propriamente são os mesmos que prevaleceriam com o seu ajustamento aos dados amostrais apenas (isto é, sem a inclusão das variáveis binárias e sem os dados de x e y correspondentes aos períodos da previsão).

5.3 — As variáveis binárias na previsão utilizando mínimos quadrados generalizados

A linha de raciocínio seguida aqui, será semelhante àquela utilizada no caso das variáveis binárias para previsão no contexto dos *MQO* (isto é, no Item 5.1), e as demonstrações aqui efetuadas confirmam os resultados apresentados por Fuller (1980).

Considere-se o modelo global escrito em notação matricial como:

$$\begin{pmatrix} y \\ 0 \end{pmatrix} = \begin{pmatrix} x & 0 \\ \tilde{x} & I \end{pmatrix} \begin{pmatrix} b \\ \tilde{y} \end{pmatrix} + \begin{pmatrix} u \\ \tilde{u} \end{pmatrix}, \quad \text{ou} \quad Y = XB_g + U, \tag{5-33}$$

onde u e $\tilde{u}$ são os vetores dos erros estocásticos correspondentes aos períodos amostral e da previsão, respectivamente. Seja, em seguida, a matriz das variâncias e covariâncias dos erros estocásticos dada por:

$$V(U) = \sigma^2 Q \neq \sigma^2 I, \quad \text{onde} \quad Q = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{5-34}$$

Em vista da estrutura da matriz em (5-34), faz-se apropriada a estimação do modelo pelo método dos MQG, cujos estimadores são dados pelo vetor {Johnston (1972: 210)}:

$$\hat{B}_g = (X'Q^{-1}X)^{-1} X'Q^{-1}Y \tag{5-35}$$

que pode obviamente ser alternativamente escrito como:

$$(X'Q^{-1}X)\hat{B}_g = X'Q^{-1}Y \ . \tag{5-36}$$

A aplicação da regra para inversão de matriz particionada {Graybill (1969: 165)} à matriz Q dada em (5-34) produz:

$$Q^{-1} = \begin{bmatrix} (A-BD^{-1}C)^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & (D-CA^{-1}B)^{-1} \end{bmatrix} = \begin{bmatrix} E & F \\ G & H \end{bmatrix} \tag{5-37}$$

que substituída em (5-35) fornece as seguintes equações:

$$(x'E + \tilde{x}'G)xb + (x'F + \tilde{x}'H)\tilde{x}b - (x'F + \tilde{x}'H)\tilde{y} = (x'E + \tilde{x}'G)y \tag{5-38}$$

e

$$(-Gx - H\tilde{x})b + H\tilde{y} = -Gy. \tag{5-39}$$

Assim, desta última relação, tem-se que:

$$\tilde{y} = \tilde{x}b - H^{-1}G(y - xb), \tag{5-40}$$

que após substituirmos $H^{-1}G$ pelos seus valores equivalentes em (5-37) temos (ver apêndice técnico):

$$\tilde{y} = \tilde{x}b - CA^{-1}(y - xb), \tag{5-41}$$

que, conforme foi demonstrado por Goldberger (1962), seria o vetor das previsões nos *MQG*.

A substituição de (5-40) em (5-38), por seu turno, fornece (ver apêndice técnico):

$$x'(E - FH^{-1}G)\ (y - xb) = 0, \tag{5-42}$$

que, quando considerados os valores equivalentes em (5-37), produz (ver apêndice técnico):

$$b = (x'A^{-1}x)^{-1}x'A^{-1}y, \tag{5-43}$$

que seriam os estimadores de *MQG* aplicados aos dados amostrais, apenas.

Quanto à matriz das variâncias e covariâncias de $\hat{B}_g = \begin{pmatrix} b \\ \tilde{y} \end{pmatrix}$, sabe-se que seria dada por {Johnston (1972: 210)}:

$$V(\hat{B}_g) = \sigma^2 (X'Q^{-1}X)^{-1} \tag{5-44}$$

com as variâncias de b e $\tilde{y}$ sendo produzidas, nesta ordem, na diagonal principal da matriz. Conforme veremos na aplicação a seguir, as variâncias de b, assim obtidas, coincidem com aquelas encontradas através da aplicação dos *MQG* aos dados amostrais apenas e, como foi demonstrado por Fuller (1980), as variâncias de $\tilde{y}$ são também as variâncias dos erros de previsão nos *MQG*.

Portanto, assim como no caso dos *MQO*, temos aqui que o ajustamento de *MQG* aplicado aos dados globais (isto é, adicionando-se variáveis binárias apropriadas), permitirá obter, como subproduto, os estimadores (com os correspondentes desvios padrão) válidos para o ajustamento do modelo amostral aos dados amostrais apenas. Por outro lado, os coeficientes das variáveis binárias produzem os valores previstos nos *MQG*, enquanto que os desvios padrão de tais coeficientes fornecem os próprios desvios padrão dos erros de previsão dos estimadores. A aplicação que se segue servirá para confirmar estes resultados.

5.3.1 — Exemplo

Para ilustrar os resultados apresentados acima, utilizaremos a estimação de um modelo caracterizado pela presença de autocorrelação positiva nos erros estocásticos. Os dados amostrais foram gerados aleatoriamente de um modelo populacional, com parâmetros conhecidos *a priori*. Mais especificamente, o modelo populacional, que é o mesmo utilizado por Kmenta (1971: 293), mas em outro contexto, tem as seguintes características:

$$y_t = 10 + 2x_t + u_t,$$

onde $\quad u_t = 0{,}8\, u_{t-1} + v_t,$

sendo $\quad v_t \sim N(0;\ \sigma^2 = 0{,}36),\ E(v_t u_{t-1}) = 0$ e $u_o \sim N(0;\ 1)$.

Estas condições implicam em:

$$E(u_t) = 0 \quad \text{e} \quad V(u_t) = 1.$$

A amostra aleatória para este modelo, foi gerada como se segue. Primeiramente, fixamos um conjunto de valores arbitrários para a variável x que para efeito dessa ilustração adotamos os mesmos 10 valores pré-fixados por Kmenta (1971) em um dos seus experimentos. A obtenção dos y resultou, portanto, da soma dos u_t (gerados segundo a espe-

cificação acima) à parte sistemática do modelo (isto é, $10 + 2x$). Note-se que os u_t, por sua vez, foram obtidos após a geração aleatória dos v_t normais. Para iniciar o processo, adotou-se o valor inicial de u_t como sendo zero (isto é, $u_o = 0$). Os dados amostrais de x e y, assim gerados, são reproduzidos abaixo, separados pelas linhas pontilhadas:

y	x	D_1	D_2
14,83	2	0	0
10,14	0	.	.
12,43	1	.	.
16,88	3	.	.
27,97	9	.	.
12,51	1	.	.
18,02	4	.	.
27,88	9	.	.
17,96	4	.	.
15,52	3	0	0
0	5	1	0
0	6	0	1

A aplicação de *MQG* aos dados amostrais forneceu:

$$\hat{y} = \underset{(0,725)}{10,369} + \underset{(0,055)}{1,945x},$$

onde os dados entre parênteses são os desvios padrão dos coeficientes.

Suponha-se agora que seja de interesse obter duas previsões de y: uma para quando $x_{n+1} = 5$ e, a outra, para $x_{n+2} = 6$. De acordo com o que já foi exposto, deveremos então ajustar (ainda pelo método dos *MQG*) o modelo global (isto é, que inclui as variáveis binárias apropriadas) aos dados apresentados no quadro acima. Isto porque os coeficientes das variáveis binárias seriam os próprios valores previstos e, os desvios padrão desses coeficientes seriam os desvios padrão dos erros de previsão correspondentes.

Com efeito, a estimação do modelo global produziu:

$$\hat{y} = \underset{(0,725)}{10,369} + \underset{(0,055)}{1,954x} - \underset{(0,624)}{19,546D_1} - \underset{(0,823)}{21,600D_2},$$

onde os dados entre parênteses são os desvios padrão dos coeficientes. Note-se que os coeficientes (com desvios padrão respectivos) do ajusta-

mento amostral, são produzidos como subproduto do ajustamento global. Quanto aos coeficientes das variáveis binárias, estes correspondem, afora o problema do sinal trocado (que seria evitado se substituíssemos I por $-I$, no quadro dos dados), aos valores previstos de y quando $x = 5$ e $x = 6$, respectivamente. Para nos certificarmos da validade desses resultados, basta estimá-los pelo método convencional. De acordo com Johnston (1972: 266), o valor previsto para o período $n + 1$, num modelo caracterizado por autocorrelação positiva (isto é, $p > 0$), seria:

$$\hat{y}_{n+1} = \hat{B}_0 + \hat{B}_1 x_{n+1} + pe_n ,$$

onde $\hat{B}_0$ e $\hat{B}_1$ são estimadores de MQG, p é o coeficiente de autocorrelação ($p = 0{,}8$ nesta análise) e e_n é o resíduo dos MQG, correspondente ao último ano do período amostral (aqui $e_n = 0{,}684$). A substituição dessas informações na equação acima para $x_{n+1} = 5$, produziu $\hat{y}_{n+1} = 19{,}546$, que é exatamente igual ao coeficiente da variável D_1 no ajustamento global, como aliás previsto pela técnica apresentada. Note-se, ainda, que o coeficiente da variável D_2 é o valor previsto para $\hat{y}_{n+2}$ quando $x_{n+2} = 6$; vale dizer, na equação acima devemos substituir e_n por e_{n+1}, que corresponderia ao resíduo obtido no ajustamento global, para o período $n + 1$; no nosso caso, $e_{n+1} = -0{,}549$ que, quando substituído na equação acima, produziu $\hat{y}_{n+2} = 21{,}600$, valor idêntico ao coeficiente da variável D_2.

Observe-se que, no modelo autorregressivo, é importante a ordem em que os x_{n+i} entram na previsão. Por exemplo, se invertessemos a ordem em que os x_{n+i} entram no quadro dos dados acima (isto é, em vez da seqüência 5 e 6, adote-se 6 e 5), então o ajustamento global forneceria:

$$\hat{y} = \underset{(0{,}725)}{10{,}369} + \underset{(0{,}055)}{1{,}945x} - \underset{(0{,}635)}{21{,}490D_1} - \underset{(0{,}815)}{19{,}655D_2}$$

Agora o coeficiente de D_1 é o valor previsto para $\hat{y}_{n+1}$ quando $x_{n+1} = 6$, e o coeficiente de D_2 é o valor previsto $\hat{y}_{n+2}$ quando $x_{n+2} = 5$. Esses valores diferem do ajustamento anterior, já que a ordem dos x_{n+i} difere também nos dois casos.

Para concluir, ressalte-se que os desvios padrão dos coeficientes das variáveis D_1 e D_2 são, na verdade, os desvios padrão dos erros de previsão produzidos pela estimação de MQG, nas correspondentes situações.

5.4 — Extensões

As técnicas aqui apresentadas, quanto ao uso das variáveis binárias em estudos de previsão, podem ser estendidas, também, para estimações envolvendo sistema de equações simultâneas e séries temporais autorregressivas {Fuller (1980)}. Porém, como os aspectos teóricos exigidos

nesses tópicos são de complexidade maior que a pretendida para este texto, nos limitaremos, aqui, a uma breve descrição da questão e, assim mesmo, apenas no que diz respeito ao problema das equações simultâneas.

Imagine-se um sistema de equações simultâneas com G variáveis endógenas e K variáveis exógenas (ou pré-determinadas), sendo as informações disponíveis para n períodos. Se desejarmos estimar a previsão para o período $n + 1$, basta adicionar uma variável binária como variável exógena (assumindo o valor 0 no período amostral e 1 no período da previsão) às várias equações e aumentar o número de observações para $n + 1$. As variáveis endógenas devem ser todas iguais a zero no período $n + 1$. Nestas circunstâncias, Fuller (1980) demonstrou que se a estimação for efetuada para a forma reduzida das equações simultâneas, então os coeficientes das variáveis binárias seriam os valores previstos do $\hat{y}_{n+1}$ nas correspondentes equações, enquanto que os desvios padrão dos coeficientes das variáveis binárias forneceriam os desvios padrão das previsões. Se, entretanto, a estimação for efetuada para a forma estrutural das equações simultâneas, então, como ainda demonstrado por Fuller (1980) cálculos adicionais seriam necessários para a obtenção dos valores previstos, com os seus respectivos desvios padrão; de qualquer maneira, tais cálculos teriam como ponto de partida os coeficientes das variáveis binárias obtidos na estimação estrutural.

6 — A REGRESSÃO COM VARIÁVEL DEPENDENTE QUALITATIVA

Há, freqüentemente, situações onde a variável dependente da regressão representa a escolha entre duas ou mais alternativas qualitativas. Diversos exemplos dessa natureza, logo vem à mente, como questões do tipo escolher essa ou aquela ocupação, participar ou não da força de trabalho, votar contra ou a favor de certa questão, viajar de trem, ônibus ou carro próprio, etc. Naturalmente, se houver apenas duas alternativas, então a variável dependente na regressão poderá ser representada por uma variável binária. Esta é, aliás, a primeira vez que surge aqui a possibilidade de tal uso para as variáveis binárias, pois nas seções anteriores elas estiveram sempre associadas ao lado direito da equação de regressão.

A escolha entre as alternativas depende, em geral, de vários fatores, que devem incluir tanto as características do indivíduo (aqui entendido num sentido genérico, podendo se tratar de entidades como empresas, cidades, etc.) como os atributos das alternativas objeto da escolha pelo indivíduo. Assim sendo, na formulação do modelo de regressão, a preocupação central deverá ser com a identificação de tais fatores, bem como a estimação das suas influências relativas na escolha dos indiví-

duos, para que, com base nesse conhecimento, possa-se prever a probabilidade de que um indivíduo (parte ou não da amostra) com dadas características venha a escolher uma alternativa em vez da outra. As dificuldades que surgem, quando o modelo usual de regressão é utilizado como especificação para problemas dessa natureza, são analisadas no próximo item, com as seções seguintes tratando de outras especificações, como o modelo logit.

6.1 — O modelo linear de probabilidade e suas limitações

Considere-se, inicialmente, que duas alternativas se apresentem ao indivíduo, a escolha entre as quais é decidida em função de vários atributos individuais, de natureza sócio-econômica, por exemplo. Uma primeira especificação para consideração seria o modelo linear dado por:

$$y_i = X_i B + \mu_i \tag{6-1}$$

onde:

X_i = vetor linha dos valores dos atributos;

B = vetor coluna de parâmetros;

μ_i = erro estocástico com $E(\mu_i) = 0$;

$$y_i = \begin{cases} 1 \text{ se a primeira alternativa for escolhida} \\ 0 \text{ se a escolha for a segunda alternativa.} \end{cases}$$

Note-se que:

$$E(y_i) = X_i B \,. \tag{6-2}$$

e supondo-se ser P_i a probabilidade de $y_i = 1$, e $1 - P_i$ a probabilidade de $y_i = 0$, segue-se então que:

$$E(y_i) = P_i \tag{6-3}$$

e, conseqüentemente:

$$P_i = X_i B \,. \tag{6-4}$$

Nestas circunstâncias, pois, a equação (6-1) poderá ser interpretada como um modelo probabilístico (ver aplicação no apêndice).

A utilização da técnica de MQO para a estimação do modelo em (6-1) é, entretanto, inadequada. Primeiramente, porque, como demonstrado por exemplo por Goldberger (1964: 249), a variância de μ_i seria:

$$V(\mu_i) = E(y_i) \; [1 - E(y_i)] \tag{6-5}$$

indicando ser o modelo heterocedástico. Em segundo lugar, não é difícil perceber, também, que μ_i não tem distribuição normal, invalidando pois a aplicação dos testes estatísticos clássicos para os parâmetros da regressão.

Naturalmente, o problema da heterocedasticidade poderia ser contornado com a estimação do modelo pelo método dos MQG — ou, no caso, Mínimos Quadrados Ponderados (MQP) —, com a variância de μ_i sendo estimada por:

$$\sigma_i^2 = \hat{y}_i \ (1 - \hat{y}_i) \tag{6-6}$$

onde:

$$\hat{y}_i = X_i\hat{B} \tag{6-7}$$

e $\hat{B}$ é o vetor dos estimadores de MQO. A maior dificuldade aqui, seria o fato de $\hat{y}_i$ poder situar-se fora do intervalo (0,1). Quando esse for o caso, talvez pudesse pensar-se na simples eliminação das observações correspondentes a tais $\hat{y}_i$ ou então ter esses $\hat{y}_i$ assumido valores do tipo 0,01 ou 0,99, conforme seja a situação. Entretanto, como ressaltado por Pindyck & Rubinfeld (1981: 276), esse procedimento poderia resultar em estimadores de MQP não-eficientes, particularmente para pequenas amostras; para amostras grandes não haveria maiores dificuldades.

A violação da condição de normalidade para a distribuição dos erros μ_i, apesar de também causar problemas para pequenas amostras, não traria dificuldades no caso de grandes amostras, dada a existência de testes estatísticos apropriados para tais circunstâncias [Pindyck & Rubinfeld, (1981: 277)].

Apesar dessas eventuais soluções, uma limitação básica do modelo linear de probabilidade continuaria sendo a possibilidade de obter-se previsões fora do intervalo (0, 1). Naturalmente, argumentar-se pelo estabelecimento de tais previsões extremas como sendo 0 ou 1, conforme o caso, não seria uma solução satisfatória, já que isto sugeriria a previsão da escolha de uma alternativa como sendo certa quando é inteiramente possível que ela não venha a ocorrer. Nem a estimação de MQO com a restrição da desigualdade $0 < \hat{y}_i < 1$ seria de todo satisfatória, pois apesar desse procedimento garantir previsões dentro do intervalo (0, 1) quando do uso dos vetores amostrais X_i, o mesmo não poderia ser dito com relação a outros valores de X [Judge, Griffiths, Hill & Lee (1980: 589)]; além disso, esses estimadores, apesar de consistentes, seriam agora tendenciosos em pequenas amostras, embora possuíssem menores variâncias [Domencich & McFadden (1975: 105)].

As dificuldades apontadas deixam patente a necessidade de uma especificação alternativa para a regressão. Como a limitação maior do modelo linear de probabilidade parece ser a da possibilidade de previsões

fora do intervalo (0, 1) cumpre indagar que tipos de transformações poderiam ser operadas no modelo para que as previsões fossem contidas nesse intervalo. Naturalmente, qualquer tentativa nesse sentido deverá concentrar-se na conversão dos valores de x em valores de y contidos dentro da escala 0 — 1. Seria desejável, ainda, que a transformação contemplada fosse do tipo monotônica — isto é, a aumentos em x deveriam corresponder aumentos (ou diminuições) nos valores da variável dependente da regressão. Há duas transformações básicas com essa preocupação: o modelo probit (com base na distribuição acumulada normal) e o modelo logit (baseado na distribuição logística). Como esses dois casos têm características bem semelhantes e, considerando-se ainda ser menor o esforço computacional no modelo logit [Pindyck & Rubinfeld (1981: 287)], concentraremos os esforços aqui à formulação deste último apenas.

6.2 — O modelo logit para escolha binária

Seja a função de probabilidade da distribuição logística dada por (onde *exp* é a função exponencial):

$$P_i = [1 + exp(-X_iB)]^{-1}; \tag{6-8}$$

note-se que agora P_i estaria contido dentro do intervalo (0, 1), embora X_iB possa variar de $-\infty$ a ∞. Desse resultado segue-se:

$$Log \frac{P_i}{1 - P_i} = X_iB \tag{6-9}$$

que é o denominado modelo logit para escolha binária.

Uma propriedade atrativa do modelo logit é que, como a função logística tem inclinação máxima quando $P_i = 0{,}5$ $\left\{\text{já que } \frac{dP_i}{dX} = P_i(1 - P_i)B\right\}$ e essa inclinação diminui gradativamente quando nos afastamos na direção das duas extremidades, $P_i = 0$ e $P_i = 1$, então isto significa, obviamente, que seriam exigidas maiores variações nas condições que influem nas decisões dos indivíduos, à medida que nos dirigimos a essas extremidades, para que estes alterem as suas atitudes (escolhas). Vale dizer, se x for a renda do indivíduo, então, nas extremidades da pobreza e da riqueza, os indivíduos estariam menos sujeitos a mudanças de atitude, quando as suas rendas variam; o que parece, aliás, ser um pressuposto bastante razoável, neste caso.

Quanto aos aspectos da estimação do modelo em (6-9), uma primeira dificuldade pareceria ser o fato de que se $P_i = 1$ ou 0, então $P_i/(1 - P_i)$ seria ∞ ou 0 e, conseqüentemente, o logaritmo dessa razão seria indefinido. Existem, entretanto, métodos apropriados de estima-

ção para esse modelo, havendo porém a necessidade de se distinguir, primeiramente, entre duas classes de dados: observações individuais e observações agrupadas.

Suponha-se, inicialmente, que os dados sejam agrupados segundo os atributos x_i. Assim sendo, se conhecermos as freqüências das escolhas de uma alternativa, para cada grupo, então poderemos substituir P_i em (6-9) por:

$$f_i = r_i/n_i \tag{6-10}$$

onde r_i seria o número de casos que selecionaram dada alternativa, no grupo i, com n_i sendo o número total de casos no grupo i. Assim, em vez do modelo em (6-9) teríamos agora a especificação:

$$Log \frac{f_i}{1 - f_i} = Log \frac{r_i}{n_i - r_i} = X_i B + \mu_i, \tag{6-11}$$

onde o erro μ_i seria naturalmente dado por $Log \frac{f_i}{1 - f_i} - Log \frac{P_i}{1 - P_i}$. Se, por um lado, o estimador de B poderá ser tendencioso para amostras pequenas, por outro lado, à medida que os n_i forem aumentando, os resultados melhorariam tornando-se o estimador, inclusive, consistente, quando esses n_i atingirem valores arbitrariamente grandes (Pindyck & Rubinfeld, 1981: 290).

Como foi notado, f_i é tão somente uma aproximação de P_i. Desse modo, pressupondo-se que cada observação individual, no grupo i, seja independente e tenha distribuição binomial, então demonstra-se [por exemplo, Theil (1967: 66); ou Theil (1971: 636)] que μ_i em (6-11) terá, para amostras grandes, distribuição aproximadamente normal, com média zero e variância estimada por (ver apêndice técnico).

$$V(\mu_i) = \frac{n_i}{r_i(n_i - r_i)} = \frac{1}{n_i f_i (1 - f_i)}. \tag{6-12}$$

Isto é, o modelo é heterocedástico, necessitando ser estimado por MQP, onde as observações sejam ponderadas por $1/\ V(\mu_i)$ (ver chamada 2).

No caso de dispor-se de observações individuais, sem que se queira adotar qualquer agregação das informações, então poderemos ter uma

---

[2] Se as amostras n_i forem pequenas, então Cox (1970) propõe os seguintes ajustamentos [ver também Pindyck & Rubinfeld (1981: 293)]: primeiramente adotar o modelo

$$Log \left(\frac{r_i + 0{,}5}{n_i - r_i + 0{,}5} \right) = X_i B$$

que evitará as indeterminações mencionadas em relação à equação em (6-11); em segundo lugar, para corrigir a heterocedasticidade, adotar a variância

$$V(u_i) = \frac{(n_i + 1)\ (n_i + 2)}{n_i (r_i + 1)\ (n_i - r_i + 1)}$$

situação onde, para cada variável (ou conjunto das variáveis) x, haja a escolha de apenas uma alternativa (isto é, $f_i = 1$ ou 0). Nestas circunstâncias, a estimação terá de ser efetuada pelo método da máxima verossimilhança que, se por um lado, é computacionalmente mais onerosa (por envolver métodos não-lineares de estimação), por outro lado, produz estimadores consistentes; as variâncias assintóticas desses estimadores são também geralmente obtidas no processo de estimação. Esse método será discutido em maiores detalhes apenas no último item desta seção, pois nos próximos itens nos concentraremos nos casos onde f_i possa substituir P_i para efeito da estimação do modelo logit.

6.3 — O modelo logit binário para variáveis qualitativas

Imagine-se, inicialmente, que a escolha de um indivíduo, por uma dada alternativa em vez da outra, seja determinada por variáveis qualitativas dos tipos: sexo do indivíduo e o local da sua origem (por exemplo, urbana e rural), as quais denotaremos de x_j e z_k, para j e $k = 1$ e 2, por exemplo. Desta forma, P_{jk} seria a probabilidade da escolha de dada alternativa diante das condições dadas por x_j e z_k.

Admitindo-se não haver interações entre as variáveis x e z (isto é, o efeito de x sobre a escolha da alternativa independe das condições de z), então, seguindo-se Theil (1970), teríamos a especificação:

$$L_{jk} = Log \frac{P_{jk}}{1 - P_{jk}} = \alpha + B_j + C_k \qquad (6\text{-}13)$$

que mede, naturalmente, o efeito das variáveis x e z no logaritmo da razão quando $x = x_j$ e $z = z_k$; sendo α, B_j e C_k os parâmetros do modelo. Para a estimação apropriada desse modelo, Theil (1970) sugere que se adote, sem que haja qualquer prejuízo quanto a generalidade dos resultados obtidos, a seguinte normalização:

$$B_1 = C_1 = 0. \qquad (6\text{-}14)$$

Assim sendo, $L_{11} = \alpha$ indicaria o log da razão favorável à escolha de uma dada alternativa, quando $x = x_1$ e $z = z_1$.

Uma maneira alternativa para a especificação desse mesmo modelo logit, seria através da adoção de variáveis binárias para representar os fatores sexo e origem do indivíduo. Com efeito, considere-se a especificação:

$$L_{jk} = Log \frac{P_{jk}}{1 - P_{jk}} = \alpha + B_j x_j + C_k z_k , \qquad (6\text{-}15)$$

com $x_1 = 0$, se o indivíduo for feminino; $x_2 = 0$, se o indivíduo for masculino; $z_1 = 0$, se o indivíduo for rural, e, $z_2 = 1$ se o indivíduo for urbano.

Assim, o modelo em detalhes seria:

$$
\begin{aligned}
L_{11} &= \alpha + B_1x_1 + C_1z_1 \\
L_{12} &= \alpha + B_1x_1 + C_2z_2 \\
L_{21} &= \alpha + B_2x_2 + C_1z_1 \\
L_{22} &= \alpha + B_2x_2 + C_2z_2 \, .
\end{aligned}
\tag{6-16}
$$

Como, porém, $x_1 = z_1 = 0$, então não haveria necessidade em estimarmos B_1 e C_1. Daí podermos adotar a normalização $B_1 = C_1 = 0$, dada em (6-14). Desta forma, necessitamos estimar apenas 3 dos 5 parâmetros do modelo (ver apêndice para x_j, com $j = 1, 2$ e 3).

Afora essas particularidades que acabamos de discutir, o modelo logit binário para variáveis qualitativas seria estimado de maneira semelhante ao caso do modelo logit binário com regressores quantitativos, apresentado em (6-11), usando, inclusive, o mesmo esquema para contornar o problema da heterocedasticidade, mostrado em (6-12).

6.4 — O modelo logit multinomial

Quando houver a possibilidade de escolha entre mais de duas alternativas, o modelo logit apropriado é então denominado de multinomial. A especificação e estimação de tal modelo são discutidas em seguida.

À primeira vista, poder-se-ia imaginar que a especificação do modelo multinomial seguiria linhas semelhantes ao caso do modelo logit binário. Vale dizer, com três alternativas, cujas probabilidades de ocorrência fossem P_1, P_2 e P_3, poder-se-ia pensar na estimação do modelo utilizando três relações do tipo:

$$L_i = Log \frac{P_i}{1 - P_i} = XB_i \, , \quad \text{para } i = 1, 2 \text{ e } 3. \tag{6-17}$$

Esse procedimento não é, entretanto, adequado, pois como salientado por Theil (1970: 118), a aplicação de tal princípio conduziria a uma estimação inconsistente. O fato é que cada uma dessas equações em (6-17) poderia ser reescrita de modo que P_i fosse uma função explícita dos parâmetros da correspondente equação logit. Porém, a soma dos P_i seria necessariamente igual a unidade, e não há nenhuma garantia de que a soma das correspondentes funções dos parâmetros seja igual a um.

Em vista da dificuldade que acabamos de apresentar, um modelo mais apropriado para o caso multinomial seria a chamada função logit condicional do tipo:

$$L_{rs} = Log \left(\frac{P_r}{P_s} \right) = XB_{rs} \quad \text{para } r \neq s \, . \tag{6-18}$$

Note-se que, por esta especificação, determinamos somente a razão das probabilidades. As probabilidades individuais podem, entretanto, ser determinadas sem qualquer dificuldade, uma vez considerado o fato de que a soma de tais probabilidades seria igual a unidade; retornaremos a este ponto mais adiante.

Um aspecto importante do modelo logit multinomial é que, como fora demonstrado por Theil (1969), não seria necessário estimarmos as equações em (6-18) para todos os r e s, pois os parâmetros dessas relações são caracterizados por certas condições de circularidade que permitiriam obter os parâmetros de uma da combinação dos parâmetros das demais. Nesse sentido, consideremos como ilustração o caso da logit condicional seguinte:

$$L_{rt} = Log\left(\frac{P_r}{P_t}\right) = XB_{rt}\ . \tag{6-19}$$

Esta relação poderia ser também escrita como:

$$L_{rt} = Log\left(\frac{P_r}{P_t}\ \frac{P_s}{P_s}\right) = Log\frac{P_r}{P_s} + Log\frac{P_s}{P_t} = \\ = X(B_{rs} + B_{st}) = XB_{rt} \tag{6-20}$$

Assim,

$$B_{rs} = B_{rt} - B_{st} = B_r - B_s \tag{6-21}$$

já que t seria comum aos dois vetores B. Desta forma, verificamos que, se fixarmos s e variarmos r, então todas as razões L_{rs} conterão B_s. Portanto, nas estimações dos diversos L_{rs}, basta adotar a normalização $B_s = 0$. Para o caso multinomial com três alternativas, por exemplo, se fixarmos $s = 1$, então basta estimar as relações L_{21} e L_{31}, com $B_s = 0$, e se desejarmos os parâmetros B_{23} correspondentes à relação L_{23}, estes seriam dados por:

$$L_{23} = Log\left(\frac{P_2}{P_1}\cdot\frac{P_1}{P_3}\right) = L_{21} + L_{13} = L_{21} - L_{31} = B_{21} - B_{31} = B_2 - B_3 \tag{6-22}$$

Para efeito de estimação, basta substituir no modelo em (6-18) $Log\left(\frac{P_r}{P_s}\right)$ por $Log(f_r/f_s)$ e adicionar o erro μ_{rs}, definido como $Log\frac{f_r}{f_s} - Log\frac{P_r}{P_s}$. Desse modo, o modelo logit multinomial para dados agrupados, de modo análogo ao modelo logit binário, é heterocedástico, já que a variância dos erros μ_{rs} é, conforme demonstrado por Theil (1970) e também no apêndice técnico deste texto, dada por:

$$V(\mu_{rs}) = \frac{1}{n_i}\left(\frac{1}{P_{ri}} + \frac{1}{P_{si}}\right), \tag{6-23}$$

para um dado grupo i. Além do mais, esses erros estocásticos são correlacionados entre as equações da logit condicional dentro de um dado grupo i; por exemplo, para o caso com três alternativas, e onde a referência é $s = 1$, a covariância entre os erros das duas equações L_{21} e L_{31} seria (ver apêndice técnico):

$$Cov\ (\mu_{21}\ \mu_{31}) = \frac{1}{n_i}\ \frac{1}{P_{1i}}. \tag{6-24}$$

Deste modo, para o caso com três alternativas, e onde $s = 1$, temos que a matriz das variâncias e covariâncias dos erros seria, para um dado grupo i:

$$V_i = \frac{1}{n_i}\begin{bmatrix} \frac{1}{P_1} + \frac{1}{P_2} & \frac{1}{P_1} \\ \frac{1}{P_1} & \frac{1}{P_1} + \frac{1}{P_3} \end{bmatrix} \tag{6-25}$$

cujo inverso seria, de acordo com Theil (1970):

$$V_i^{-1} = n_i \begin{bmatrix} P_2(1 - P_2) & -P_2P_3 \\ -P_2P_3 & P_3(1 - P_3) \end{bmatrix}, \tag{6-26}$$

onde o subíndice i foi deliberadamente omitido nas várias probabilidades das duas matrizes para simplificar a apresentação. Assim sendo, para obtermos os estimadores de MQG basta aplicar a fórmula:

$$B = (X'V^{-1}X)^{-1}X'V^{-1}Y \tag{6-27}$$

onde V^{-1} é uma matriz bloco diagonal composta de várias submatrizes do tipo dado em (6-26), uma para cada grupo i. A variância dos estimadores de MQG, por outro lado, seria dada por:

$$V(B) = (X'V^{-1}X)^{-1} \tag{6-28}$$

6.5 — O modelo logit multinomial para variáveis qualitativas

A adaptação do modelo para o caso com regressores qualitativos segue orientação semelhante aquela utilizada com relação ao modelo binário, discutida anteriormente. Desta forma, imaginemos o modelo multinomial com três alternativas e onde as variáveis explicativas sejam as mesmas do caso binário, isto é, sexo (x_j) e local de origem (z_k), para j e $k = 1$ e 2. Assim, temos agora duas equações básicas para a logit condicional, em vez de uma equação como no caso binário. Mais precisamente, se a referência for $s = 1$, temos então as equações:

$$\begin{aligned} L_{21/jk} &= \alpha_2 + B_{j2}\ x_j + C_{k2}\ z_k \\ L_{31/jk} &= \alpha_3 + B_{j3}\ x_j + C_{k3}\ z_k \end{aligned} \tag{6-29}$$

que quando escritas em detalhe fornecem os seguintes pares de equações:

$$
\begin{aligned}
L_{21/11} &= \alpha_2 + B_{12}\, x_1 + C_{12}\, z_1 \\
L_{31/11} &= \alpha_3 + B_{13}\, x_1 + C_{13}\, z_1 \\
L_{21/12} &= \alpha_2 + B_{12}\, x_1 + C_{22}\, z_2 \\
L_{31/12} &= \alpha_3 + B_{13}\, x_1 + C_{23}\, z_2 \\
L_{21/21} &= \alpha_2 + B_{22}\, x_2 + C_{12}\, z_1 \\
L_{31/21} &= \alpha_3 + B_{23}\, x_2 + C_{13}\, z_1 \\
L_{21/22} &= \alpha_2 + B_{22}\, x_2 + C_{22}\, z_2 \\
L_{31/22} &= \alpha_3 + B_{23}\, x_2 + C_{23}\, z_2
\end{aligned}
\tag{6-30}
$$

Se adotarmos agora o mesmo procedimento do caso binário e tratarmos as variáveis x e z como variáveis binárias onde $x_1 = 0$ se o indivíduo for feminino, $x_2 = 1$ se o indivíduo for masculino, $z_1 = 0$ se o indivíduo for rural e $z_2 = 1$ se o indivíduo for urbano, então como $x_1 = z_1 = 0$ os coeficientes associados a essas variáveis não precisam ser estimados. Daí podermos adotar a seguinte normalização:

$$B_{12} = C_{12} = B_{13} = C_{13} = 0, \tag{6-31}$$

havendo portanto que se estimar apenas 6 dos 10 parâmetros em (6-30). Com essas considerações, temos que $L_{21/11} = \alpha_2$ mostra a logit condicional favorável a escolha da alternativa 2, relativamente à escolha 1, quando se é mulher e de origem rural, enquanto que $L_{31/11} = \alpha_3$ mostra a escolha da alternativa 3, relativamente à escolha 1, diante dessa mesma combinação de variáveis, etc.

Observe-se que as equações em (6-30), combinadas com as restrições em (6-31), podem ser escritas em forma matricial como $Y = XB$, onde:

$$
Y = \begin{bmatrix} L_{21/11} \\ L_{31/11} \\ L_{21/12} \\ L_{31/12} \\ L_{21/21} \\ L_{31/21} \\ L_{21/22} \\ L_{31/22} \end{bmatrix} ;\ X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \text{ e } B = \begin{bmatrix} \alpha_2 \\ \alpha_3 \\ B_{22} \\ B_{23} \\ C_{22} \\ C_{23} \end{bmatrix}
\tag{6-32}
$$

É interessante notar, ainda, que a matriz X é composta de uma série de submatrizes unitárias e submatrizes nulas, todas de ordem 2×2; isto é:

$$X = \begin{bmatrix} I & 0 & 0 \\ I & 0 & I \\ I & I & 0 \\ I & I & I \end{bmatrix} \tag{6-33}$$

A lei de formação dessa matriz é completamente geral. Por exemplo, se no mesmo modelo logit, aqui considerado, 4 fossem as alternativas em vez das 3 utilizadas, então a única alteração na matriz X seria quanto à ordem das submatrizes I e 0, que passariam a ter ordem 3×3, em vez de 2×2; naturalmente se houvesse apenas duas alternativas, como no caso binário, então os escalares 0 e 1 assumiriam os lugares das submatrizes 0 e I.

Afora as particularidades discutidas acima, o modelo logit multinomial para variáveis qualitativas seria estimado de modo semelhante àquele com regressores quantitativos, isto é, considerando-se o mesmo esquema de heterocedasticidade apresentado em (6-26).

Uma vez estimadas as logit condicionais, podemos determinar facilmente as probabilidades individuais. Com efeito, da relação em (6-18), se $s = 1$, então para $r \neq s$ temos:

$$P_r = P_1 \, exp(XB_r) \qquad \text{para } r = 2, 3, \ldots, k. \tag{6-34}$$

Considerando-se ainda que as probabilidades devem somar igual a 1, segue-se que:

$$P_1 = \left[1 + \sum_{r=2}^{k} exp(X \, B_r) \right]^{-1} \tag{6-35}$$

6.6 — Desenvolvimentos mais recentes para a logit com dados agrupados

Conforme foi salientado nas seções anteriores, as variâncias dos erros em (6-12), para a logit binária, e em (6-25), para a logit multinomial, resultam do fato de usarmos a razão de freqüências amostrais, em vez da razão de probabilidades, como variável dependente na regressão. Mais especificamente, esses erros são do tipo:

$$\mu_i = Log \frac{f_i}{1 - f_i} - Log \frac{P_i}{1 - P_i} \text{ (caso binário)} \tag{6-36}$$

e

$$\mu_{rs} = Log \frac{f_r}{f_s} - Log \frac{P_r}{P_s} \text{ (caso multinomial).} \tag{6-37}$$

Entretanto, como proposto por Amemiya & Nold (1975), em relação ao modelo logit binário, além do erro μ, haveria ainda um outro componente de erro devido a uma possível má especificação do próprio modelo populacional (por exemplo, omissão de variável explicativa). Nestas circunstâncias, em vez da especificação em (6-11), teríamos:

$$Log \frac{f_i}{1 - f_i} = X_i B + \mu_i + v_i \,. \tag{6-38}$$

Se adotarmos agora os pressupostos $E(v_i) = 0$; $V(v_i) = \sigma^2$ e $Cov(\mu_i, v_i) = 0$, e se denominarmos $\varepsilon_i = \mu_i + v_i$, temos então:

$$V(\varepsilon_i) = \frac{1}{n_i P_i(1 - P_i)} + \sigma^2 \,. \tag{6-39}$$

Assim, Amemiya & Nold (1975) propõem estimar o parâmetro B em (6-38), através do seguinte procedimento: primeiramente, aplicamos MQO diretamente em (6-38) para obter o somatório dos resíduos quadráticos Σe_i^2 que então forneceria uma estimativa da variância de ε_i como:

$$\hat{V}(\varepsilon_i) = \frac{\Sigma e_i^2}{n} \,, \tag{6-40}$$

onde n seria o tamanho da amostra, no caso número de grupos. Em seguida uma estimativa para σ^2 seria obtida através de:

$$s^2 = \frac{1}{n}\left[\Sigma e_i^2 - \Sigma \frac{1}{n_i f_i(1 - f_i)}\right]. \tag{6-41}$$

Finalmente, o modelo logit seria estimado após a ponderação das observações pelos fatores:

$$1/V(\varepsilon_i) = \left[s^2 + \frac{1}{n_i f_i(1 - f_i)}\right]^{-1} \,. \tag{6-42}$$

Os estimadores assim obtidos são denominados pelos autores de Mínimos Quadrados Ponderados Modificados (MQM), e o modelo como especificado em (6-38) será então simplesmente denominado aqui de logit modificado. Na aplicação efetuada por Amemiya & Nold (1975) fica demonstrado que esses estimadores são mais eficientes do que os estimadores de MQO e os estimadores de MQP discutidos no item 6.2 [isto é, com $s^2 = 0$, em (6-42)]. Isso, naturalmente, quando as variâncias daqueles estimadores forem comparadas com as variâncias não tendenciosas desses dois últimos estimadores, pois se a especificação em (6-38) for correta, então a fórmula padrão para estimar as variâncias dos estimadores de MQP tende a subestimá-las consideravelmente. Na verdade, essa subestimação seria tanto maior quanto maior for a razão entre a variância de v_i e a variância de μ_i [Amemiya & Nold (1975)].

Os resultados obtidos por Amemiya & Nold (1975) que discutimos anteriormente, foram recentemente generalizados por Parks (1980) para o modelo logit multinomial. Assim, seja a especificação da logit condicional dada por:

$$Y_{it} = Log\left(\frac{P_{it}}{P_{ot}}\right) = X_{it}\, B_i + v_{it}\ , \qquad (6\text{-}43)$$

onde v_{it} seria um erro devido, por exemplo, à omissão de alguma variável independente na regressão; os subíndices i e t referem-se a alternativa escolhida e grupo do indivíduo, respectivamente. Como, entretanto, para efeito de estimação do modelo as freqüências amostrais devem substituir as probabilidades, então a especificação apropriada seria:

$$y_{it} = Log\left(\frac{f_{it}}{f_{ot}}\right) = X_{it}\, B_i + v_{it} + \mu_{it}\ , \qquad (6\text{-}44)$$

onde $\mu_{it} = Log\left(\frac{f_{it}}{f_{ot}}\right) - Log\left(\frac{P_{it}}{P_{ot}}\right)$, como já referido anteriormente.

Observe-se que no caso logit multinomial temos um sistema de equações dentro de cada grupo, em vez de uma equação apenas como no modelo binário. A estrutura dos erros μ_{it} em tais situações já foi discutida em (6-25) no contexto de um modelo logit para três alternativas (resultados aqueles, aliás, que podem ser facilmente generalizados). Naquela ocasião, ficou patente que os erros são correlacionados entre as equações dentro de um grupo, mas não há qualquer correlação entre as equações de diferentes grupos. Pressuposto semelhante se aplica para os erros v_{it}. Mais precisamente, temos:

$$E(v_{it}, v_{js}) = \begin{cases} \sigma_{ij} \text{ para } t = s, \text{ para todo } i \text{ e } j, \\ 0 \text{ para } t \neq s, \text{ para todo } i \text{ e } j. \end{cases} \qquad (6\text{-}45)$$

Isto é, como v_{it} seria o resultado de um erro de especificação, então o erro cometido com relação a uma probabilidade de escolha deverá ser contrabalançado com as alternativas remanescentes do mesmo grupo, mas esses erros não são correlacionados entre os grupos.

Além disso, assim como no caso do modelo logit binário, temos também aqui que por pressuposto os erros μ_{it} são independentes dos erros v_{js} para todo i, j, t e s.

Com essas considerações, temos as seguintes generalizações dos resultados de Amemiya & Nold (1975):

1 — em vez do escalar em (6-40) teríamos agora a matriz de covariância dos resíduos

$$S = \left[\left(\frac{1}{n}\sum_{t=1}^{n} e_{it}\, e_{jt}\right)\right], \qquad (6\text{-}46)$$

onde e_{it} e e_{jt} representam respectivamente os resíduos de MQO, relativos às alternativas i e j no grupo t; assim, com três alternativas dentro de cada grupo, esta matriz seria de ordem 2 x 2;

2 — uma estimativa consistente da matriz das covariâncias dos erros contemporâneos (isto é, entre as equações dentro de um mesmo grupo) seria, generalizando o resultado em (6-41),

$$\Delta = S - \frac{1}{n} \sum_{t=1}^{n} V_t \quad , \tag{6-47}$$

onde V_t seria naturalmente a matriz do tipo dada em (6-25), para o caso com três alternativas; e, finalmente;

3 — o resultado em (6-42) seria generalizado como

$$\Omega = \begin{bmatrix} V_1 + \Delta & & & O \\ & V_2 + \Delta & & \\ & & \ddots & \\ O & & & V_n + \Delta \end{bmatrix} . \tag{6-48}$$

Desse modo, teríamos que os estimadores de *MQO, MQP* e *MQM*, para o modelo logit multinomial, seriam dados respectivamente por:

a — $b_{MQO} = (X'X)^{-1} X'Y$; b — $b_{MQP} = (X'V^{-1}X)^{-1} X'V^{-1}Y$

e

$$\text{c} — b_{MQM} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y \tag{6-49}$$

cujas variâncias respectivas seriam {Parks (1980)}:

$$\text{a} — V(b_{MQO}) = (X'X)^{-1} X'\Omega X (X'X)^{-1} \tag{6-50}$$

em vez da fórmula usual $s^2(X'X)^{-1}$, onde $s^2 = e'e/J(n-k)$ e onde J é o número de alternativas dentro de cada grupo, k o número de parâmetros estimados, e sendo e o vetor dos resíduos;

$$\text{b} — V(b_{MQP}) = (X'V^{-1}X)^{-1} X'V^{-1} \Omega V^{-1} X (X'V^{-1}X)^{-1} \tag{6-51}$$

em vez de $(X'V^{-1}X)$; e, finalmente,

$$\text{c} — V(B_{MQM}) = (X'\Omega^{-1}X)^{-1}. \tag{6-52}$$

Assim como em Amemiya & Nold (1975), Parks (1980) mostra, com uma aplicação da logit multinomial, que as fórmulas tradicionais para a variância dos estimadores de *MQO* e *MQP*, dadas lado a lado com as variâncias apropriadas em (6-50) e (6-51), produzem valores viesados

para as verdadeiras variâncias. De qualquer maneira, os estimadores de *MQM* produzem variâncias menores que aquelas em (6-50) e (6-51), conforme será mostrado no exemplo no fim desta seção, sendo, portanto, mais eficientes que aqueles dois outros estimadores.

6.7 — Testes para a precisão do ajustamento

Um teste que poderá ser utilizado para avaliar a precisão do ajustamento do modelo logit para dados agregados seria, inicialmente para o caso binário {Pindyck & Rubinfeld (1981: 293) ou Amemiya (1981: 1506)}:

$$\sum_{i=1}^{G} n_i \frac{(f_i - \hat{P}_i)^2}{\hat{P}_i (1 - \hat{P}_i)}, \tag{6-53}$$

onde n_i é o número de observações no grupo i; f_i é a freqüência observada e $\hat{P}_i$ a freqüência estimada no modelo. Esta relação tem distribuição χ^2_{G-k} (isto é, qui-quadrado com $G-k$ graus de liberdade), onde k é o número de parâmetros estimados no modelo. Naturalmente, se f_i diferir significativamente de $\hat{P}_i$, indicando portanto um mau ajustamento, o valor desta relação cairá dentro da área crítica da distribuição qui-quadrado.

Um teste equivalente àquele em (6-53) é obtido através de:

$$(y - X\hat{B})' V^{-1}(y - X\hat{B}) = \chi^2_{G-k}, \tag{6-54}$$

onde y seria o vetor da logit observada; B o vetor de estimadores de *MQP* discutidos no item 6.2 e V^{-1} seria a matriz diagonal, cujos elementos são o inverso daqueles dados em (6-12).

A adaptação do teste em (6-53) para o caso multinomial seria simplesmente {Amemiya (1981)}:

$$\sum_{t=1}^{T} n_t \sum_{j=0}^{m_t} \frac{(f_{tj} - \hat{P}_{tj})^2}{\hat{P}_{tj}} = \chi^2_{n-K}, \tag{6-55}$$

onde $n = \sum_{t=1}^{T} n_t$, e K é o número total de parâmetros no modelo; note-se que m_t representa o número de alternativas dentro do grupo t. Naturalmente, a fórmula em (6-54) ainda seria válida como um teste equivalente ao dado em (6-55); haveria agora apenas a necessidade de considerar-se a matriz V^{-1} apropriada {isto é, do tipo dada em (6-27)} e também considerar os graus de liberdade adequados para a nova situação.

Ainda outros testes, nessas mesmas linhas, poderiam ser utilizados conforme enfatizado por Amemiya (1981). Também testes com base na

teoria da informação são às vezes empregados para avaliar a precisão do ajustamento {Theil (1971)}, mas para um entendimento apropriado dessa questão seria necessário muito mais espaço do que o pretendido neste trabalho.

6.8 — O modelo logit para dados individuais

6.8.1 — Estimação

Conforme já foi enfatizado anteriormente (ver Item 6.2), se dispomos de observações individuais, sem que se queira agregá-las (aliás, não se deve agregar observações, pois isto significaria perder informação), então a estimação do modelo logit na sua formulação em (6-11) apresentaria uma séria dificuldade, já que exigiria a avaliação de logaritmo de zero. A estimação pelo método da máxima verossimilhança contorna tal dificuldade. Assim, consideremos inicialmente o modelo logit binário onde $y_i = 1$ sempre que for escolhida a primeira alternativa e $y_i = 0$ no caso da escolha da segunda alternativa. Seja agora P_i a probabilidade de escolha da primeira alternativa e $1 - P_i$ a probabilidade de escolha da segunda alternativa. Desse modo, se pretendermos estimar o vetor B do modelo logit pelo método da máxima verossimilhança, isto significará encontrar o valor para esse vetor, que maximize a probabilidade de obtenção da amostra observada na escolha, dos indivíduos. Como os P_i são independentes entre os indivíduos então, se considerarmos que os primeiros n_1 indivíduos selecionaram a primeira alternativa e os demais n_2 indivíduos selecionaram a segunda alternativa (note-se que $N = n_1 + n_2$), a função de verossimilhança seria:

$$L = \prod_{i=1}^{n_1} P_i \prod_{i=n_1+1}^{N} (1 - P_i) = \prod_{i=1}^{N} P_i^{y_i} (1 - P_i)^{1-y_i} \qquad (6\text{-}56)$$

já que $y_i = 1$ para os primeiros n_1 indivíduos e $y_i = 0$ para os demais.

Os estimadores de máxima verossimilhança seriam então obtidos maximizando-se a função dada em (6-56). Alternativamente poderíamos maximizar o logaritmo de tal função, já que uma transformação monotônica da função não afetaria o seu ponto de máximo. Desse modo, como

$$Log\ L = \sum_{i=1}^{N} y_i\ Log\ P_i + \sum_{i=1}^{N} (1 - y_i)\ Log\ (1 - P_i) \qquad (6\text{-}57)$$

vem

$$\frac{\partial\ Log\ L}{\partial\ B} = \sum_{i=1}^{N} y_i\ X_i - \sum_{i=1}^{N} P_i\ X_i = 0\ , \qquad (6\text{-}58)$$

que se admitirmos ser $B = (a, b)'$, onde a é a interseção linear, e $X_i = (1, x_i)$, então:

$$\frac{\partial \; Log \; L}{\partial \; a} = \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} P_i = 0$$

e

$$\frac{\partial \; Log \; L}{\partial \; b} = \sum_{i=1}^{N} y_i x_i - \sum_{i=1}^{N} P_i x_i = 0 \, . \qquad (6\text{-}59)$$

Observe-se que, como essas equações são não-lineares nos parâmetros, estes terão de ser estimados por métodos iterativos. Conforme foi demonstrado por Dhrymes (1978: 344-6), a função em (6-57) terá um único máximo. Assim, qualquer procedimento iterativo que convirja para um ponto estacionário também convergirá para um máximo global {Amemiya (1981: 1495)}.

Na estimação do vetor B pelo método iterativo, e também para obter a matriz da variância assintótica desse vetor de estimadores, necessitamos calcular inicialmente a matriz das derivadas de segunda ordem da função em (6-57), bem como obter o valor esperado de tal matriz. Assim sendo, temos {Dhrymes (1978: 345), ou Amemiya (1981: 1495)}:

$$\left(E \; \frac{\partial^2 \; Log \; L}{\partial \; B \; \partial \; B'} \right) = - \sum_{i=1}^{N} f \; X_i' X_i \; , \qquad (6\text{-}60)$$

onde f é a derivada de P_i com relação ao vetor B. A variância assintótica de B seria dada pela matriz:

$$V(\hat{B}) = - \left(E \; \frac{\partial^2 \; Log \; L}{\partial \; B \; \partial \; B'} \right)^{-1} . \qquad (6\text{-}61)$$

Quanto ao estimador de máxima verossimilhança, que é consistente e assintoticamente normal, o processo iterativo para a sua estimação seria efetuado através da seguinte fórmula básica:

$$\hat{B}_2 = B_1 - \left(\frac{\partial^2 \; Log \; L}{\partial \; B \; \partial \; B'} \bigg|_{B_1} \right)^{-1} \frac{\partial \; Log \; L}{\partial \; B} \bigg|_{B_1} \; , \qquad (6\text{-}62)$$

onde $\hat{B}_1$ é um valor inicial qualquer. O estimador B_3 da terceira iteração seria obtido após a substituição de B_1 por B_2 nesta expressão. Esse processo de iteração seria repetido até que houvesse a convergência de B, que no caso resultaria no estimador de máxima verossimilhança.

Quanto ao modelo logit multinomial, se P_j for a probabilidade de escolha da alternativa j e se y_j for 0 ou 1, conforme a alternativa j seja escolhida ou não, então diante de J alternativas temos $\sum_{j=1}^{J} P_j = \sum_{j=1}^{J} y_j = 1$. Se admitirmos, ainda, a existência de n indivíduos e que

cada um desses confronta-se com as mesmas J alternativas temos, em analogia com a expressão em (6-56), a seguinte função de verossimilhança:

$$L = \prod_{i=1}^{n} P_{i1}^{y_{i1}} P_{i2}^{y_{i2}} \ldots\ldots\ldots P_{ij}^{y_{ij}}, \tag{6-63}$$

que considerando-se P_{i1} como a probabilidade de referência para o modelo, vem:

$$L = \prod_{i=1}^{n} \left[1 + \sum_{j=2}^{J} \exp\,(X_i B_j) \right]^{-1} \prod_{j=2}^{J} \prod_{i=1}^{n} \exp\,(X_i B_j y_{ij}) \,. \tag{6-64}$$

Naturalmente, na forma logarítmica, esta expressão fica

$$Log\ L = -\sum_{i=1}^{n} Log \left[1 + \sum_{j=2}^{J} exp(X_i B_j) \right] + \sum_{j=2}^{J} \sum_{i=1}^{n} X_i B_j y_{ij} \,, \tag{6-65}$$

cujo máximo seria obtido igualando-se as seguintes derivadas parciais a zero:

$$\frac{\partial\ Log\ L}{\partial\ B_j} = -\sum_{i=1}^{n} P_{ij}\ X_i - \sum_{i=1}^{n} X_i\ y_{ij} = 0, \text{ para } j = 2, \ldots, J \,. \tag{6-66}$$

Note-se que B_j é um vetor contendo tantos elementos quanto o número de variáveis explicativas no modelo mais um.

O processo iterativo de obtenção dos parâmetros do modelo multinomial, bem como a estimação das suas variâncias assintóticas, seguiria linha de raciocínio semelhante ao do caso do modelo logit binário, mas as correspondentes fórmulas seriam obviamente bem mais complexas. Por isso mesmo elas foram omitidas aqui, podendo, entretanto, ser encontradas em Dhrymes (1978: 348), por exemplo. De qualquer modo, uma comparação entre os resultados obtidos pela estimação de máxima verossimilhança e a estimação por MQG é apresentada mais adiante nos "Exemplos" desta seção, para o caso da logit multinomial.

6.8.2 — Precisão do ajustamento e testes de hipóteses

Como medida do grau de precisão do ajustamento do modelo logit estimado pelo método da máxima verossimilhança pode-se utilizar {Pindyck & Rubinfeld (1981: 312)}:

$$z = 1 - L_o/L_{max} \,, \tag{6-67}$$

onde L_o seria o valor da função de verossimilhança quando os seus parâmetros, à exceção da constante, são igualados a zero; L_{max} seria o valor máximo da função de verossimilhança (isto é, sem qualquer

restrição imposta aos seus parâmetros). Naturalmente, os valores desta medida variam entre zero e um, com valores próximos de zero indicando obviamente um mau ajustamento. Na verdade a relação

$$\lambda = \frac{L_o}{L_{max}} \tag{6-68}$$

permite testar a significância estatística do modelo como um todo, já que sob a hipótese nula de que o modelo nada explica temos que

$$-2 \; Log \; \lambda = -2(Log \; L_o - Log \; L_{max}) \tag{6-69}$$

terá distribuição qui-quadrado com k graus de liberdade, onde k é o número de parâmetros na equação (à exceção da constante). Diversos outros testes, relativos a subconjuntos de parâmetros do modelo, poderiam ser ainda efetuados com essa mesma relação, bastando, obviamente, apenas avaliar L_o com as restrições da hipótese nula impostas e adaptando k para representar o número de tais restrições.

6.8.3 — Base teórica para o modelo logit

Até agora consideramos o modelo logit do ponto de vista das suas propriedades puramente matemáticas e, também, do ponto de vista de sua estimação estatística. O modelo é, entretanto, fundamentado em certos pressupostos de comportamento do indivíduo no seu processo de tomada de decisão, que vale a pena explorar. Com efeito, considere-se que o indivíduo i seja dotado do vetor de características s_i (por exemplo, sua renda, idade, etc.), o qual encontra-se diante de J_i alternativas com atributos x_{ij}. Se considerarmos ainda a existência de uma função de utilidade para um indivíduo representativo, dada por

$$\overline{U}_{ij} = \overline{U}(x_{ij}, s_i) = X_{ij} \; B \tag{6-70}$$

então, para um indivíduo qualquer, poder-se-ia imaginar a seguinte função de utilidade:

$$U_{ij} = \overline{U}_{ij} + e_{ij}. \tag{6-71}$$

Se, agora, o indivíduo, diante de J alternativas, maximiza a sua função de utilidade, então, a probabilidade de que a primeira alternativa venha a ser por ele escolhida, é dada por:

$$\begin{aligned} P_{i1} &= Prob(U_{i1} > U_{i2} \text{ e } U_{i1} > U_{i3} \; .... \text{ e } U_{i1} > U_{iJ}) \\ &= Prob(e_{i2} < \overline{U}_{i1} - \overline{U}_{i2} + e_{i1} \; .... \text{ e } e_{iJ} < \overline{U}_{i1} - \overline{U}_{iJ} + e_{i1}) \\ &= Prob(e_{i2} - e_{i1} < \overline{U}_{i1} - \overline{U}_{i2} \text{ e } .... \; e_{iJ} - e_{i1} < \overline{U}_{i1} - \overline{U}_{iJ}) \end{aligned} \tag{6-72}$$

Obviamente, expressões semelhantes seriam obtidas para os demais P_{ij}.

Se adotarmos, em seguida, o pressuposto de que esses e_{ij} sejam variáveis estocásticas independentes entre si e identicamente distribuídas com uma função de densidade de Weibull (que guarda alguma semelhança com a distribuição normal mas que é ligeiramente assimétrica) então a diferença entre tais variáveis terá uma função de distribuição logística, originando portanto o modelo logit binário {ver detalhes em Domencich & McFadden (1975)}. Isso significa, aliás, que também no modelo logit multinomial, apenas comparações binárias são efetuadas. Essa é, na verdade, uma limitação importante do modelo logit multinomial. Tal pressuposto é denominado de independência das alternativas irrelevantes e o seu sentido prático é que a razão entre duas alternativas não ficaria afetada com a adição de novas alternativas. Assim, o modelo parece não ser plausível para aplicações onde as alternativas possam ser consideradas como substitutas próximas. Para exemplificar, se, seguindo Judge, Griffiths, Hill & Lee (1980: 597), a população prefere o transporte à cavalo ao transporte por bicicleta, na razão 2 por 1, então com a inclusão de bicicletas com uma cor diferente da anterior parece razoável imaginarmos que a preferência pelo cavalo permanecesse na razão 2 por 1. No entanto, o pressuposto da independência entre alternativas irrelevantes implica em que a probabilidade de escolher o transporte eqüino deve cair à metade, para que a relação das preferências cavalo-bicicleta continue na razão de 2 por 1. Parece claro que a questão aqui está na impropriedade em se considerar bicicletas de diferentes cores como alternativas distintas. Quando as alternativas forem efetivamente distintas, o modelo logit produzirá resultados adequados.

Para contornar essa limitação do modelo logit multinomial Hausman & Wise (1978) propuseram como alternativa o modelo condicional probit. Este não tem as restrições daquele. De qualquer modo não parece apresentar resultados tão superiores, conforme sugerido por Amemiya (1981: 1523-4), além de ser computacionalmente bem mais oneroso. Naturalmente, para dados agregados e em se tratando de duas alternativas apenas, faz muito pouca diferença qual desses modelos é utilizado, como já foi devidamente enfatizado.

6.9 — Exemplos

a — Apresentamos, primeiramente, uma aplicação do modelo linear de probabilidade. Nesta aplicação, tanto as variáveis explicativas como a variável dependente, da regressão, são do tipo variáveis binárias. Os resultados ora mostrados, provêm de Johnston (1972) que, entretanto, os atribui a uma pesquisa efetuada por Orcutt & Rivlin. Naquela aplicação, esses autores tinham dois objetivos: primeiramente estimar, com base em várias características individuais descritas abaixo, uma equação que pudesse prever a probabilidade de uma unidade familiar de dispêndio nos USA vir a ter débito relativo a financiamento de casa própria;

em segundo lugar, procurar estimar, para as famílias com algum débito deste tipo, uma equação que fornecesse a quantidade do débito. As duas equações obtidas foram, respectivamente:

$$\hat{Y}_1 = -0{,}08 + F_{1.1} + F_{1.2} + F_{1.3} + F_{1.4}$$

e

$$\hat{Y}_2 = 3040 + F_{2.1} + F_{2.2} + F_{2.3} + F_{2.4}$$

onde os valores desses *Fs* encontram-se no quadro 4 e são definidos adiante. Considerando os valores desses *Fs* temos, por exemplo, que a probabilidade estimada de que uma unidade familiar, representada por um indivíduo negro de 32 anos de idade, tendo ginásio incompleto e 3 anos de casado, venha a ter débito relativo a financiamento de casa própria, seria (ver Quadro 4):

$$\hat{Y}_1 = -0{,}08 + 0{,}19 + 0{,}16 + 0{,}05 - 0{,}10 = 0{,}22.$$

Agora, considerando-se apenas os indivíduos com algum débito do tipo mencionado, temos que para o mesmo indivíduo descrito acima o débito esperado seria:

$$\hat{Y}_2 = 3040 + 1530 + 1560 + 1190 - 1040 = \$\ 6280.$$

Quanto aos valores dos *Fs* foram assim obtidos. Primeiramente, os coeficientes dos F_1s foram estimados efetuando-se uma regressão linear múltipla entre o *Y* binário e as diversas variáveis explicativas binárias do quadro 4. Por exemplo, para a variável estado civil do indivíduo, as variáveis binárias foram definidas como:

$$X_1 = \begin{cases} 1 \text{ se } 1 \text{ ano de casado} \\ 0 \text{ de outro modo} \end{cases}$$

$$X_2 = \begin{cases} 1 \text{ se } 2 \text{ anos de casado} \\ 0 \text{ de outro modo} \end{cases}$$

$$\vdots$$

$$X_7 = \begin{cases} 1 \text{ se mais de } 20 \text{ anos de casado} \\ 0 \text{ de outro modo.} \end{cases}$$

As outras variáveis explicativas foram definidas de modo análogo e então efetuada a regressão, pelo método de *MQO*, entre os *Ys* e todas essas variáveis, obtendo-se os coeficientes listados no quadro 4 para os F_1s (ver a chamada 3). Os coeficientes para os F_2s foram obtidos de modo

[3] Note-se que como o modelo linear de probabilidade é, conforme visto antes, heterocedástico, essa estimação deveria ter sido efetuada pelo método dos MQP, já que isto permitiria obter estimadores mais eficientes.

semelhante, com a única diferença que agora os *Ys* não são mais variáveis binárias, e sim tendo os valores dos débitos das diversas combinações das características individuais descritas no quadro 4.

QUADRO 4

APLICAÇÃO DO MODELO LINEAR DE PROBABILIDADE

ESPECIFICAÇÃO	VARIÁVEIS							
	$F_{1.1}$	$F_{2.1}$	$F_{1.2}$	$F_{2.2}$	$F_{1.3}$	$F_{2.3}$	$F_{1.4}$	F_{24}
ESTADO CIVIL								
Solteiro	0	0						
Casado								
1 ano	0,07	4 550						
2 anos	0,15	1 380						
3 anos	0,19	1 530						
4 anos	0,27	167						
5 a 9 anos	0,31	49						
10 a 20 anos	0,25	744						
Mais de 20 anos	0,14	215						
IDADE DO CHEFE DA FAMÍLIA								
18 a 20 anos			0	0				
21 a 24 anos			—0,02	526				
25 a 29 anos			0,09	1 290				
30 a 34 anos			0,16	1 560				
35 a 39 anos			0,25	1 200				
40 a 44 anos			0,22	964				
45 a 49 anos			0,24	134				
50 a 54 anos			0,16	166				
55 a 59 anos			0,15	—1 270				
60 a 64 anos			0,12	—1 140				
Mais de 65 anos			0,07	—1 000				
EDUCAÇÃO DO CHEFE								
Nenhuma					0	0		
Primário					0,02	835		
Ginásio incompleto					0,05	1 190		
Ginásio					0,08	1 980		
Superior incompleto					0,12	3 610		
Superior					0,13	4 930		
TIPO RACIAL DO CHEFE								
Branco							0	0
Negro							—0,10	—1 040
Outros							—0,14	—1 010

FONTE — Adaptado de Johnston (1972: 183-5).

Como comentário final, note-se que as variáveis binárias entraram aqui apenas de modo aditivo, mas efeitos de interações entre essas poderiam ser ainda incluídos na especificação; observe-se, entretanto, que isso iria exigir a estimação de um maior número de parâmetros e talvez não se dispusesse de graus de liberdade suficientes para tal.

b — Apresentamos, como segunda ilustração, uma aplicação da logit binária efetuada por Theil (1971: 632-6), onde os regressores são qualitativos. Tal aplicação refere-se a certo setor da indústria alemã, sendo a natureza dos dados utilizados assim definida. Aos empresários foi perguntado, no começo do mês, "se planejavam, para aquele mês, aumentar (a ser indicado pelo sinal +), não alterar (representado pelo

valor zero) ou diminuir (indicado pelo sinal —) suas produções". Um mês mais tarde, foi verificado se os planos haviam ou não sido realizados. A realização do plano seria então representada, conforme o caso, pelas seguintes seqüências: ++, 00 ou ——. Por outro lado, as revisões positivas os planos de produção assumiriam as seqüências 0+, —0 ou —+. As revisões negativas teriam as seqüências 0—, +0 ou +—.

A questão seguinte é investigar quais variáveis seriam capazes de explicar as revisões nos planos de produção. Neste sentido, duas variáveis foram consideradas. A primeira diz respeito às encomendas pelo produto dos fabricantes. Estas foram definidas como esperadas (no início do mês) e atuais (observadas no fim do mês). As seqüências possíveis são, então, assim representadas:

1 — surpresa negativa nas encomendas quando ocorrer a seqüência 0—, +0 ou +—;

2 — surpresa positiva quando ocorrer a seqüência 0+, —0 ou —+;

3 — ausência de surpresa quando tivermos ++, 00 ou ——.

A outra variável considerada, refere-se à avaliação dos fabricantes quanto à situação dos seus estoques de produtos finais. Aos fabricantes foi perguntado se os estoques estavam muito baixos, normais ou muito altos.

Como o objetivo seria explicar a probabilidade de revisões nos planos de produção, em termos das surpresas nas encomendas e das avaliações nos estoques, foram eliminados todos os casos onde:

a — os planos de produção foram realizados;

b — não houve surpresa quanto as encomendas recebidas; ou

c — os estoques foram considerados normais.

Desse modo, as variáveis podem assim ser descritas:

$$Y = \begin{cases} Y_1 \text{ Revisão positiva no plano de produção} \\ Y_2 \text{ Revisão negativa no plano de produção} \end{cases}$$

$$X = \begin{cases} X_1 \text{ Surpresa positiva nas encomendas recebidas} \\ X_2 \text{ Surpresa negativa nas encomendas recebidas} \end{cases}$$

$$Z = \begin{cases} Z_1 \text{ Estoques considerados muito baixos} \\ Z_2 \text{ Estoques considerados muito elevados.} \end{cases}$$

Os dados do estudo permitiram levantar 379 observações que estão tabuladas nas três primeiras linhas do quadro 5. Como era de se esperar, a freqüência de revisões positivas nos planos de produção é maior sob a combinação (X_1Z_1) e menor sob a combinação (X_2Z_2).

QUADRO 5

DADOS SOBRE AS REVISÕES NOS PLANOS DE PRODUÇÃO POR SURPRESAS NAS ENCOMENDAS RECEBIDAS E AVALIAÇÕES DOS ESTOQUES

ESPECIFICAÇÃO	OBSERVAÇÕES				
	Total	Combinações			
		$X_1 Z_1$	$X_1 Z_2$	$X_2 Z_1$	$X_2 Z_2$
Total de casos	379	89	83	43	164
Revisões positivas	151	68	50	14	19
Freqüência relativa observada	0,40	0,76	0,60	0,33	0,12
Freqüência relativa estimada para o modelo logit	—	0,79	0,58	0,28	0,13

FONTE — Adaptado de Theil (1971).

Se representarmos a probabilidade condicional de uma revisão positiva para dada combinação (XZ) por $P_{jk} = P(Y = Y_1 \mid X = X_j, Z = Z_k)$, para $j, k = 1$ e 2, e a correspondente probabilidade condicional de uma revisão negativa por $1 - P_{jk}$, então o modelo pode assim ser especificado:

$$Log \frac{P_{jk}}{1 - P_{jk}} = B_0 + B_1 D_1 + B_2 D_2 + \mu$$

onde $$D_1 = \begin{cases} 0 \text{ se } X = X_1 \\ 1 \text{ se } X = X_2 \end{cases} \quad \text{e} \quad D_2 = \begin{cases} 0 \text{ se } Z = Z_1 \\ 1 \text{ se } Z = Z_2 \end{cases}.$$

Para efeito da estimação desta relação, basta fazer

$$P_{jk} = f_{jk} = r_{jk}/n_{jk}$$

onde f_{jk} é a freqüência relativa dos casos favoráveis a dada alternativa no grupo jk.

A substituição dos valores das variáveis X e Z no modelo, na sua versão dada em (6-16), fornece:

$$\begin{bmatrix} Log(68/21) \\ Log(50/33) \\ Log(14/29) \\ Log(19/145) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ B_2 \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix}$$

Supondo-se que as freqüências sejam baseadas em amostras independentes de populações binomiais implica que os μ_i são também variáveis independentes, tendo média zero {Theil (1971)} e variância como dada em (6-12). Assim, em vista da presença de heterocedasticidade, foi aplicado o método dos MQP, com os pesos proporcionais ao inverso da variância estimada para os μ_i. O resultado obtido foi:

$$Log \frac{r_i}{n_i - r_i} = \underset{(0,22)}{1,30} - \underset{(0,26)}{2,23D_1} - \underset{(0,26)}{0,98D_2}$$

com os valores entre parênteses indicando os desvios padrão dos estimadores. Note-se que o efeito da variável "surpresa nas encomendas" (D_1) é maior que o da "avaliação dos estoques" (D_2).

As freqüências relativas nas revisões positivas nos planos de produção, produzidas com a função estimada para o modelo logit, são apresentados na última linha do quadro 5. Como pode-se observar, há boa concordância entre essas freqüências e as freqüências efetivamente observadas, constatando-se, portanto, a utilidade da especificação do modelo logit, neste caso.

c — Como outra ilustração, utilizaremos uma aplicação da logit multinomial com uma variável explicativa apenas e do tipo quantitativa. Essa ilustração é tomada do artigo de Parks (1980), que utiliza dados sobre o número de veículos possuídos ao longo das várias faixas de renda em que fora agregada uma amostra de famílias americanas, em 1969; esses dados são apresentados no quadro 6.

QUADRO 6

FREQÜÊNCIA DA POSSE DE VEÍCULOS NOS ESTADOS UNIDOS — 1969

RENDA MÉDIA FAMILIAR DA CLASSE US$	NÚMERO DE FAMÍLIAS	PERCENTAGENS DE FAMÍLIAS POR NÚMERO DE VEÍCULOS POSSUÍDOS		
		Nenhum	Um	Dois ou mais
700	36	75	22	3
1 500	139	59	40	1
2 500	178	50	43	7
3 500	170	40	45	6
4 500	134	30	61	9
5 500	134	25	66	9
6 750	270	14	71	15
8 750	412	8	66	26
12 500	631	4	55	41
24 313	472	4	36	60

FONTE — Adaptado de Parks (1980).

Parks (1980) utilizou o logaritmo da renda como variável explicativa da logit multinomial, e, como ponto de referência para as logits condicionais, fora adotado "nenhum" veículo. Assim, temos a seguinte apresentação matricial do modelo:

$$\begin{bmatrix} Log\ 22/75 \\ Log\ 3/75 \\ Log\ 40/59 \\ \vdots \\ Log\ 36/4 \\ Log\ 60/4 \end{bmatrix} = \begin{bmatrix} 1 & Log\ 700 & 0 & 0 \\ 0 & 0 & 1 & Log\ 700 \\ 1 & Log\ 1500 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & Log\ 24313 & 0 & 0 \\ 0 & 0 & 1 & Log\ 24313 \end{bmatrix} \begin{bmatrix} a \\ b \\ a_1 \\ b_1 \end{bmatrix}$$

Apresentamos, no quadro 7, os resultados da aplicação com estes dados, dos três métodos de estimação discutidos acima, mais precisamente *MQO, MQP* e *MQM*. Além dos estimadores, mostramos ainda os seus respectivos desvios padrão, sendo que para os estimadores de *MQO* e *MQP* apresentamos esses desvios tanto na sua versão tendenciosa (isto é, como dados pelas fórmulas usuais) como na sua formulação não tendenciosa; os primeiros são apresentados diretamente abaixo dos coeficientes das variáveis e os segundos encontram-se nas duas últimas linhas. Observe-se que, como foi notado anteriormente, os estimadores de *MQM* são mais eficientes que os outros dois métodos de estimação, quando os seus desvios padrão são comparados com os desvios não tendenciosos dos outros dois estimadores. Naturalmente, a comparação entre os desvios padrão dos *MQM* com aqueles fornecidos pelas fórmulas usuais para os outros dois estimadores não é apropriada, dado serem esses desvios tendenciosos. Ressalte-se ainda que as fórmulas usuais subestimam os verdadeiros valores dos desvios nos dois casos. Entretanto, para os estimadores de *MQO* poder-se-ia obter também a superestimação, como poderá ser verificado tanto na ilustração seguinte como em Amemiya & Nold (1975). Para os estimadores de *MQP*, porém, parece haver apenas a possibilidade da subestimação dos desvios que, diga-se de passagem, é bastante acentuada nesta ilustração e também naquela de Amemiya & Nold (1975), bem como no próximo exemplo.

Como observação final, note-se que, por ser a variável explicativa da regressão o logaritmo da renda, então o seu coeficiente indica diretamente a elasticidade-renda das chances favoráveis a posse tanto de um como de dois ou mais veículos, conforme seja o caso, relativamente a nenhum veículo. Desse modo, temos que a elasticidade-renda para a posse de dois ou mais veículos é bem maior que aquela para um veículo, indicando tal elasticidade um caso típico de dispêndio em bem superior.

QUADRO 7

RESULTADOS DA REGRESSÃO COM OS TRÊS MÉTODOS DE ESTIMAÇÃO

MÉTODOS DE ESTIMAÇÃO	RESULTADOS DA REGRESSÃO				
	Possuidores de 1 veículo		Possuidores de 2 ou mais veículos		
	Constante	Logaritmo da renda	Constante	Logaritmo da renda	R^2
b_{MQO}	—8,842 (1,072)	1,166 (0,128)			0,912
			—17,870 (2,209)	2,058 (0,263)	0,884
b_{MQP}	—9,405 (0,639)	1,232 (0,076)	—21,959 (0,921)	2,535 (0,104)	0,929
b_{MQM}	—8,942 (1,041)	1,179 (0,123)	—19,906 (2,091)	2,295 (0,242)	0,911
ESTIMATIVAS CONSISTENTES PARA OS DESVIOS PADRÃO DE:					
MQO	1,113	0,131	2,473	0,285	
MQP	1,107	0,131	2,722	0,307	

FONTE — Adaptado de Parks (1980).
NOTA — Os desvios padrão estimados estão entre parênteses.

d — Nesta ilustração, apresentamos o caso da logit multinomial com regressores qualitativos, sendo utilizados na estimação os dados de renda da população de 10 e mais anos de idade, fornecidos pelo Censo de 1980 [4]. Essas informações são classificadas segundo 8 faixas de renda (também um grupo sem qualquer remuneração), segundo o sexo do indivíduo e o seu local de origem (rural ou urbana). Para efeito da análise deste estudo, foram excluídos os indivíduos sem qualquer remuneração, e os grupos de renda foram agregados em três classes: até 1 salário-mínimo (SM); maior que 1 SM e até 3 SM; e, maior que 3 SM. Tais informações são apresentadas, de maneira simplificada, no quadro 8.

A aplicação dos três métodos de estimação, representados por MQO, MQP e MQM, aos dados do quadro 8, permitiu obter os resultados do quadro 9. Cabem aqui os seguintes comentários sobre esses resultados: primeiramente, e de modo análogo ao exemplo anterior com base em Parks (1980), as estimativas dos MQO e MQP são mais

[4] Ver IBGE. *Tabulações Avançadas do Censo Demográfico 1980. Brasil.* Rio de Janeiro. p. 24.

próximas entre si do que com as estimativas de MQM [embora em outras aplicações as estimativas de MQM pudessem estar mais próximas das estimativas de MQP do que daqueles de MQO; ver, por

QUADRO 8

DADOS DE RENDA PARA A POPULAÇÃO COM 10 E MAIS ANOS DE IDADE — BRASIL — 1980

ORIGEM	POPULAÇÃO DE 10 ANOS E MAIS			
	Número de pessoas (milhões)	Freqüência de pessoas (1)		
		f_1	f_2	f_3
		Mulheres		
Rural.....	n_{11} = 3,21	0,857	0,126	0,016
Urbana.....	n_{12} = 12,34	0,473	0,378	0,149
		Homens		
Rural.....	n_{21} = 10,10	0,564	0,352	0,084
Urbana.....	n_{22} = 22,50	0,215	0,455	0,330

FONTE — Ver o texto.

NOTA — Excluida a população sem qualquer remuneração.

1) As freqüências f_1, f_2 e f_3 correspondem, respectivamente, as classes "até 1 SM", "maior que 1 e até 3 SM" e 'maior que 3 SM".

exemplo, Amemiya & Nold (1975)]. Em segundo lugar, e diferentemente do resultado obtido por Parks, mas de modo semelhante a Amemiya & Nold, temos que a fórmula usual para estimar o erro padrão dos coeficientes de MQO produziu aqui valores com "viés" positivo, enquanto que a fórmula comumente utilizada para o cálculo do erro padrão dos coeficientes de MQP apresentou valores com "viés" negativo; resultado este último aliás que está de acordo com aqueles encontrados por Parks, e Amemiya & Nold. Devemos reconhecer, neste estágio, que a subestimação apresentada pela fórmula convencional para o erro padrão dos coeficientes de MQP é particularmente acentuada na nossa aplicação. A explicação para este fato parece ser a seguinte. Conforme foi notado no item 6.6, tal subestimação seria tanto maior quanto maior fosse a razão entre a variância do erro v_{it} e a variância do erro μ_{it}, e esta última é muito reduzida, na nossa aplicação, em vista da amostra se confundir com a própria população, pois os dados, por serem censitários, incluiriam presumivelmente toda a população. Naturalmente, poder-se-ia argumentar que, em qualquer censo, teremos sempre indivíduos não contados e, assim, os dados representariam ainda uma espécie de amostra, mesmo que tendendo a população. Este é, de fato, o ponto de vista que assumimos nesta análise.

Ainda com relação aos erros padrão dos coeficientes, notamos, do quadro 9, que quando estes erros são estimados de modo consistente, então os valores obtidos são menores para o caso dos MQO do que no caso dos MQP, resultado, aliás, que é semelhante ao encontrado por

QUADRO 9

ESTIMATIVAS PRODUZIDAS PELOS TRÊS MÉTODOS

PARÂMETROS	ESTIMATIVAS PRODUZIDAS		
	Estimativas	Erro padrão	
		Consistente	Fórmula
	Método MQO		
α_2	—1,7964	0,10226	0,20591
α_3	—3,8492	0,10369	0,20591
β_{22}	1,2069	0,11807	0,23776
β_{23}	1,8188	0,11967	0,23776
γ_{22}	1,4556	0,11807	0,23776
γ_{23}	2,5765	0,11967	0,23776
	Método MQP		
α_2	—1,5074	0,16031	0,00094
α_3	—3,5686	0,17432	0,00148
β_{22}	1,1363	0,10447	0,00077
β_{23}	1,6259	0,15374	0,00100
γ_{22}	1,2762	0,14292	0,00081
γ_{23}	2,3918	0,15111	0,00126
	Método MQM		
α_2	—1,7971	0,10215	
α_3	—3,8498	0,10381	
β_{22}	1,2079	0,11816	
β_{23}	1,8197	0,11986	
γ_{22}	1,4567	0,11817	
γ_{23}	2,5776	0,11987	

FONTE — Ver o texto.

Parks, mas o oposto do encontrado por Amemiya & Nold. Observe-se, adicionalmente, que além dos valores dos estimadores de MQO estarem próximos dos estimadores MQM também os seus correspondentes erros padrão estimados (quando os primeiros são considerados na sua forma consistente) apresentam valores bastante próximos uns dos outros.

Com essas considerações, verificamos portanto haver certa discordância entre esses três estudos, quanto à direção em que se alteram os valores de alguns desses estimadores, quando os diversos métodos de estimação são empregados. De qualquer modo, é interessante notar-se

que, apesar de um maior erro padrão para os estimadores de MQP (quando esses erros são estimados consistentemente), tal método parece apresentar um melhor desempenho preditivo que os dois outros métodos, como pode-se atestar dos resultados do quadro 10. Na verdade, se adotarmos como medida de precisão do ajustamento a relação proposta por Theil (1970), como [já mencionada antes na expressão em (6-55)]

$$\sum_k \sum_j n_{jk} \sum_r \left(\frac{f_{r/jk} - \hat{P}_{r/jk}}{\hat{P}_{r/jk}} \right)^2$$

(onde $\hat{P}$ é a freqüência prevista), que será assintomaticamente distribuída como χ^2_{n-k} (onde n é o número de células no Quadro 8 e k é o número de parâmetros estimados no modelo) quando houver concordância entre as freqüências observadas e previstas, então temos que os estimadores de MQP produzem melhor resultado do que os estimadores de MQO e MQM, pois, enquanto para os primeiros o valor desta relação fora de 4,6 para os dois outros estimadores esse valor foi cerca de 12,8 (ver chamada 5).

Com relação à interpretação dos coeficientes obtidos para as diversas variáveis, limitar-nos-emos aqui aos estimadores de MQP, já que apresentaram melhor desempenho preditivo, como já salientamos. Assim sendo, se considerarmos x_j como a variável sexo (com $x_1 = 0$ se feminino e $x_2 = 1$ se masculino) e z_k a variável local de origem (com $z_1 = 0$ se rural e $z_2 = 1$ se urbano), então, levando-se em conta as restrições em

QUADRO 10

FREQÜÊNCIAS OBSERVADAS VERSUS FREQÜÊNCIAS ESTIMADAS

PROBABILIDADES	FREQÜÊNCIAS			
	Observadas	Estimadas		
		MQO	MQP	MQM
$P_{1/11}$	0,8574	0,8423	0,8058	0,8424
$P_{1/12}$	0,4726	0,5022	0,4803	0,5021
$P_{1/21}$	0,5643	0,5931	0,5742	0,5931
$P_{1/22}$	0,2153	0,1959	0,2109	0,1957
$P_{2/11}$	0,1264	0,1397	0,1716	0,1397
$P_{2/12}$	0,3723	0,3571	0,3726	0,3572
$P_{2/21}$	0,3522	0,3290	0,3436	0,3290
$P_{2/22}$	0,4549	0,4658	0,4593	0,4659
$P_{3/11}$	0,0162	0,0179	0,0226	0,0179
$P_{3/12}$	0,1492	0,1407	0,1475	0,1407
$P_{3/21}$	0,0835	0,0779	0,0822	0,0779
$P_{3/22}$	0,3298	0,3383	0,3298	0,3384

FONTE — Ver o texto.

[5] Esses valores são na verdade apresentados aqui apenas com propósitos ilustrativos, já que foram obtidos após dividirmos os vários n_{jk} por 1.000.

(6-31) e o fato de que a referência para a logit condicional é P_1 (isto é, probabilidade de situar-se na primeira faixa de renda), temos que:

1 — $\alpha_2 = -1{,}5074$ é a estimativa para a logit condicional favorável a que um indivíduo situe-se na segunda faixa de renda, relativamente à primeira (isto é, a mais baixa) quando este for mulher e residir no campo;

2 — $\beta_{22} = 1{,}1363$ é o efeito sobre esta mesma logit quando o indivíduo for homem, em vez de mulher;

3 — $\gamma_{22} = 1{,}2762$ é o efeito sobre esta mesma logit quando o indivíduo residir na cidade, em vez do campo;

4 — $\alpha_3 = -3{,}5686$ é o efeito sobre a logit condicional favorável a que um indivíduo situe-se na classe de renda mais elevada, relativamente à classe mais baixa quando este for mulher e residir no campo;

5 — $\beta_{23} = 1{,}6259$ é o efeito sobre esta mesma logit quando o indivíduo for homem, em vez de mulher; e

6 — $\gamma_{23} = 2{,}3918$ é o efeito sobre esta mesma logit quando o indivíduo residir na cidade, em vez do campo.

Desses resultados, cabem as seguintes observações:

a — dos itens 1 e 4 temos que as chances favoráveis a que um indivíduo situe-se entre os grupos de renda mais elevada, diminuem apreciavelmente quando se é mulher residindo no meio rural;

b — dos itens 2 e 5, por outro lado, temos que as chances favoráveis a que um indivíduo situe-se entre as faixas de renda mais elevadas, aumentam quando se é homem; e

c — dos itens 3 e 6, temos que as chances favoráveis à melhores rendas, aumentam quando passa-se do meio rural para o meio urbano.

Esses resultados parecem ainda sugerir que a dicotomia rural versus urbana representa uma barreira maior para o indivíduo alcançar faixas de renda mais elevadas, do que as diferenças de sexo.

Outras conclusões podem ainda ser extraídas dos valores dos coeficientes acima. Conforme foi visto anteriormente, os coeficientes da logit condicional $L_{32/jk} = Log\ (P_{3/jk}/P_{2/jk})$ podem ser obtidos da relação $L_{32} = L_{31} - L_{21}$. Assim, temos os seguintes resultados:

$L_{32/11} = -2{,}0526$; $L_{32/12} = -0{,}9317$; $L_{32/21} = -1{,}4408$; e $L_{32/22} = -0{,}3199$.

Isso mostra que as dificuldades na ascensão da segunda para a terceira classe de renda permanecem imensas para a mulher rural. Tal ascensão é também muito difícil tanto para o homem rural como para a mulher urbana, sendo o grau dessa dificuldade maior para o homem rural; resultado este, aliás, que parece corroborar as conclusões apresentadas no parágrafo anterior, onde as ascensões são medidas em relação a primeira classe de renda. Note-se ainda que a ascensão da segunda para a terceira classe de renda é também difícil para o homem urbano, mas um pouco mais viável que nas outras situações aqui consideradas.

Como observação final, devemos acrescentar que o fato de muitos dos resultados que acabamos de apresentar parecerem estar inteiramente de acordo com as nossas expectativas sobre o assunto só poderia contribuir para demonstrar a utilidade do modelo logit em aplicações semelhantes à aqui efetuada.

d — Como última ilustração, apresentamos uma aplicação de Wrigley (1975) para um modelo logit multinomial, onde os parâmetros são estimados tanto pelo método da máxima verossimilhança como pelos MQP. No estudo de Wrigley o comportamento do consumidor é analisado quanto à sua estratégia de compra. Para tanto foram inicialmente coletados cuidadosamente, ao longo de seis meses, dados sobre os dispêndios de 140 famílias no Município de Bradford (Inglaterra). Como um aspecto importante da estratégia de compra é representado pelo nível de diversificação dessas compras, então 4 categorias distintas de diversificação foram identificadas dos dados levantados. Para explicar a escolha desses vários níveis de diversificação, as três seguintes variáveis foram utilizadas:

1 — o tempo disponível à dona de casa para as compras, representado por uma variável binária com valor zero, caso a dona de casa trabalhasse fora do lar, e, valor um se não trabalhasse;

2 — dispêndio total em alimentação; e

3 — distância média entre a residência da família e quatro tipos distintos de estabelecimentos varejistas para produtos de alimentação.

Os resultados da estimação estão no quadro 11. Como observado por Wrigley (1975), ambos os métodos de estimação identificam a influência significativa da variável "dona de casa não trabalha fora", na escolha da categoria "estratégia de compra mais diversificada". Entretanto, as esperadas influências significativas das variáveis "dispêndio total em alimentação" e "distância" na escolha da estratégia menos diversificada só são captadas pelo método da máxima verossimilhança.

A razão para o desempenho menos satisfatório do método dos MQP decorre do fato de ter-se aqui uma amostra de pequeno tamanho. Recorde-se que o método dos MQP requer o agrupamento dos dados, para

QUADRO 11

RESULTADOS DA ESTIMAÇÃO

VARIÁVEL	MÉTODO	ESTIMATIVAS		
		Menos diversificado		Mais diversificado
		Nível 0 de diversificação	Nível 1 de diversificação	Nível 3 de diversificação
Constante.....	MQP	—2,23 (—1,95)	—0,19 (—0,22)	—1,06 (—0,96)
	MV	—3,27 (—2,35)	—0,23 (—0,34)	—1,77 (—1,77)
Efeito "dona de casa não trabalha".....	MQP	0,56 (0,85)	—0,14 (—0,32)	1,27 (2,04)
	MV	0,24 (0,38)	—0,44 (—1,13)	1,51 (2,41)
Efeito "dispêndio total em alimentação".....	MQP	0,17 (0,91)	0,04 (0,29)	—0,09 (—0,51)
	MV	0,35 (2,02)	0,09 (0,81)	—0,02 (—0.10)
Efeito "distância".....	MQP	0,15 (0,29)	0,10 (0,26)	—0,17 (—0,35)
	MV	0,55 (1,36)	0,19 (0,68)	—0,11 (—0,28)

FONTE — Adaptado de Wrigley (1975).
NOTA — Os dados entre parênteses são valores de t.

que as freqüências relativas observadas possam substituir as probabilidades nas relações logit. Como neste estudo, os dados foram agrupados segundo quatro níveis da variável "dispêndio em alimentação", dois níveis para a variável "distância", e, naturalmente, dois níveis para a variável "tempo disponível", então obtivemos 16 grupos, muitos dos quais tendo freqüência observada muito baixa. Conforme foi visto anteriormente, em situações desse tipo não é recomendada a utilização do método dos MQP. Nestas circunstâncias, o desempenho menos satisfatório desse último método não deve causar nenhuma surpresa. Para amostras grandes, com poucas variáveis explicativas, seria possível obter-se um número razoável de grupos e ainda cada um desses tendo freqüências observadas apropriadas para substituir as probabilidades nas relações logit. Quando este for o caso, os métodos da máxima verossimilhança e dos MQP produziriam então resultados semelhantes.

7 — APÊNDICES

7.1 — Apêndice da seção 2 (relativo ao Item 2.3.1)

Demonstraremos inicialmente, que a estimação de Mínimos Quadrados para os parâmetros do modelo em (2-7) equivale àquela obtida com duas funções separadas: uma do tipo (2-3) para os anos de paz e outra do tipo (2-8) para os anos de guerra. Entretanto, como é demonstrado aqui também, as variâncias parecem ficar afetadas por esses dois procedimentos (isto é, em (2-7) ou em (2-3)).

Suponhamos que os primeiros n anos sejam relativos ao período de paz, com os k anos seguintes correspondendo aos anos de guerra. O modelo em (2-7), em detalhes, seria então:

$$\left.\begin{array}{l} y_1 = a + bx_1 + 0 + 0 + \mu_1 \\ \vdots \\ y_n = a + bx_n + 0 + 0 + \mu_n \end{array}\right\} \textit{PAZ}$$

$$\left.\begin{array}{l} y_{n+1} = a + bx_{n+1} + c + dx_{n+1} + \mu_{n+1} \\ \vdots \\ y_{n+k} = a + bx_{n+k} + c + dx_{n+k} + \mu_{n+k} \end{array}\right\} \textit{GUERRA}$$

ou, em forma matricial e considerando-se as partições acima:

$$\begin{pmatrix} Y \\ \tilde{Y} \end{pmatrix} = \begin{pmatrix} X & 0 \\ \tilde{X} & \tilde{X} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} + \begin{pmatrix} u \\ \tilde{u} \end{pmatrix}$$

A aplicação de Mínimos Quadrados neste último modelo produziria:

$$\hat{B} = \begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \\ \hat{d} \end{bmatrix} = \left[\begin{pmatrix} X' & \tilde{X}' \\ 0' & \tilde{X}' \end{pmatrix} \begin{pmatrix} X & 0 \\ \tilde{X} & \tilde{X} \end{pmatrix} \right]^{-1} \begin{bmatrix} X' & \tilde{X}' \\ 0' & \tilde{X}' \end{bmatrix} \begin{bmatrix} Y \\ \tilde{Y} \end{bmatrix} =$$

$$= \begin{bmatrix} X' X + \tilde{X}' \tilde{X} & \tilde{X}' \tilde{X} \\ \tilde{X}' \tilde{X} & \tilde{X}' \tilde{X} \end{bmatrix}^{-1} \begin{bmatrix} X' Y + \tilde{X}' \tilde{Y} \\ \tilde{X}' \tilde{Y} \end{bmatrix}$$

Para simplificar a notação, seja

$$\hat{B} = A^{-1} Z$$

onde, observando-se as partições acima, temos

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{e} \quad Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}.$$

Assim, de acordo com Johnston (1972: 93), vem:

$$A^{-1} = \begin{bmatrix} E & -EA_{12}\, A_{22}^{-1} \\ -A_{22}^{-1}\, A_{21}E & A_{22}^{-1} + A_{22}^{-1}\, A_{21}\, E\, A_{12}\, A_{22}^{-1} \end{bmatrix}$$

onde $E = (A_{11} - A_{12}\, A_{22}^{-1}\, A_{21})^{-1}$.

Segue-se então que:

$$\hat{B} = \begin{bmatrix} (X'\,X)^{-1} & -(X'\,X)^{-1} \\ -(X'\,X)^{-1} & (\tilde{X}'\,\tilde{X})^{-1} + (X'\,X)^{-1} \end{bmatrix} \begin{bmatrix} X'\,Y + \tilde{X}'\,\tilde{Y} \\ \tilde{X}'\,\tilde{Y} \end{bmatrix}$$

ou

$$\hat{B} = \begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \\ \hat{d} \end{bmatrix} = \begin{bmatrix} (X'\,X)^{-1}\,X'\,Y \\ -(X'\,X)^{-1}\,X'\,Y + (\tilde{X}'\,\tilde{X})^{-1}\,\tilde{X}'\,\tilde{Y} \end{bmatrix} = \begin{bmatrix} \hat{a} \\ \hat{b} \\ -\hat{a}+\hat{a}' \\ -\hat{b}+\hat{b}' \end{bmatrix},$$

isto após reescrevermos a função em (2-8) como:

$$y_i = a' + b'x_i + \mu_i\,, \quad \text{onde} \quad a' = a + c \quad \text{e} \quad b' = b + d\,.$$

Assim, fica demonstrado que os estimadores obtidos em (2-7) equivalem àqueles obtidos com duas regressões separadas: uma para os anos de paz e outra para os anos de guerra. Naturalmente, o mesmo raciocínio poderia ser estendido para os casos envolvendo maior número de variáveis independentes e/ou maior número de períodos.

Quanto às variâncias desses estimadores, elas seriam obtidas através da seguinte matriz:

$$V(\hat{B}) = \sigma^2\, A^{-1} = \sigma^2 \begin{bmatrix} (X'X)^{-1} & -(X'X)^{-1} \\ -(X'X)^{-1} & (\tilde{X}'\tilde{X})^{-1} + (X'X)^{-1} \end{bmatrix}$$

onde $\sigma^2 = \Sigma e_i^2/n-k$, sendo n o número de observações; k o número de parâmetros e e_i um resíduo de *MQO*.

Portanto, as variâncias de $\hat{a}$ e $\hat{b}$ seriam dadas nesta ordem, pelos elementos da diagonal principal em $\sigma^2(X'X)^{-1}$, e as variâncias de $\hat{c}$ e $\hat{d}$ por aquelas em $\sigma^2\{(\tilde{X}'\tilde{X})^{-1} + (X'X)^{-1}\}$.

Conclui-se então que as variâncias dos estimadores $\hat{a}$ e $\hat{b}$, em (2-7), só seriam iguais àquelas obtidas com um ajustamento separado de (2-3), aos dados relativos apenas ao período de paz, se as estimativas de σ^2 forem idênticas nas duas equações estimadas; uma possibilidade certamente muito remota.

7.2 — Apêndice da seção 5

7.2.1 — *Justificativa para a formulação em* (5-4)

Seja o modelo linear geral a ajustar dado por:

1 - $y = xb + \mu$, onde μ representa o vetor dos resíduos.

Chamemos 1 o modelo a ser ajustado aos dados amostrais (ou simplesmente modelo amostral) que, uma vez estimado, servirá para efetuar previsões para um ou mais períodos.

A substituição das variáveis dos períodos das previsões, no modelo 1, permite representá-lo como:

2 - $\overset{*}{y} = \tilde{x}b + \tilde{\mu}$ (onde $\overset{*}{y}$ seria a previsão), podendo ser, alternativamente, escrito como:

3 - $0 = \tilde{x}b - \overset{*}{y} + \tilde{\mu}$, ou

4 - $\tilde{y} = \tilde{x}b - \overset{*}{y} + \tilde{y} + \tilde{\mu}$.

Nesta última formulação, $\tilde{y}$ representa um vetor de valores quaisquer.

A combinação de 1 e 4 fornece, por outro lado:

5 - $\begin{bmatrix} y \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} x & 0 \\ \tilde{x} & I \end{bmatrix} \begin{bmatrix} b \\ -\overset{*}{y} + \tilde{y} \end{bmatrix} + \begin{bmatrix} u \\ \tilde{u} \end{bmatrix}$, que fazendo $c = \tilde{y} - \overset{*}{y}$ produz a expressão (5-4) do texto.

7.2.2 — *Prova alternativa do resultado em* (5-19)

Da expressão (5-4), no texto, temos:

$$\begin{bmatrix} e \\ \tilde{e} \end{bmatrix} = \begin{bmatrix} y \\ \tilde{y} \end{bmatrix} - \begin{bmatrix} x & 0 \\ \tilde{x} & I \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix}, \quad \text{ou}$$

$$e = y - x\hat{b}$$

e

$\tilde{e} = \tilde{y} - \tilde{x}\hat{b} - \hat{c}I = 0$, após substituir $\hat{c}$ pelo seu valor dado em (5-11), no texto. Portanto, não só são os resíduos correspondentes ao período da previsão iguais a zero, mas os resíduos amostrais obtidos no bojo do ajustamento global são exatamente os mesmos que seriam obtidos do ajustamento do modelo amostral aos dados amostrais apenas.

7.2.3 — *Prova de que* $H^{-1}G = -CA^{-1}$ *em* (5-40) e (5-41)

De (5-37), no texto, temos:

$$\begin{aligned} H^{-1}G &= (D-CA^{-1}B)\ \{-D^{-1}C(A-BD^{-1}C)^{-1}\} \\ &= -C(A-BD^{-1}C)^{-1} + CA^{-1}BD^{-1}C(A-BD^{-1}C)^{-1} \\ &= -C(I-A^{-1}BD^{-1}C)(A-BD^{-1}C)^{-1} \end{aligned}$$

$$C^{-1}H^{-1}G = -(I-A^{-1}BD^{-1}C)\ (A-BD^{-1}C)^{-1}$$

$$\begin{aligned} AC^{-1}H^{-1}G &= -A\ (I-A^{-1}BD^{-1}C)\ (A-BD^{-1}C)^{-1} \\ &= -(A-BD^{-1}C)\ (A-BD^{-1}C)^{-1} \end{aligned}$$

$$\begin{aligned} AC^{-1}H^{-1}G &= -I \\ C^{-1}H^{-1}G &= -A^{-1} \\ H^{-1}G &= -CA^{-1} \end{aligned}$$

7.2.4 — *Prova do resultado em* (5-42)

A substituição da equação (5-40) em (5-38), no texto, produz:

$$(x'E+\tilde{x}'G)\,xb+(x'F+\tilde{x}'H)\,\tilde{x}b-(x'F+\tilde{x}'H)\,\tilde{x}b-H^{-1}G(y-xb)=(x'E+\tilde{x}'G)y;$$

$$x'Exb+\tilde{x}'Gxb+x'F\tilde{x}b+\tilde{x}'H\tilde{x}b-x'F\tilde{x}b-x'H\tilde{x}b+x'FH^{-1}G(y-xb)-$$

$$-\tilde{x}'G(y-xb)=x'Ey+\tilde{x}'Gy;$$

$$x'E(y-xb) = x'FH^{-1}G(y-xb)$$

$$[x'E-x'FH^{-1}G]\ (y-xb) = 0$$

$$x'[E-FH^{-1}G]\ (y-xb) = 0$$

7.2.5 — *Prova de que* $E-FH^{-1}G = A^{-1}$ *em* (5-42)

Chamemos $L = E-FH^{-1}G$ que, quando substituídos pelos seus valores em (5-37), no texto, viria:

$$\begin{aligned} L = {} & (A-BD^{-1}C)^{-1} - [-A^{-1}B(D-CA^{-1}B)^{-1}]. \\ & (D-CA^{-1}B).\ [-D^{-1}C(A-BD^{-1}C)^{-1}] = \\ = {} & (A-BD^{-1}C)^{-1} - \{-A^{-1}B\ .\ [-D^{-1}C(A-BD^{-1}C)^{-1}]\}. \end{aligned}$$

Seja $N = (A - BD^{-1}C)^{-1}$, então temos:

$$\begin{aligned} L &= (I - A^{-1}BD^{-1}C)\ N \quad \text{ou} \\ AL &= A(I - A^{-1}BD^{-1}C)\ N \\ AL &= (A - BD^{-1}C)\ N \\ AL &= I \Longrightarrow L = A^{-1} = E - FH^{-1}G\ . \end{aligned}$$

7.3 — Apêndice da seção 6

7.3.1 — *Prova do resultado em (6-12)*

Seja f a freqüência amostral em n retiradas independentes de uma distribuição binomial com probabilidade p. Como é sabido, a variância de f seria então:

$$V(f) = \frac{1}{n} p(1-p).$$

Considere-se em seguida a expansão de Taylor (retendo-se apenas a parte linear) da função $Log \frac{f}{1-f}$ em torno do ponto p; isto é:

$$Log \frac{f}{1-f} = Log \frac{p}{1-p} + \left(\frac{\partial \, Log \frac{f}{1-f}}{\partial \, f} \right) \Bigg|_{f=p} (f-p)$$

ou

$$u = Log \frac{f}{1-f} - Log \frac{p}{1-p} = \frac{1}{f(1-f)} (f-p)$$

Assim,

$$V(u) = \frac{1}{np(1-p)} = \frac{1}{nf(1-f)}$$

para $n \to \infty$, já que $\underset{n \to \infty}{plim} \, f = p$ (onde *plim* significa probabilidade no limite).

7.3.2 — *Adaptação do modelo em (6-15) quando a variável explicativa do tipo qualitativo assumir mais de dois níveis distintos*

Seja o modelo como em (6-15); isto é:

$$L_{jk} = \alpha + B_j x_j + C_k z_k \, ,$$

mas onde agora $j = 1, 2$ e 3 e $k = 1$ e 2.

Redefina-se a variável x_j como:

$$x_j = \begin{cases} 0 & \text{se } j = 1 \\ D_1 = 1 & \text{se } j = 2 \\ D_2 = 1 & \text{se } j = 3 \, . \end{cases}$$

Assim temos agora as equações

$$L_{1k} = \alpha + C_k z_k$$

$$L_{2k} = \alpha + B_1 D_1 + C_k z_k$$

$$L_{3k} = \alpha + B_2 D_2 + C_k z_k$$

que permitirão estimar quatro parâmetros, em vez dos três parâmetros do caso onde x_j assume apenas dois níveis.

7.3.3 — Prova dos resultados em (6-25)

Sejam $f_1, f_2, \ldots, f_k$ as freqüências relativas obtidas em n retiradas independentes, provindas de uma distribuição multinomial com probabilidade $P_1, P_2, \ldots, P_k$, respectivamente. Assim sendo, a expansão de Taylor (retendo-se apenas a sua parte linear) da logit condicional $Log\ f_r/f_1$, em torno dos pontos P_r e P_1, seria:

$$Log\ f_r/f_1 = Log\ P_r/P_1 + \left[\frac{\partial\ Log\ f_r/f_1}{\partial\ f_r}\right]_{f_r = P_r} (f_r - P_r) +$$

$$+ \left[\frac{\partial\ Log\ f_r/f_1}{\partial\ f_1}\right]_{f_1 = P_1} (f_1 - P_1)$$

para $r = 2, 3, \ldots, k$

ou

$$u_{r1} = Log\ f_r/f_1 - Log\ P_r/P_1 = (f_r - P_r)/P_r - (f_1 - P_1)/P_1.$$

Desta forma temos:

$$u_{21} = (f_2 - P_2)/P_2 - (f_1 - P_1)/P_1$$

$$u_{31} = (f_3 - P_3)/P_3 - (f_1 - P_1)/P_1$$

$$\vdots$$

$$u_{k1} = (f_k - P_k)/P_k - (f_1 - P_1)/P_1$$

que em notação matricial seria:

$$u = \begin{bmatrix} -1/P_1 & 1/P_2 & 0 & \ldots\ldots\ldots\ldots & 0 \\ -1/P_1 & 0 & 1/P_3 & 0 \ldots\ldots\ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -1/P_1 & 0 & \ldots\ldots\ldots\ldots\ldots\ldots & 0 & 1/P_k \end{bmatrix} \begin{bmatrix} f_1 - P_1 \\ f_2 - P_2 \\ \vdots \\ f_k - P_k \end{bmatrix} = PF$$

Segue-se então que a matriz das covariâncias par u seria do tipo:

$$V(u) = PV(F)P' .$$

Como $V(f_r) = P_r(1 - P_r)/n$ e $Cov(f_r, f_s) = -P_rP_s/n$ {Theil (1971: 65)}, vem

$$V(u) = \frac{1}{n}\begin{bmatrix} 1/P_1 + 1/P_2 & 1/P_1 \ldots\ldots & 1/P_1 \\ 1/P_1 & 1/P_1 + 1/P_3 \ldots & 1/P_1 \\ \vdots & & \\ 1/P_1 & 1/P_1 & 1/P_1 + 1/P_k \end{bmatrix} \Longrightarrow$$

$$V^{-1} = n\begin{bmatrix} P_2(1-P_2) & -P_2P_3 \ldots\ldots\ldots\ldots & -P_2P_k \\ -P_3P_2 & P_3(1-P_3) \ldots\ldots\ldots & -P_3P_k \\ \vdots & & \\ -P_kP_2 & -P_kP_3 \ldots\ldots\ldots\ldots & P_k(1-P_k) \end{bmatrix}$$

{Theil (1970)}.

8 — BIBLIOGRAFIA

AIGNER, D. J. — *Basic Econometrics*. New York, Prentice-Hall, 1971.

AMEMIYA, T. — Qualitative response models: a survey. *Journal of Economic Literature, XIX*: 1483-1536, dec. 1981.

——— & NOLD, F. — A modified logit model. *The Review of Economics and Statistics*: 255-57, may 1978.

CHRIST, C. — *Econometrics Models and Methods*. New York, John Wiley, 1966.

COSTA NETO, P. L. O. — *Estatística*. São Paulo, Blucher, 1977.

COX, D. R. — *Analysis of Binary Data*. London, Methuen, 1970.

DHRYMES, P. J. — *Introductory Econometrics*. New York, Springer — Verlag, 1978.

DOMENCICH, T. & McFADDEN, D. — *Urban Travel Demand: a Behavioral Analysis*. Amsterdam, North-Holland, 1975.

FULLER, W. A. — The use of indicator variables in computing prediction. *Journal of Econometrics, 12*: 231-43, 1980.

GOLDBERGER, A. S. — Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association, 57*: 231-43, 1962.

———. — *Econometric Theory*. New York, John Wiley, 1964.

———. — *Topics in Regression Analysis*. New York, MacMillan, 1968.

GRAYBILL, F. A. — *Introduction to Matrices with Applications in Statistics*. Califórnia, Wadsworth Pub. Co., 1969.

HAUSMAN, J. A. & WISE, D. A. — A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica, 46* (2): 403-26, mar. 1978.

HOFFMANN, R. — Elasticidades de Engel para dispêndios familiares na Cidade do Rio de Janeiro: outro método de estimação. *Pesquisa e Planejamento Econômico*.

——— & VIEIRA, S. — *Análise de Regressão: uma Introdução à Econometria*. São Paulo, Editora Hucitec, 1983.

JOHNSTON, J. — *Econometric Methods*. New York, McGraw-Hill, 1972.

JUDGE, G. G.; GRIFFITHS, W. E.; HILL, R. C. & LEE, T. C. — *The Theory and Practice of Econometrics*. New York, John Wiley, 1980.

KMENTA, J. — *Elements of Econometrics*. New York, MacMillan, 1971.

KOOYMAN, M. A. — *Dummy Variables in Econometrics*. Holland, Tilburg University Press, 1976.

LOVELL, M. C. — Seasonal adjustment of economics series and multiple regression. *Journal of the American Statistics Association, 58* (304): 993-1010, dec. 1963.

PARKS, R. W. — On the estimation of multinomial logit models from relative frequency data. *Journal of Econometrics, 13*: 293-303, 1980.

PINDYCK, R. S. & RUBINFELD, D. L. — *Econometric Models and Economic Forecast*. New York, McGraw-Hill, 1981.

RASCHE, R. H.; GAFFNEY, J.; KOO, A. Y. & OBST, N. — Functional forms for estimating the Lorenz curve. *Econometrica, 48*: 1016-62, may 1980.

ROSSI, J. W. — Income distribution in Brazil; a regional approach. *The Journal of Development Studies, 17* (2): 226-34, jan. 1981.

———. — *Índices de Desigualdade de Renda e Medidas de Concentração Industrial: aplicação a casos brasileiros.* Rio de Janeiro, Zahar, 1982.

———. — O desempenho de duas funções de Lorenz com dados de renda do Brasil. *Revista Brasileira de Economia,* Rio de Janeiro, *36* (3): 247-58, jul./set. 1982.

SALKEVER, D. S. — The use of dummy variables to computer predictions, prediction errors and confidence intervals. *Journal of Econometrics,* (4); 392-97, 1976.

SAPSFORD, D. — A note on the use of spline functions for seasonal adjustment. *Economics Letters,* (6): 145-51, 1980.

THEIL, H. — *Economics and Information Theory.* Amsterdam, North-Holland, 1967.

———. — A multinomial extension of the linear logit model. *International Economic Review,* (10): 251-59, 1969.

———. — On the estimation of relationships involving qualitative variables. *American Journal of Sociology, 67*: 103-54, july 1970.

———. — *Principles of Econometrics.* New York, John Wiley, 1971.

THOMAS, J. J. — *Introdução à Análise Estatística para economistas.* Rio de Janeiro, Zahar, 1978.

——— & WALLIS, K. F. — Seasonal variations in regression analysis. *Journal of the Royal Statistical Society,* Series A, *134* (I): 57-72, 1971.

WESOLOWSKY, G. O. — *Multiple Regression and Analysis of Variance.* New York, John Wiley, 1976.

WRIGLEY, N. — Analyzing multiple alternative dependent variables. *Geographical Analysis, 7*: 187-95, apr. 1975.

RESUMO

Neste estudo discutimos os usos mais importantes das variáveis binárias em análise de regressão, reunindo numa única publicação tópicos que acham-se dispersos na literatura. A matéria é exposta de modo tanto quanto possível didático e cobre desde os assuntos mais familiares como o uso das variáveis binárias para captar mudanças nos parâmetros da regressão entre dois ou mais períodos, passando pelo seu uso no ajustamento sazonal de séries históricas ou como instrumento para efetuar previsões, e indo até a sua utilização como variável dependente da regressão, onde uma extensão é efetuada para cobrir também o chamado modelo logit multinomial. Cada tópico é, em geral, amplamente ilustrado com exemplos, havendo ainda um apêndice técnico onde os principais resultados são derivados.

AJUSTE SAZONAL PARA OS INDICADORES DE PRODUÇÃO FÍSICA*

Departamento de Indicadores Conjunturais (IBGE/DEICO)

SUMÁRIO

1 — Introdução
2 — Método X-11
3 — O índice da produção industrial
4 — Aplicação do método
5 — Resultados
6 — Anexos
7 — Bibliografia

1 — INTRODUÇÃO

Este relatório apresenta uma síntese dos resultados obtidos nos testes de ajustamento sazonal nas séries de produção industrial realizados pela equipe do Departamento de Indicadores Conjunturais (DEICO). Os estudos nesse sentido iniciaram-se em 1983 e continuarão não só para o aperfeiçoamento das técnicas utilizadas, mas também com vistas à sua extensão às outras séries econômicas divulgadas pelo Departamento.

A sazonalidade é definida como um comportamento cíclico que ocorre em base regular no calendário anual, isto é, em ciclos cuja periodicidade é anual, semestral, mensal, ou qualquer outra unidade do calendário anual.

* Este trabalho foi elaborado pelo Departamento de Indicadores Conjunturais (IBGE/DEICO).

No caso da indústria brasileira, vários fatores concorrem para um comportamento sazonal, entre eles: a existência de uma ampla gama de produtos que processam matérias-primas agrícolas, e cujo comportamento é fortemente influenciado pela safra agrícola; o "pique" na produção de uma série de bens de consumo nos meses anteriores as festas de fim de ano; a existência de produtos cujo consumo é marcadamente estacional (por exemplo: sorvetes, sucos, refrigerantes e cerveja, no verão, e alguns remédios, no inverno); a prática estabelecida por várias empresas de conceder férias coletivas nos primeiros meses do ano, etc.

Em conjunto, esses e outros fatores provocam movimentos regulares que dificultam a interpretação do comportamento da indústria ao longo do ano.

A importância de se proceder ao ajuste sazonal — prática estabelecida internacionalmente pela maioria dos organismos responsáveis pela divulgação de indicadores mensais de atividade econômica — deriva da necessidade de permitir uma visão mais clara das tendências de comportamento do produto industrial. Além disso, e como conseqüência, as séries dessazonalizadas se prestam melhor a exercícios de previsão sobre o comportamento futuro da indústria.

A escolha do método a ser utilizado baseou-se no estudo das metodologias empregadas internacionalmente, em especial, naquelas aplicadas nos Estados Unidos e Canadá. Esse estudo reduziu a opção ao método X-11 e a uma variante deste, o X-11 ARIMA [1]. A diferença entre os dois está na forma de obtenção dos pesos para o cálculo das médias móveis, passo intermediário e necessário para se retirar a componente sazonal, enquanto o X-11 repete os pesos dos anos anteriores para obter os de anos futuros, o X-11 ARIMA os estima por um modelo linear de série de tempo.

Apesar da aparente superioridade técnica da variante X-11 ARIMA, optamos por iniciar os trabalhos de ajustamento sazonal utilizando o método X-11. Na medida em que avançarem os estudos sobre métodos de previsão no Departamento, poderá ou não ser incorporada a variante. A segunda seção deste trabalho traz uma descrição do método X-11.

Os problemas relacionados à aplicação do método aos índices de produção industrial são apresentados nas seções 3 e 4 deste trabalho. A seção 3 descreve sucintamente as características dos indicadores de produção física. A forma como as especificidades dos índices atualmente

[1] A expressão ARIMA *(Auto-Regressive Integrated Moving Average)* designa a combinação de um processo auto-regressivo (isto é, pelo qual uma variável em um determinado período depende de seu comportamento no passado) com um processo de médias móveis (pelo qual as variações da mesma variável são aplicadas pelos elementos aleatórios anteriores), aplicada a uma série de diferenças (ou seja, à uma série que é obtida pela diminuição de cada elemento pelo elemento que se segue).

calculados foram tratadas na aplicação do método, o nível de agregação e o estágio de elaboração do dado escolhido para se efetuar a correção sazonal são apresentadas na seção 4.

Finalmente, o relatório é concluído com a apresentação das séries dessazonalizadas para o período 1975-84, gráficos e comentários sobre o padrão de comportamento sazonal da indústria neste período.

Cabe ressaltar que, como mencionado anteriormente, esse relatório e a divulgação regular dos indicadores sazonalmente ajustados, não encerram os estudos da equipe nessa área. Pelo contrário, inicia-se agora uma nova fase cujos principais objetivos são:

— extensão da aplicação de fatores de correção sazonal aos demais indicadores conjunturais (emprego, salário, etc.);

— aprofundamento dos estudos sobre processos de previsão autogeradores visando o aperfeiçoamento da técnica de dessazonalização das séries, bem como a obtenção de previsões sobre o comportamento futuro das variáveis;

— obtenção de fatores de correção de variações de calendário, através da coleta de informações sobre a semana padrão de trabalho para os diversos grupos de indústria.

2 — O MÉTODO X-11

Em ciência e tecnologia, vários problemas ocorrem em que podemos idealizar modelos de análise onde uma variável X é descrita por suas observações ao longo do tempo. Vários métodos matemáticos vêm sendo desenvolvidos para o estudo das séries de tempo. As flutuações existentes nestas séries podem ser devidas a vários fatores e os esforços no sentido de sua compreensão caminham na direção da decomposição da série em algumas componentes para maior entendimento da importância de cada uma delas na série original. Em geral, a série original é decomposta em quatro componentes: de tendência (T), de sazonalidade (S), de dias trabalhos (TD = *trading-days*) e componente irregular (I). No nosso caso, deseja-se a obtenção de uma série que exclua a componente sazonal.

O método X-11 é a última variante de uma sucessão de métodos de ajustamentos sazonais construídos no *Bureau of the Census* americano, conhecidos pela letra "*X*", numa seqüência que se iniciou em 1954 (com o método X-1). O método X-11 foi introduzido em 1965, sendo agora amplamente utilizado internacionalmente para ajustar séries de índices agregados a nível nacional.

Este método é um refinamento do método de médias móveis, que tem a vantagem de medir mais precisamente os componentes da série e permitir maior flexibilidade do que outros métodos usados até o mo-

mento. Além disso, ele permite a análise de cada um dos estágios sucessivos dos processos de ajustamento sazonal. Por estas razões, ele foi adotado por quase todos os grupos engajados no trabalho de ajustamento sazonal em grande escala, a despeito do fato de ser relativamente trabalhoso.

O primeiro passo do método de médias móveis é obter uma estimativa dos fatores de tendência e ciclo pelo uso de uma simples média móvel de 12 meses que elimina as flutuações sazonais. Tal média móvel é conhecida como curva de "tendência e ciclo". O que resta na série original excluindo-se esta média móvel de 12 meses, são as componentes sazonal e irregular. Existem dois procedimentos para obtenção destas componentes. O primeiro, consiste em subtrair a média móvel (tendência e ciclo) da série original. Trata-se do método aditivo, onde a soma de todas as componentes é igual à série original. Outro procedimento é o de se dividir a série original pela média móvel. É chamado método multiplicativo, onde cada componente é dada em termos percentuais e onde o produto das componentes é igual à série original.

Note-se que, uma vez eliminada a componente de tendência e ciclo, resta-nos ainda uma série onde se somam (método aditivo) ou se multiplicam (multiplicativo) as componentes sazonal e irregular. Com efeito, a média móvel ao somar os 12 meses, elimina o efeito sazonal, mas ao mesmo tempo aplaina as irregularidades da série. É necessário então separar a componente sazonal da componente irregular. Uma estimativa do fator de ajustamento sazonal para um dado mês é obtido pela média do conjunto sazonalidade-irregularidade daquele mês para um certo número de anos. Está implícita a suposição de que o fator irregular será cancelado no processo de médias. Finalmente, a série sazonalmente ajustada será obtida dividindo-se (multiplicativo) cada observação mensal da série original pelo fator sazonal do mês correspondente.

De modo formal a decomposição da série original pode ser representada por:

— $O = T * S * TD * I$ pelo procedimento multiplicativo, ou

— $O = T + S + TD + I$ pelo procedimento aditivo.

A componente sazonal é, muitas vezes, proporcional ao valor da série original, isto é, nos anos de queda a componente sazonal é menor, nos anos de alta a componente é maior. Deste ponto de vista o método multiplicativo é superior ao método aditivo, uma vez que, pelo método multiplicativo, a componente sazonal é calculada como uma percentagem da série original. A experiência vem indicando que as componentes se relacionam de modo multiplicativo na maioria das séries econômicas nacionais.

Certas considerações devem ser feitas com respeito à obtenção dos fatores de sazonalidade. A utilização do mesmo padrão de sazonalidade

para todos os anos da série, vale dizer, um padrão obtido com base na média de todos os anos, não é conveniente. Utiliza-se então, um padrão móvel de sazonalidade, isto é, um padrão baseado em uma média móvel com um número de anos menor do que o total. Por um lado, se lançarmos mão de poucos anos para a sua obtenção, o padrão de sazonalidade fica muito sensível às variações conjunturais ou irregulares. Por outro lado, um número excessivo de anos pode não levar em consideração mudanças neste padrão. Isto nos remete à necessidade de decidir o número ótimo de anos a ser utilizado. O procedimento usual do método X-11 é o de utilizar uma média móvel ponderada de 5 anos (dupla média móvel 3 x 3). Esta média móvel funciona de modo simétrico para os anos centrais da série. Para as observações dos primeiros e últimos 2 anos, o programa aplica filtros assimétricos manejando apenas a sazonalidade estável. Ao mesmo tempo, para suavizar qualquer problema que possa advir da sensibilidade do padrão de sazonalidade às mudanças conjunturais, é dado um tratamento estatístico de correção de irregularidade extrema.

Além disso, é usual algum tipo de correção das variações de calendário. Em problemas econômicos torna-se necessário não só ajustar as diferenças em número de dias de cada mês, mas também o número de dias de trabalho nos diversos meses e no mesmo mês em diversos anos. Por exemplo: num ano, janeiro pode ter cinco domingos e no seguinte, quatro; o carnaval pode cair em fevereiro ou março; etc. O modo ideal de se fazerem comparações mensais seria o de se verificar o número de dias trabalhados em cada um dos estabelecimentos para cada mês e ponderar a produção de maneira a tornar os números comparáveis. Como em geral isto não é possível, o método X-11 utiliza, além da correção do número de dias do mês, um ajuste por meio de uma regressão múltipla onde as variáveis independentes são os dias da semana em cada mês, ou seja:

$$Y = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6 + a_7X_7$$

onde, Y é a variável a ser ajustada, X_1 é o número de segundas-feiras de cada mês, X_2 é o número de terças-feiras e assim por diante até X_7, o número de domingos de cada mês.

Em linhas gerais, o procedimento padrão do método X-11 consiste nos seguintes passos:

1 — obtenção de estimativas preliminares da componente tendência-ciclo aplicando uma média móvel centrada de 12 meses (dupla média móvel 2 x 12);

2 — obtenção de estimativa preliminar da componente sazonal através de média móvel ponderada de 5 anos (isto é, dupla média móvel 3 x 3) em cada mês, de forma a obter os fatores de correção sazonal a serem aplicados ao mês;

3 — substituição dos valores da componente sazonalidade-irregularidade com valores de irregularidade extrema. Os critérios de substituição são os abaixo:

a — obtenção de desvios padrões móveis de 5 anos (60 meses);

b — atribuição de pesos zero aos valores cuja irregularidade esteja acima de 2,5 σ;

c — atribuição de pesos proporcionais (entre 1 e 0) aos valores em que a componente irregular esteja situada entre 1,5 σ 2,5 σ;

d — atribuição de peso 1 aos valores com irregularidade abaixo de 1,5 σ;

e — substituição dos valores de peso inferior a 1, levando em consideração o peso e os valores de mesmo mês próximos do dado a ser substituído.

4 — Após a substituição dos valores com irregularidade extrema, recalcular os passos 1 e 2;

5 — ajustamento da variação correspondente aos dias de trabalho (*trading-days*) por regressão entre a série obtida em quatro e as variáveis $X_1, X_2, \ldots, X_7$, em que X_1 = n.º de segundas-feiras, ..., X_7 = n.º de domingos no mês. Em seguida recalcular o passo 2;

6 — ajustar a série pelo número de dias do mês (a menos que, um comando opcional atribuindo diferentes pesos para os meses sejam dados *a priori* pelo operador);

7 — recalcular os passos de 1 a 4 para obtenção das séries sazonalmente ajustadas finais;

8 — aplicação de uma média móvel centrada de 13 meses a série sazonalmente ajustada para obtenção das componentes tendência-ciclo e irregular finais;

9 — testar a existência de sazonalidade estável e existência da componente *trading-days;*

10 — calcular estimativas de previsão do ano seguinte aos dados;

11 — traçar gráficos, dar um sumário das medidas obtidas, etc.

Como opções, o programa apresenta as possibilidades: dar peso *a priori* a cada dia da semana; dar o número de dias do mês; dar outros limites diferentes de 1,5 σ 2,5 σ para o cálculo dos irregulares extremos; dar pesos diferentes para cada mês da série; aplicar médias móveis diferentes da média móvel ponderada de 5 anos; aplicar médias móveis diferentes para cada mês; aplicar médias móveis de 9, 13 ou 23 meses para a estimativa da componente cíclica de acordo com a relação irregular/tendência; ajustar a série para greves, etc.

3 — O ÍNDICE DA PRODUÇÃO INDUSTRIAL

Desde o início da década de 70, o Instituto Brasileiro de Geografia e Estatística (IBGE) divulga mensalmente indicadores da produção industrial, baseados em informações de quantidades produzidas de um amplo conjunto de estabelecimentos industriais selecionados pela sua participação no valor da produção da indústria, segundo os Censos Industriais.

O objetivo dos indicadores conjunturais de quantum é fornecer uma estimativa do movimento do produto industrial real, vale dizer, da evolução do valor agregado real da indústria. Como índices conjunturais, têm qualidade definida por sua capacidade de indicar a tendência efetiva da variável em estudo e pela disponibilidade dos dados com um mínimo de defasagem do período de referência.

A opção pela utilização de um índice de produção física como *proxy* da evolução do valor agregado se deu, em função de considerações sobre as dificuldades de obtenção do valor agregado real por dupla deflação, entre elas, a inadequabilidade dos índices de preços disponíveis. Supõe-se, portanto, que às variações observadas no volume da produção física correspondem variações no valor — admitindo-se implicitamente a constância na relação valor agregado/valor da produção —, e para tanto o sistema de pesos do índice tem por referência o valor agregado de cada um dos produtos que o integram.

O índice de quantum pode então ser visualizado como resultado da combinação de um conjunto de informações sobre as quantidades produzidas com um sistema de pesos, através da fórmula de cálculo que sintetiza as hipóteses acima mencionadas.

A série de índices da Produção Industrial Mensal (PIM) do IBGE se inicia em janeiro de 1975 incorporando cerca de 600 produtos, 2.500 empresas informantes e ponderações do Censo de 1970 posteriormente atualizadas para uma ponderação com base no Censo Industrial de 1975. Em 1983, foi feita uma reformulação da PIM retroagindo a dados de 1981 e envolvendo as seguintes modificações:

— reavaliação do conjunto de produtos que compõe o índice e atualização do painel de informantes considerando informações levantadas pela Pesquisa Industrial Anual (PIA) de 1978 e dados preliminares do Censo Industrial de 1980, totalizando cerca de 736 produtos e 5 mil empresas informantes;

— incorporação da estrutura de ponderação referente à PIA de 1978;

— adoção da fórmula de Laspeyres em cadeia, base de ponderação fixa e atualização de pesos via variação de quantidades.

A fórmula utilizada, por satisfazer a propriedade de circularidade, vale dizer, aquelas para as quais $I_{0,t} = I_{0,t-1} * I_{t-1,t}$, permite a deriva-

ção de todos os indicadores a partir do número índice, ou índice de base fixa mensal.

O indicador da produção industrial, na sua nova versão, possui um painel bimestral onde o informante entra no cálculo bastando existir informação para dois meses consecutivos. Sendo assim, existem duas quantidades para cada mês, uma comparando com o mês anterior, e a outra com o mês subseqüente.

Para ilustrarmos, a fórmula se segue:

$$I_{0,t} = \sum_i W_{0,i} \frac{q_{1,i}}{q_{0,i}} \frac{q_{2,i}}{q^*_{1,1}} \cdots \frac{q_{t,i}}{q^*_{t-1,i}}$$

onde:

$q_{0,i}$ é a quantidade produzida do produto i no período base de comparação do índice;

$q_{t,i}$ é a quantidade produzida do produto i no mês t, no painel comparável ao mês $t-1$;

$q^*_{t,i}$ é a quantidade produzida do produto i no mês t, no painel comparável ao mês $t+1$;

$W_{0,i}$ é o peso do produto i no ano base de ponderação do índice.

Tal formulação pode ser identificada como Laspeyres modificado base fixa, cálculo em cadeia. Preferimos a denominação em cadeia por melhor qualificar o procedimento do cálculo do índice a partir de relativos mês/mês anterior.

A citada reformulação faz com que tenhamos duas séries a estudar para obtenção de índices sazonalmente ajustados. A primeira (velha PIM) se inicia em janeiro de 1975 e termina em dezembro de 1982. A segunda (nova PIM), de janeiro de 1981 em diante. As séries de índices de produção física-Brasil resultam portanto do encadeamento de índices oriundos da antiga PIM (janeiro de 1975 a dezembro de 1980) com os da PIM reformulada que se iniciam em 1981. Tal encadeamento se fez pela relação do índice de janeiro de 1981 da série reformulada, e o índice de janeiro de 1981 da série antiga.

4 — APLICAÇÃO DO MÉTODO

As séries que temos a dessazonalizar se referem ao período 1975-84. Na realidade devemos ressaltar que, durante este período houve uma reformulação na PIM tendo de ser consideradas duas séries para este estudo: a da "velha" PIM com 600 produtos e a "nova" PIM com 736 produtos. Isto nos leva à necessidade de tomarmos em consideração na dessazonalização das séries problemas de compatibilização, encadeamento, etc. [2]

[2] Ver: *Os Novos Índices da Produção Industrial.* Rio de Janeiro, DINEC, 1984.

Devido às particularidades da sazonalidade de cada produto, consideramos a situação ideal a aplicação do método à cada produto que compõe o indicador. Esta aplicação pode se dar em dois níveis: o de quantidade produzida ou no índice base fixa do produto. Os testes iniciais de utilização do método X-11 se fizeram nas séries de quantidade produzida de cada produto no período da "velha" PIM, isto é, 1975-82. Uma análise mais detalhada foi realizada sobre os produtos de grande peso e alta variabilidade de padrão sazonal. São produtos, na sua maioria, ligados à produção agrícola, tais como: açúcar, fumo, óleo de soja, etc. Embora os resultados obtidos na dessazonalização de cada série de quantidade em particular tenham sido satisfatória, resta o problema de que os painéis podem ser diferentes levando a que as quantidades não sejam comparáveis. Isto nos fez abandonar este procedimento.

As séries trabalhadas poderiam ser as de índice base fixa (produto indicador) já que estas são obtidas a partir de relativos de quantidade com painéis comparáveis. O método foi então aplicado a cada produto indicador para a série de janeiro de 1975 a dezembro de 1982 (PIM antiga). Para cada produto deveriam ser obtidas estimativas dos fatores de ajustamento sazonal para os meses do ano posterior a fim de entrar com estes fatores a nível de produto e obter o indicador sazonalmente ajustado. No entanto, com a reformulação da PIM, alguns produtos foram incluídos, alguns agregados e outros desagregados, sendo a situação mais complexa aquela em que parte de um produto foi agregado a parte de outro. Sendo assim, para obtermos os fatores posteriores a dezembro de 1982, tornar-se-ia necessária a compatibilização dos produtos da PIM antiga com a PIM reformulada, uma vez que, o método X-11 tem como requisito um mínimo de 7 anos. Mesmo que a compatibilização fosse possível, o método não poderia ser aplicado aos novos produtos incluídos na PIM reformulada.

Dadas as dificuldades apontadas de utilização do método a nível de produto, a alternativa passa ser, a de aplicação na série de índices de base fixa a nível de subgênero.

A aplicação do método às séries de subgêneros é feita de forma direta às séries encadeadas. O problema se passa quando da obtenção do indicador geral sazonalmente ajustado. Esta obtenção é possível por dois caminhos:

a — encadeando-se as séries de índices de base fixa a nível de subgênero, aplicando-se o método, encontrando-se a série sazonalmente ajustada para cada um destes subgêneros para posterior obtenção do indicador geral ajustado;

b — aplicando-se diretamente à série encadeada do indicador geral o método X-11.

Evidentemente, dada a especificidade da componente sazonal de cada um dos subgêneros, o primeiro processo nos permite uma análise mais refinada. De todo modo, uma comparação entre os dois processos deve ser feita. O gráfico 1.1 (Anexo 1) nos apresenta as duas séries de índices de base fixa. É de se notar que os resultados obtidos são semelhantes. Isto se deve ao método que é aproximadamente linear, a menos de algumas passagens tais como a de correção de irregularidades extremas. No entanto, quando examinamos o gráfico 1.2 (Anexo 1) das taxas de crescimento mensal (indicador mês contra mês anterior) já se nota uma maior discrepância entre os dois procedimentos na obtenção de um indicador sazonalmente ajustado. Isto nos leva a uma tomada de decisão entre os dois processos. Como já dissemos, a análise por subgênero nos permite uma maior segurança quanto aos padrões de sazonalidade, sendo portanto a opção mais adequada.

O processo de encadeamento utilizado nas séries originais foi de se multiplicar toda a série da PIM antiga, pela razão entre os índices de janeiro de 1981 da série reformulada e o índice de janeiro de 1981 da série antiga. A aplicação do método X-11 nestas séries encadeadas nos permite a obtenção de séries ajustadas a nível de subgênero. O indicador geral é então obtido por meio dos pesos da PIA de 1978 como na série reformulada. No entanto, para os anos anteriores a 1981 deveriam ser aplicados os pesos fornecidos pelo Censo Industrial de 1975 como na série antiga. Isto nos remete ao problema de encadear as séries do indicador geral ajustado, vale dizer, encadear a série ajustada da PIM antiga (1975-80) com a ajustada para a PIM reformulada (1981-84). Utilizamos dois processos de encadeamento como alternativas:

a — multiplicação da série anterior a janeiro de 1981 pela razão entre janeiro de 1981 da série reformulada e janeiro de 1981 da série antiga, tal como foi feito no encadeamento da série original;

b — multiplicação da série antiga pela razão entre as médias das séries ajustadas para o ano de 1981, comum a ambas.

Para análise e opção entre os processos acima descritos, aplicamos na série antiga estes processos ao período comum às duas séries (1981-82). O resultado se encontra no gráfico 1.3 (Anexo 1). Pode-se notar que o encadeamento via razão entre os valores de janeiro das duas séries é aquele que mais se distancia dos valores ajustados para a série nova. Isto é, se usarmos a série nova como padrão, verificaremos que a multiplicação da série antiga pela razão entre as médias das duas séries para o ano de 1981 apresenta um melhor resultado.

Em resumo, o método de ajustamento para as séries do indicador geral da produção física industrial que propomos ser adotado compreende os seguintes passos:

a — aplicação do método X-11 a cada uma das séries dos subgêneros. Esta aplicação feita nas séries antigas (1975-82);

b — aplicação do método X-11 às séries dos subgêneros encadeadas (1975-84) para obtenção das séries ajustadas no período 1981-84;

c — obtenção do indicador geral ajustado sazonalmente por meio dos pesos obtidos no Censo de 1975 para o período de janeiro de 1975 a dezembro de 1980 e dos pesos da PIA de 1978 para o período de janeiro de 1981 em diante;

d — encadeamento dos dois períodos pela multiplicação da série anterior a dezembro de 1980 pela razão entre as médias do ano de 1981.

5 — RESULTADOS

O método X-11 foi aplicado a cada uma das séries de subgêneros e o indicador geral da produção industrial sazonalmente ajustado foi obtido por agregação. As tabelas 5.1 à 5.20 apresentam respectivamente as séries originais e sazonalmente ajustadas. Uma visão gráfica dos dados se encontra no anexo 2. Os gráficos correspondentes aos padrões de sazonalidade de cada um dos subgêneros podem ser encontrados no anexo 3.

Apresentaremos a seguir uma análise resumida dos resultados.

Embora o teste *F* para hipótese de existência de sazonalidade estável seja significativo para todos os subgêneros, em alguns deles a variabilidade do padrão de sazonalidade é pequena. A significância nestes casos (subgêneros: extrativa mineral, papel e papelão, autoveículos) resulta de diferenças de calendário (Semana Santa em abril, Natal em dezembro, etc.), apresentando queda de produção nos meses de verão, abril e setembro.

Na maioria dos subgêneros se nota uma produção maior no segundo semestre (julho e outubro). Isto se reflete no padrão de sazonalidade do indicador geral da indústria.

A análise dos gráficos de padrão de sazonalidade nos conduz a uma divisão da estrutura industrial já encontrada quando dos trabalhos de análise da Matriz Interindustrial de 1970 [3], onde os setores metal-mecânicos se diferenciam dos demais. Assim encontramos um padrão de sazonalidade semelhante para esses setores devido a sua interação na produção.

[3] Fürst, Patricia. — Um estudo das relações intersetoriais — Brasil/1970; Silva, Paulo Fontenelle. — *Aspectos Tecnológicos da Estrutura Industrial Brasileira.*

Dentre os demais subgêneros podemos destacar:

1 — subgêneros ligados à produção agropecuária com alta variabilidade no padrão sazonal:

— produtos alimentares, onde existe uma marcante diferença de níveis de produção entre o primeiro e o segundo semestres. Destacam-se alguns produtos pela alta participação na estrutura de pesos no subgênero como também por serem fortemente sazonais. Por exemplo: açúcar e abate e preparação de carnes;

— fumo, onde a sazonalidade depende da safra agrícola de um único produto.

2 — Subgêneros cuja produção se volta para as vendas de fim de ano, tais como: vestuário, calçados e artefatos de tecidos, produtos de matérias plásticas e perfumaria onde o pico da produção se dá nos meses de outubro e novembro.

3 — Subgêneros com comportamento sazonal singular:

— bebidas: destaca-se do seu padrão de sazonalidade o fato de que a produção se concentra nos meses de verão devido principalmente ao consumo de refrigerantes e cervejas;

— borracha: mostrando pelo padrão de sazonalidade que existe maior nível da produção entre março e julho ao contrário dos outros subgêneros. Os produtos deste subgênero são insumos na produção dos demais subgêneros;

— farmacêutica: subgênero onde a maior produção se dá nos meses de inverno, conseqüência da produção de antibióticos, antigripais, analgésicos, etc.

TABELA 5.1

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA — BRASIL — 1975

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	72,92	68,43	74,10	78,09	78,36	80,95
Extrativa mineral	72,41	68,77	76,94	78,09	81,85	77,77
Indústrias de transformação	72,93	68,42	74,03	78,08	78,27	81,03
Minerais não-metálicos	72,26	64,62	71,40	72,61	74,54	72,45
Metalúrgica	78,50	70,67	76,45	80,58	81,82	80,63
Metalúrgica básica	70,12	64,34	71,04	76,38	78,57	75,81
Outros produtos metalúrgicos	94,45	82,86	87,06	89,09	88,66	90,27
Mecânica	82,41	83,44	94,08	95,62	92,03	101,20
Material elétrico e de comunicações	61,52	56,48	61,07	66,57	66,64	71,07
Material de transporte	90,42	86,94	97,45	104,41	98,33	102,83
Autoveículos	96,44	92,13	105,83	111,04	105,45	108,08
Outros produtos de transporte	73,60	71,71	76,36	85,48	79,12	86,18
Papel e papelão	63,39	55,85	60,47	59,23	62,03	62,22
Borracha	77,43	82,32	80,10	81,73	84,16	87,65
Química	64,92	62,66	61,51	64,65	67,36	68,48
Petroquímica, refino e destilação do carvão	82,07	78,11	75,64	77,83	78,15	77,22
Outros produtos químicos	55,27	53,95	53,53	57,18	61,21	62,45
Farmacêutica	74,88	74,14	83,05	82,87	86,00	98,17
Perfumaria, sabões e velas	56,51	55,42	55,98	62,77	61,72	66,30
Produtos de matérias plásticas	73,02	67,58	71,38	73,87	74,68	71,94
Têxtil	76,29	68,85	83,36	100,93	102,12	95,34
Vestuário, calçados e artefatos de tecidos	66,56	57,75	62,42	67,10	67,35	67,72
Produtos alimentares	67,20	63,29	66,30	71,86	72,00	82,26
Bebidas	82,01	71,61	85,13	68,07	64,73	72,51
Fumo	81,72	83,37	97,51	101,12	89,68	81,30

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	84,01	82,47	84,54	87,47	82,00	79,56
Extrativa mineral	77,82	78,07	73,74	75,78	73,62	77,74
Indústrias de transformação	84,16	82,59	84,81	87,77	82,21	79,61
Minerais não-metálicos	77,11	76,86	75,46	78,01	74,01	76,89
Metalurgica	85,84	79,35	80,13	83,32	80,22	76,52
Metalúrgica básica	80,76	75,05	75,25	78,31	75,56	70,70
Outros produtos metalúrgicos	95,99	88,05	89,86	93,34	89,58	87,88
Mecânica	100,17	101,68	107,91	113,59	100,56	96,89
Material elétrico e de comunicações	77,30	76,48	80,45	88,98	80,72	73,38
Material de transporte	105,27	99,88	104,08	112,50	101,57	88,57
Autoveículos	108,96	100,77	106,43	115,19	101,30	88,01
Outros produtos de transporte	90,86	90,29	91,86	99,07	93,66	82,17
Papel e papelão	67,41	63,12	63,89	67,17	64,91	67,30
Borracha	88,66	89,33	90,38	87,09	85,50	83,11
Química	72,70	75,58	79,93	78,76	75,57	73,47
Petroquímica, refino e destilação do carvão	88,50	87,50	87,61	77,97	86,41	87,27
Outros produtos químicos	63,76	68,77	75,45	78,95	69,36	65,64
Farmacêutica	99,19	88,73	93,90	95,85	80,92	74,17
Perfumaria, sabões e velas	66,79	62,45	65,64	69,09	69,06	61,03
Produtos de matérias plásticas	80,47	86,10	86,45	94,40	88,31	79,83
Têxtil	93,13	87,59	88,92	94,87	87,19	85,14
Vestuário, calçados e artefatos de tecidos	75,68	75,03	81,70	90,21	83,52	75,65
Produtos alimentares	86,37	84,63	81,04	77,16	75,49	78,06
Bebidas	64,44	70,95	80,18	85,71	84,76	97,01
Fumo	63,80	62,92	67,94	74,93	66,12	71,53

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.2

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA — BRASIL — 1976

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	78,75	80,79	86,30	84,39	89,12	91,87
Extrativa mineral	77,92	72,97	76,48	72,59	77,58	77,08
Indústrias de transformação	78,76	80,98	86,55	84,69	89,41	92,25
Minerais não-metálicos	76,43	74,45	80,51	78,17	82,04	82,58
Metalúrgica	83,01	79,17	85,76	82,18	86,79	89,28
Metalúrgica básica	79,99	72,53	78,66	74,98	79,01	80,67
Outros produtos metalúrgicos	89,48	92,03	99,53	96,06	101,77	105,77
Mecânica	90,89	103,51	102,68	103,54	106,63	113,00
Material elétrico e de comunicações	66,63	77,12	80,33	79,89	84,91	84,13
Material de transporte	91,00	97,58	106,71	102,59	110,85	111,95
Autoveículos	96,46	103,36	112,14	105,18	113,52	114,39
Outros produtos de transporte	74,99	80,55	89,48	90,11	97,58	98,94
Papel e papelão	69,49	68,74	76,03	74,55	78,56	79,50
Borracha	83,85	87,05	91,86	87,76	92,49	97,45
Química	67,92	71,61	75,37	75,87	82,03	84,10
Petroquímica, refino e destilação do carvão	78,03	79,32	86,13	81,42	91,98	92,07
Outros produtos químicos	62,14	67,14	69,20	72,58	76,29	79,45
Farmacêutica	70,46	91,26	93,05	99,27	108,99	113,83
Perfumaria, sabões e velas	63,99	68,51	71,74	75,97	72,69	74,35
Produtos de matérias plásticas	85,34	83,49	97,67	86,69	92,44	95,89
Têxtil	86,02	83,79	99,83	102,19	102,73	97,48
Vestuário, calçados e artefatos de tecidos	69,53	74,56	76,46	72,67	82,37	76,22
Produtos alimentares	77,32	73,23	78,13	76,38	80,41	93,94
Bebidas	95,83	84,68	93,11	72,45	75,50	74,29
Fumo	89,43	94,71	104,90	89,95	88,37	82,51

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	93,60	94,54	93,00	95,97	90,29	87,69
Extrativa mineral	81,60	83,10	78,39	82,96	78,17	78,85
Indústrias de transformação	93,90	94,83	93,37	96,30	90,60	87,92
Minerais não-metálicos	85,97	86,78	85,49	89,91	86,34	87,53
Metalúrgica	92,07	92,82	90,39	90,64	86,63	87,08
Metalúrgica básica	84,77	85,92	85,05	85,03	79,76	83,09
Outros produtos metalúrgicos	106,27	106,30	101,07	101,80	99,99	95,30
Mecânica	112,07	114,08	111,12	114,09	102,51	102,93
Material elétrico e de comunicações	87,27	92,31	97,39	91,82	89,60	81,50
Material de transporte	115,52	116,06	112,41	117,01	112,29	101,45
Autoveículos	118,11	119,43	117,58	121,59	119,00	104,99
Outros produtos de transporte	101,98	101,26	95,11	100,26	92,58	87,58
Papel e papelão	79,36	76,56	77,09	80.70	78,40	76,70
Borracha	98,88	100,71	98,78	98,67	96,61	95,86
Química	85,11	88,33	89,32	92,18	85,83	84,43
Petroquímica, refino e destilação do carvão	96,90	94,38	93,42	89,61	86,83	96,43
Outros produtos químicos	78,35	84,72	86,78	93,31	85,01	77,55
Farmacêutica	113,63	109,95	97,29	96,36	89,87	83,79
Perfumaria, sabões e velas	79,99	73,55	69,93	77,84	70,62	68,31
Produtos de matérias plásticas	101,27	105,07	99,43	105,15	100,74	91,21
Têxtil	95,00	94,89	91,85	93,44	85,72	82,72
Vestuário, calçados e artefatos de tecidos	81,58	85,64	85,25	91,35	88,19	77,85
Produtos alimentares	94,68	90,83	88,95	95,86	87,89	83,79
Bebidas	76,97	86,96	90,43	95,98	96,93	106,59
Fumo	81,06	79,22	72,50	83,26	76,66	85,92

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.3

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA — BRASIL — 1977

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	81,12	78,55	92,07	85,49	94,13	94,07
Extrativa mineral	71,65	68,26	80,68	76,10	79,03	77,50
Indústrias de transformação	81,36	78,81	92,36	85,73	94,52	94,50
Minerais não-metálicos	86,47	80,14	90,96	85,68	91,73	88,15
Metalúrgica	86,63	82,60	95,12	87,72	95,76	94,17
Metalúrgica básica	84,97	80,64	93,99	86,72	93,87	92,53
Outros produtos metalúrgicos	90,70	87,15	98,36	90,64	100,35	98,31
Mecânica	83,48	90,00	103,44	93,99	103,56	102,33
Material elétrico e de comunicações	68,83	73,37	83,67	76,40	90,86	87,77
Material de transporte	92,73	98,95	113,03	92,83	115,68	111,79
Autoveículos	97,83	109,56	121,63	97,77	119,76	113,77
Outros produtos de transporte	77,14	74,25	90,31	77,50	99,82	99,53
Papel e papelão	74,94	68,67	77,95	74,65	80,09	78,57
Borracha	92,23	87,93	87,24	86,41	90,69	92,35
Química	77,30	69,84	79,08	80,73	84,93	84,77
Petroquímica, refino e destilação do carvão	92,70	83,40	90,02	84,68	89,50	83,82
Outros produtos químicos	68,58	62,15	72,81	78,30	82,15	85,04
Farmacêutica	63,57	70,85	93,01	80,89	97,25	92,07
Perfumaria, sabões e velas	64,30	62,43	70,21	67,12	75,70	70,70
Produtos de matérias plásticas	92,36	88,56	99,64	86,72	98,14	96,05
Têxtil	80,74	78,45	111,78	101,07	108,81	101,30
Vestuário, calçados e artefatos de tecidos	68,08	65,90	82,09	72,29	81,90	78,64
Produtos alimentares	79,07	71,81	85,56	83,93	89,43	105,93
Bebidas	100,38	91,00	101,10	83,08	95,75	91,02
Fumo	90,92	93,03	121,17	102,80	108,41	100,33

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	93,70	97,42	93,56	96,48	93,47	89,15
Extrativa mineral	75,56	75,83	72,71	78,65	73,98	75,17
Indústrias de transformação	94,16	97,97	94,09	96,94	93,97	89,51
Minerais não-metálicos	92,36	92,23	89,01	92,09	89,32	89,20
Metalúrgica	95,94	98,22	94,52	96,87	94,92	92,26
Metalúrgica básica	94,65	93,83	90,79	93,05	91,75	91,28
Outros produtos metalúrgicos	99,48	107,30	102,37	104,92	101,80	95,19
Mecânica	98,39	109,93	108,24	106,15	99,08	92,82
Material elétrico e de comunicações	85,25	93,37	88,36	91,64	91,25	84,90
Material de transporte	111,02	118,96	107,27	112,40	114,32	102,52
Autoveículos	112,23	122,48	107,86	114,59	119,95	103,91
Outros produtos de transporte	100,03	103,69	97,55	99,74	96,12	91,95
Papel e papelão	79,63	82,53	80,21	78,52	79,33	81,76
Borracha	94,06	97,44	94,32	96,94	94,72	92,87
Química	90,96	96,69	94,19	96,51	93,56	85,49
Petroquímica, refino e destilação do carvão	93,81	98,88	91,25	90,88	97,92	91,41
Outros produtos químicos	89,10	95,18	95,50	99,31	90,86	81,96
Farmacêutica	84,67	85,49	78,59	81,64	80,29	70,52
Perfumaria, sabões e velas	72,98	73,17	69,25	74,67	70,99	67,08
Produtos de matérias plásticas	94,38	99,46	95,75	101,83	96,43	98,46
Têxtil	94,41	94,42	91,14	94,80	92,00	89,66
Vestuário, calçados e artefatos de tecidos	78,97	83,78	83,56	90,03	89,03	81,81
Produtos alimentares	101,32	99,97	94,50	98,23	91,50	87,73
Bebidas	95,03	100,15	99,91	109,11	109,36	109,77
Fumo	85,99	83,32	78,85	80,86	77,42	90,13

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.4

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA — BRASIL — 1978

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	86,72	82,33	94,60	91,41	96,63	97,31
Extrativa mineral	73,12	71,52	81,92	81,42	83,52	81,43
Indústrias de transformação	87,07	82,60	94,93	91,66	96,96	97,72
Minerais não-metálicos	88,31	82,41	92,92	91,43	94,48	92,48
Metalúrgica	91,86	84,67	97,83	93,86	100,00	99,32
Metalúrgica básica	90,24	87,00	96,23	94,22	98,72	96,91
Outros produtos metalúrgicos	95,93	81,59	101,94	94,42	103,57	104,89
Mecânica	84,47	87,82	103,25	97,26	102,44	99,10
Material elétrico e de comunicações	78,85	78,41	97,57	93,26	100,21	100,77
Material de transporte	103,23	101,99	116,88	108,06	117,96	122,82
Autoveículos	106,84	110,17	127,68	113,76	125,99	135,63
Outros produtos de transporte	89,12	80,84	90,42	90,30	95,74	92,72
Papel e papelão	81,32	74,95	84,12	81,82	85,06	85,63
Borracha	92,63	93,36	90,72	97,55	95,39	100,26
Química	82,20	77,22	85,10	84,74	90,78	92,41
Petroquímica, refino e destilação do carvão	97,53	88,38	94,08	94,01	99,45	93,44
Outros produtos químicos	73,51	70,83	79,88	79,36	85,72	91,56
Farmacêutica	65,35	66,16	88,87	78,10	87,65	88,78
Perfumaria, sabões e velas	71,06	64,72	71,40	75,30	79,38	80,35
Produtos de matérias plásticas	94,17	91,40	102,10	99,25	108,80	101,57
Têxtil	91,05	89,74	112,55	109,89	108,37	102,67
Vestuário, calçados e artefatos de tecidos	77,37	69,63	81,66	78,23	84,52	84,43
Produtos alimentares	85,20	74,77	80,75	81,62	86,65	95,17
Bebidas	102,29	95,18	109,89	89,86	96,97	97,16
Fumo	108,32	105,89	125,84	106,63	104,59	103,41

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	98,84	105,99	100,44	106,83	100,25	94,74
Extrativa mineral	85,76	87,19	82,58	85,97	80,47	78,15
Indústrias de transformação	99,18	106,47	100,89	107,36	100,75	95,16
Minerais não-metálicos	96,48	102,25	96,27	100,12	95,93	93,95
Metalúrgica	99,51	102,97	99,87	105,03	101,39	99,02
Metalúrgica básica	98,18	100,31	99,32	105,86	101,89	101,76
Outros produtos metalúrgicos	103,18	109,03	102,14	104,89	101,80	95,42
Mecânica	99,09	112,84	102,89	112,50	109,33	100,38
Material elétrico e de comunicações	105,71	114,27	104,95	111,40	107,47	95,08
Material de transporte	122,71	135,68	117,64	132,28	126,74	119,93
Autoveículos	129,05	143,27	117,43	139,36	131,96	124,84
Outros produtos de transporte	102,74	112,68	108,31	110,37	108,17	102,43
Papel e papelão	90,00	93,83	88,49	93,20	92,35	92,18
Borracha	103,24	107,86	101,98	103,65	101,15	103,46
Química	98,08	106,68	103,27	108,95	94,12	88,41
Petroquímica, refino e destilação do carvão	96,83	111,45	102,20	104,01	95,09	104,81
Outros produtos químicos	98,46	103,71	103,53	111,32	93,30	79,10
Farmacêutica	91,17	96,39	87,57	92,21	82,88	67,59
Perfumaria, sabões e velas	82,83	81,69	78,81	85,58	83,95	79,24
Produtos de matérias plásticas	107,32	115,56	108,91	115,06	108,59	102,19
Têxtil	98,00	102,71	99,87	105,21	99,05	93,78
Vestuário, calçados e artefatos de tecidos	84,07	92,94	91,56	100,66	99,03	85,28
Produtos alimentares	94,82	100,64	98,00	101,84	90,82	86,80
Bebidas	97,41	106,21	106,84	124,34	121,46	122,12
Fumo	85,17	91,49	82,20	85,76	85,39	92,49

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.5

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA — BRASIL — 1979

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	96,00	89,57	99,87	96,29	104,69	104,70
Extrativa mineral	83,95	63,46	86,20	85,28	92,33	95,15
Indústrias de transformação	96,31	90,24	100,22	96,57	105,01	104,94
Minerais não-metálicos	98,84	88,20	103,72	96,31	102,91	99,28
Metalúrgica	103,48	91,07	103,15	99,44	108,58	105,34
Metalúrgica básica	103,87	89,12	105,49	102,61	111,30	108,93
Outros produtos metalúrgicos	104,11	95,72	100,30	95,07	105,15	100,28
Mecânica	94,87	96,33	104,20	103,86	110,27	108,25
Material elétrico e de comunicações	91,07	95,90	103,14	95,77	113,10	103,29
Material de transporte	125,37	113,19	105,30	123,36	134,44	129,59
Autoveículos	135,69	116,73	99,46	130,14	147,14	138,38
Outros produtos de transporte	98,93	98,38	105,78	102,64	103,55	105,20
Papel e papelão	96,58	88,39	98,13	91,85	99,18	96,99
Borracha	95,85	102,35	102,27	104,12	102,26	101,20
Química	87,61	85,23	94,29	92,30	101,42	107,63
Petroquímica, refino e destilação do carvão	99,28	95,36	103,45	99,26	108,50	105,12
Outros produtos químicos	80,91	79,39	88,95	88,17	97,20	108,67
Farmacêutica	66,95	70,27	85,07	90,91	97,12	96,52
Perfumaria, sabões e velas	86,91	75,31	88,01	85,25	89,13	87,46
Produtos de matérias plásticas	104,48	100,79	111,40	98,77	108,91	106,42
Têxtil	98,47	99,41	127,41	119,14	124,14	110,52
Vestuário, calçados e artefatos de tecidos	84,03	77,41	85,87	83,66	88,21	84,78
Produtos alimentares	87,10	74,11	79,77	73,66	80,55	98,40
Bebidas	122,77	99,38	115,31	85,75	103,15	96,34
Fumo	114,37	116,13	141,22	122,40	126,28	120,36

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	106,33	112,90	103,91	116,27	107,01	99,10
Extrativa mineral	95,29	100,78	94,52	100,84	95,67	96,85
Indústrias de transformação	106,61	113,21	104,14	116,66	107,30	99,15
Minerais não-metálicos	101,32	103,98	95,87	104,82	98,19	98,84
Metalúrgica	108,06	114,13	106,72	116,54	109,56	106,09
Metalúrgica básica	112,91	116,53	111,61	118,94	112,40	111,17
Outros produtos metalúrgicos	100,81	111,31	99,37	113,75	105,90	98,40
Mecânica	108,63	121,23	111,64	128,41	113,64	102,85
Material elétrico e de comunicações	111,11	120,39	106,40	126,86	115,06	97,48
Material de transporte	123,67	142,40	118,14	143,55	138,60	123,67
Autoveículos	126,60	153,73	120,84	154,09	146,05	128,33
Outros produtos de transporte	108,95	112,99	104,23	115,29	115,58	106,22
Papel e papelão	98,26	103,55	100,22	108,04	99,88	99,44
Borracha	109,64	113,71	109,15	112,71	112,22	111,68
Química	112,35	114,89	108,87	113,46	101,60	96,42
Petroquímica, refino e destilação do carvão	104,84	113,64	105,57	106,55	106,05	105,81
Outros produtos químicos	116,13	115,22	110,34	116,91	98,83	90,94
Farmacêutica	98,17	98,93	90,18	94,80	88,50	70,21
Perfumaria, sabões e velas	94,93	96,83	87,15	99,83	95,88	88,30
Produtos de matérias plásticas	108,09	120,65	115,40	129,91	121,99	110,13
Têxtil	108,69	111,55	102,71	112,30	105,94	95,43
Vestuário, calçados e artefatos de tecidos	91,13	97,59	92,57	108,90	104,67	83,48
Produtos alimentares	96,25	102,18	93,09	107,49	93,85	86,44
Bebidas	96,44	112,41	107,89	135,49	129,44	124,19
Fumo	104,48	95,42	76,29	88,62	78,99	81,36

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.6

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA — BRASIL — 1980

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	102,32	100,30	110,75	103,03	113,24	114,72
Extrativa mineral	93,22	96,66	102,70	101,52	107,38	104,08
Indústrias de transformação	102,55	100,39	110,96	103,07	113,39	114,99
Minerais não-metálicos	99,93	96,21	105,76	100,09	105,33	104.42
Metalúrgica	113,11	108,64	118,22	109,49	122,56	119,94
Metalúrgica básica	116,26	112,67	122,89	112,11	127,17	123,81
Outros produtos metalúrgicos	108,94	102,84	111,42	106,22	115,90	114,57
Mecânica	107,22	110,78	130,81	103,65	126,36	128,15
Material elétrico e de comunicações	94,26	98,13	110,49	112,01	119,42	118,67
Material de transporte	133,32	128,51	131,44	93,61	132,29	136,28
Autoveículos	144,10	135,00	139,62	75,32	135,55	140,53
Outros produtos de transporte	105,54	107,83	107,86	114,51	116,32	118,46
Papel e papelão	106,60	101,96	109,57	109,16	112,80	109,32
Borracha	103,36	110,68	112,25	109,46	111,89	118,00
Química	93,08	91,40	101,14	102,97	111,64	115,33
Petroquímica, refino e destilação do carvão	102,75	90,96	102,47	102,28	112,37	97,98
Outros produtos químicos	87,46	91,35	100,09	103,02	110,89	124,52
Farmacêutica	78,55	80,07	93,13	93,84	99,75	111,58
Perfumaria, sabões e velas	98,40	95,92	94,98	91,38	94,71	93,99
Produtos de matérias plásticas	119,07	116,91	128,19	122,11	126,03	125,90
Têxtil	104,23	112,38	134,94	126,47	130,29	114,86
Vestuário, calçados e artefatos de tecidos	89,51	83,16	88,71	92,18	92,72	93,79
Produtos alimentares	88,01	78,60	86,35	79,96	87,57	104,00
Bebidas	122,95	112,17	112,44	102,35	101,90	94,47
Fumo	106,61	112,70	135,97	129,95	124,81	99,05

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	120,31	118,99	122,03	124,32	115,30	104,97
Extrativa mineral	108,61	110,73	105,84	106,87	100,26	92,44
Indústrias de transformação	120,61	119,19	122,44	124,77	115,69	105,29
Minerais não-metálicos	113,04	112,67	112,70	115,70	111,42	108.31
Metalúrgica	127,36	122,53	124,05	128,93	123,08	113.05
Metalúrgica básica	129,16	125,75	123,41	129,71	125,52	119,25
Outros produtos metalúrgicos	125,78	118,36	126,80	129,18	120,32	103,46
Mecânica	134,96	130,72	140,65	139,91	131,64	108,23
Material elétrico e de comunicações	131,08	129,54	138,83	147,34	130,02	107,14
Material de transporte	135,20	134,32	147,30	150,98	138,80	127,74
Autoveículos	139,29	139,35	158,13	167,37	148,36	130,62
Outros produtos de transporte	117,70	115,42	118,29	112,99	102,46	112,74
Papel e papelão	114,24	114,45	110,63	114,93	107,16	102,11
Borracha	120,72	122,15	123,28	124,61	121,92	118,41
Química	118,75	118,97	114,24	107,31	102,82	99,43
Petroquímica, refino e destilação do carvão	109,40	108,88	106,55	97,66	93,31	101,30
Outros produtos químicos	123,52	124,15	118,12	112,28	107,73	98,09
Farmacêutica	108,44	101,38	110,56	103,51	102,70	86,23
Perfumaria, sabões e velas	108,90	98,50	100,62	107,25	95,65	92,04
Produtos de matérias plásticas	131,17	133,33	140,29	136,96	132,21	117,89
Têxtil	116,37	114,18	115,55	121,06	111,02	100,06
Vestuário, calçados e artefatos de tecidos	102,46	104,97	115,06	122,59	118,23	94,41
Produtos alimentares	109,33	110,09	111,92	117,22	97,48	92,31
Bebidas	102,90	112,61	119,01	128,98	122,51	123,27
Fumo	92,12	79,04	85,88	90,36	80,39	80,05

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.7

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA — BRASIL — 1981

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	102,52	101,73	102,62	95,86	98,97	101,91
Extrativa mineral	96,91	99,13	103,04	95,40	103,13	97,76
Indústrias de transformação	102,66	101,80	102,61	95,87	98,86	102,02
Minerais não-metálicos	105,60	99,26	104,67	96,54	96,61	95,81
Metalúrgica	114,77	109,18	112,59	101,94	101,12	97,99
Metalúrgica básica	119,41	112,25	117,55	103,96	101,47	97,04
Outros produtos metalúrgicos	107,97	104,70	105,33	98,98	100,62	99,38
Mecânica	103,45	111,08	112,09	100,24	103,18	102,48
Material elétrico e de comunicações	95,01	107,14	97,84	97,50	101,31	103,21
Material de transporte	120,06	122,16	111,62	99,11	109,26	107,03
Autoveículos	128,89	127,83	114,98	98,10	113,16	110,19
Outros produtos de transporte	96,39	106,95	102,63	101,83	98,80	98,55
Papel e papelão	108,03	101,58	105,29	94,72	97,58	94,86
Borracha	105,57	114,33	117,51	106,44	104,04	103,06
Química	96,50	89,36	93,15	93,61	96,55	101,26
Petroquímica, refino e destilação do carvão	105,17	98,48	96,35	92,32	97,13	94,27
Outros produtos químicos	91,44	84,04	91,29	94,37	96,20	105,34
Farmacêutica	90,31	95,76	91,98	97,64	107,99	109,84
Perfumaria, sabões e velas	95,23	92,24	96,64	98,16	106,96	96,88
Produtos de matérias plásticas	111,11	111,07	106,70	91,52	91,40	94,85
Têxtil	104,15	100,41	101,06	93,43	96,96	100,18
Vestuário, calçados e artefatos de tecidos	91,06	94,62	95,40	93,39	94,60	95,94
Produtos alimentares	89,48	82,17	86,38	80,95	86,28	112,62
Bebidas	112,39	107,12	112,60	99,78	100,56	92,71
Fumo	89,76	125,64	144,94	142,07	125,87	101,32

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	105,58	102,21	99,78	102,66	96,79	89,35
Extrativa mineral	96,60	98,60	99,03	107,70	102,51	100,19
Indústrias de transformação	105,81	102,31	99,80	102,54	96,65	89,07
Minerais não-metálicos	102,76	101,33	100,30	103,77	97,79	95,57
Metalúrgica	102,00	94,37	90,04	95,19	93,66	87,14
Metalúrgica básica	100,06	92,62	86,04	92,29	89,58	87,73
Outros produtos metalúrgicos	104,85	96,94	95,89	99,43	99,63	86,28
Mecânica	102,64	97,67	97,90	97,56	90,11	81,59
Material elétrico e de comunicações	109,95	101,66	101,34	101,90	102,64	80,50
Material de transporte	103,39	89,93	80,63	93,94	87,85	75,01
Autoveículos	102,06	84,33	73,74	90,96	83,42	72,35
Outros produtos de transporte	106,95	104,96	99,10	101,93	99,75	82,15
Papel e papelão	99,42	100,52	96,76	100,61	98,86	101,76
Borracha	103,70	100,87	75,72	88,83	86,95	92,99
Química	108,57	111,26	106,95	109,25	98,68	94,85
Petroquímica, refino e destilação do carvão	107,11	107,77	101,73	103,28	96,45	99,94
Outros produtos químicos	109,42	113,30	110,00	112,73	99,99	91,88
Farmacêutica	116,46	107,74	103,97	98,67	101,39	78,25
Perfumaria, sabões e velas	101,01	94,59	100,98	111,69	109,82	95,81
Produtos de matérias plásticas	99,07	98,22	99,86	104,41	99,31	92,47
Têxtil	106,03	101,84	100,13	104,68	98,00	93,13
Vestuário, calçados e artefatos de tecidos	102,19	103,16	107,11	114,65	114,98	92,90
Produtos alimentares	117,48	120,50	119,14	111,89	98,98	94,11
Bebidas	88,28	88,24	95,14	100,88	93,09	109,20
Fumo	89,25	70,56	78,86	78,50	72,59	80,65

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.8

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA — BRASIL — 1982

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	86,83	85,71	101,51	95,33	100,90	105,26
Extrativa mineral	106,06	100,83	108,76	105,12	110,27	104,43
Indústrias de transformação	86,34	85,32	101,32	95,08	100,66	105,28
Minerais não-metálicos	95,27	89,54	103,21	99,32	103,38	99,21
Metalúrgica	91,49	89,08	103,17	96,80	102,28	100,01
Metalúrgica básica	91,34	88,58	100,90	95,98	100,44	99,41
Outros produtos metalúrgicos	91,70	89,81	106,50	98,00	104,97	100,88
Mecânica	74,21	77,99	92,45	82,77	87,63	86,40
Material elétrico e de comunicações	78,19	81,67	108,35	103,15	109,32	107,19
Material de transporte	78,77	81,28	101,69	96,98	95,95	100,56
Autoveículos	79,23	81,65	101,68	96,79	94,26	101,50
Outros produtos de transporte	77,55	80,28	101,73	97,48	100,47	98,02
Papel e papelão	101,53	98,03	108,99	106,87	110,40	109,96
Borracha	91,84	86,55	94,15	100,53	103,15	102,53
Química	89,60	84,48	95,59	97,39	110,13	123,34
Petroquímica, refino e destilação do carvão	103,31	91,24	95,58	99,25	101,63	103,13
Outros produtos químicos	81,59	80,53	95,59	96,29	115,10	135,15
Farmacêutica	75,92	95,39	112,21	92,67	111,81	102,03
Perfumaria, sabões e velas	92,38	91,35	113,92	108,79	110,22	104,45
Produtos de matérias plásticas	95,28	98,76	116,77	112,06	113,89	115,56
Têxtil	95,04	93,56	110,89	103,02	109,46	110,25
Vestuário, calçados e artefatos de tecidos	84,97	84,28	104,24	99,26	103,30	104,18
Produtos alimentares	82,40	75,92	89,70	76,30	82,76	110,77
Bebidas	98,20	92,07	101,57	92,20	85,40	90,68
Fumo	89,53	109,53	159,71	139,10	126,09	108,29

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	108,39	110,97	106,90	105,40	99,91	90,26
Extrativa mineral	110,54	102,79	94,13	112,10	114,02	108,49
Indústrias de transformação	108,34	111,18	107,23	105,23	99,55	89,79
Minerais não-metálicos	101,64	101,99	98,21	98,85	91,74	87,67
Metalúrgica	105,04	105,62	100,56	100,46	95,33	83,29
Metalúrgica básica	106,74	106,14	100,56	98,22	92,26	81,25
Outros produtos metalúrgicos	102,56	104,85	100,55	103,73	99,83	86,28
Mecânica	89,38	95,08	90,15	82,92	79,03	69,56
Material elétrico e de comunicações	106,70	117,85	108,72	108,87	105,98	84,71
Material de transporte	99,74	107,20	101,10	103,43	101,77	89,63
Autoveículos	100,25	106,91	101,03	104,36	105,85	91,48
Outros produtos de transporte	98,39	107,95	101,28	100,94	90,82	84,66
Papel e papelão	112,37	112,52	106,74	109,87	106,01	103,35
Borracha	101,77	96,35	82,49	87,95	83,91	84,67
Química	131,66	131,97	126,92	118,46	105,57	90,00
Petroquímica, refino e destilação do carvão	108,85	111,38	105,21	103,81	102,73	107,34
Outros produtos químicos	144,98	144,00	139,60	127,02	107,24	79,87
Farmacêutica	104,06	107,52	100,73	99,65	105,96	93,89
Perfumaria, sabões e velas	101,08	99,51	102,27	100,81	104,26	98,29
Produtos de matérias plásticas	118,05	120,97	112,46	109,07	104,28	88,49
Têxtil	114,13	114,05	105,84	105,99	98,96	93,29
Vestuário, calçados e artefatos de tecidos	109,12	116,04	114,77	114,71	111,54	93,53
Produtos alimentares	113,33	116,87	122,33	119,87	109,99	108,82
Bebidas	93,03	90,11	99,00	102,70	104,13	110,11
Fumo	91,25	74,84	77,06	72,50	80,59	94,37

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.9

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA — BRASIL — 1983

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	84,22	81,35	96,15	86,54	95,54	96,30
Extrativa mineral	116,37	104,30	118,36	110,18	122,70	118,37
Indústrias de transformação	83,40	80,77	95,59	85,93	94,85	95,74
Minerais não-metálicos	83,96	77,28	84,98	78,87	85,19	79,62
Metalúrgica	84,00	81,69	97,49	92,42	99,23	96,06
Metalúrgica básica	80,40	77,75	93,51	93,38	97,83	96,27
Outros produtos metalúrgicos	89,27	87,45	103,32	91,02	101,28	95,75
Mecânica	61,25	63,20	80,24	67,95	79,45	76,24
Material elétrico e de comunicações	73,96	82,43	102,18	86,68	95,93	95,51
Material de transporte	79,91	78,45	95,99	82,96	102,13	93,91
Autoveículos	84,72	84,29	102,18	89,38	111,11	102,01
Outros produtos de transporte	66,99	62,77	79,38	65,75	78,04	72,18
Papel e papelão	106,02	94,22	113,70	106,75	113,32	110,87
Borracha	73,61	84,57	95,77	98,65	103,65	95,43
Química	82,38	78,72	90,31	91,45	98,00	103,33
Petroquímica, refino e destilação do carvão	99,81	91,95	96,64	97,14	101,46	94,37
Outros produtos químicos	72,21	71,00	86,62	83,14	95,98	108,56
Farmacêutica	78,21	89,99	110,85	89,01	111,41	98,39
Perfumaria, sabões e velas	90,78	92,09	107,55	107,18	103,41	101,65
Produtos de matérias plásticas	93,09	92,01	106,87	96,05	101,13	96,95
Têxtil	92,87	85,03	102,45	91,04	99,89	96,90
Vestuário, calçados e artefatos de tecidos	81,20	76,67	98,05	86,42	92,56	89,35
Produtos alimentares	98,27	85,85	93,02	75,88	85,58	113,62
Bebidas	100,54	94,32	101,61	80,66	81,58	82,05
Fumo	94,67	114,88	159,27	143,29	137,45	95,74

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	96,29	104,01	100,19	102,54	99,34	91,78
Extrativa mineral	124,42	119,02	126,06	129,84	134,08	141,29
Indústrias de transformação	95,57	103,63	99,53	101,84	98,46	90,52
Minerais não-metálicos	82,67	86,03	79,80	82,32	80,34	81,12
Metalúrgica	99,03	103,35	100,09	102,81	102,13	99,07
Metalúrgica básica	101,53	104,52	104,68	105,64	105,59	106,48
Outros produtos metalúrgicos	95,38	101,63	93,37	98,65	97,05	88,21
Mecânica	73,26	81,94	84,38	81,92	82,02	75,83
Material elétrico e de comunicações	93,11	97,58	94,24	96,79	89,79	71,81
Material de transporte	80,01	94,87	90,97	86,55	86,96	86,95
Autoveículos	84,41	102,17	95,05	90,83	91,76	93,90
Outros produtos de transporte	68,20	75,28	80,01	75,06	74,09	68,28
Papel e papelão	108,69	115,09	112,36	117,97	114,00	110,91
Borracha	106,30	100,97	93,16	102,66	81,34	85,43
Química	112,06	123,30	119,22	124,01	111,19	95,49
Petroquímica, refino e destilação do carvão	95,34	106,16	99,52	105,53	97,41	98,93
Outros produtos químicos	121,83	133,31	130,72	134,81	119,24	93,48
Farmacêutica	91,12	99,54	91,99	86,09	98,07	77,02
Perfumaria, sabões e velas	98,69	104,28	98,70	107,69	107,27	100,13
Produtos de matérias plásticas	96,38	98,07	98,88	99,35	96,39	85,52
Têxtil	95,91	101,99	98,35	98,27	90,34	80,30
Vestuário, calçados e artefatos de tecidos	84,85	100,65	100,38	100,57	104,14	91,65
Produtos alimentares	116,83	127,20	115,00	124,14	121,18	110,58
Bebidas	80,82	87,39	81,23	89,05	98,19	104,42
Fumo	69,07	75,33	75,20	71,86	73,11	77,93

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.10

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA — BRASIL — 1984

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)						
	Janeiro	Fevereiro	Março	Abril	Maio	Junho	Julho
Indústria geral	87,74	91,70	93,80	89,93	101,55	103,67	106,86
Extrativa mineral	149,09	147,81	147,33	145,01	156,19	146,13	158,70
Indústrias de transformação	86,18	90,28	92,44	88,53	100,16	102,59	105,54
Minerais não-metálicos	78,73	78,06	79,15	73,23	82,03	78,57	84,05
Metalúrgica	96,66	101,42	104,19	101,71	110,16	105,36	111,34
Metalúrgica básica	103,88	106,48	114,54	109,00	119,19	116,38	119,66
Outros produtos metalúrgicos	86,09	94,01	89,05	91,03	96,95	89,21	99,15
Mecânica	66,20	82,88	83,92	83,91	90,39	89,40	89,60
Material elétrico e de comunicações	75,55	86,61	87,28	81,60	94,74	90,57	97,76
Material de transporte	82,66	91,52	91,05	80,62	107,78	96,15	95,94
Autoveículos	92,33	101,13	99,04	85,32	116,99	103,54	101,84
Outros produtos de transporte	56,70	65,74	69,62	68,00	83,08	76,35	80,12
Papel e papelão	113,81	109,81	114,86	113,45	117,51	115,45	118,06
Borracha	81,08	98,43	97,43	97,90	102,19	108,41	109,94
Química	90,67	88,87	97,64	100,17	112,57	120,67	127,89
Petroquímica, refino e destilação do carvão	105,37	98,45	106,33	103,05	105,27	108,26	110,87
Outros produtos químicos	82,09	83,27	92,57	98,49	116,83	127,92	137,84
Farmacêutica	85,70	93,39	93,27	85,82	120,06	94,56	102,33
Perfumaria, sabões e velas	95,67	97,32	95,11	96,87	101,98	92,39	103,70
Produtos de matérias plásticas	87,44	96,78	92,86	89,78	99,07	94,75	97,98
Têxtil	83,18	85,87	90,19	85,50	96,31	95,08	98,84
Vestuário, calçados e artefatos de tecidos	83,06	85,95	87,33	88,53	97,37	90,78	97,31
Produtos alimentares	95,10	87,05	83,73	75,79	84,75	126,90	123,09
Bebidas	97,55	96,46	98,78	65,93	84,91	85,48	79,98
Fumo	86,37	113,31	145,36	141,12	151,61	102,13	78,84

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.11

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA COM AJUSTAMENTO SAZONAL BRASIL — 1975

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	77,43	76,47	75,08	79,75	76,51	78,92
Extrativa mineral	76,87	76,17	75,54	77,64	78,89	76,87
Indústrias de transformação	77,43	76,46	75,02	79,79	76,44	78,96
Minerais não-metálicos	72,87	70,91	71,04	74,37	72,88	73,52
Metalúrgica	79,45	77,46	77,94	82,80	79,65	80,08
Metalúrgica básica	70,22	70,45	70,46	78,72	76,10	75,89
Outros produtos metalúrgicos	96,22	91,09	89,05	91,14	87,31	88,60
Mecânica	94,50	93,56	94,33	96,66	92,10	98,51
Material elétrico e de comunicações	70,31	64,18	64,58	68,67	65,84	69,97
Material de transporte	97,30	95,03	97,32	110,03	93,72	98,59
Autoveículos	101,72	97,52	105,24	121,52	99,86	105,92
Outros produtos de transporte	82,81	84,77	81,29	88,25	79,01	82,17
Papel e papelão	64,19	61,30	60,22	60,04	60,20	61,88
Borracha	81,95	85,03	84,04	85,06	85,10	85,61
Química	72,25	72,96	68,10	69,29	68,80	70,90
Petroquímica, refino e destilação do carvão	82,37	85,10	78,57	81,84	76,98	81,67
Outros produtos químicos	63,50	65,81	59,29	60,54	61,55	61,71
Farmacêutica	93,22	89,34	82,10	81,81	79,62	86,06
Perfumaria, sabões e velas	58,75	60,22	58,02	61,23	61,08	64,92
Produtos de matérias plásticas	76,25	74,35	70,53	78,46	73,57	74,14
Têxtil	82,93	78,65	76,00	93,50	89,22	91,17
Vestuário, calçados e artefatos de tecidos	71,31	68,48	67,28	71,51	67,76	70,46
Produtos alimentares	71,51	75,21	73,67	77,77	74,48	75,99
Bebidas	73,59	74,38	82,44	77,39	69,15	84,66
Fumo	75,28	77,61	79,55	89,56	78,86	79,21

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	78,79	78,66	81,24	80,18	82,52	81,89
Extrativa mineral	75,51	74,94	74,62	72,87	74,92	78,02
Indústrias de transformação	78,86	78,75	81,40	80,37	82,70	81,98
Minerais não-metálicos	74,01	74,47	75,10	74,72	75,32	77,17
Metalúrgica	81,06	77,69	79,33	79,82	81,72	78,63
Metalúrgica básica	76,33	73,85	75,10	75,59	77,01	72,45
Outros produtos metalúrgicos	90,48	85,81	88,13	87,74	91,64	90,62
Mecânica	96,08	96,70	101,38	101,52	102,73	102,34
Material elétrico e de comunicações	71,69	72,37	75,45	78,27	79,82	77,00
Material de transporte	97,14	95,14	102,16	100,60	101,08	94,17
Autoveículos	101,66	99,15	107,00	101,94	103,80	98,03
Outros produtos de transporte	82,22	85,11	88,01	90,02	93,44	84,03
Papel e papelão	64,83	61,76	64,03	65,06	66,07	68,22
Borracha	85,47	84,35	87,39	84,54	85,14	83,70
Química	70,92	71,13	75,28	73,98	75,86	77,19
Petroquímica, refino e destilação do carvão	83,73	81,80	84,41	80,85	85,09	83,42
Outros produtos químicos	60,87	62,23	66,73	66,81	66,70	70,12
Farmacêutica	84,41	85,24	90,30	89,04	87,44	87,24
Perfumaria, sabões e velas	61,76	62,42	66,43	63,61	71,23	63,25
Produtos de matérias plásticas	77,06	82,03	83,39	86,44	88,32	83,47
Têxtil	91,47	89,18	91,77	92,73	94,44	94,31
Vestuário, calçados e artefatos de tecidos	72,09	73,00	76,26	76,10	78,05	76,60
Produtos alimentares	77,19	77,85	76,71	68,49	77,88	78,86
Bebidas	68,65	74,15	79,35	76,97	80,84	84,56
Fumo	66,29	72,88	81,26	81,71	82,83	78,80

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.12

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA COM AJUSTAMENTO SAZONAL BRASIL — 1976

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	84,48	88,14	85,48	86,62	88,27	88,19
Extrativa mineral	82,68	80,74	75,06	72,26	74,79	76,22
Indústrias de transformação	84,51	88,31	85,77	86,96	88,60	88,49
Minerais não-metálicos	77,39	79,24	79,00	80,59	80,77	82,70
Metalúrgica	84,72	84,12	84,04	8[illegible],54	85,45	87,38
Metalúrgica básica	80,65	76,75	75,95	77,00	77,60	79,75
Outros produtos metalúrgicos	92,93	98,62	98,02	99,06	101,25	102,26
Mecânica	106,71	111,71	99,97	104,69	107,72	108,93
Material elétrico e de comunicações	79,94	84,34	80,99	82,57	84,32	82,49
Material de transporte	100,06	103,63	103,10	106,22	109,14	106,76
Autoveículos	103,98	108,63	105,42	110,16	113,09	111,15
Outros produtos de transporte	85,52	91,00	92,53	92,66	98,55	93,90
Papel e papelão	70,27	72,89	74,80	75,98	76,91	78,10
Borracha	88,78	89,83	96,20	90,96	93,75	95,25
Química	75,36	83,33	83,13	83,14	84,66	84,98
Petroquímica, refino e destilação do carvão	76,21	84,79	88,16	86,92	92,54	93,38
Outros produtos químicos	71,59	81,91	76,54	77,28	76,89	76,97
Farmacêutica	90,96	105,77	87,48	97,69	102,06	99,21
Perfumaria, sabões e velas	67,60	72,77	71,63	73,96	73,01	72,16
Produtos de matérias plásticas	88,58	90,02	93,93	93,30	92,56	96,05
Têxtil	93,60	92,85	89,21	92,81	92,22	93,19
Vestuário, calçados e artefatos de tecidos	76,25	85,77	78,82	76,95	85,16	77,73
Produtos alimentares	82,65	85,67	83,14	85,48	84,85	83,08
Bebidas	83,97	86,69	86,33	83,34	83,62	82,38
Fumo	82,87	85,97	81,58	78,38	82,32	76,49

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	89,49	88,64	89,18	89,85	88,91	90,41
Extrativa mineral	79,11	79,54	79,27	79,76	79,61	79,42
Indústrias de transformação	89,74	88,87	89,41	90,10	89,13	90,67
Minerais não-metálicos	83,11	83,91	85,03	86,60	87,24	87,96
Metalúrgica	88,47	89,56	89,05	88,59	86,73	89,39
Metalúrgica básica	81,17	84,00	84,42	83,17	80,52	84,61
Outros produtos metalúrgicos	102,93	101,06	98,55	99,51	98,73	98,59
Mecânica	111,45	105,13	103,31	105,87	101,96	108,30
Material elétrico e de comunicações	84,34	84,36	89,75	85,10	85,49	84,80
Material de transporte	109,64	107,51	110,09	109,86	107,14	106,34
Autoveículos	112,81	112,84	118,17	116,67	112,82	113,35
Outros produtos de transporte	95,78	92,34	90,97	94,29	89,76	89,43
Papel e papelão	77,41	73,99	77,11	79,01	79,16	77,96
Borracha	95,40	95,04	95,81	95,86	96,23	96,45
Química	82,85	82,94	84,79	85,77	86,28	89,70
Petroquímica, refino e destilação do carvão	91,96	87,84	93,38	90,83	84,27	94,33
Outros produtos químicos	74,12	76,31	76,88	78,95	82,04	83,22
Farmacêutica	102,51	99,09	93,45	94,33	91,59	99,27
Perfumaria, sabões e velas	75,08	72,52	70,90	74,08	70,40	70,50
Produtos de matérias plásticas	98,52	98,24	97,48	97,76	97,49	97,02
Têxtil	93,65	95,54	94,88	93,52	90,42	90,88
Vestuário, calçados e artefatos de tecidos	80,30	80,85	79,51	80,33	78,78	78,57
Produtos alimentares	85,41	83,34	85,10	86,74	86,90	86,24
Bebidas	83,62	90,30	89,95	88,07	89,29	93,04
Fumo	86,15	89,77	87,56	94,68	92,69	93,88

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.13

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA COM AJUSTAMENTO SAZONAL BRASIL — 1977

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	87,68	87,66	90,46	89,29	92,10	90,53
Extrativa mineral	76,10	75,22	79,14	75,97	76,33	76,74
Indústrias de transformação	87,96	87,95	90,73	89,62	92,46	90,88
Minerais não-metálicos	88,32	88,11	88,80	88,54	89,91	88,80
Metalúrgica	89,35	90,13	92,49	91,45	93,24	92,17
Metalúrgica básica	87,01	87,95	90,95	90,05	91,23	91,41
Outros produtos metalúrgicos	95,18	95,38	96,95	95,90	98,31	95,14
Mecânica	98,57	100,60	99,45	98,65	101,45	98,48
Material elétrico e de comunicações	83,04	83,62	83,73	82,80	87,17	85,46
Material de transporte	102,73	107,42	108,33	100,41	111,01	105,56
Autoveículos	109,30	116,42	113,65	109,55	115,00	107,70
Outros produtos de transporte	88,02	86,37	91,73	82,53	97,73	94,40
Papel e papelão	76,04	75,11	75,82	76,83	78,82	77,57
Borracha	97,80	90,56	90,91	89,16	92,16	90,32
Química	86,82	85,64	86,35	87,80	87,48	85,89
Petroquímica, refino e destilação do carvão	93,19	90,95	91,09	88,18	88,53	87,94
Outros produtos químicos	79,66	76,01	80,26	83,91	82,99	81,74
Farmacêutica	82,66	85,40	86,05	84,11	86,55	80,44
Perfumaria, sabões e velas	68,11	68,36	70,06	66,82	74,71	69,31
Produtos de matérias plásticas	97,43	96,83	95,49	93,35	97,21	97,53
Têxtil	89,81	88,11	98,38	93,90	97,08	97,07
Vestuário, calçados e artefatos de tecidos	76,48	77,83	83,57	79,01	82,07	80,44
Produtos alimentares	85,36	85,73	93,06	93,31	95,17	96,04
Bebidas	94,71	95,04	93,24	95,49	104,06	102,70
Fumo	89,24	86,59	90,58	91,72	95,48	94,14

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	90,68	90,31	89,91	90,18	91,54	93,01
Extrativa mineral	73,13	72,29	73,31	75,48	75,34	76,24
Indústrias de transformação	91,12	90,75	90,33	90,54	91,93	93,43
Minerais não-metálicos	89,61	88,51	88,44	88,89	89,52	90,07
Metalúrgica	93,57	93,94	93,54	94,23	94,74	95,71
Metalúrgica básica	91,83	91,40	90,12	91,19	92,10	93,50
Outros produtos metalúrgicos	98,04	99,58	100,60	101,47	100,02	101,15
Mecânica	98,93	100,73	101,51	98,55	97,45	100,31
Material elétrico e de comunicações	82,68	84,07	82,13	84,75	86,09	93,09
Material de transporte	108,47	108,84	105,24	105,62	109,27	109,52
Autoveículos	113,85	112,97	107,34	109,49	115,95	114,21
Outros produtos de transporte	95,73	95,41	93,37	94,34	93,27	95,17
Papel e papelão	78,44	79,49	80,51	76,77	79,87	83,23
Borracha	90,82	91,96	92,00	94,17	94,29	93,25
Química	88,33	89,93	89,44	90,04	94,21	90,51
Petroquímica, refino e destilação do carvão	90,63	91,47	90,17	92,47	95,60	87,05
Outros produtos químicos	82,99	84,98	85,03	84,11	88,08	88,51
Farmacêutica	78,11	77,23	75,08	79,69	81,59	86,64
Perfumaria, sabões e velas	69,19	70,84	69,81	71,51	70,56	71,03
Produtos de matérias plásticas	93,87	91,93	93,60	94,41	93,31	103,97
Têxtil	95,55	94,55	92,83	94,77	97,44	98,90
Vestuário, calçados e artefatos de tecidos	79,33	78,39	77,33	79,68	79,01	84,85
Produtos alimentares	92,96	87,41	90,18	88,33	89,00	90,60
Bebidas	105,26	101,09	99,25	101,10	98,49	97,09
Fumo	93,83	94,74	93,12	95,93	94,23	100,71

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.14

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA COM AJUSTAMENTO SAZONAL BRASIL — 1978

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	92,46	91,93	93,42	96,43	93,93	94,13
Extrativa mineral	77,81	78,53	80,13	81,70	80,52	80,63
Indústrias de transformação	92,81	92,24	93,73	96,79	94,25	94,48
Minerais não-metálicos	89,95	90,86	90,62	95,26	92,07	93,49
Metalúrgica	93,69	92,17	95,57	9[illegible],30	96,68	97,86
Metalúrgica básica	91,53	94,83	93,32	98,11	95,88	95,78
Outros produtos metalúrgicos	98,80	89,28	100,28	101,91	99,34	102,11
Mecânica	96,61	98,02	100,49	102,21	99,77	96,39
Material elétrico e de comunicações	92,42	89,80	96,33	102,04	94,91	98,90
Material de transporte	109,77	110,09	111,40	120,12	112,02	115,25
Autoveículos	113,27	117,05	118,55	134,66	118,09	125,41
Outros produtos de transporte	97,87	92,84	92,78	95,19	94,61	88,63
Papel e papelão	81,35	81,81	82,67	84,03	82,89	85,37
Borracha	98,98	96,03	93,58	100,20	96,98	98,02
Química	92,73	91,25	93,90	92,52	93,44	92,69
Petroquímica, refino e destilação do carvão	97,41	96,49	97,72	98,89	97,88	97,02
Outros produtos químicos	86,18	87,09	88,20	85,65	86,92	87,04
Farmacêutica	80,33	79,17	83,92	80,78	78,63	78,10
Perfumaria, sabões e velas	72,95	71,00	70,05	77,15	77,14	79,00
Produtos de matérias plásticas	98,18	99,59	100,00	108,06	106,85	103,27
Têxtil	99,55	99,70	98,54	104,52	97,47	98,24
Vestuário, calçados e artefatos de tecidos	84,06	82,19	83,59	86,38	84,65	85,97
Produtos alimentares	91,92	89,89	87,99	96,00	89,73	87,09
Bebidas	95,35	99,74	102,98	105,09	102,97	109,05
Fumo	102,14	98,93	94,59	94,41	90,92	94,77

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	95,78	97,50	97,87	98,41	98,01	100,33
Extrativa mineral	83,21	83,19	83,15	82,37	82,00	79,95
Indústrias de transformação	96,10	97,85	98,27	98,81	98,39	100,84
Minerais não-metálicos	94,03	97,90	95,45	95,91	96,40	95,38
Metalúrgica	97,11	97,89	100,45	100,78	101,01	104,36
Metalúrgica básica	95,96	96,68	100,04	102,37	102,04	105,42
Outros produtos metalúrgicos	101,11	101,42	102,30	99,32	99,22	103,89
Mecânica	99,41	101,80	99,83	101,89	106,51	110,28
Material elétrico e de comunicações	102,40	102,45	102,05	99,31	100,10	105,97
Material de transporte	121,68	123,74	121,08	121,00	119,34	129,88
Autoveículos	132,05	131,01	126,26	127,88	124,07	143,27
Outros produtos de transporte	98,92	102,28	107,77	101,85	104,18	106,70
Pap l e papelão	88,68	89,26	89,65	90,34	93,19	94,78
Borracha	99,55	101,79	100,39	101,09	101,06	103,91
Química	94,70	98,00	97,71	101,99	96,43	99,96
Petroquímica, refino e destilação do carvão	94,94	101,78	99,22	103,95	96,25	101,87
Outros produtos químicos	89,94	91,48	92,49	94,66	91,05	86,10
Farmacêutica	84,21	85,79	88,16	85,98	83,04	84,84
Perfumaria, sabões e velas	79,77	79,45	80,64	80,22	83,28	84,93
Produtos de matérias plásticas	106,35	106,76	106,06	105,50	106,13	110,49
Têxtil	99,81	101,52	103,53	103,36	103,16	105,77
Vestuário, calçados e artefatos de tecidos	85,38	85,80	87,59	85,77	87,41	90,57
Produtos alimentares	87,07	88,98	90,34	91,08	89,46	90,60
Bebidas	110,28	105,92	106,77	112,19	110,85	110,23
Fumo	95,96	101,93	101,15	98,49	107,18	108,87

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.15

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA COM AJUSTAMENTO SAZONAL BRASIL — 1979

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	101,41	99,91	99,41	100,77	101,24	102,79
Extrativa mineral	88,89	69,01	84,12	85,89	88,89	94,58
Indústrias de transformação	101,70	100,65	99,78	101,14	101,54	103,01
Minerais não-metálicos	100,23	97,37	101,30	100,82	101,21	100,64
Metalúrgica	104,55	98,81	101,32	103,29	104,25	105,55
Metalúrgica básica	105,22	96,90	102,60	106,54	107,18	109,31
Outros produtos metalúrgicos	104,71	104,74	100,79	100,58	101,03	101,11
Mecânica	107,68	107,46	103,88	107,72	105,53	109,57
Material elétrico e de comunicações	105,94	110,44	106,16	100,66	106,81	105,80
Material de transporte	129,46	121,44	104,02	133,17	127,51	126,75
Autoveículos	138,88	123,86	95,48	146,74	136,64	135,52
Outros produtos de transporte	108,91	111,77	110,48	105,38	100,70	105,40
Papel e papelão	96,11	96,04	96,32	94,27	96,03	98,19
Borracha	103,44	105,25	104,26	105,89	103,33	98,54
Química	94,24	101,03	102,66	102,33	104,11	105,73
Petroquímica, refino e destilação do carvão	98,95	104,24	104,32	106,71	105,72	107,05
Outros produtos químicos	95,98	98,52	98,53	96,00	98,90	101,85
Farmacêutica	82,11	82,52	83,28	90,70	85,86	91,12
Perfumaria, sabões e velas	87,29	82,55	88,01	85,43	86,58	88,18
Produtos de matérias plásticas	107,73	109,48	108,46	107,18	106,98	108,32
Têxtil	106,35	109,16	111,98	113,68	112,52	109,03
Vestuário, calçados e artefatos de tecidos	91,09	91,41	89,99	89,87	87,86	89,59
Produtos alimentares	89,90	89,70	87,89	88,68	86,93	88,50
Bebidas	113,21	104,02	109,42	99,47	108,74	108,08
Fumo	109,97	108,20	104,75	103,90	103,54	115,07

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	101,41	104,37	101,64	106,34	104,70	105,58
Extrativa mineral	92,56	96,79	95,01	96,52	97,39	99,73
Indústrias de transformação	101,63	104,57	101,81	106,56	104,85	105,71
Minerais não-metálicos	98,23	99,65	95,28	99,48	98,38	101,13
Metalúrgica	104,42	109,41	107,88	110,83	109,73	112,39
Metalúrgica básica	109,11	112,59	112,80	114,56	112,21	116,26
Outros produtos metalúrgicos	97,09	103,62	100,74	104,94	103,51	107,86
Mecânica	105,38	110,19	107,49	116,48	110,98	113,64
Material elétrico e de comunicações	103,82	106,65	103,65	111,51	107,57	110,83
Material de transporte	120,49	129,46	124,53	129,78	129,43	133,44
Autoveículos	125,58	139,24	136,89	138,09	134,29	146,87
Outros produtos de transporte	102,56	103,94	102,60	107,33	111,09	109,87
Papel e papelão	95,90	99,51	101,37	104,69	101,16	102,26
Borracha	105,13	107,34	108,93	110,63	112,92	112,29
Química	106,98	105,94	103,00	106,59	105,00	104,25
Petroquímica, refino e destilação do carvão	101,81	106,42	104,04	105,56	106,53	104,89
Outros produtos químicos	103,75	100,09	98,45	99,73	97,21	100,02
Farmacêutica	86,56	89,70	89,74	89,57	87,46	88,06
Perfumaria, sabões e velas	90,75	93,56	90,54	92,25	93,83	95,54
Produtos de matérias plásticas	105,54	113,88	113,15	118,20	119,16	119,22
Têxtil	109,17	109,15	107,21	109,20	108,02	107,62
Vestuário, calçados e artefatos de tecidos	89,38	90,09	89,15	92,19	91,30	90,02
Produtos alimentares	88,14	89,15	87,67	91,43	92,33	89,67
Bebidas	107,00	113,41	110,56	119,67	117,78	114,43
Fumo	112,73	110,92	96,67	104,22	98,02	101,25

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.16

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA COM AJUSTAMENTO SAZONAL BRASIL — 1980

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	107,52	108,30	111,32	106,35	111,72	112,11
Extrativa mineral	97,63	104,34	99,89	102,81	102,92	103,76
Indústria de transformação	107,73	108,34	111,59	106,43	111,93	112,36
Minerais não-metálicos	101,44	102,53	104,04	103,58	103,12	106,87
Metalúrgica	113,03	113,56	116,90	112,25	119,37	120,14
Metalúrgica básica	116,93	118,28	120,86	114,96	123,86	124,69
Outros produtos metalúrgicos	109,82	108,49	112,83	110,54	113,57	114,83
Mecânica	119,92	119,91	131,19	107,15	124,90	126,43
Material elétrico e de comunicações	110,40	108,66	113,73	116,96	117,37	116,28
Material de transporte	135,07	130,90	131,69	99,72	128,87	130,78
Autoveículos	145,07	136,60	141,09	84,39	126,59	134,95
Outros produtos de transporte	115,26	117,76	113,32	116,88	116,45	116,47
Papel e papelão	104,80	107,10	108,02	110,60	110,95	110,77
Borracha	113,56	114,09	112,79	109,86	111,47	114,33
Química	105,10	107,55	111,34	112,07	115,20	118,41
Petroquímica, refino e destilação do carvão	101,44	95,99	106,22	104,69	109,42	104,19
Outros produtos químicos	104,93	115,09	111,81	113,21	113,30	115,28
Farmacêutica	94,36	88,74	92,20	93,91	92,63	101,43
Perfumaria, sabões e velas	99,64	100,06	95,86	90,17	92,54	94,94
Produtos de matérias plásticas	122,93	123,00	126,51	127,57	125,80	129,85
Têxtil	110,53	117,88	122,68	122,82	120,30	114,14
Vestuário, calçados e artefatos de tecidos	96,84	94,31	95,12	97,04	95,44	97,95
Produtos alimentares	93,15	91,56	97,15	94,36	97,68	97,29
Bebidas	113,22	111,92	110,46	111,61	109,37	108,62
Fumo	101,74	99,81	104,20	100,82	101,65	97,51

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	112,97	112,22	116,70	113,70	114,71	110,06
Extrativa mineral	105,88	107,91	106,57	102,35	101,95	95,06
Indústrias d transformação	113,14	112,36	116,97	113,95	115,06	110,41
Minerais não-metálicos	108,73	108,54	110,96	109,64	112,09	109,68
Metalúrgica	121,15	120,00	123,61	122,14	125,27	118,26
Metalúrgica básica	122,76	122,90	123,16	124,27	127,33	123,40
Outros produtos metalúrgicos	118,91	114,90	124,28	119,03	120,72	110,82
Mecânica	128,12	123,17	131,33	126,82	132,30	116,77
Material elétrico e de comunicações	120,49	120,77	130,63	127,84	126,52	119,64
Material de transporte	131,06	129,07	149,26	134,23	136,49	132,46
Autoveículos	135,47	137,72	165,48	147,14	152,09	140,98
Outros produtos de transporte	110,75	109,07	112,83	104,42	110,57	113,46
Papel e papelão	110,31	111,17	111,03	111,55	109,25	104,12
Borracha	114,60	115,56	124,96	123,24	124,04	119,55
Química	111,13	108,33	108,85	101,96	106,43	107,97
Petroquímica, refino e destilação do carvão	105,04	99,91	103,05	98,43	93,60	100,19
Outros produtos químicos	108,19	106,01	104,49	95,94	106,27	109,43
Farmacêutica	95,56	96,82	104,01	99,27	105,27	103,63
Perfumaria, sabões e velas	104,87	98,76	100,71	98,73	96,37	96,13
Produtos de matérias plásticas	125,56	127,85	132,78	127,40	131,01	125,60
Têxtil	114,73	113,00	116,22	115,50	116,40	110,81
Vestuário, calçados e artefatos de tecidos	98,60	100,29	105,48	102,37	107,82	98,26
Produtos alimentares	97,33	96,96	100,25	101,03	96,84	90,99
Bebidas	110,80	117,35	118,23	115,19	115,35	108,43
Fumo	97,23	98,29	105,42	106,64	107,23	97,13

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.17

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA COM AJUSTAMENTO SAZONAL BRASIL — 1981

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	111,29	114,30	102,72	101,35	99,54	98,87
Extrativa mineral	98,51	105,97	100,44	97,16	98,90	98,10
Indústrias de transformação	111,62	114,51	102,77	101,45	99,56	98,89
Minerais não-metálicos	108,27	109,72	102,63	100,43	94,92	97,42
Metalúrgica	118,14	118,60	110,47	104,80	99,59	97,78
Metalúrgica básica	122,18	121,76	114,24	106,64	99,68	97,02
Outros produtos metalúrgicos	112,22	113,97	104,95	102,11	99,46	98,88
Mecânica	119,78	123,92	108,58	103,58	102,79	100,13
Material elétrico e de comunicações	117,06	123,12	97,22	100,44	99,78	101,15
Material de transporte	126,34	132,05	109,60	106,05	106,79	101,08
Autoveículos	132,20	136,37	111,52	106,76	109,83	102,67
Outros produtos de transporte	110,63	120,48	104,46	104,14	98,62	96,79
Papel e papelão	107,78	109,54	102,71	95,76	96,84	95,25
Borracha	118,16	118,38	116,58	105,47	101,79	99,23
Química	108,27	107,80	101,66	101,92	97,43	96,26
Petroquímica, refino e destilação do carvão	103,23	107,70	99,35	96,98	96,15	95,53
Outros produtos químicos	111,22	107,86	103,01	104,81	98,18	96,70
Farmacêutica	113,64	106,32	86,19	98,99	99,72	99,84
Perfumaria, sabões e velas	98,86	100,33	95,39	96,35	105,38	96,93
Produtos de matérias plásticas	117,23	120,57	103,35	96,12	92,90	94,15
Têxtil	110,19	109,46	91,98	92,38	92,99	98,06
Vestuário, calçados e artefatos de tecidos	102,95	112,31	98,40	98,79	98,57	98,41
Produtos alimentares	95,64	100,61	97,46	100,26	99,63	101,85
Bebidas	105,44	110,54	107,98	110,08	110,24	102,25
Fumo	90,46	115,89	103,29	107,57	101,64	97,27

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	99,26	96,16	94,67	94,88	95,60	94,12
Extrativa mineral	94,44	97,81	99,90	103,64	103,91	102,65
Indústrias de transformação	99,38	96,12	94,53	94,65	95,39	93,91
Minerais não-metálicos	99,00	97,94	97,99	98,68	98,25	96,68
Metalúrgica	96,66	91,82	88,75	91,05	93,28	91,78
Metalúrgica básica	95,20	90,45	85,20	89,16	90,73	90,35
Outros produtos metalúrgicos	98,78	93,83	93,95	93,81	97,02	93,87
Mecânica	97,73	92,38	90,15	90,78	89,00	87,49
Material elétrico e de comunicações	100,11	94,83	94,18	92,84	95,28	91,86
Material de transporte	99,76	87,81	82,53	86,36	85,13	78,52
Autoveículos	98,65	83,77	78,26	83,09	81,15	77,41
Outros produtos de transporte	102,74	98,66	94,00	95,11	95,81	81,51
Papel e papelão	97,15	97,69	97,10	97,64	100,78	103,02
Borracha	97,21	95,70	77,79	88,50	89,45	94,45
Química	98,45	97,19	97,01	98,10	99,05	101,74
Petroquímica, refino e destilação do carvão	105,19	100,33	98,17	101,14	100,05	98,29
Outros produtos químicos	94,51	95,35	96,33	96,33	98,46	103,76
Farmacêutica	105,77	102,70	99,27	98,90	98,42	92,86
Perfumaria, sabões e velas	97,54	96,02	100,74	105,30	106,93	99,96
Produtos de matérias plásticas	96,44	93,84	94,43	97,16	98,93	99,26
Têxtil	103,28	99,59	100,36	99,39	100,79	102,05
Vestuário, calçados e artefatos de tecidos	98,44	98,39	97,49	99,57	101,95	96,56
Produtos alimentares	103,66	105,23	104,76	96,78	97,79	94,15
Bebidas	96,04	94,09	92,75	92,23	88,85	94,90
Fumo	97,10	93,39	97,67	94,58	96,67	96,62

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.18

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA COM AJUSTAMENTO SAZONAL BRASIL — 1982

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	95,84	96,64	100,77	100,97	101,41	101,68
Extrativa mineral	105,76	107,41	105,69	107,49	105,48	105,35
Indústrias de transformação	95,59	96,36	100,65	100,80	101,31	101,59
Minerais não-metálicos	98,19	98,86	100,96	103,44	101,71	100,38
Metalúrgica	95,97	96,67	100,32	99,61	100,38	99,38
Metalúrgica básica	95,02	96,16	97,63	98,41	99,02	98,84
Outros produtos metalúrgicos	97,36	97,42	104,27	101,37	102,38	100,17
Mecânica	87,79	86,98	88,29	86,48	86,76	83,68
Material elétrico e de comunicações	97,56	93,29	106,53	107,16	107,10	103,07
Material de transporte	87,79	87,88	98,95	102,27	91,93	96,30
Autoveículos	86,32	86,80	97,11	102,86	89,40	96,21
Outros produtos de transporte	91,75	90,75	103,91	100,68	98,70	96,55
Papel e papelão	102,28	105,54	106,10	108,15	109,31	109,88
Borracha	104,90	90,14	92,43	98,50	99,01	98,06
Química	101,06	103,42	105,01	105,80	111,60	116,37
Petroquímica, refino e destilação do carvão	102,22	99,82	98,26	102,73	101,88	104,37
Outros produtos químicos	100,38	105,53	109,11	107,59	117,27	123,38
Farmacêutica	96,65	103,33	106,14	95,26	100,95	94,33
Perfumaria, sabões e velas	98,40	99,01	110,49	105,53	107,84	104,70
Produtos de matérias plásticas	103,28	107,24	111,09	117,06	114,42	114,35
Têxtil	102,36	102,15	102,29	103,20	106,25	108,43
Vestuário, calçados e artefatos de tecidos	98,80	100,26	106,46	104,51	107,70	106,49
Produtos alimentares	89,42	93,41	97,52	96,69	97,89	98,95
Bebidas	93,61	93,95	94,83	101,35	95,25	97,59
Fumo	95,11	100,38	111,55	99,10	102,39	104,91

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	102,57	102,50	101,52	98,62	97,23	95,51
Extrativa mineral	108,23	103,78	95,27	108,08	115,35	110,46
Indústrias de transformação	102,42	102,46	101,68	98,38	96,76	95,12
Minerais não-metálicos	98,10	98,01	96,47	94,24	91,47	88,95
Metalúrgica	100,43	101,27	98,95	97,34	94,11	88,46
Metalúrgica básica	101,60	102,14	99,21	96,09	92,77	84,26
Outros produtos metalúrgicos	98,72	99,99	98,58	99,16	96,07	94,60
Mecânica	86,98	87,83	82,21	78,62	77,53	75,34
Material elétrico e de comunicações	101,58	106,46	100,35	99,72	98,82	97,83
Material de transporte	99,17	101,29	101,97	99,42	97,09	93,61
Autoveículos	100,16	102,51	104,39	101,29	101,06	96,76
Outros produtos de transporte	96,53	98,02	95,48	94,41	86,44	85,15
Papel e papelão	109,95	108,58	107,44	107,43	107,06	106,00
Borracha	94,06	91,75	85,82	88,20	87,25	86,42
Química	116,47	113,49	114,93	106,52	104,18	97,54
Petroquímica, refino e destilação do carvão	103,77	102,74	104,33	103,09	102,48	108,67
Outros produtos químicos	123,89	119,77	121,12	108,53	105,17	91,03
Farmacêutica	98,51	97,26	96,78	102,15	100,80	113,40
Perfumaria, sabões e velas	100,84	99,71	102,29	96,05	99,82	101,44
Produtos de matérias plásticas	113,63	114,50	107,66	104,05	100,97	97,95
Têxtil	110,15	108,67	104,37	102,30	100,06	102,22
Vestuário, calçados e artefatos de tecidos	107,44	105,37	104,54	101,49	97,45	97,52
Produtos alimentares	100,00	101,26	109,81	104,46	104,64	107,85
Bebidas	101,89	94,69	98,24	96,94	96,61	96,18
Fumo	102,22	96,83	97,77	90,41	104,07	114,12

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.19

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA COM AJUSTAMENTO SAZONAL BRASIL — 1983

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Janeiro	Fevereiro	Março	Abril	Maio	Junho
Indústria geral	93,37	90,85	95,38	93,35	94,35	93,04
Extrativa mineral	114,26	110,87	115,01	112,81	117,44	120,30
Indústrias de transformação	92,84	91,36	94,88	92,86	93,77	92,35
Minerais não-metálicos	87,13	85,19	83,24	82,24	83,18	80,95
Metalúrgica	88,88	38,54	94,42	96,35	95,82	95,22
Metalúrgica básica	84,68	84,41	89,61	96,95	95,22	95,48
Outros produtos metalúrgicos	95,02	94,59	101,47	95,47	96,70	94,83
Mecânica	72,89	70,47	75,34	73,35	76,52	73,36
Material elétrico e de comunicações	92,64	93,66	100,08	94,21	90,55	90,37
Material de transporte	89,75	84,38	92,02	92,63	92,81	87,95
Autoveículos	93,20	89,22	96,52	100,72	100,06	94,32
Outros produtos de transporte	80,50	71,41	79,94	70,94	73,36	70,85
Papel e papelão	106,97	101,36	109,77	108,71	111,07	110,66
Borracha	85,23	88,61	93,71	95,82	98,17	90,66
Química	93,26	96,83	99,50	98,97	98,82	98,48
Petroquímica, refino e destilação do carvão	99,13	100,69	99,00	98,50	101,15	97,98
Outros produtos químicos	89,83	94,58	99,79	99,25	97,46	98,76
Farmacêutica	99,19	95,70	102,52	98,55	93,73	91,84
Perfumaria, sabões e velas	98,04	99,77	104,96	105,42	98,43	101,70
Produtos de matérias plásticas	100,96	99,92	101,03	99,94	99,43	96,85
Têxtil	100,20	92,95	94,82	94,99	96,14	94,09
Vestuário, calçados e artefatos de tecidos	96,10	91,39	99,25	94,25	92,63	91,35
Produtos alimentares	107,43	106,10	103,62	95,13	102,25	103,92
Bebidas	97,39	95,41	93,66	89,00	89,42	89,18
Fumo	106,03	104,48	107,38	103,87	104,95	94,38

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Indústria geral	91,78	94,90	95,08	96,17	96,62	98,36
Extrativa mineral	121,66	121,31	127,72	125,14	135,39	143,58
Indústrias de transformação	91,02	94,23	94,25	95,44	95,63	97,20
Minerais não-metálicos	79,89	82,20	78,47	78,77	79,65	82,76
Metalúrgica	95,44	98,13	98,78	99,88	100,53	106,61
Metalúrgica básica	97,28	100,08	102,96	104,03	105,74	111,32
Outros produtos metalúrgicos	92,75	95,29	92,66	93,81	92,90	99,71
Mecânica	72,28	75,21	77,29	77,70	80,11	84,19
Material elétrico e de comunicações	88,76	87,27	87,86	88,88	83,04	87,95
Material de transporte	82,82	88,87	90,61	83,44	84,46	94,07
Autoveículos	88,45	96,46	96,33	88,51	89,64	103,00
Outros produtos de transporte	67,75	68,48	75,26	69,84	70,57	70,11
Papel e papelão	107,17	110,99	113,74	115,24	114,71	114,05
Borracha	97,34	96,26	97,70	103,49	85,19	87,36
Química	99,17	105,42	107,06	111,63	109,77	103,54
Petroquímica, refino e destilação do carvão	92,03	97,33	97,40	105,62	97,70	97,83
Outros produtos químicos	103,34	110,14	112,71	115,14	116,82	106,88
Farmacêutica	87,97	91,02	88,85	88,16	93,10	96,14
Perfumaria, sabões e velas	100,16	103,55	98,45	102,73	102,47	105,52
Produtos de matérias plásticas	94,41	91,77	94,85	94,61	94,00	94,62
Têxtil	93,56	96,21	95,40	94,89	92,23	88,19
Vestuário, calçados e artefatos de tecidos	84,53	91,30	90,75	89,90	91,07	98,20
Produtos alimentares	103,56	105,28	103,41	108,67	114,09	110,21
Bebidas	89,78	90,26	80,44	86,10	90,26	91,77
Fumo	80,28	98,56	93,41	93,16	94,04	95,47

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

TABELA 5.20

INDICADORES DA PRODUÇÃO INDUSTRIAL POR CLASSES E GÊNEROS DE INDÚSTRIA SÉRIE ENCADEADA COM AJUSTAMENTO SAZONAL BRASIL — 1984

CLASSES E GÊNEROS	ÍNDICE BASE FIXA (Base: média de 1981 = 100)						
	Janeiro	Fevereiro	Março	Abril	Maio	Junho	Julho
Indústria geral	95,92	100,19	94,57	97,05	98,75	101,36	100,47
Extrativa mineral	145,20	157,20	143,15	148,73	149,66	148,97	155,15
Indústrias de transformação	94,67	98,74	93,33	95,73	97,45	100,14	99,07
Minerais não-metálicos	81,29	82,85	78,22	76,72	79,41	80,08	81,11
Metalúrgica	101,22	105,63	102,44	105,46	104,74	106,40	105,97
Metalúrgica básica	108,58	111,61	110,95	113,44	114,33	116,71	113,90
Outros produtos metalúrgicos	90,43	96,89	89,98	93,77	90,71	91,31	94,36
Mecânica	77,13	88,81	81,10	88,76	85,44	89,43	86,06
Material elétrico e de comunicações	91,67	93,24	89,18	85,01	88,20	90,25	90,12
Material de transporte	89,54	95,45	88,92	88,01	94,92	94,16	95,91
Autoveículos	98,09	103,94	95,13	94,10	101,63	100,27	102,90
Outros produtos de transporte	66,62	72,66	72,26	71,66	76,91	77,78	77,16
Papel e papelão	113,91	114,11	112,46	115,24	113,93	116,64	115,17
Borracha	94,51	103,53	95,28	94,51	96,12	102,83	100,34
Química	102,95	109,47	107,78	110,48	113,21	113,44	112,88
Petroquímica, refino e destilação do carvão	103,30	105,38	109,05	108,78	104,07	108,84	107,01
Outros produtos químicos	102,75	111,85	107,03	111,47	118,55	116,12	116,31
Farmacêutica	103,08	95,48	91,00	90,71	99,22	94,69	93,62
Perfumaria, sabões e velas	102,05	101,59	94,25	94,15	95,84	94,02	103,71
Produtos de matérias plásticas	93,78	101,76	88,96	94,56	95,43	94,85	94,10
Têxtil	88,45	91,37	84,28	90,08	91,70	93,10	94,67
Vestuário, calçados e artefatos de tecidos	95,39	98,82	90,72	95,54	95,74	95,29	94,24
Produtos alimentares	103,87	102,42	94,05	99,28	99,78	113,92	109,27
Bebidas	92,28	93,35	92,49	75,38	90,37	92,68	88,38
Fumo	93,11	100,06	99,43	103,86	112,22	101,69	90,82

FONTE — IBGE, Departamento de Indicadores Conjunturais (DEICO).

6 — ANEXOS

Anexo 1

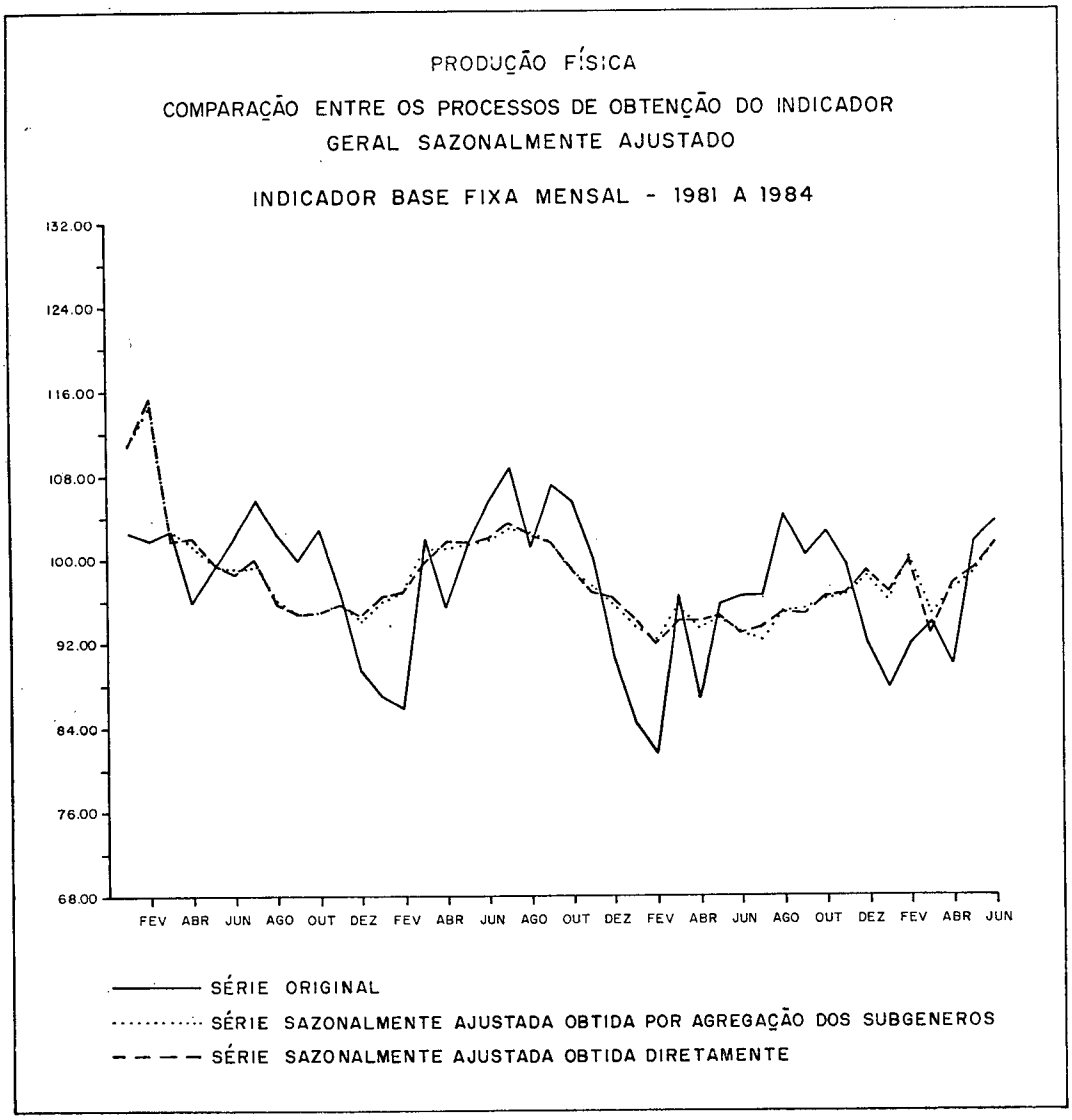

GRÁFICO 1.1

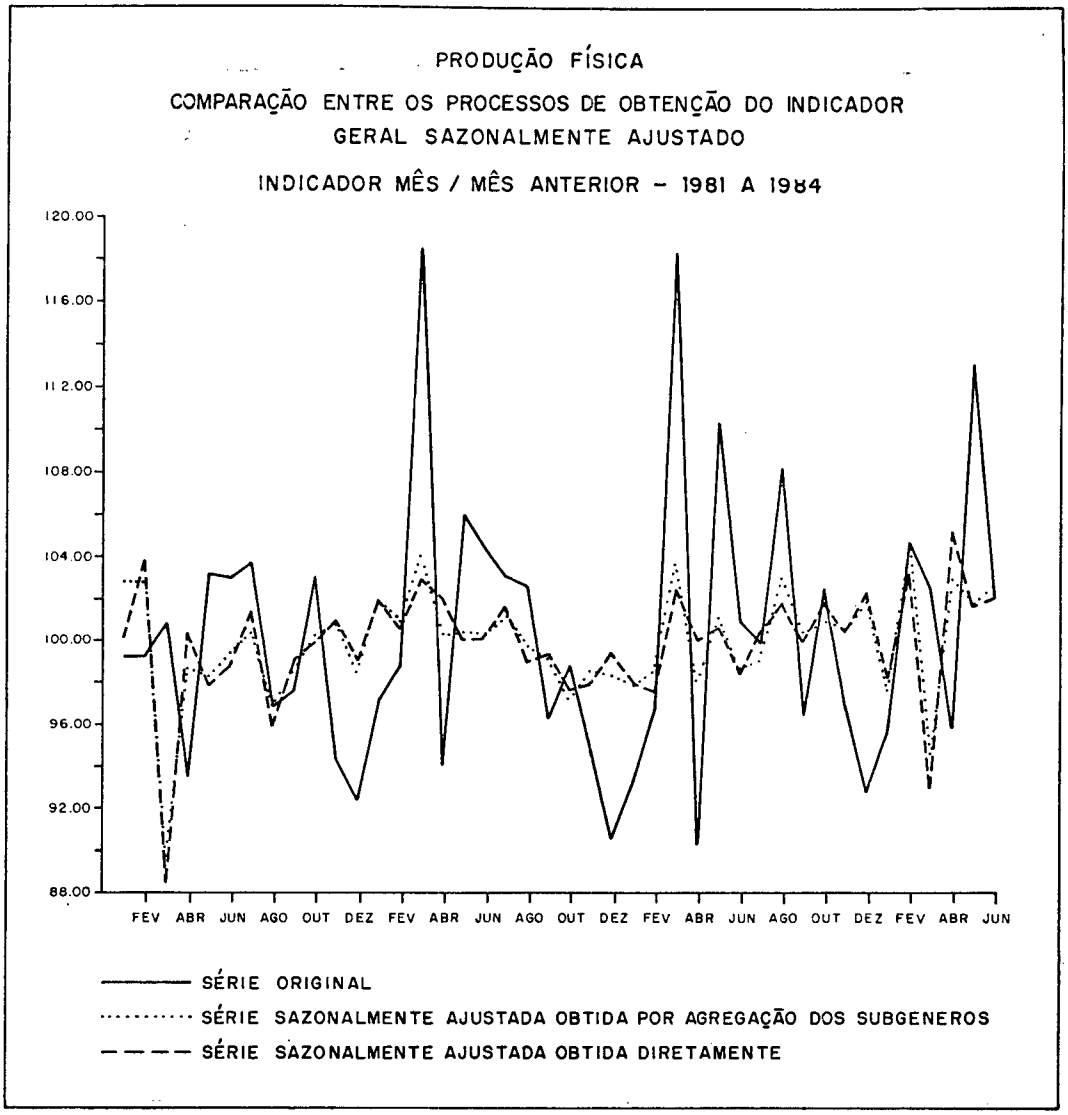
PRODUÇÃO FÍSICA
COMPARAÇÃO ENTRE OS PROCESSOS DE OBTENÇÃO DO INDICADOR GERAL SAZONALMENTE AJUSTADO
INDICADOR MÊS / MÊS ANTERIOR - 1981 A 1984
120.00
116.00
112.00
108.00
104.00
100.00
96.00
92.00
88.00
FEV ABR JUN AGO OUT DEZ FEV ABR JUN AGO OUT DEZ FEV ABR JUN AGO OUT DEZ FEV ABR JUN
SÉRIE ORIGINAL
SÉRIE SAZONALMENTE AJUSTADA OBTIDA POR AGREGAÇÃO DOS SUBGENEROS
SÉRIE SAZONALMENTE AJUSTADA OBTIDA DIRETAMENTE

GRAFICO 1.2

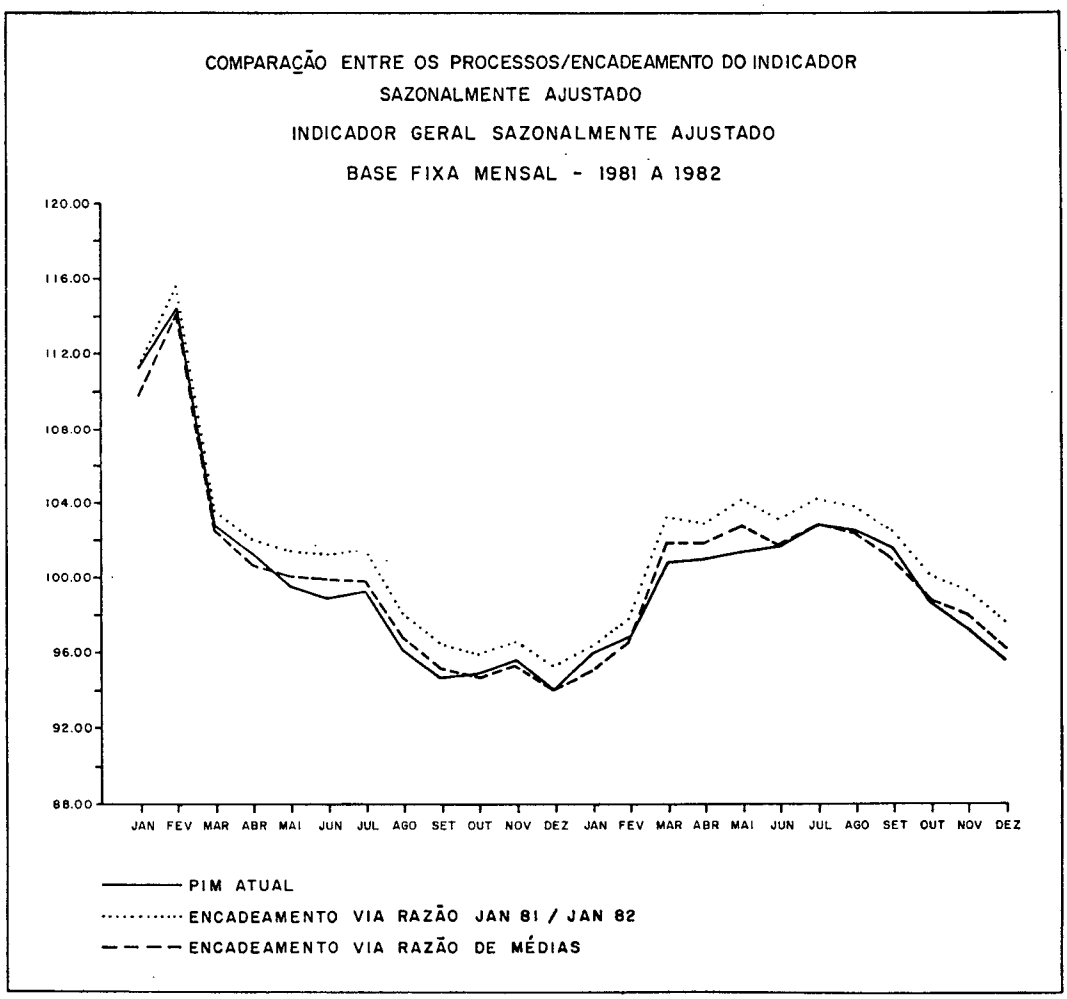
COMPARAÇÃO ENTRE OS PROCESSOS/ENCADEAMENTO DO INDICADOR
SAZONALMENTE AJUSTADO
INDICADOR GERAL SAZONALMENTE AJUSTADO
BASE FIXA MENSAL - 1981 A 1982
120.00
116.00
112.00
108.00
104.00
100.00
96.00
92.00
88.00
JAN FEV MAR ABR MAI JUN JUL AGO SET OUT NOV DEZ JAN FEV MAR ABR MAI JUN JUL AGO SET OUT NOV DEZ
PIM ATUAL
ENCADEAMENTO VIA RAZÃO JAN 81 / JAN 82
ENCADEAMENTO VIA RAZÃO DE MÉDIAS

GRÁFICO 1.3

Anexo 2

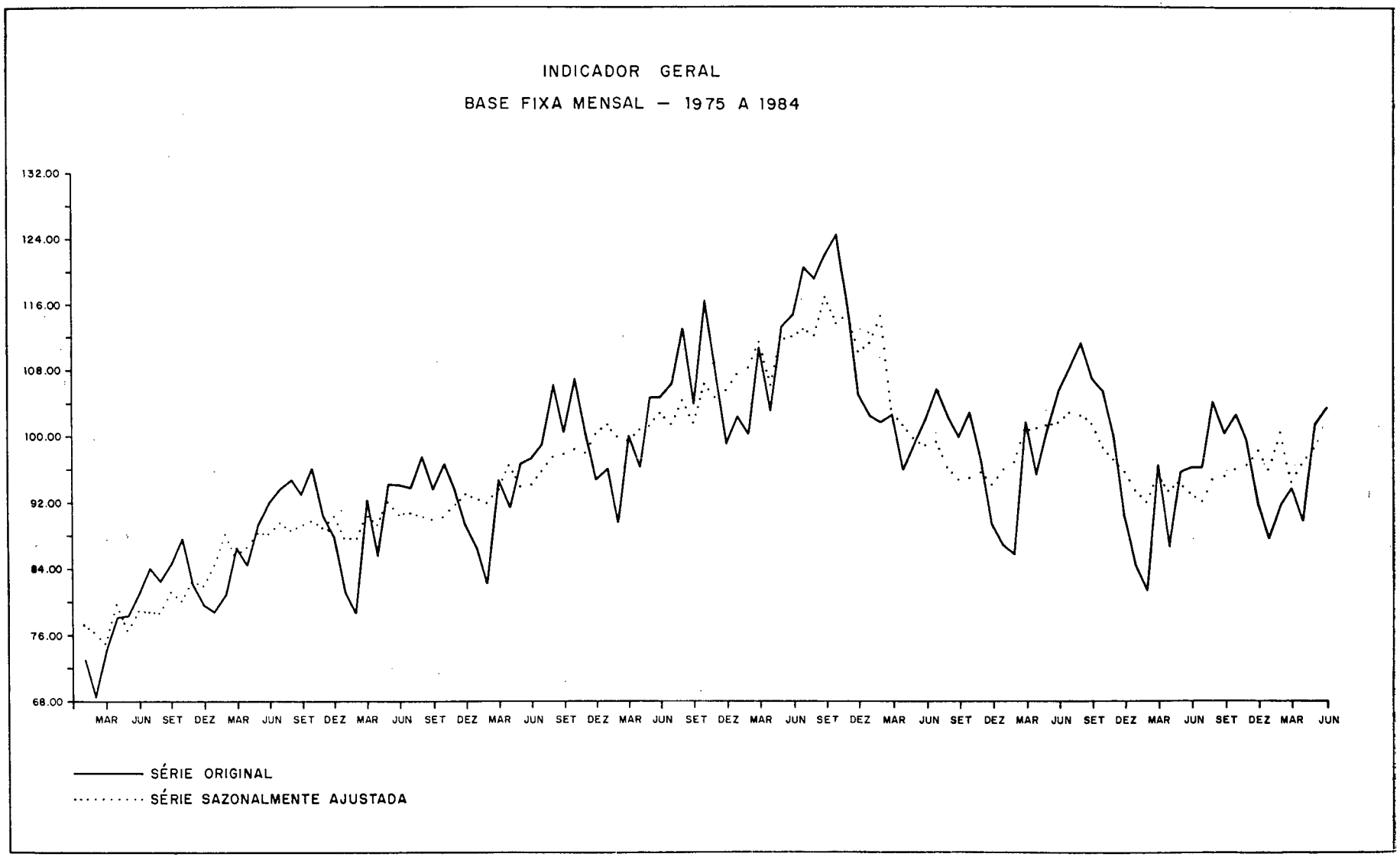

GRÁFICO 2.1

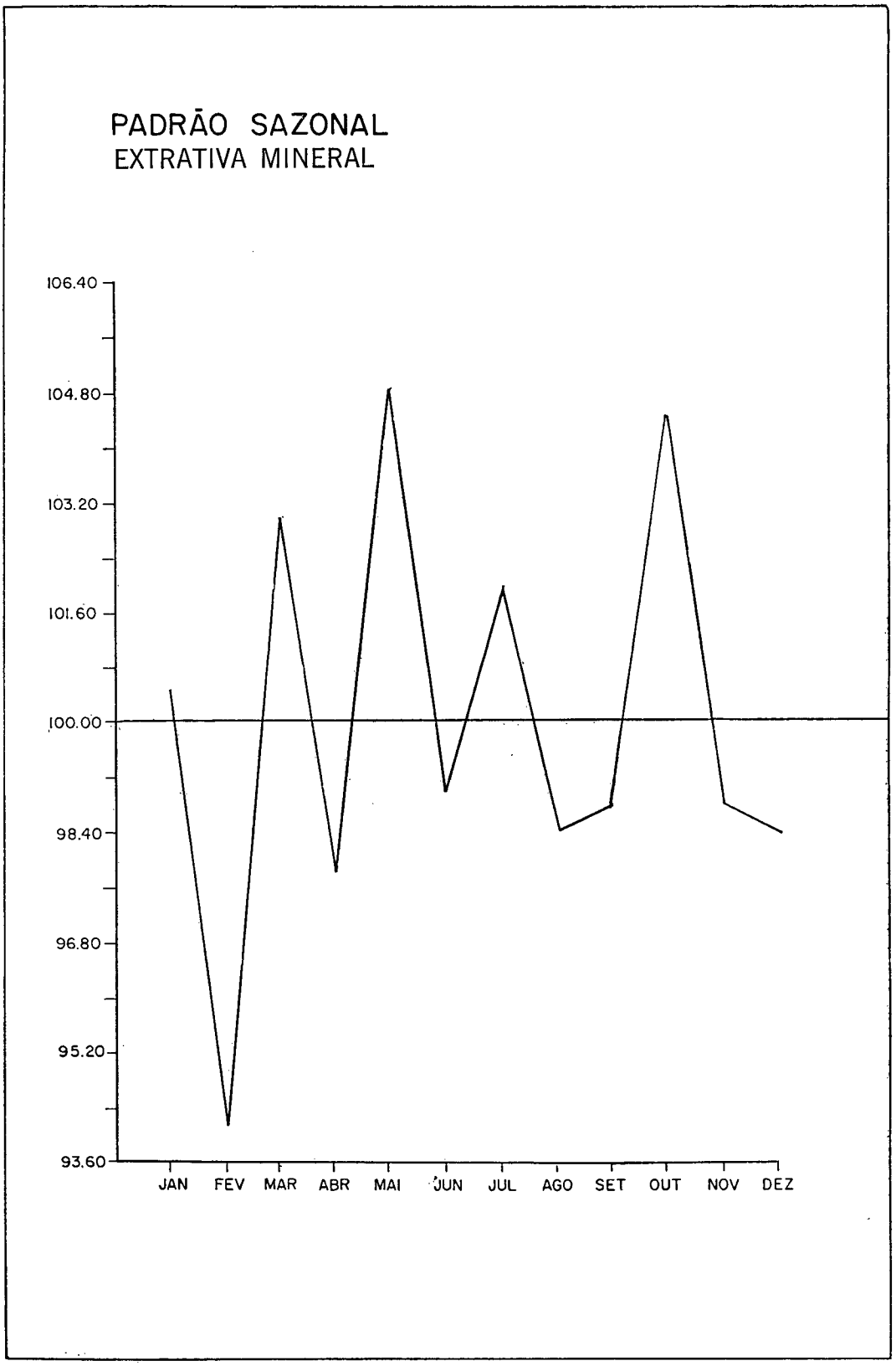
PADRÃO SAZONAL
EXTRATIVA MINERAL
106.40
104.80
103.20
101.60
100.00
98.40
96.80
95.20
93.60
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRAFICO 3.1

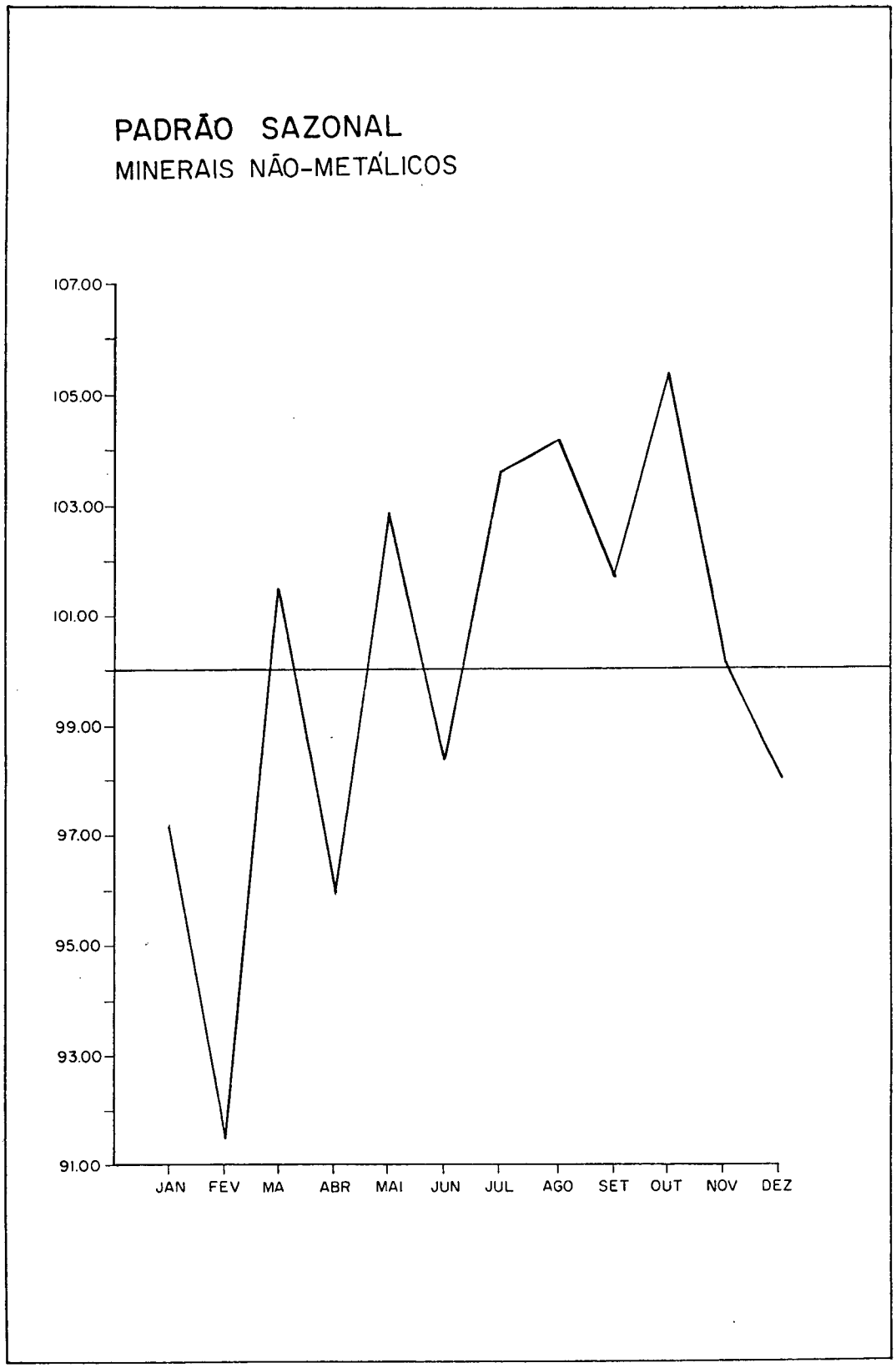
PADRÃO SAZONAL
MINERAIS NÃO-METÁLICOS
107.00
105.00
103.00
101.00
99.00
97.00
95.00
93.00
91.00
JAN
FEV
MA
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.2

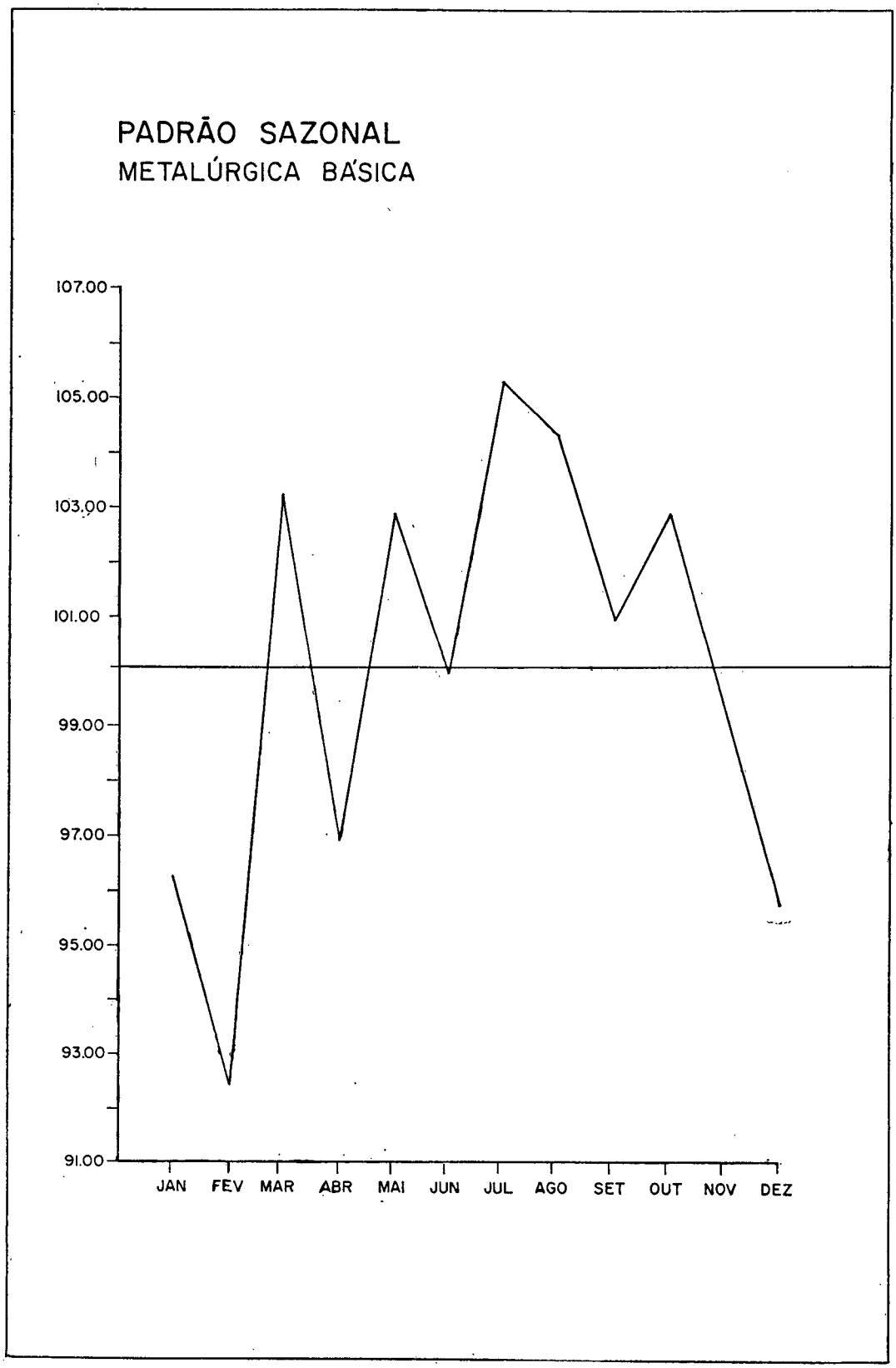
PADRÃO SAZONAL
METALÚRGICA BÁSICA
107.00
105.00
103.00
101.00
99.00
97.00
95.00
93.00
91.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.3

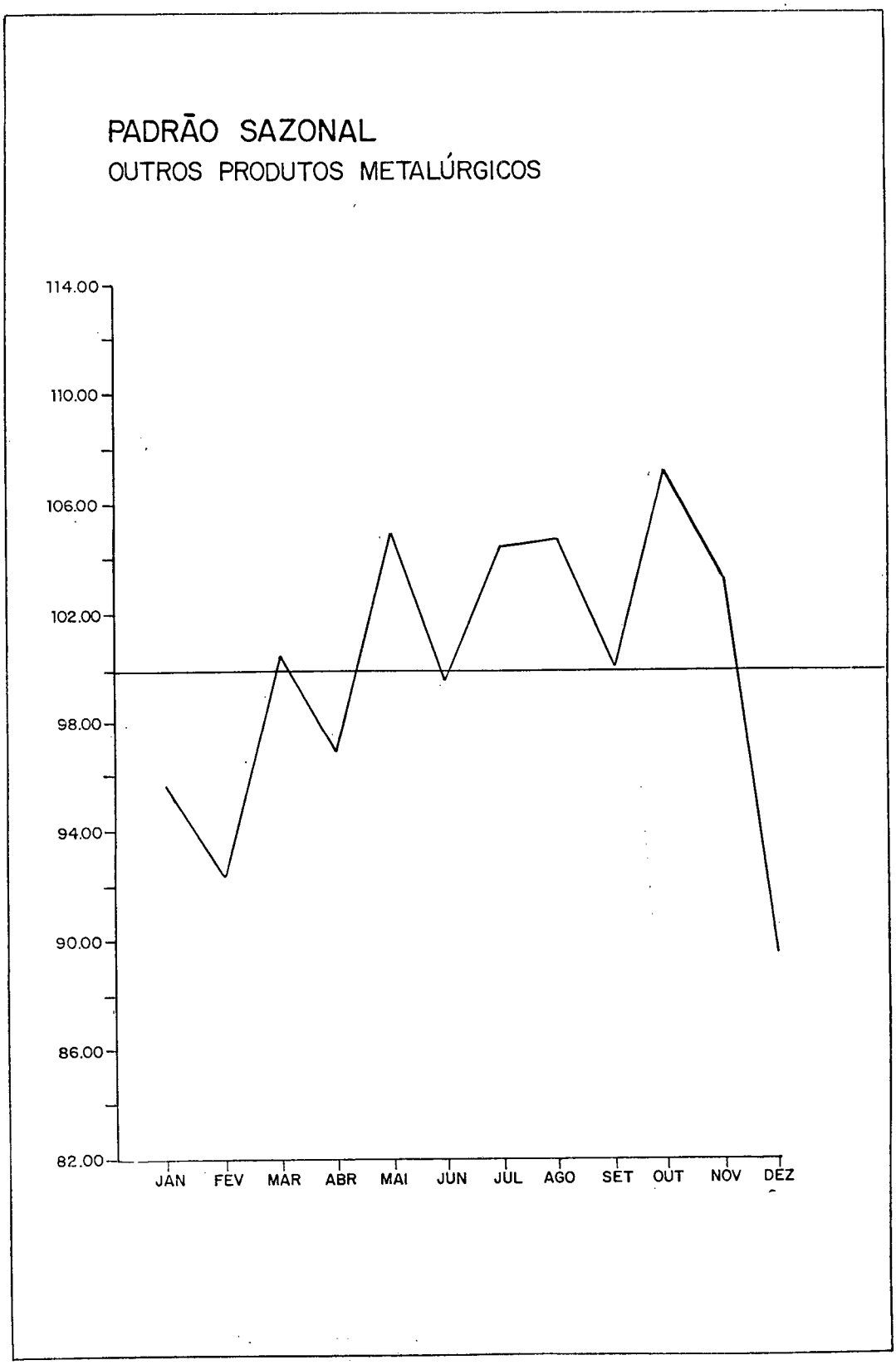
PADRÃO SAZONAL
OUTROS PRODUTOS METALÚRGICOS
114.00
110.00
106.00
102.00
98.00
94.00
90.00
86.00
82.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.4

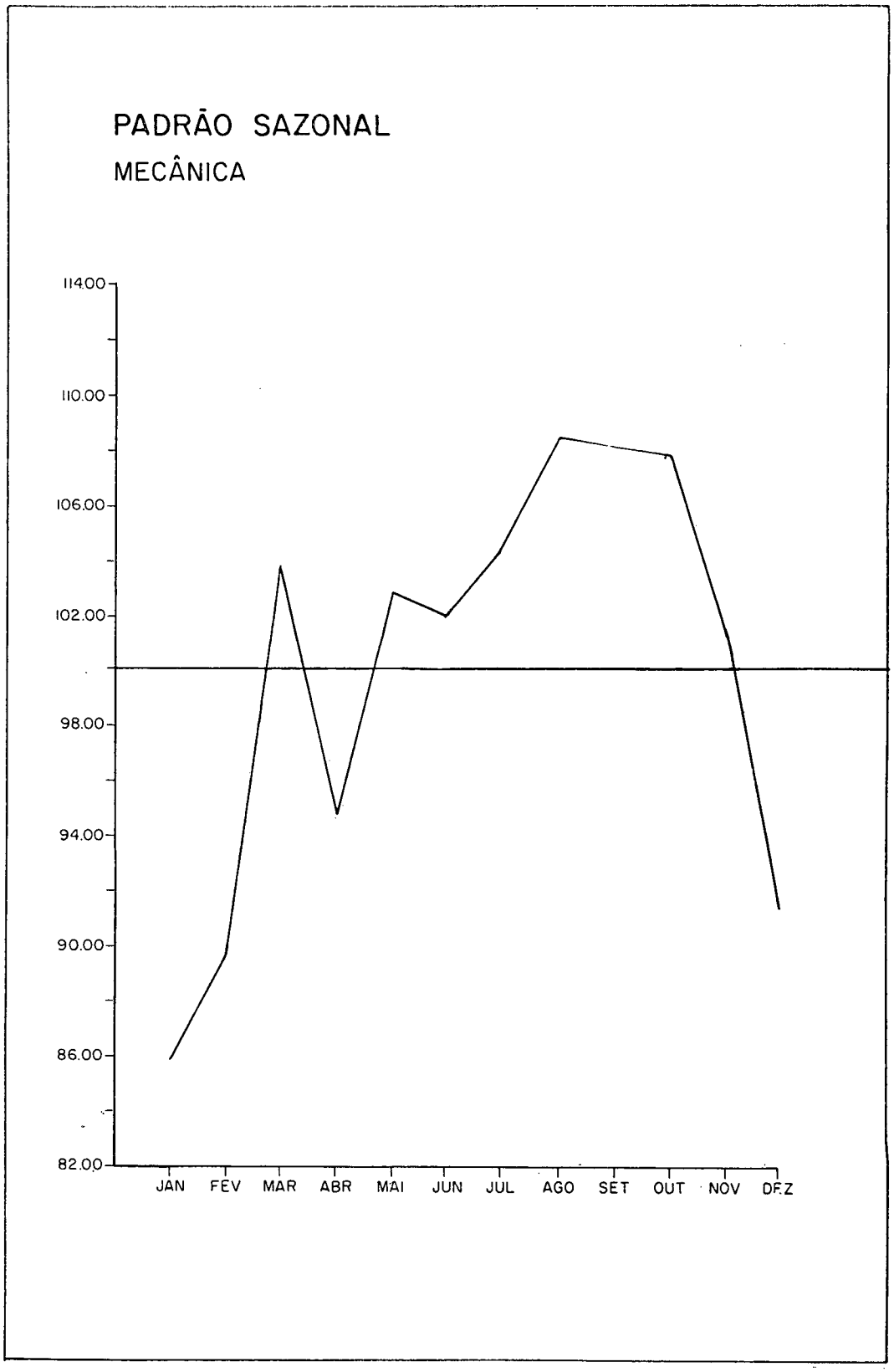
PADRÃO SAZONAL
MECÂNICA
114.00
110.00
106.00
102.00
98.00
94.00
90.00
86.00
82.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.5

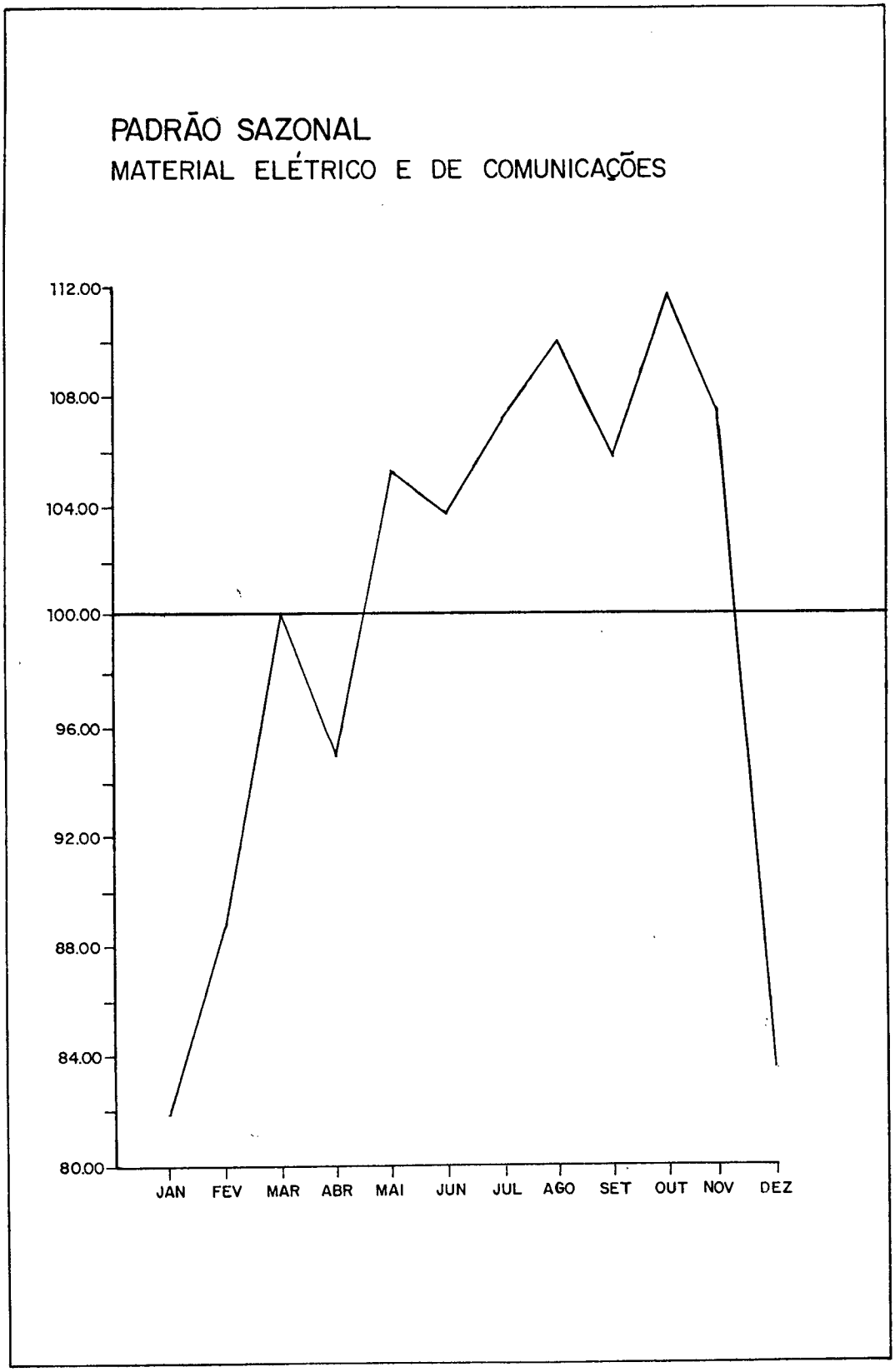
PADRÃO SAZONAL
MATERIAL ELÉTRICO E DE COMUNICAÇÕES
112.00
108.00
104.00
100.00
96.00
92.00
88.00
84.00
80.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.6

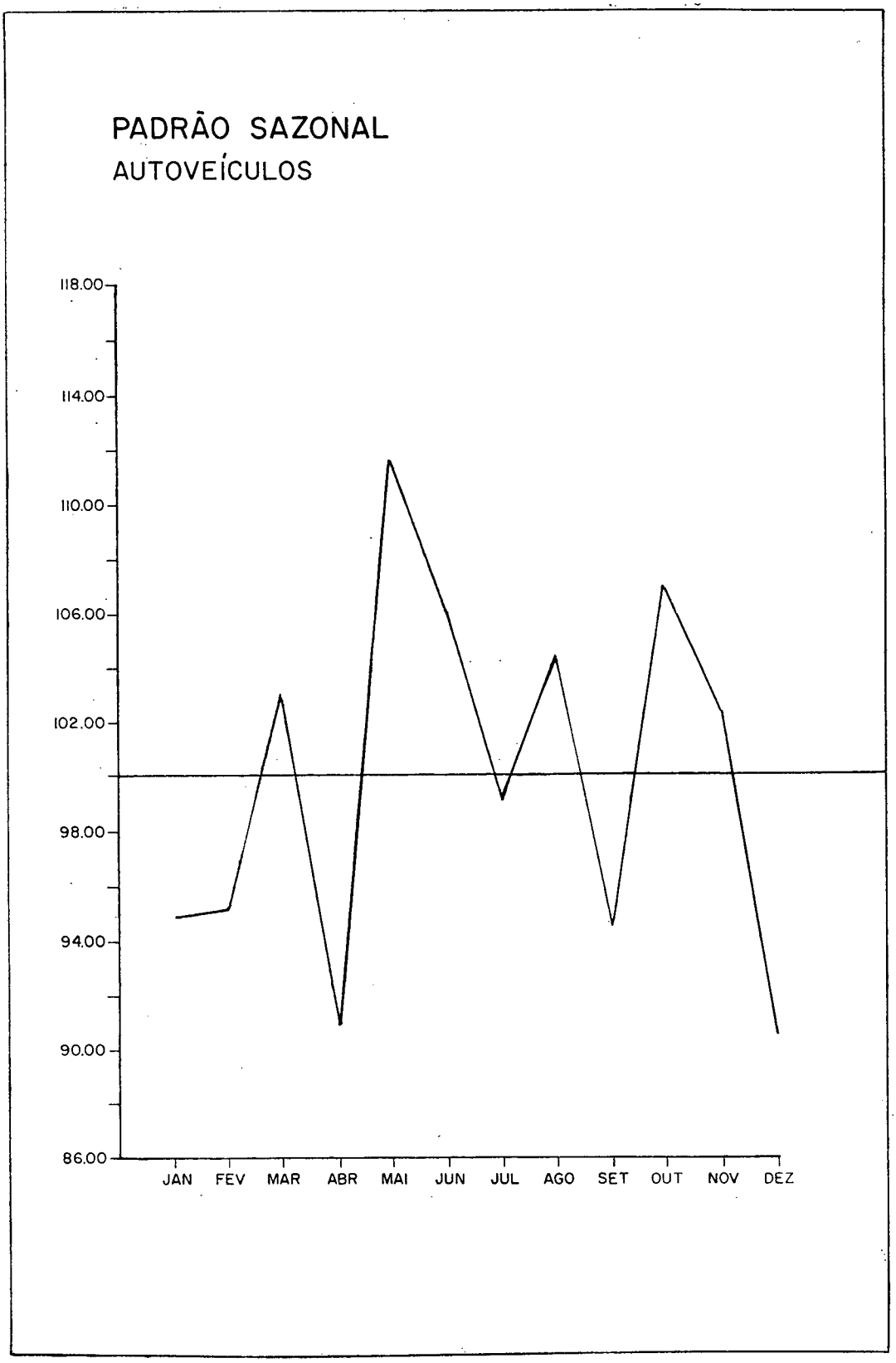
PADRÃO SAZONAL
AUTOVEÍCULOS
118.00
114.00
110.00
106.00
102.00
98.00
94.00
90.00
86.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.7

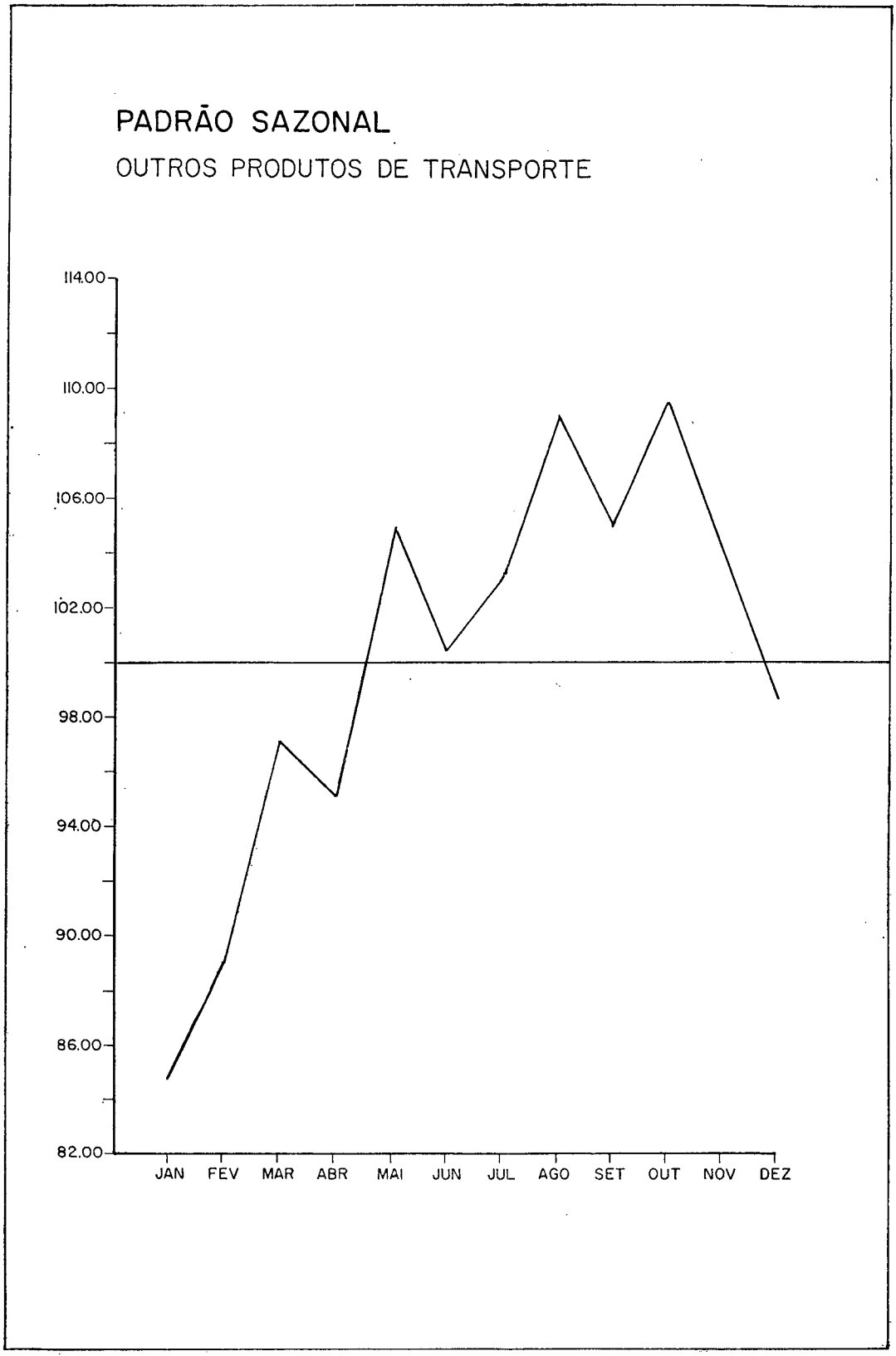
PADRÃO SAZONAL
OUTROS PRODUTOS DE TRANSPORTE
114.00
110.00
106.00
102.00
98.00
94.00
90.00
86.00
82.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.8

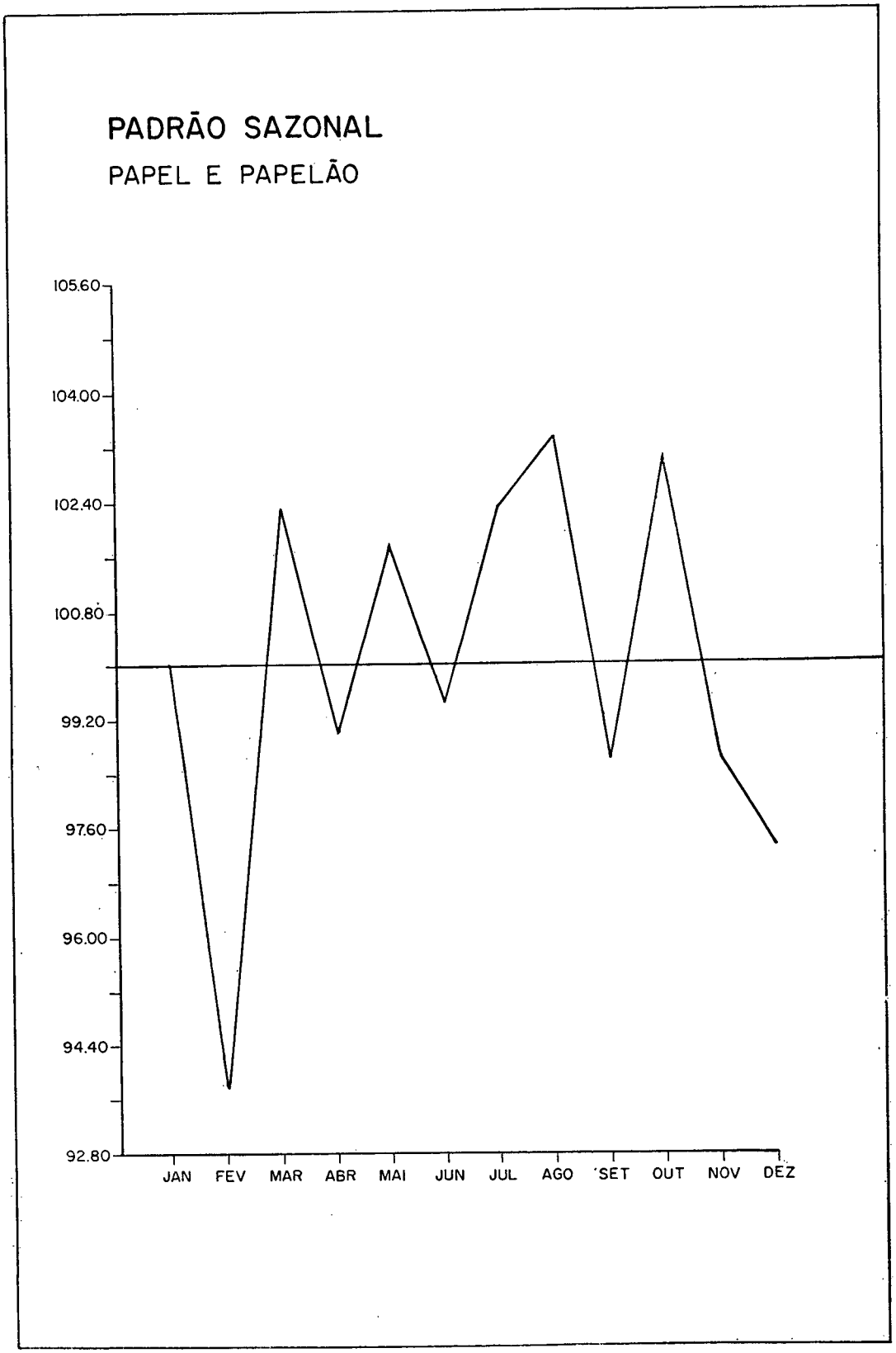
PADRÃO SAZONAL
PAPEL E PAPELÃO
105.60
104.00
102.40
100.80
99.20
97.60
96.00
94.40
92.80
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.9

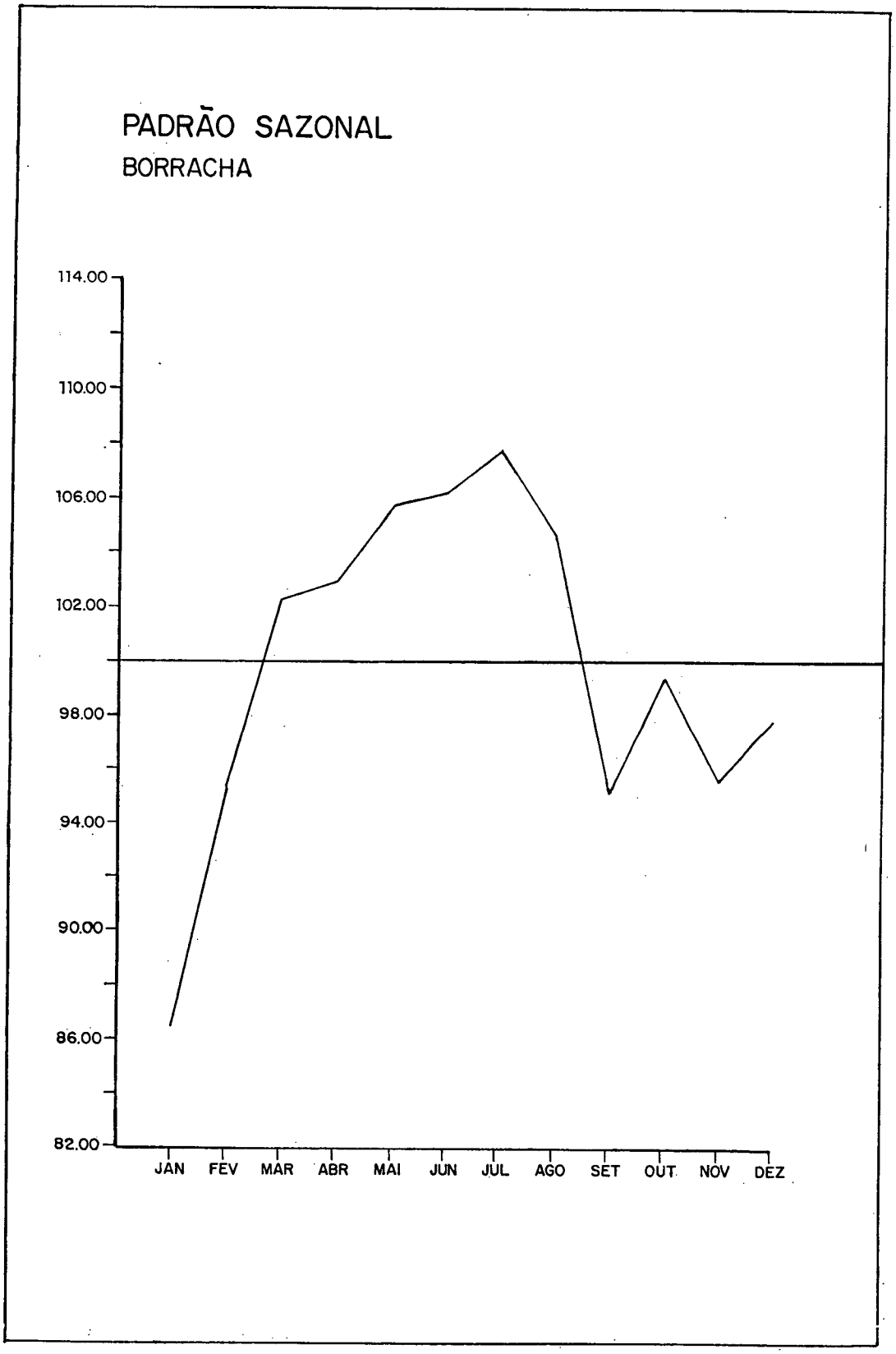
PADRÃO SAZONAL
BORRACHA
114.00
110.00
106.00
102.00
98.00
94.00
90.00
86.00
82.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.10

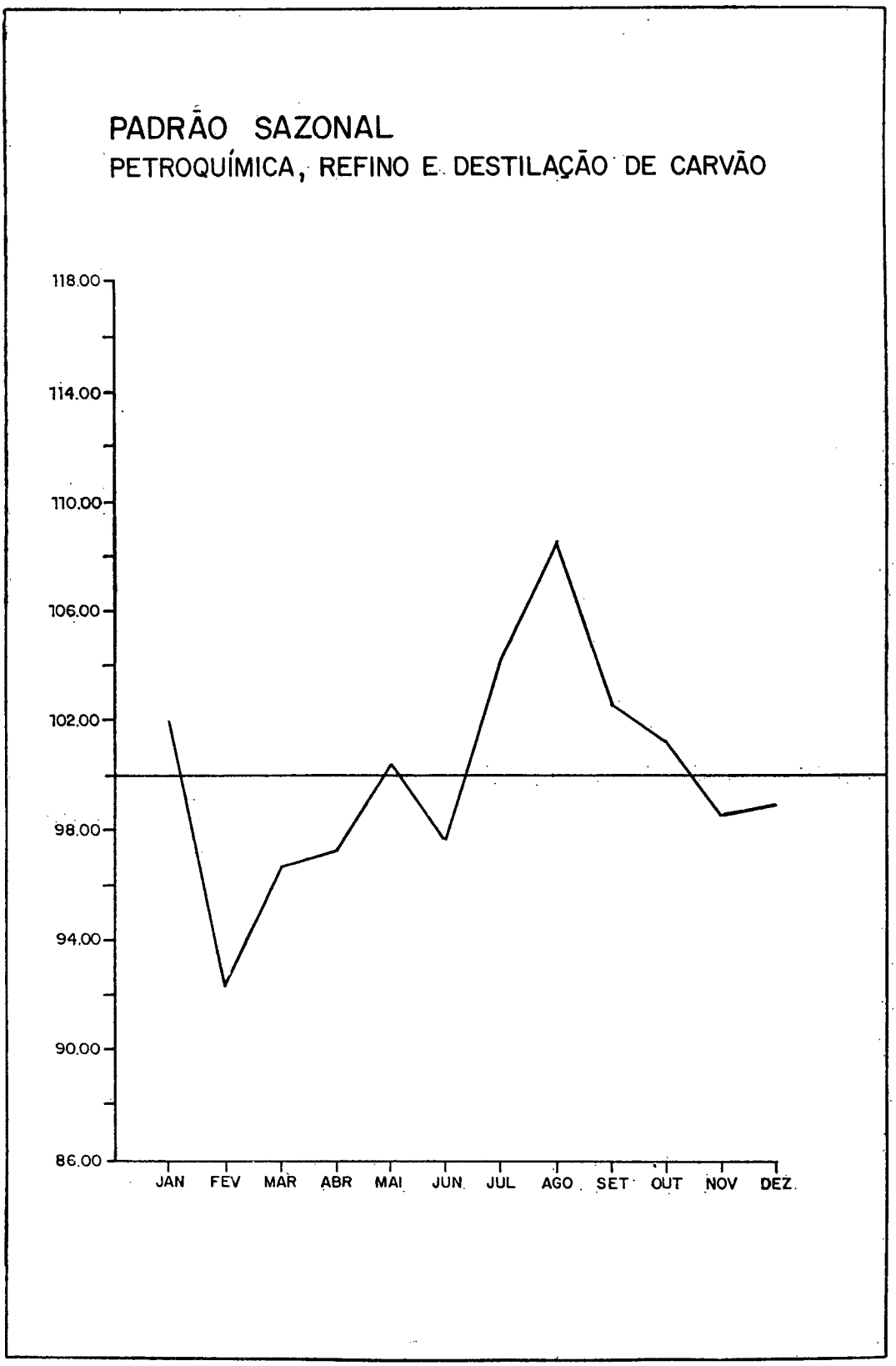
PADRÃO SAZONAL
PETROQUÍMICA, REFINO E DESTILAÇÃO DE CARVÃO
118.00
114.00
110.00
106.00
102.00
98.00
94.00
90.00
86.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRAFICO 3.11

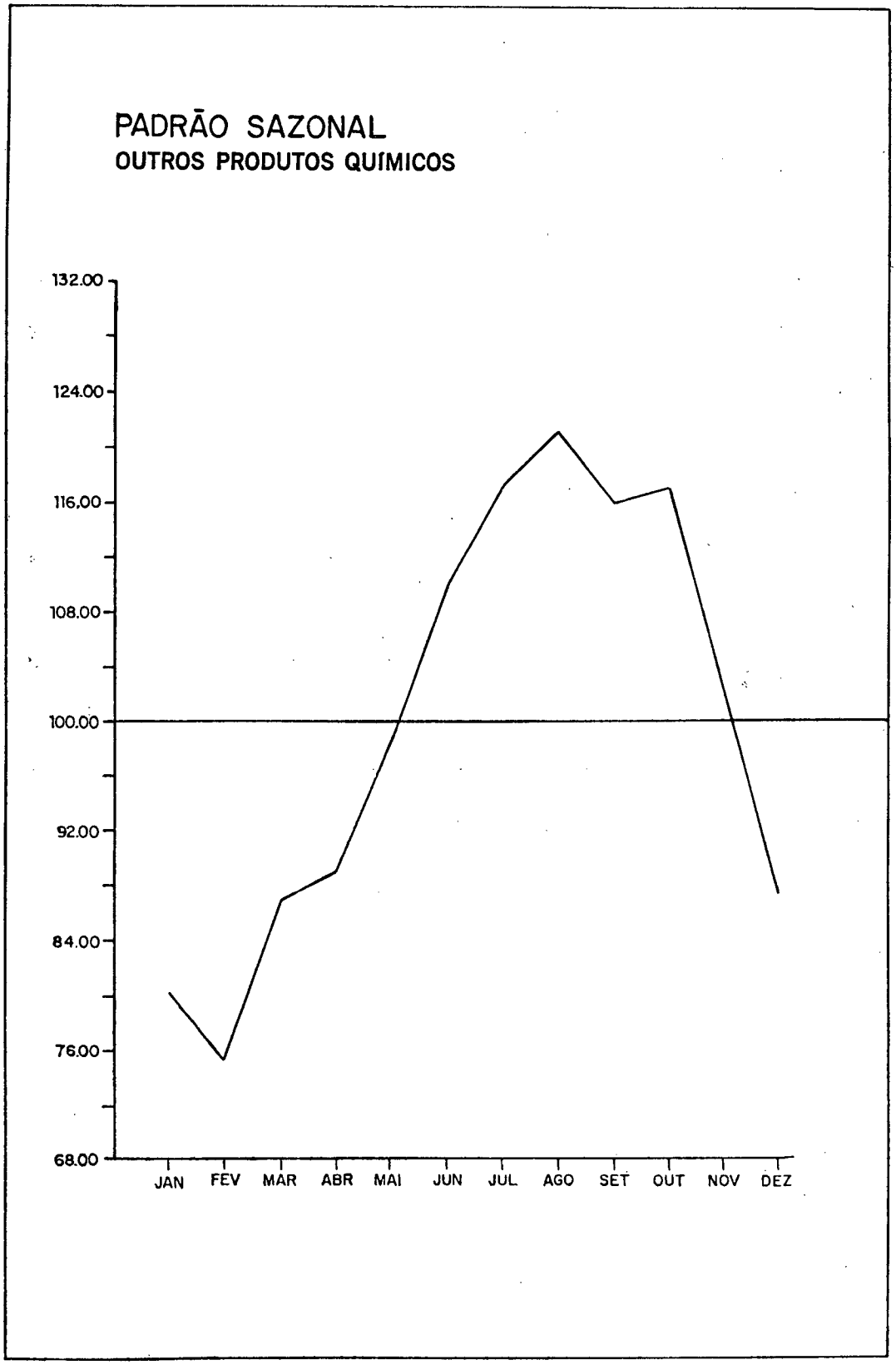
PADRÃO SAZONAL
OUTROS PRODUTOS QUÍMICOS
132.00
124.00
116.00
108.00
100.00
92.00
84.00
76.00
68.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.12

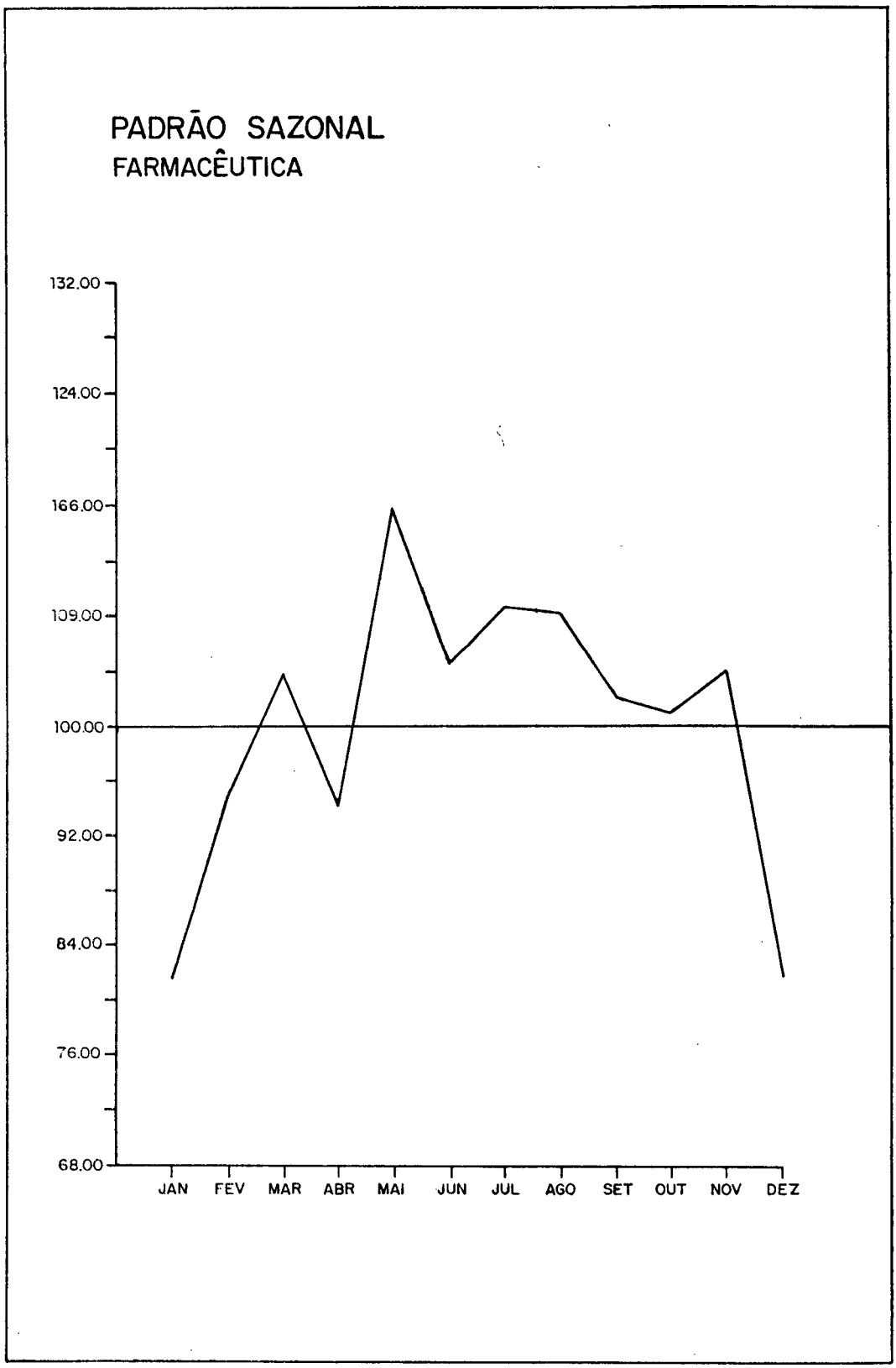
PADRÃO SAZONAL
FARMACÊUTICA
132.00
124.00
166.00
109.00
100.00
92.00
84.00
76.00
68.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.13

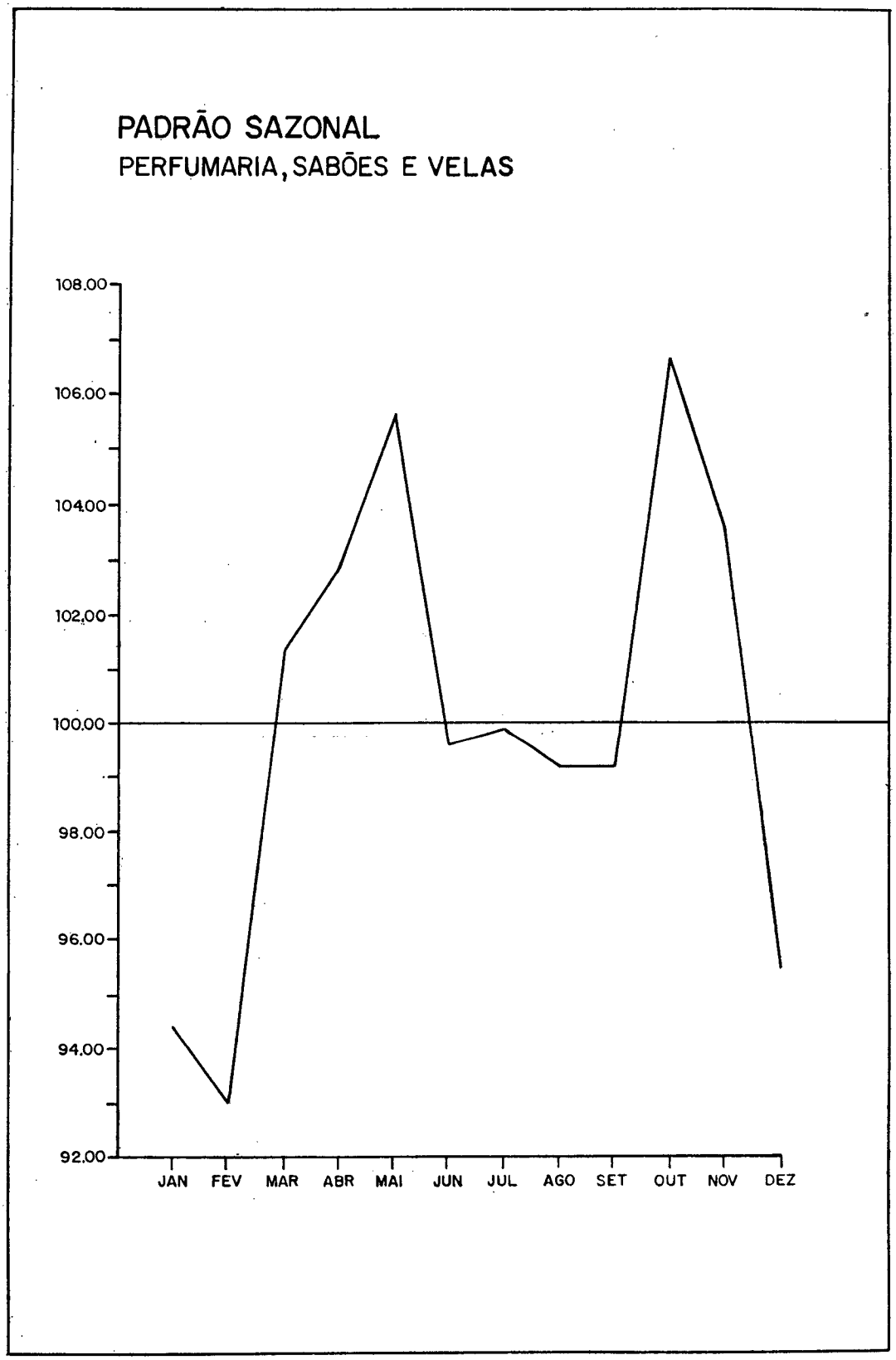
PADRÃO SAZONAL
PERFUMARIA, SABÕES E VELAS
108.00
106.00
104.00
102.00
100.00
98.00
96.00
94.00
92.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRAFICO 3.14

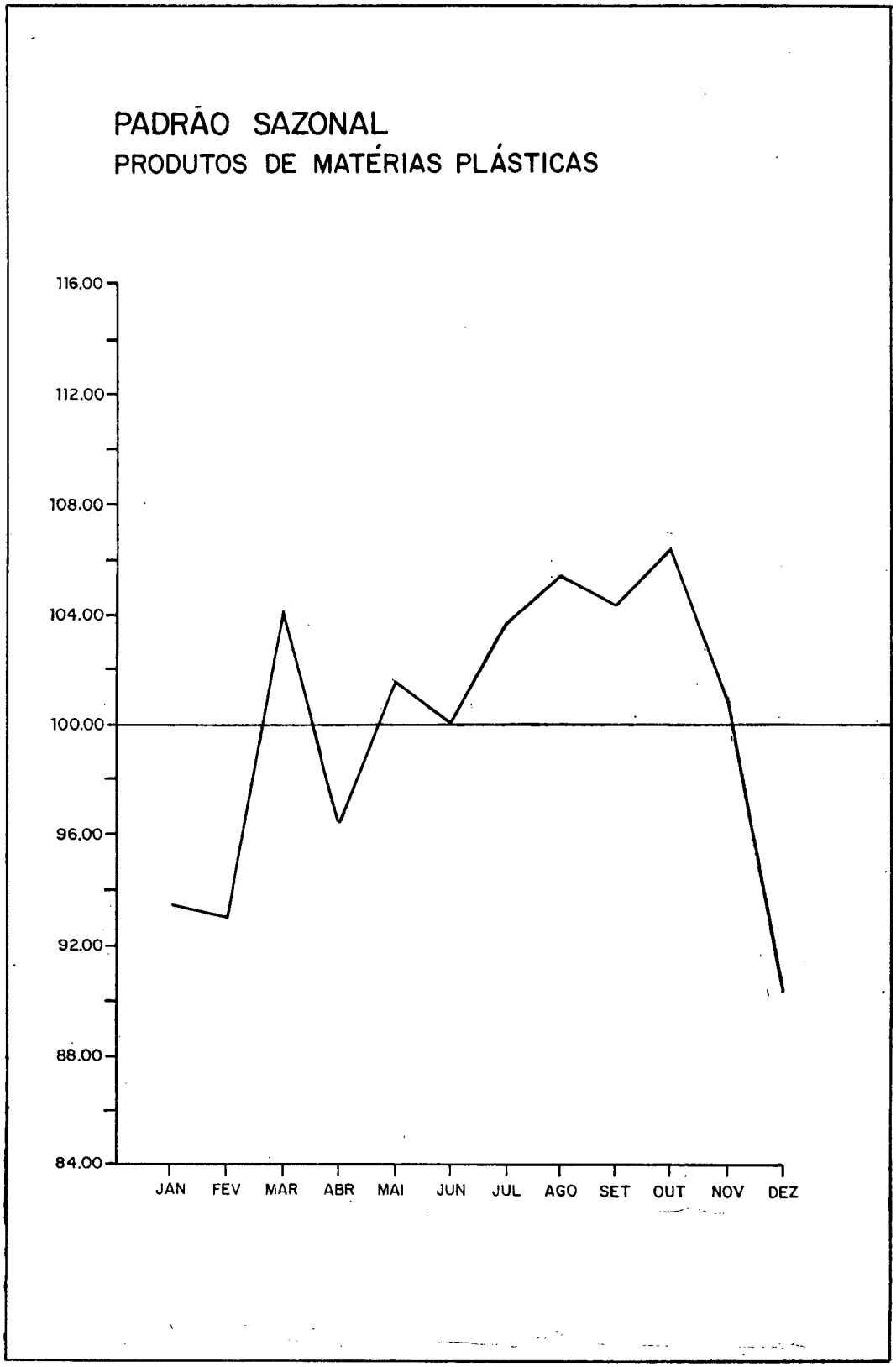
PADRÃO SAZONAL
PRODUTOS DE MATÉRIAS PLÁSTICAS
116.00
112.00
108.00
104.00
100.00
96.00
92.00
88.00
84.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.15

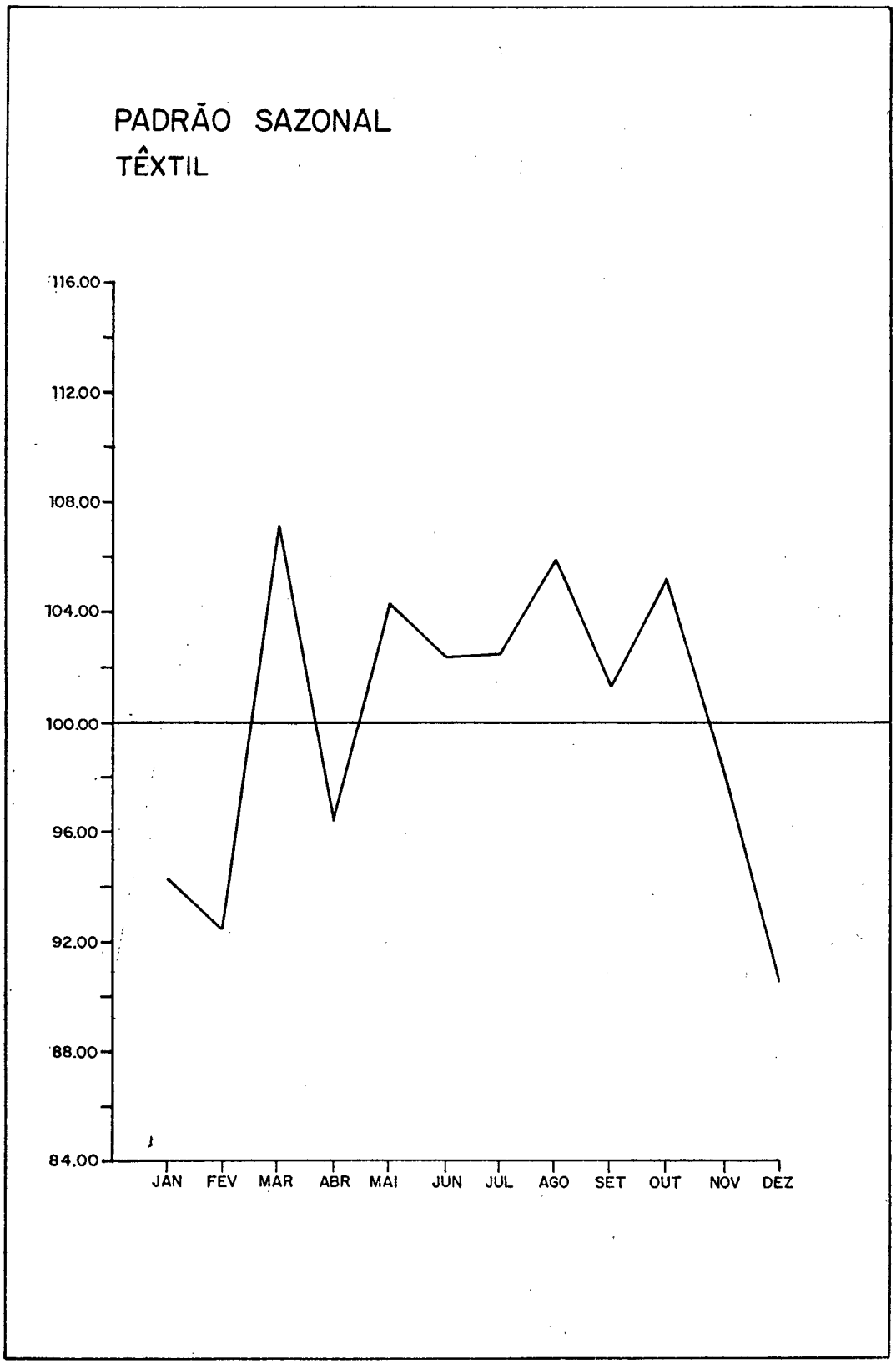
PADRÃO SAZONAL
TÊXTIL
116.00
112.00
108.00
104.00
100.00
96.00
92.00
88.00
84.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.16

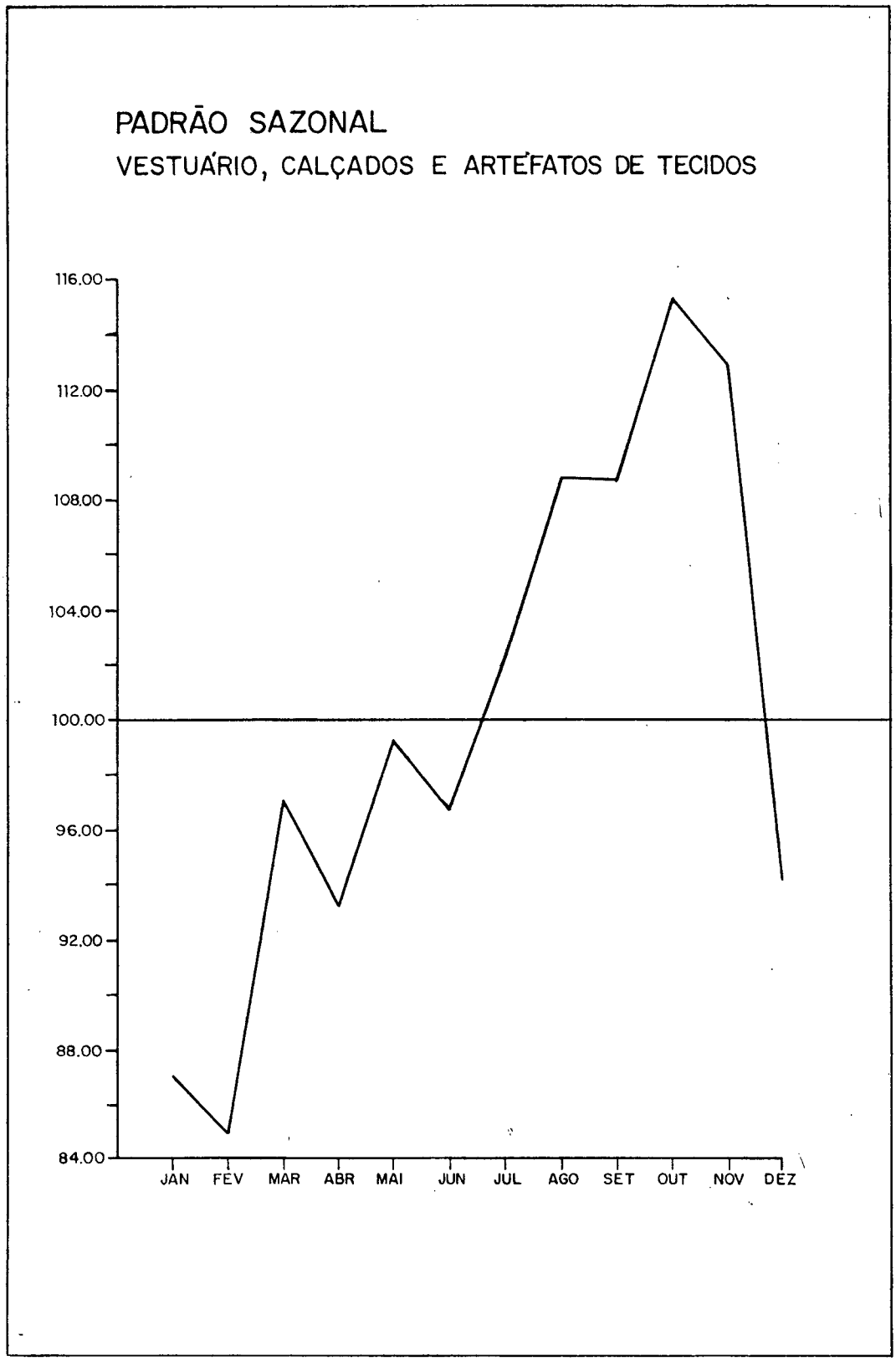
PADRÃO SAZONAL
VESTUÁRIO, CALÇADOS E ARTEFATOS DE TECIDOS
116.00
112.00
108.00
104.00
100.00
96.00
92.00
88.00
84.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.17

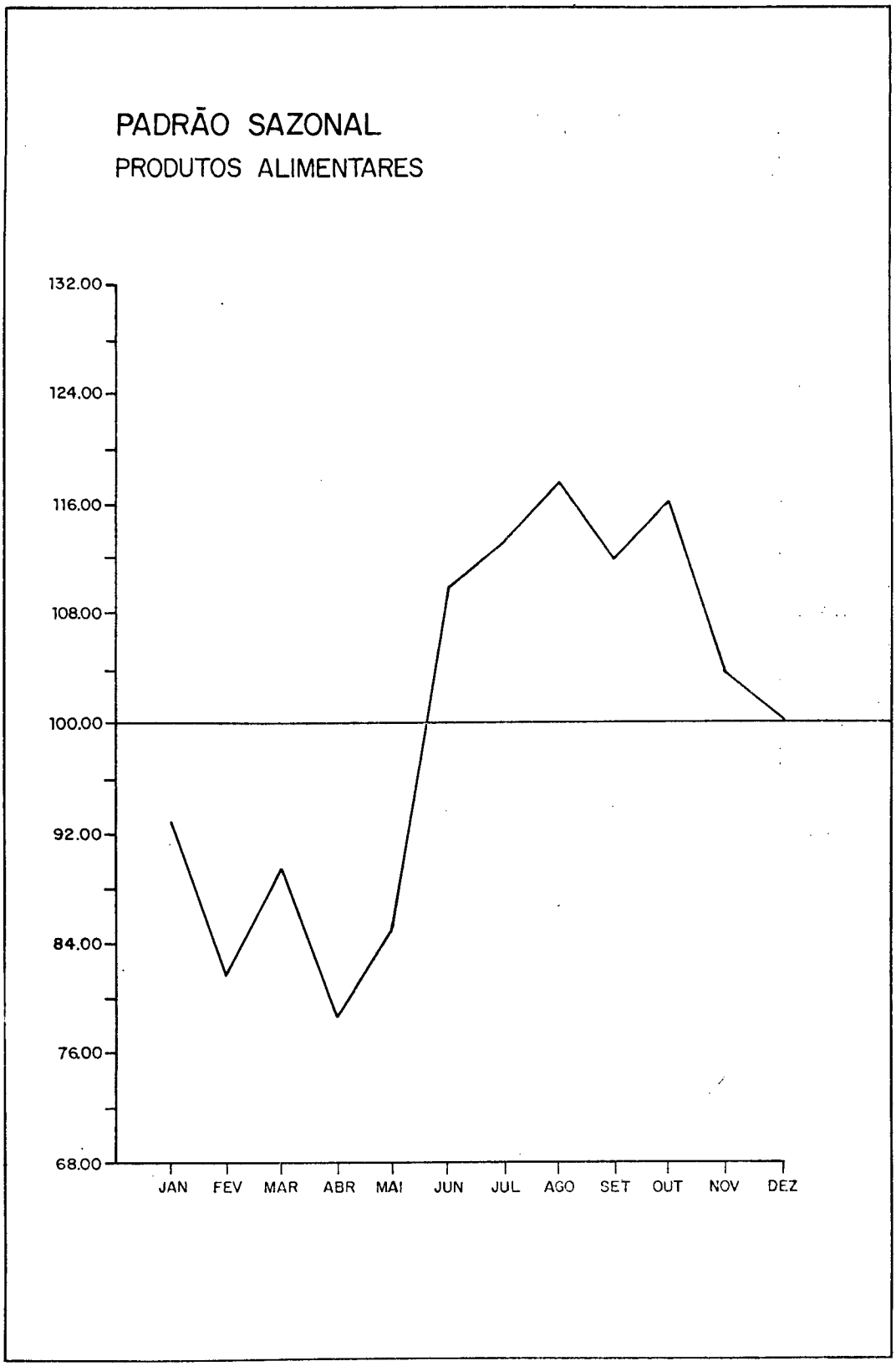
PADRÃO SAZONAL
PRODUTOS ALIMENTARES
132.00
124.00
116.00
108.00
100.00
92.00
84.00
76.00
68.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.18

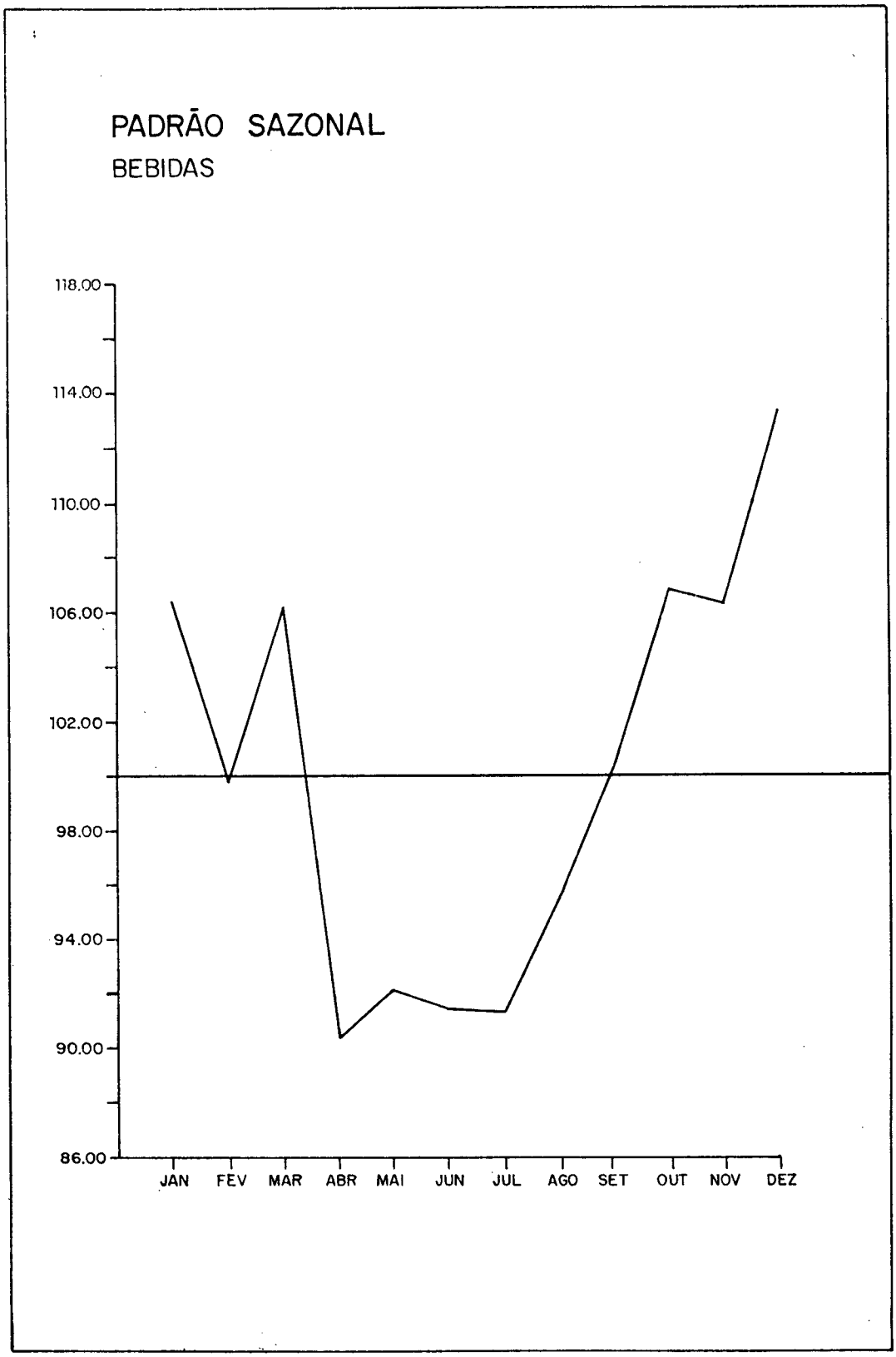
PADRÃO SAZONAL
BEBIDAS
118.00
114.00
110.00
106.00
102.00
98.00
94.00
90.00
86.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.19

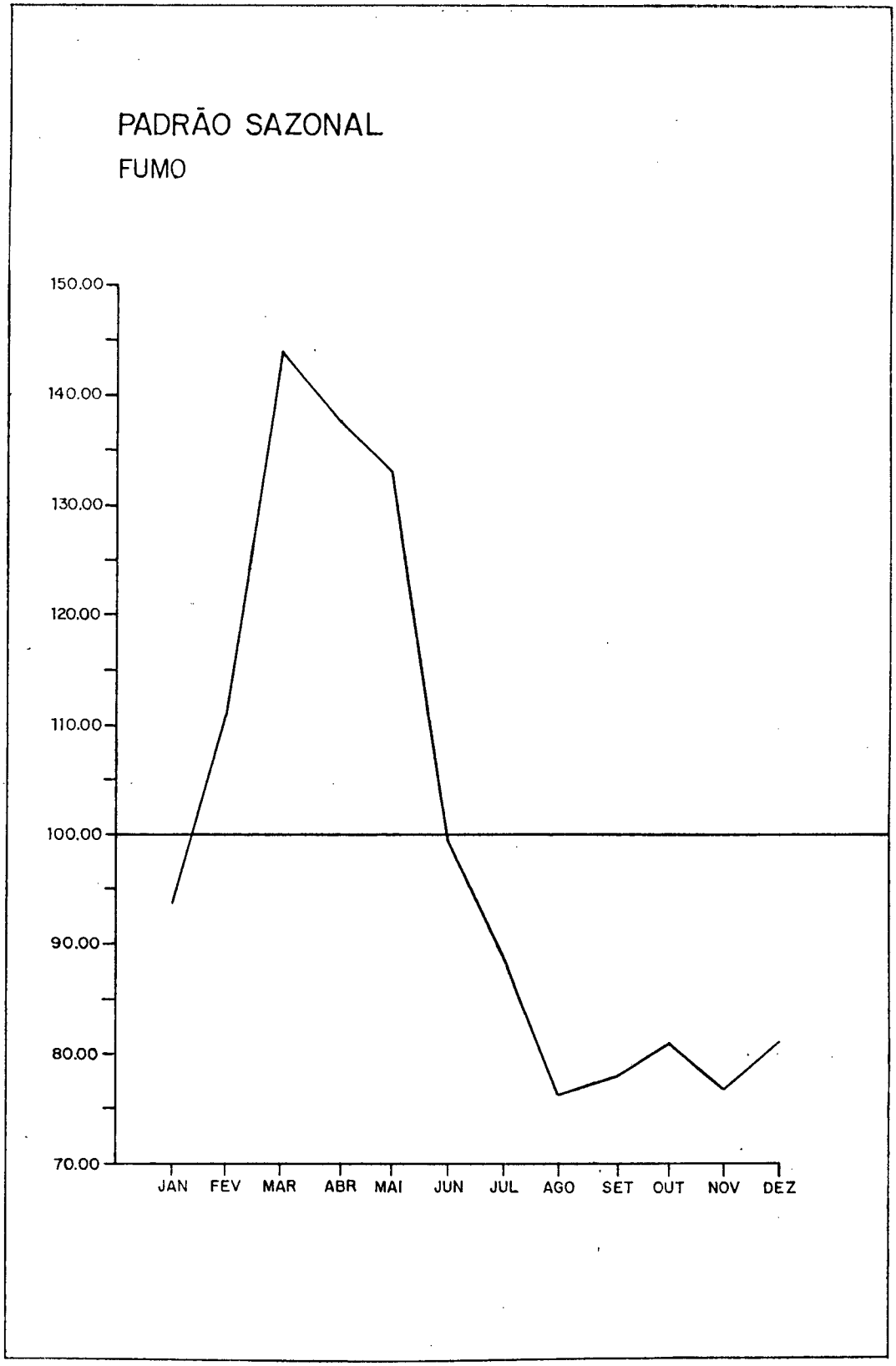
PADRÃO SAZONAL
FUMO
150.00
140.00
130.00
120.00
110.00
100.00
90.00
80.00
70.00
JAN
FEV
MAR
ABR
MAI
JUN
JUL
AGO
SET
OUT
NOV
DEZ

GRÁFICO 3.20

7 — BIBLIOGRAFIA

DAGUM, Estela B. — *The Magic Box and the four Golden Rules.* Technical paper, Ottawa, Statistics Canada, 1980.

———. *The Effect of Time Varying Filters on Seasonal Factor Revisions.* Technical paper, Ottawa, Statistics Canada, 1981.

FONTENELLE E SILVA, Paulo. — *Aspectos Tecnológicos da Estrutura Industrial Brasileira.* Rio de Janeiro, BNDE, 1980.

FÜRST, Patrícia. — Um estudo das relações intersetoriais — Brasil — 1975. Aplicação do método de componentes principais, tese de mestrado, Eng. produção, COPPE/UFRJ, 1984.

KENNY, P. B. & DURBIN, S. — Local trend estimation and seasonal adjustment of economic and social time series. *Journal of the Royal Statical Society,* séries A, *145* (1): 1-41, 1982.

MCKENZIE, Sandra. — *An Evoluation of Concurrent Adjustment on Census Bureau Time Series.* Technical report, Washington, D.C. United States Department of commerce, bureau of the census, 1982.

SHISKIN, J.; YOUNG, A. & MUSGRAVE, S. — *The X-II variant of the Census Method II Seasonal Adjustment Program.* United States Department of commerc, bureau of the census, 1967.

RESUMO

O texto apresenta os principais aspectos do estudo de dessazonalização das séries de produto industrial que deverá resultar na divulgação sistemática de índices sazonalmente ajustados. Utilizou-se o método X-11 do Bureau of Census Americano, que consiste na aplicação de médias móveis para identificação das diversas componentes de uma série de tempo. Os fatores de correção sazonal foram obtidos para cada subgênero, sendo o índice geral da indústria obtido por composição. A continuação dos estudos prevê a extensão da aplicação do ajustamento sazonal aos demais indicadores conjunturais da indústria bem como a utilização de métodos alternativos.

Bibliografia

PUBLICAÇÕES DE INTERESSE PARA A ESTATÍSTICA EDITADAS PELO IBGE NO PERÍODO DE ABRIL A JUNHO DE 1983 *

ANUÁRIO ESTATÍSTICO DO BRASIL. Rio de Janeiro, 1982.

ARMAZENAGEM E ESTOCAGEM A SECO E A FRIO — 1979. Rio de Janeiro, 1983, v. 4, t. 10 — Brasil.

Brasil: sintesis informativa 1982. Rio de Janeiro, 1982. Espanhol.

Brasília, DF. Rio de Janeiro, 1983. (Coleção de monografias, 1).

Campos Novos, SC. Rio de Janeiro, 1983. (Coleção de monografias, 5).

CENSO DEMOGRÁFICO 1980; dados distritais. Rio de Janeiro, 1983. n. 3 — Amazonas.

———. n. 23 — Goiás.

CENSO DEMOGRÁFICO 1980; dados gerais: migração, instrução, fecundidade, mortalidade. Rio de Janeiro, 1983. n. 7 — Maranhão.

———. n. 9 — Ceará.

———. n. 19 — São Paulo.

———. n. 20 — Paraná.

———. n. 21 — Santa Catarina.

CENSO DEMOGRÁFICO 1980; mão-de-obra. Rio de Janeiro, 1983. n. 17 — Espírito Santo.

ESTATÍSTICAS DA SAÚDE; assistência médico-sanitária — 1980. Rio de Janeiro, 1982. v. 5.

ESTATÍSTICAS ECONÔMICAS DO GOVERNO ESTADUAL E MUNICIPAL; balanços estaduais, versão analítica da despesa realizada — 1977. Rio de Janeiro, 1982. v. 3, t. 2.

INDICADORES IBGE. Rio de Janeiro, 1983. v. 2, n. 2/3, fev./mar. 1983.

Santa Luzia, MA. Rio de Janeiro, 1983. (Coleção de monografias, 3).

* Preparado no Departamento de Documentação e Referência da Biblioteca Central do IBGE por Hesperia Zuma de Rosso.

PESQUISA INDUSTRIAL 1979; dados gerais. Rio de Janeiro, 1983. t. 2 — Região Nordeste.

———. t. 3 — Região Sudeste.

PESQUISA NACIONAL POR AMOSTRA DE DOMICÍLIOS — 1981. Rio de Janeiro, 1983, t. 1 — Ceará.

———. t. 2 — Pernambuco.

———. t. 9 — Goiás.

———. t. 11 — Brasil e Grandes Regiões.

———. t. 12 — Regiões Metropolitanas: Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba, Porto Alegre.

Silvânia, GO. Rio de Janeiro, 1983. (Coleção de monografias, 6).

SILVICULTURA — 1980. Rio de Janeiro, 1983. v. 3 — Brasil, Grandes Regiões, Unidades da Federação, Mesorregiões, Microrregiões Homogêneas, Municípios.

Sinopse estatística da Região Centro-Oeste. Rio de Janeiro, 1982.

Xapuri, AC. Rio de Janeiro, 1983. (Coleção de monografias, 2).