

ANÁLISE DE TRAJETÓRIAS

JORGE ALBERTO ACHCAR

DISSERTAÇÃO APRESENTADA AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ORIENTADOR:

PROF. DR. CLÓVIS DE ARAUJO PERES

- SÃO PAULO, ABRIL DE 1976 -



À

Alina

ÍNDICE

CAP.0 - PRELIMINARES SOBRE O ESTUDO DA ANÁLISE DE TRAJETÓRIAS . . .	1
CAP.1 - INTRODUÇÃO	5
1.1 - O Método da Análise de Trajetórias	5
1.2 - Estimação dos Coeficientes de Trajetória	7
1.3 - O Teorema Fundamental da Análise de Trajetórias. . .	10
1.4 - Uma Interpretação para a Fórmula da Determinação Com pleta.	12
1.5 - Decomposição dos Efeitos na Análise de Trajetórias .	13
1.6 - Estimador do Coeficiente de Trajetórias Residual . .	14
1.7 - Comentários.	15
1.8 - Um Diagrama Adequado	15
CAP.2 - ANÁLISE DE TRAJETÓRIAS EM MODELOS DE REGRESSÃO SIMPLES . .	17
2.1 - O Modelo	17
2.2 - Estimação dos Coeficientes de Trajetórias.	18
2.3 - Distribuição Amostral de \hat{p}_{21} quando X_1 e X_2 são va- riáveis Aleatórias	19
2.3.1 - Distribuição Amostral de \hat{p}_{21} quando $p_{21}=0$. .	19
2.3.2 - Distribuição de \hat{p}_{21} , quando $p_{21} \neq 0$	20
2.4 - Distribuição do Estimador de p_{21} Quando X_1 não é Va- riável Aleatória	21
CAP.3 - ANÁLISE DE TRAJETÓRIAS EM MODELOS DE REGRESSÃO MÚLTIPLAS .	25
3.1 - O Modelo	25
3.2 - Estimação dos Coeficientes de Trajetórias.	27
3.3 - Relação entre os Coeficientes de Trajetória e os Coe- ficientes de Correlação Parciais	28
3.4 - Modelos de Trajetórias quando as Variáveis Indepen- dentes são não Correlacionadas	31
3.5 - Distribuição Amostral de \hat{p}_{01}	34
CAP.4 - APLICAÇÕES	37
4.1 - Aplicações da Análise de Trajetórias em Genética . .	37
4.2 - Aplicação da Análise de Trajetórias no Estudo de In- fluência de alguns Fatores sobre a Taxa Salarial de Pessoal Especializado das Indústrias de S.Paulo. . .	38
4.3 - Aplicação da Análise de Trajetórias no Estudo da Se- gregação Residencial nas Cidades Norte-Americanas. .	43
APÊNDICE 1	47

APÊNDICE 2 - PROPRIEDADES DE INVARIÂNCIA DOS ESTIMADORES DE MÍNIMOS QUADRADOS DOS COEFICIENTES DE TRAJETÓRIAS	51
APÊNDICE 3	55
A3.1 - Relação entre Coeficientes de Correlação Parciais e Coeficientes de Correlação de Ordem Zero	55
A3.2 - Regressão Parcial Linear	56
A3.3 - Relação entre a Variância Residual com o Coeficiente de Regressão Parcial.	58
A3.4 - Relações entre as Variâncias, as Regressões e as Correlações de Diferentes Ordens.	59
A3.5 - Outras Relações.	60
APÊNDICE 4	62
A4.1 - Coeficiente de Correlação Múltipla	62
APÊNDICE 5	64
A5.1 - O Coeficiente de Trajetórias e o Coeficiente de Correlação de Ordem Zero, quando as Variáveis Independentes não são intercorrelacionadas	64
REFERÊNCIAS BIBLIOGRÁFICAS	67

AGRADECIMENTOS

Inicialmente agradeço ao Prof. Dr. Clóvis de Araujo Peres do Instituto de Matemática e Estatística da Universidade de São Paulo, orientador deste trabalho, pela sua valiosa dedicação e pelos incentivos dados a mim durante o decorrer da pesquisa relativa a esta dissertação.

Agradeço aos Professores do Instituto de Matemática e Estatística da Universidade de São Paulo responsáveis pela minha formação, sem a qual eu não estaria concluindo estas notas. Em especial agradeço ao Prof. Carlos Alberto de Bragança Pereira pela colaboração dada no início de meus cursos de Pós-Graduação.

Agradeço aos colegas do Departamento de Ciências Exatas da Faculdade de Filosofia, Ciências e Letras de São José do Rio Preto da Universidade Estadual Paulista Dr. Júlio de Mesquita Filho, pelo ambiente de camaradagem durante a elaboração deste trabalho.

Finalmente quero agradecer ao Sr. João Baptista Esteves de Oliveira pela sua eficiência no trabalho de datilografia desta dissertação.

CAPÍTULO 0

PRELIMINARES SOBRE O ESTUDO

DA ANÁLISE DE TRAJETÓRIAS

O objetivo deste trabalho é apresentar de uma forma clara, os conceitos básicos da análise de trajetórias ("path analysis"). Para tanto, nos primeiros capítulos daremos um enfoque das idéias gerais e das contribuições existentes até agora na literatura. Num capítulo final daremos de forma detalhada alguns exemplos de aplicação da análise de trajetórias em problemas das mais diversas áreas, para que o leitor possa avaliar concretamente a importância deste método estatístico. Todavia, antes de iniciarmos a apresentação do método, faremos uma breve apresentação de alguns dados, assinalando o desenvolvimento da análise de trajetórias desde seus dias iniciais, até o momento.

O geneticista Sewall Wright foi o precursor da análise de trajetórias, com trabalhos apresentados a partir de de 1918.

Wright introduziu a técnica da análise de trajetórias a partir de diagramas genéticos, envolvendo sistemas de cruzamentos. Em 1921, publicou um artigo sintetizando todas suas idéias e nesse mesmo ano, publicou um artigo com aplicações na genética. De uma forma inédita, Wright confirmou nessas aplicações, a utilidade da análise de trajetórias ao verificar que alguns resultados previamente conhecidos sobre sistemas de endocruzamento obtidos por procedimentos aritméticos muito complicados coincidiam com os resultados ob

tidos a partir da aplicação do método por ele proposto.

Além disso, Wright encontrou resultados ainda não conhecidos nessa época sobre outros sistemas de endocruzamento, por serem impossíveis de se obter sem a técnica da análise de trajetórias.

Apesar de um grande êxito inicial, a análise de trajetórias sofreu algumas críticas por parte de alguns geneticistas e estatísticos.

Uma das críticas formuladas, partiu de Niles em 1922 e 1923, que não concordava com as idéias envolvendo algumas variáveis como causas e outras variáveis como efeitos. Esta crítica foi respondida por Wright em 1923. Em 1934 Wright apresentou um artigo englobando todas contribuições existentes, e apresentou algumas correções das aplicações.

Uma outra crítica apareceu em 1954, por parte de Tukey. Tukey iniciou seu trabalho, dividindo a análise de dados em análise descritiva, tangencial e funcional. Com essa classificação, Tukey considerou o coeficiente de correlação como uma medida descritiva, e o coeficiente de regressão como uma medida funcional. A crítica formulada por Tukey, foi relativa à utilização do coeficiente de correlação na aplicação do método da análise de trajetórias. Em algumas aplicações da análise, era calculado o coeficiente de correlação de Pearson entre as variáveis, e a partir das leis da análise de trajetórias eram encontrados os efeitos diretos ou indiretos. Tukey considerou incorreto determinar a partir de uma medida descritiva uma medida funcional, porque os efeitos diretos e indiretos são coeficientes de regressão padronizados e produtos destes. Tukey sugeriu uma análise de trajetórias formulada em termos da análise de regressão, tal como é considerada atualmente.

Em 1956, Li apresentou na literatura uma exposição

detalhada da análise de trajetórias, juntamente com aplicações na genética.

A partir da metade da década de 50, a análise de trajetórias começou a ser utilizada com frequência pelos pesquisadores de Ciências Humanas, para interpretar sistemas causais. Assim, temos o trabalho de Simon em 1954, e os trabalhos de Blalock em 1960, 1961 e 1962.

Em 1959, Turner e Stevens consideraram os sistemas causais interpretados pela análise de trajetórias como casos especiais de regressão múltipla.

Em sociologia, também devemos citar as contribuições de Duncan em 1966, de Blau e Duncan em 1967 e de Land em 1969.

REFERÊNCIAS:

- [1] - BLALOCK, H.M. "Correlational Analysis and Causal Inference", American Antropologist, 62 (1960), 624-31.
- [2] - BLALOCK, H.M. "Correlation and Causality: The Multivariate Case", Social Force, 39 (1961), 1246-51.
- [3] - BLALOCK, H.M. "Evaluating the Relative Importance of variables", American Sociological Review, 26 (1961), 866-74.
- [4] - BLALOCK, H.M. "Further observations on asymmetrical causal models", American Sociological Review, 27 (1962), nº 4.
- [5] - BLAU, P.M. e DUCAN, O.D. (1967) - vide citações bibliográficas.
- [6] - DUNCAN, O.D. (1966) - vide citações bibliográficas.
- [7] - LAND, K.C. (1969) - vide citações bibliográficas.
- [8] - LI, C.C. (1956) - vide citações bibliográficas.

- [9] - NILES, H.E. "Correlation, Causation and Wright's Theory of Path Coefficients", Genetics, 7 (1922), 258.
- [10] - NILES, H.E. "The method of path coefficients. An answer to Wright", Genetics, (1923), 261.
- [11] - SIMON, H.A. "Spurious Correlation: a causal interpretation". Amer. Statistical Ass. Jour., (setembro 1954), 467-79.
- [12] - TUKEY, J.W. (1954) - vide citações bibliográficas.
- [13] - TURNER, R and STEVENS, C.D. (1959) - vide citações bibliográficas.
- [14] - WRIGHT, S. "On the nature of size factors", Genetics, 3 (1918), 367-74.
- [15] - WRIGHT, S. "The relative importance of hereditary and environment in determining the piebald pattern of guinea pigs", Proc. Nat. Acad. Sci., (1920), 320-32.
- [16] - WRIGHT, S. "Causation and Correlation", Jour. Ag. Res., 20 (1921), 557-85.
- [17] - WRIGHT, S. "The Theory of path coefficients. A reply to Niles Criticism", Genetics, 8 (1923), 239-55.
- [18] - WRIGHT, S. (1943) - vide citações bibliográficas.
- [19] - WRIGHT, S. (1960) - vide citações bibliográficas.

CAPÍTULO I

INTRODUÇÃO

1.1 - O MÉTODO DA ANÁLISE DE TRAJETÓRIAS

Para a aplicação da análise de trajetórias, nós precisamos inicialmente considerar um diagrama de trajetórias, associado às variáveis do problema em questão.

Este diagrama é constituído pelo pesquisador considerando algumas variáveis como efeitos e outras variáveis como causas, baseado nas considerações teóricas da área de aplicação.

Na figura (1.1), temos um exemplo de diagrama de trajetórias.

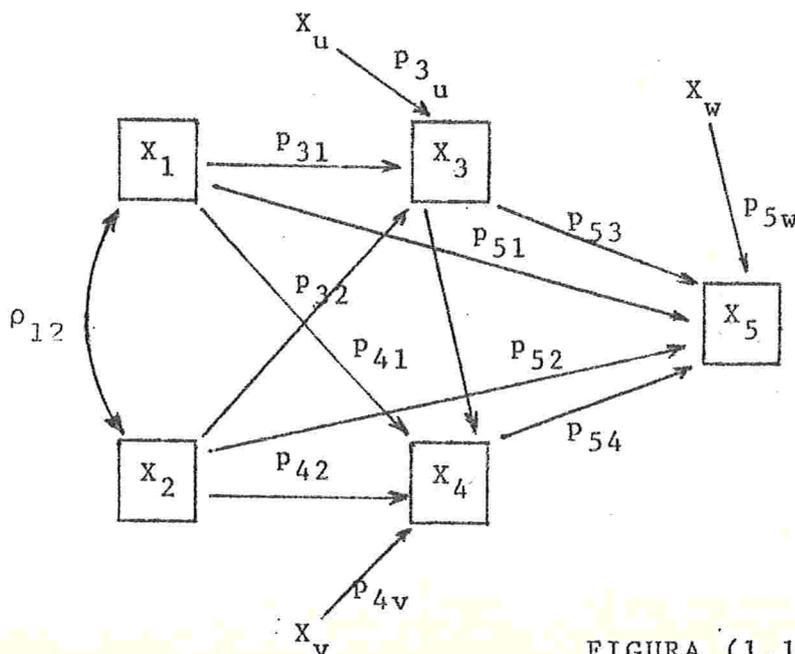


FIGURA (1.1)

Com uma notação que levaremos até ao final deste trabalho, denotaremos os coeficientes de trajetórias por p_{ij} , onde o índice i está associado à variável efeito e o índice j está associado à variável causa.

Em correspondência ao diagrama de trajetórias, associamos a cada seta unidirecional um coeficiente de trajetória e a cada seta bidirecional um coeficiente de correlação.

A cada diagrama de trajetórias associamos um conjunto de equações de regressão linear, onde cada variável e feito será uma variável dependente no modelo de regressão.

A relação entre os coeficientes de trajetórias e os coeficientes de regressão é obtida escrevendo cada equação de regressão do modelo em duas formas equivalentes.

Isto é:

$$(1.1) \quad X_0 = \alpha_1 + \sum_{i=1}^m \beta_{0i} X_i + \epsilon$$

$$(1.2) \quad Z_0 = \sum_{i=1}^m p_{0i} + p_{0u} Z_u,$$

onde

$$(1.3) \quad Z_i = \frac{X_i - u_i}{\sigma_i}, \quad i=1, \dots, m$$

$$(1.4) \quad \epsilon = p_{0u} Z_u$$

$$(1.5) \quad p_{0i} = \frac{\beta_{0i} \sigma_i}{\sigma_0}, \quad i=1, 2, \dots, m.$$

No exemplo dado pelo diagrama de trajetórias da figura (1.1), temos:

$$\begin{aligned} Z_3 &= p_{32}Z_2 + p_{31}Z_1 + p_{3u}Z_u \\ (1.6) \quad Z_4 &= p_{43}Z_3 + p_{42}Z_2 + p_{41}Z_1 + p_{4v}Z_v \\ Z_5 &= p_{54}Z_4 + p_{53}Z_3 + p_{52}Z_2 + p_{51}Z_1 + p_{5w}Z_w \end{aligned}$$

Todas as suposições envolvidas são conhecidas da teoria geral de Modelos Lineares.

Os coeficientes de trajetória definidos em (1.5) medem o efeito direto de uma variável sobre outra variável.

A interpretação de sistemas causais por meio de análise de trajetórias se torna mais eficiente porque podemos decompor a influência de uma variável sobre outra variável em efeitos diretos, efeitos indiretos e outros.

Algumas dificuldades podem aparecer na aplicação desta técnica quando o sistema casual não é bem determinado.

Podemos interpretar o coeficiente de trajetória a partir da expressão (1.5). Desta maneira o quadrado do coeficiente de trajetória mede a proporção da variância da variável dependente para a qual a variável independente é diretamente responsável. Neste sentido pode ser chamado coeficiente de determinação.

1.2 - ESTIMAÇÃO DOS COEFICIENTES DE TRAJETÓRIA

Um estimador sugerido para o coeficiente de trajetória é dado por:

$$(1.7) \quad \hat{p}_{0i} = \frac{S_i}{S_0} \hat{\beta}_{0i}, \quad i=1, \dots, m.$$

Notamos que este estimador não depende da unidade física na qual a variável X_i está medida, sendo esta uma diferença básica com o estimador do coeficiente de regressão.

Com o exemplo numérico dado aqui, queremos chamar a atenção do leitor, para a diferença entre a análise de trajetórias, e a análise de regressão linear múltipla comum, as sinalando as vantagens da análise de trajetórias na interpretação e medidas das relações causais.

Vamos considerar os dois conjuntos de dados apresentados nas tabelas 1 e 2, e vamos supor que X_0 seja combinação linear de X_1 e X_2 nos dois casos:

TABELA 1			TABELA 2		
X_0 combinação linear de X_1 e X_2					
X_0	X_1	X_2	X_0	X_1	X_2
4,36	17	3	2,26	10	3
3,20	11	5	2,60	9	5
5,62	18	6	6,22	20	6
6,26	18	8	5,96	17	8
4,82	10	11	5,12	11	11
5,48	9	14	8,18	18	14
9,10	20	15	8,50	18	15
9,16	17	18	9,16	17	18
$\sum X_0 = 48$	$\sum X_1 = 120$	$\sum X_2 = 80$	$\sum X_0 = 48$	$\sum X_1 = 120$	$\sum X_2 = 80$
$\bar{X}_0 = 6$	$\bar{X}_1 = 15$	$\bar{X}_2 = 10$	$\bar{X}_0 = 6$	$\bar{X}_1 = 15$	$\bar{X}_2 = 10$
$S_0^2 = 4$	$S_1^2 = 16$	$S_2^2 = 25$	$S_0^2 = 5,92$	$S_1^2 = 16$	$S_2^2 = 25$
$S_0 = 2$	$S_1 = 4$	$S_2 = 5$	$S_0 = 2,433$	$S_1 = 4$	$S_2 = 5$
$S_{01} = 4,8$	$S_{12} = 8,0$	$S_{12} = 0$	$S_{01} = 8$	$S_{02} = 11$	$S_{12} = 10$
$r_{01} = 0,6$	$r_{12} = 0,8$	$r_{12} = 0$	$r_{01} = 0,822$	$r_{02} = 0,904$	$r_{12} = 0,5$

Temos:

$$(1.8) \quad X_0 = \alpha_1 + \beta_{01}X_1 + \beta_{02}X_2$$

Os estimadores dos mínimos quadrados dos coeficientes de regressão para os dados das tabelas 1 e 2 coincidem e são:

$$\hat{\alpha}_1 = -1,7; \hat{\beta}_{01} = 0,30; \hat{\beta}_{02} = 0,32.$$

Uma conclusão inicial é que as influências de X_1 e X_2 em X_0 são as mesmas para os dois casos.

Mas na forma reduzida, encontramos as seguintes equações para os dados das tabelas 1 e 2, respectivamente.

$$(1.9) \quad \begin{aligned} Z_0 &= 0,60Z_1 + 0,80Z_2 \\ Z_0 &= 0,4932Z_1 + 0,6576Z_2 \end{aligned}$$

Os estimadores dos coeficientes de trajetórias, diferem para os dois conjuntos de dados.

A diferença é dada pelo fato de que na primeira tabela, as variáveis X_1 e X_2 não são correlacionadas, e na segunda tabela X_1 e X_2 são variáveis correlacionadas.

As figuras (1.2) e (1.3) representam as relações causais dadas nas tabelas 1 e 2, respectivamente.

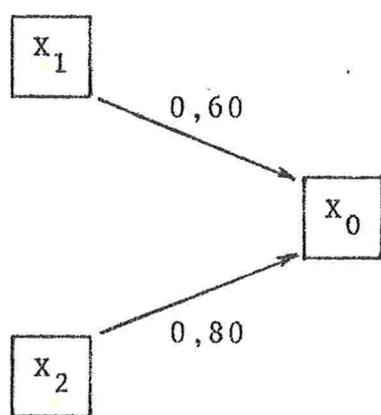


FIGURA (1.2)

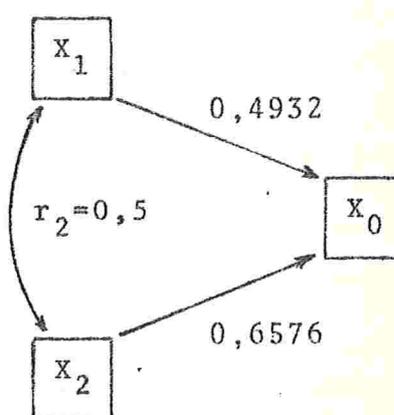


FIGURA (1.3)

Para os dados da tabela 1, os estimadores dos coeficientes de trajetória coincidem com os coeficientes de cor-

relação de Pearson entre as variáveis:

$$(1.10) \quad \begin{aligned} r_{01} &= \hat{p}_{01} = 0,60 \\ r_{02} &= \hat{p}_{02} = 0,80 \end{aligned}$$

Para os dados da tabela 2, temos as relações:

$$(1.11) \quad r_{01} = \hat{p}_{01} + r_{12}\hat{p}_{02} = 0,4932 + (0,5)(0,6576)$$

ou

$$(1.12) \quad \begin{aligned} r_{01} &= \underline{0,82} \\ r_{02} &= \hat{p}_{02} + r_{02}\hat{p}_{01} = 0,6576 + (0,5)(0,4932) \end{aligned}$$

ou

$$r_{02} = \underline{0,9042}$$

Em (1.11), verificamos que:

$\hat{p}_{01} = 0,4932$ mede a influência direta de X_1 em X_0 .

$r_{12}\hat{p}_{02} = (0,5)(0,6576) = 0,3288$ mede a influência da variável X_1 em X_0 através de X_2 .

1.3 - O TEOREMA FUNDAMENTAL DA ANÁLISE DE TRAJETÓRIAS

Podemos escrever o coeficiente de correlação de Pearson entre X_i e X_j , por:

$$(1.13) \quad r_{ij} = \frac{1}{n} \sum_{k=1}^n z_{ik}z_{jk}$$

onde n representa o número de observações.

De acordo com as equações de regressão associadas a um diagrama, temos:

$$(1.14) \quad r_{ij} = \sum_q \hat{p}_{iq}r_{jq}$$

onde q cobre todas as variáveis associadas às possíveis trajetórias entre as variáveis X_i e X_j .

Assim, num diagrama de trajetória, o coeficiente de correlação de Pearson entre duas variáveis é obtido tomando-se todos os trajetos possíveis entre as variáveis consideradas e tomando-se os produtos entre os valores numéricos dos estimadores dos coeficientes envolvidos ao longo de cada trajeto.

Por exemplo, no diagrama (1.1), temos:

$$(1.15) \quad r_{35} = \hat{p}_{54}r_{34} + \hat{p}_{53} + \hat{p}_{52}r_{23} + \hat{p}_{51}r_{13},$$

lembrando que o resíduo não é correlacionado com as variáveis independentes da equação de regressão.

Também:

$$(1.16) \quad r_{35} = \hat{p}_{53} + \hat{p}_{51}\hat{p}_{31} + \hat{p}_{51}r_{12}\hat{p}_{32} + \hat{p}_{52}\hat{p}_{32} + \hat{p}_{52}r_{12}\hat{p}_{31} + \\ + \hat{p}_{54}\hat{p}_{42}\hat{p}_{32} + \hat{p}_{54}\hat{p}_{42}r_{12}\hat{p}_{31} + \hat{p}_{54}\hat{p}_{43} + \hat{p}_{54}\hat{p}_{41}\hat{p}_{32}r_{12} + \\ + \hat{p}_{54}\hat{p}_{41}\hat{p}_{31}.$$

Uma observação importante a considerar é que se tivermos mais do que uma correlação em um dos diagramas, somente uma pode ser usada em um trajeto particular.

Um caso importante de (1.14) é considerado quando $i = j$ e é conhecido como fórmula para a determinação completa de X_i :

$$(1.17) \quad r_{ii} = 1 = \sum_q \hat{p}_{iq}r_{iq}$$

Desenvolvendo:

$$(1.18) \quad r_{ii} = 1 = \sum_q \hat{p}_{iq}^2 + 2 \sum_{q, q'} \hat{p}_{iq}r_{qq'}\hat{p}_{iq'}$$

onde q e q' ($q' > q$) incluem todas as variáveis aleatórias envolvidas, sejam ou não residuais.

Por (1.18) determinamos o estimador do coeficiente de trajetória residual.

1.4 - UMA INTERPRETAÇÃO PARA A FÓRMULA DA DETERMINAÇÃO COMPLETA

Seja o diagrama:

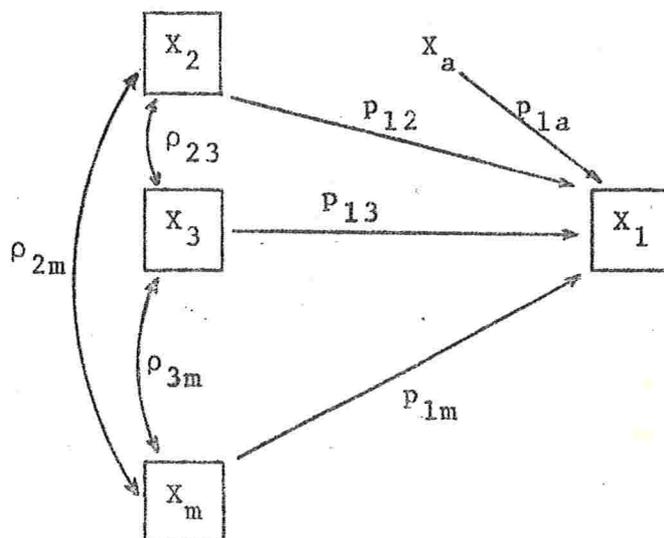


FIGURA (1.4)

$$(1.19) \quad Z_1 = \sum_{i=2}^m p_{1i} Z_i + p_{1a} Z_a$$

Por (1.18), temos:

$$(1.20) \quad 1 = r_{11} = \sum_{j=2}^m \hat{p}_{1j}^2 + 2 \sum_{j=2}^m \sum_{k=j+1}^m \hat{p}_{1j} \hat{p}_{1k} r_{jk} + \hat{p}_{1a}^2$$

Multiplicando-se por $\text{var}(X_1)$, temos:

$$(1.21) \quad \text{var}(X_1) = \text{var}(X_1) \sum_{j=2}^m \hat{p}_{1j}^2 + \\ + 2 \text{var}(X_1) \sum_{j=2}^m \sum_{k=j+1}^m \hat{p}_{1j} \hat{p}_{1k} r_{jk} + \text{var}(X_1) \hat{p}_{1a}^2.$$

Então:

$$(1.22) \quad \begin{pmatrix} \text{VARIACÃO} \\ \text{TOTAL DE} \\ X_1 \end{pmatrix} = \begin{pmatrix} \text{PROPORÇÃO} \\ \text{DADA DI-} \\ \text{RETAMENTE} \\ \text{PELOS ES-} \\ \text{TIMADORES} \\ \text{DOS COEFI} \\ \text{CIENTES DE} \\ \text{TRAJETÓRIA} \end{pmatrix} + \begin{pmatrix} \text{PROPORÇÃO} \\ \text{DEVIDO ÀS} \\ \text{INTERCOR-} \\ \text{-RELAÇÕES} \\ \text{ENTRE AS} \\ \text{VARIÁVEIS} \\ \text{INDEPEN} \\ \text{DENTES} \end{pmatrix} + \begin{pmatrix} \text{PROPORÇÃO} \\ \text{DEVIDO AO} \\ \text{RESÍDUO} \end{pmatrix}$$

1.5 - DECOMPOSIÇÃO DOS EFEITOS NA ANÁLISE DE TRAJETÓRIAS

A partir do teorema fundamental da análise de trajetórias, nós podemos decompor a correlação entre uma variável dependente e uma variável independente em efeito direto e efeitos indiretos. O primeiro é medido pela estimativa do coeficiente de trajetória correspondente e a soma dos outros é obtido por diferença.

Assim, no diagrama dado em (1.4), o efeito direto da variável X_2 em X_1 é dado por \hat{p}_{12} . O efeito indireto será dado por $r_{12} - \hat{p}_{12}$, onde r_{12} é determinado a partir do teorema fundamental.

Na expressão (1.16), temos:

$$(1.23) \quad \hat{p}_{53} \text{ é o efeito direto da variável } X_3 \text{ na variável } X_5$$

$$(1.24) \quad r_{35} - \hat{p}_{53} = \hat{p}_{51} \hat{p}_{31} + \hat{p}_{51} r_{12} \hat{p}_{32} + \hat{p}_{52} \hat{p}_{32} + \hat{p}_{52} r_{12} \hat{p}_{31} + \\ + \hat{p}_{54} \hat{p}_{42} \hat{p}_{32} + \hat{p}_{54} \hat{p}_{42} r_{12} \hat{p}_{31} + \hat{p}_{54} \hat{p}_{43} + \\ + \hat{p}_{54} \hat{p}_{41} \hat{p}_{32} r_{12} + \hat{p}_{54} \hat{p}_{41} \hat{p}_{31} \text{ será o efeito indi}$$

reto de X_3 em X_5 através de outras variáveis.

A expressão (1.24) mostra que o efeito indireto total é a soma de efeitos indiretos correspondentes às trajetórias entre X_3 e X_5 .

Além dos efeitos indiretos e efeitos diretos, temos os efeitos residuais determinados pelos estimadores dos coeficientes de trajetória residuais obtidos a partir de (1.18).

1.6 - ESTIMADOR DO COEFICIENTE DE TRAJETÓRIA RESIDUAL

Consideremos novamente o modelo de regressão dado em (1.19) associado ao diagrama da figura (1.4).

Neste exemplo, podemos escrever o estimador do coeficiente de trajetória residual por:

$$(1.25) \quad \hat{p}_{1u}^2 = 1 - \sum_{j=2}^m \hat{p}_{1j} r_{1j}$$

ou $\hat{p}_{1u} = \sqrt{1-R^2}$, onde R^2 é o coeficiente de correlação múltipla.

Aqui, aparece o coeficiente de correlação múltipla R^2 , porque:

$$(1.26) \quad R^2 = \frac{\sum_{j=2}^m \hat{p}_{1j} \sum_{i=1}^n z_{ji} z_{1i} - n \bar{z}_1^2}{\sum_{i=1}^n z_{1i}^2 - n \bar{z}_1^2}$$

onde n é o número de observações, e

$$(1.27) \quad R^2 = \frac{\sum_{j=2}^m \hat{\beta}_{1j} \sum_{i=1}^n (X_{ji} - \bar{X}_j)(X_{1i} - \bar{X}_1)}{nS_1^2}$$

e daí,

$$(1.28) \quad R^2 = \sum_{j=2}^m \hat{p}_{1j} r_{j1}$$

1.7 - COMENTÁRIOS

O coeficiente de trajetórias é igual ao coeficiente de correlação quando se trata de um modelo de regressão simples.

Assim, vale a desigualdade:

$$(1.29) \quad -1 \leq p_{ij} \leq 1$$

Quando temos modelos de regressão múltiplas, os estimadores dos coeficientes de trajetórias podem estar fora do intervalo $[-1,+1]$.

Quando temos casos como este, o efeito direto de uma variável X_j em uma variável X_i , deve ser compensado pelos efeitos indiretos através de outros fatores, isto é, esses efeitos devem ser de sinal oposto ao sinal de \hat{p}_{ij} .

1.8 - UM DIAGRAMA ADEQUADO

Para um pesquisador verificar se um diagrama de trajetórias é adequado, basta comparar os valores das correlações calculadas a partir dos dados, com as correlações calculadas a partir do teorema fundamental da análise de trajetórias.

O diagrama será adequado se as correlações forem iguais.

Tomamos como exemplo, o diagrama de trajetórias abaixo:

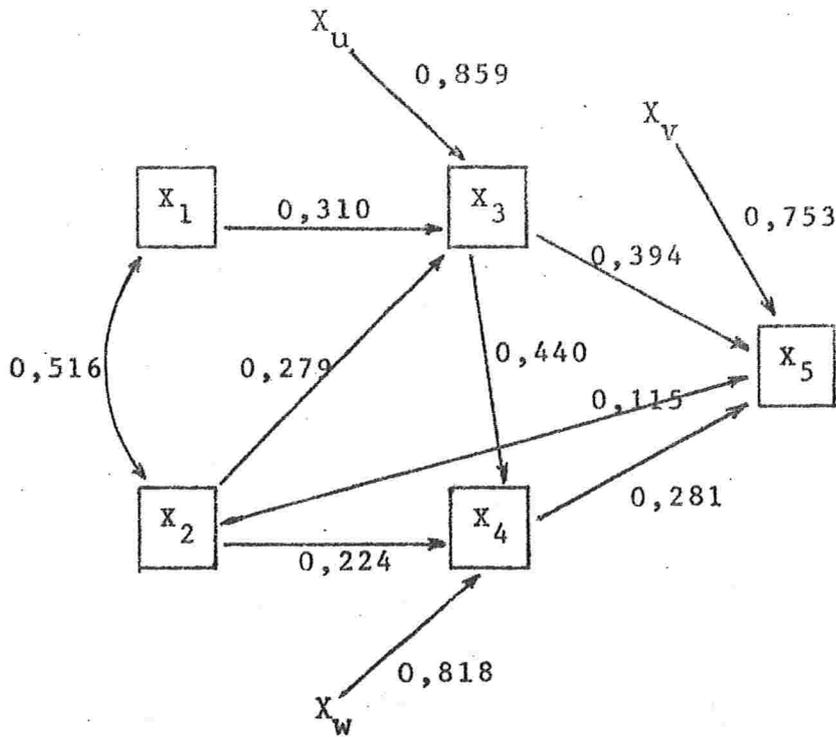


FIGURA (1.5)

O coeficiente de correlação de Pearson r_{14} é igual a 0,332.

Pelo teorema fundamental, podemos escrever:

$$(1.30) \quad r_{14} = \hat{p}_{42}r_{12} + \hat{p}_{43}r_{13} = (0,224)(0,516) + (0,440)(0,455)$$

ou

$$r_{14} = 0,315$$

Analogamente $r_{15} = 0,326$ pelo teorema fundamental e $r_{15} = 0,322$ determinado a partir dos dados.

Outras comparações, por exemplo r_{25} , revelam discrepâncias muito pequenas (não maiores que 0,001).

Notamos então que o diagrama é adequado.

CAPÍTULO II

ANÁLISE DE TRAJETÓRIAS EM MODELOS DE REGRESSÃO SIMPLES

2.1 - O MODELO

Vamos considerar agora um modelo de regressão simples, associado ao diagrama da figura (2.1)

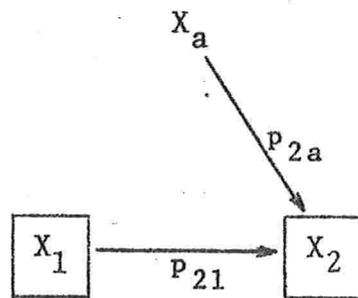


FIGURA (2.1)

$$(2.1) \quad X_2 = \alpha + \beta X_1 + \epsilon$$

ao qual fazemos corresponder o seguinte modelo da trajetória:

$$(2.2) \quad Z_2 = p_{21}Z_1 + p_{2a}Z_a$$

onde

$$Z_i = \frac{X_i - \bar{X}_i}{S_i}, \quad i=1,2 \text{ e } p_{2a} Z_a = \epsilon$$

2.2 - ESTIMAÇÃO DOS COEFICIENTES DE TRAJETÓRIAS

Os estimadores dos mínimos quadrados dos parâmetros α e β são dados por:

$$(2.3) \quad \hat{\alpha} = \bar{X}_2 - \hat{\beta} \bar{X}_1$$

$$(2.4) \quad \hat{\beta} = \frac{\sum_{i=1}^n (X_{2i} - \bar{X}_2)(X_{1i} - \bar{X}_1)}{\sum_{i=1}^n (X_{1i} - \bar{X}_1)^2}$$

Pela expressão dada em (1.7), podemos escrever um estimador do coeficiente de trajetória p_{21} da seguinte forma:

$$(2.5) \quad \hat{p}_{21} = \frac{\sum_{i=1}^n (X_{2i} - \bar{X}_2)(X_{1i} - \bar{X}_1)}{\sqrt{\sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 \sum_{i=1}^n (X_{2i} - \bar{X}_2)^2}}$$

Vemos que $\hat{p}_{21} = r_{21}$, isto é, o estimador do coeficiente de trajetória coincide numericamente com a estimativa do coeficiente de correlação de Pearson entre as variáveis X_1 e X_2 .

Podemos também estimar p_{21} aplicando diretamente o método dos mínimos quadrados na equação de trajetória (2.2).

Neste caso obtemos para estimador de p_{21} a expressão abaixo:

$$(2.6) \quad \hat{p}_{21} = \frac{\sum_{i=1}^n Z_{1i} Z_{2i}}{\sum_{i=1}^n Z_{1i}^2}$$

Observamos que as expressões (2.6) e (2.5) são idênticas. Isto mostra que neste caso os estimadores de mínimos quadrados gozam da propriedade de invariância a qual é satisfeita pelos estimadores de máxima verossimilhança.

O estimador do coeficiente de trajetória residual é obtido usando a fórmula da determinação completa. Portanto,

$$(2.7) \quad \hat{p}_{2a} = \sqrt{1 - \hat{p}_{21}^2}$$

2.3 - DISTRIBUIÇÃO AMOSTRAL DE \hat{p}_{21} QUANDO X_1 E X_2 SÃO VARIÁVEIS ALEATÓRIAS

Vamos supor que as variáveis aleatórias X_1 e X_2 tenham uma distribuição normal bivariada com média

$$\underline{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

e matriz de covariância

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho_{12} \\ \sigma_2 \sigma_1 \rho_{21} & \sigma_2^2 \end{pmatrix}$$

Desde que \hat{p}_{21} coincide com o estimador do coeficiente de correlação de Pearson, a distribuição amostral de \hat{p}_{21} é aquela do estimador do coeficiente de correlação.

Devemos considerar dois casos: $p_{21} = 0$ ou $p_{21} \neq 0$.

2.3.1 - DISTRIBUIÇÃO AMOSTRAL DE \hat{p}_{21} quando $p_{21} = 0$

Sejam N observações independentes dadas por $(X_{11}, X_{21}), \dots, (X_{1N}, X_{2N})$, ou na forma vetorial $\underline{X}_1^*, \dots, \underline{X}_N^*$, onde cada vetor \underline{X}_i^* , $i=1, \dots, N$ tem uma distribuição normal bivariada $N(\underline{u}, \Sigma)$.

Segundo Anderson [1], se $p_{21} = \rho_{21} = 0$, a função densidade de probabilidade de \hat{p}_{21} é dada por:

$$(2.8) \quad f(\hat{p}_{21}) = \frac{\Gamma\left[\frac{1}{2}(N-1)\right]}{\Gamma\left[\frac{1}{2}(N-2)\right]\sqrt{\pi}} (1-\hat{p}_{21}^2)^{\frac{(N-4)}{2}}$$

A estatística T, dada abaixo, tem uma distribuição t de Student com N-2 graus de liberdade.

$$(2.9) \quad T = \frac{\sqrt{N-2} \hat{p}_{21}}{\sqrt{1-\hat{p}_{21}^2}}$$

Assim, hipótese do tipo $H_0: p_{21} = 0$ contra alternativas do tipo $H_a: p_{21} \neq 0$, $p_{21} > 0$ ou $p_{21} < 0$, podem ser testadas da maneira usual.

2.3.2 - Distribuição de \hat{p}_{21} , quando $p_{21} \neq 0$

Da mesma maneira anterior, sejam N observações independentes X_1^* , ..., X_N^* , cada com uma distribuição normal bivariada $N(\underline{\mu}, \Sigma)$.

Também por Anderson [1], \hat{p}_{21} quando $p_{21} \neq 0$, tem uma distribuição dada pela densidade:

$$(2.10) \quad f(\hat{p}_{21}) = \frac{2^{n-2} (1-p_{21}^2)^{\frac{1}{2}n} (1-\hat{p}_{21}^2)^{\frac{1}{2}(n-3)}}{(n-2)! \pi} \sum_{\alpha=0}^{\infty} \frac{(2p_{21}\hat{p}_{21})^{\alpha}}{\alpha!} \Gamma\left(\frac{n+\alpha}{2}\right)$$

onde $n = N-1$.

A função distribuição é dada por:

$$(2.11) \quad F(r_0) = P\{r_{12} \leq r_0\} = F(r_0 | N, \rho_{12})$$

Essa função distribuição foi tabulada por F.N. David (1938), para

$$p_{21} = \rho_{12} = 0 \quad (0,1) \quad 0,9$$

$$N = 3(1)25, 50, 100, 200, 400$$

$$r_0 = -1(0,05)1.$$

Assim para testar $H_0: p_{21} = \rho_0$ contra $H_a: p_{21} > \rho_0$, rejeitamos H_0 se $\hat{p}_{21} > r_0$, e r_0 é determinado a partir da ta bela de David por meio da relação

$$1 - F(r_0 | N, \rho_0) = \alpha.$$

Para alternativas do tipo $p_{21} \neq \rho_0$ e $p_{21} < \rho_0$, o procedimento é análogo.

A partir da tabela de David, também podemos determinar a função poder do teste e construir regiões de confiança para p_{21} .

2.4 - DISTRIBUIÇÃO DO ESTIMADOR DE p_{21} QUANDO X_1 NÃO É VARIÁVEL ALEATÓRIA

Vamos supor que X_2 tem distribuição normal com média μ e variância σ^2 , e X_1 é uma variável não aleatória.

Se usarmos como estimador de p_{21} , aquele definido em (2.5), não podemos determinar a sua distribuição amostral da mesma maneira exposta na secção (2.3), porque neste caso \hat{p}_{21} não pode ser considerado um estimador do coeficiente de correlação de uma distribuição normal bivariada.

A distribuição amostral de \hat{p}_{21} , neste caso deve ser obtida usando apenas o fato de que X_2 é uma variável aleatória com distribuição normal.

Para este caso, Klemmack, D.L., e outros [10] sugeriram como estimador de p_{21} , a expressão abaixo:

$$(2.12) \quad \hat{p}_{21}^* = \frac{S_1}{\sqrt{\frac{SSE}{n-2}}} \hat{\beta}$$

onde

$$S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_{1i} - \bar{X}_1)^2,$$

$\hat{\beta}$ é o estimador dos mínimos quadrados de β , SSE é a soma dos quadrados residual e n é o número de observações.

Embora neste artigo não seja discutido a distribuição amostral de \hat{p}_{21}^* , esta pode ser obtida por meio de transformação de variáveis.

Podemos escrever estimador do coeficiente de trajetória \hat{p}_{21}^* , na forma:

$$(2.13) \quad \hat{p}_{21}^* = \frac{\sqrt{n-2} \bar{X}_1 \cdot (\bar{X}_2 - X_{21})}{n S_1 \sqrt{SSE}}$$

onde

$$\underline{X}_1 = \begin{pmatrix} X_{11} \\ X_{12} \\ \cdot \\ \cdot \\ \cdot \\ X_{1n} \end{pmatrix}, \quad \underline{X}_2 = \begin{pmatrix} X_{21} \\ X_{22} \\ \cdot \\ \cdot \\ \cdot \\ X_{2n} \end{pmatrix}, \quad \underline{1} = \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}$$

e

$$\underline{X}_1 \cdot (\underline{X}_2 - \bar{X}_2 \underline{1}) = \sum_{i=1}^n X_{1i} (X_{2i} - \bar{X}_2). \quad (*)$$

Considere os seguintes fatos:

Se \underline{X}_2 tem distribuição $N(u_2 \underline{1}, I\sigma^2)$, $u_2 = \alpha + \beta X_1$, então $\underline{X}_2 - \bar{X}_2 \underline{1}$ tem distribuição $N(0, \Sigma)$, onde

$$\Sigma = (\sigma_{ij}), \quad i, j = 1, 2, \dots, n, \quad \sigma_{ij} = \frac{\sigma^2(n-1)}{n},$$

para $i = j$ e $\sigma_{ij} = -\sigma^2/n$ para $i \neq j$. (Apêndice 1).

Se \underline{X} tem distribuição $N(\underline{u}, \Sigma)$, então $\underline{Z} = D\underline{X}$ tem distribuição $N(D\underline{u}, D\Sigma D')$ onde D é uma matriz $q \times p$ de rank $q \leq n$ (Searle [14]).

Então $\underline{X}_1 \cdot (\underline{X}_2 - \bar{X}_2 \underline{1})$ tem distribuição $N(0, \underline{X}_1' \Sigma \underline{X}_1)$ onde:

(*) - Usaremos no decorrer deste trabalho, a notação acima para produto escalar de vetores.

$$\underline{X}_1' \Sigma \underline{X}_1 = \sum_{j=1}^n X_{1j} \cdot \sum_{i=1}^n X_{1i} \sigma_{ij}$$

Como

$$\sum_{i=1}^n X_{1i} \sigma_{ij} = \sigma^2 (X_{1j} - \bar{X}_1), \quad j=1, 2, \dots, n,$$

temos:

$$(2.14) \quad \sum_{j=1}^n X_{1j} \sum_{i=1}^n X_{1i} \sigma_{ij} = n\sigma^2 S_1^2$$

Logo, $\underline{X}_1 \cdot (\underline{X}_2 - \bar{X}_2 \underline{1})$ tem distribuição $N(0, n\sigma^2 S_1^2)$.

Como $U = \underline{X}_1 \cdot (\underline{X}_2 - \bar{X}_2 \underline{1}) / \sqrt{n} \sigma S_1$ tem distribuição $N(0, 1)$ e $V = SSE / \sigma^2$ tem distribuição χ_{n-2}^2 não central (Searle [14]) e U e V são independentes (Apêndice 1), então

$$T = U / \sqrt{V / \sigma^2}$$

tem uma distribuição t de Student com $(n-2)$ graus de liberdade.

Portanto,

$$(2.15) \quad T = \sqrt{n} \hat{p}_{21}^*$$

tem distribuição de Student com $(n-2)$ graus de liberdade.

Assim para testar $H_0: p_{21} = 0$ contra alternativas da forma $H_a: p_{21} = 0, p_{21} > 0$ ou $p_{21} < 0$, usamos a estatística T dada acima.

CAPÍTULO 3

ANÁLISE DE TRAJETÓRIAS EM MODELOS
DE REGRESSÃO MÚLTIPLA

3.1 - O MODELO

Seja o modelo de regressão linear múltipla associado ao diagrama da figura (3.1):

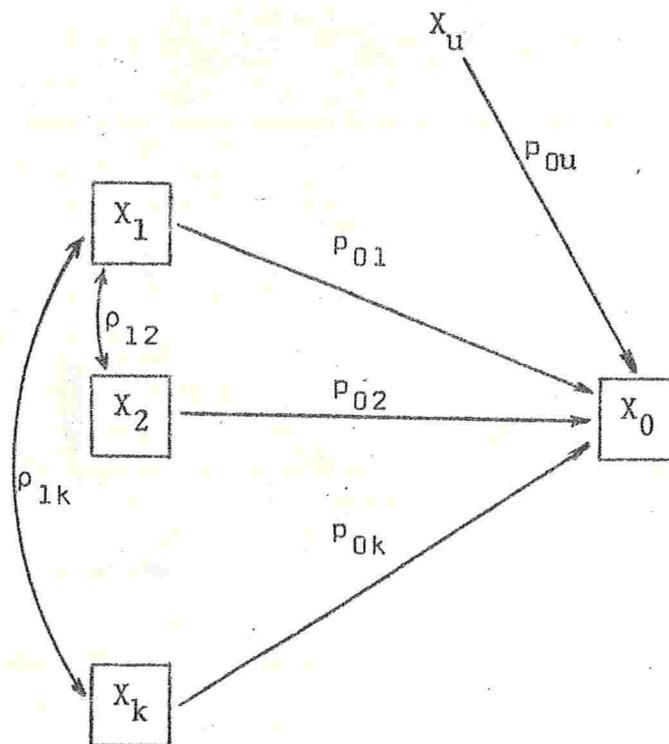


FIGURA (3.1)

$$(3.1) \quad X_0 = \alpha_1 + \sum_{i=1}^k \beta_{0i} X_i + \epsilon,$$

ou

$$\underline{X}_0 = X\underline{\beta} + \underline{\epsilon},$$

onde

$$\underline{X}_0 = \begin{pmatrix} X_{01} \\ X_{02} \\ \cdot \\ \cdot \\ \cdot \\ X_{0n} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_{11} & \dots & X_{k1} \\ 1 & X_{12} & \dots & X_{k2} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 1 & X_{1n} & \dots & X_{kn} \end{pmatrix}, \quad \underline{\beta} = \begin{pmatrix} \alpha_1 \\ \beta_{01} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{0k} \end{pmatrix}, \quad \underline{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{pmatrix}$$

e n é o número de observações.

À equação de regressão (3.1) fazemos corresponder o seguinte modelo de trajetória:

$$(3.2) \quad Z_0 = \sum_{i=1}^k p_{0i} Z_i + p_{0u} Z_u,$$

onde

$$Z_i = \frac{X_i - \bar{X}_i}{S_i}, \quad i=0,1,\dots,k$$

e

$$p_{0u} Z_u = \epsilon$$

Ou,

$$\underline{z}_0 = Z\underline{p} + \underline{\varepsilon},$$

onde

$$\underline{z}_0 = \begin{pmatrix} z_{01} \\ \vdots \\ z_{0n} \end{pmatrix}, \quad Z = \begin{pmatrix} z_{11} & \cdots & z_{k1} \\ \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots \\ z_{1n} & \cdots & z_{kn} \end{pmatrix}, \quad \underline{p} = \begin{pmatrix} p_{01} \\ \vdots \\ p_{0k} \end{pmatrix}, \quad \underline{\varepsilon} = \begin{pmatrix} p_{0u} z_{u1} \\ \vdots \\ p_{0u} z_{un} \end{pmatrix}$$

3.2 - ESTIMAÇÃO DOS COEFICIENTES DE TRAJETÓRIA

Os estimadores de mínimos quadrados dos coeficientes de regressão em (3.1) são as soluções das equações normais abaixo:

$$(3.3) \quad \begin{pmatrix} n & \sum_1^n X_{1j} & \sum_1^n X_{2j} & \cdots & \sum_1^n X_{kj} \\ \sum_1^n X_{1j} & \sum_1^n X_{1j}^2 & \sum_1^n X_{1j} X_{2j} & \cdots & \sum_1^n X_{1j} X_{kj} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_1^n X_{kj} & \sum_1^n X_{kj} X_{1j} & \cdots & \sum_1^n X_{kj}^2 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\beta}_{01} \\ \vdots \\ \hat{\beta}_{0k} \end{pmatrix} = \begin{pmatrix} \sum_1^n X_{0j} \\ \sum_1^n X_{1j} X_{0j} \\ \vdots \\ \sum_1^n X_{kj} X_{0j} \end{pmatrix}$$

Usando a expressão (1.7) os estimadores dos coeficientes de trajetória p_{0i} tem a seguinte forma:

$$(3.4) \quad \hat{p}_{0i} = \frac{S_i}{S_0} \hat{\beta}_{0i}, \quad i=1, \dots, k.$$

De maneira análoga aquela considerada no capítulo II, nós obtemos o estimador do coeficiente de trajetória residual por meio da expressão:

$$(3.5) \quad \hat{p}_{0u} = \sqrt{1 - \left(\sum_{j=1}^k \hat{p}_{0j}^2 + 2 \sum_{j=1}^k \sum_{\ell=j+1}^k \hat{p}_{0j} \hat{p}_{0\ell} r_{j\ell} \right)}$$

Podemos, também estimar os coeficientes de trajetória p_{0i} pelo método de mínimos quadrados aplicado diretamente ao modelo de trajetória (3.2).

De acordo com demonstrações feitas no apêndice 2, observamos que coincidem com aqueles obtidos pelo método acima exposto.

Portanto os estimadores de mínimos quadrados gozam neste caso da propriedade de invariância que é satisfeita pelos estimadores de máxima verossimilhança.

O estimador de trajetória residual é obtido da mesma maneira anterior.

3.3 - RELAÇÃO ENTRE OS COEFICIENTES DE TRAJETÓRIA E OS COEFICIENTES DE CORRELAÇÃO PARCIAIS

Consideremos agora, o modelo de regressão na forma:

$$(3.6) \quad E(X_0 | X_1, \dots, X_k) = \sum_{i=1}^k \beta_{0i.qi} X_i,$$

onde as variáveis X_0, X_1, \dots, X_k são variáveis aleatórias com uma distribuição normal multivariada dada por:

$$(3.7) \quad f(x_0, x_1, \dots, x_k) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{1/2p}} e^{-1/2(\underline{x}-\underline{u})' \Sigma^{-1} (\underline{x}-\underline{u})}$$

onde

$$\underline{X} = \begin{pmatrix} X_0 \\ \vdots \\ X_k \end{pmatrix}, \quad \underline{u} = \begin{pmatrix} u_0 \\ \vdots \\ u_k \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{00} & \sigma_{01} & \dots & \sigma_{0k} \\ \sigma_{10} & \sigma_{11} & \dots & \sigma_{1k} \\ \dots & \dots & \dots & \dots \\ \sigma_{k0} & \sigma_{k1} & \dots & \sigma_{kk} \end{pmatrix}$$

e

$$\beta_{0i.qi} = \beta_{0i.12\dots(i-1)(i+1)\dots k}, \quad i=1, \dots, k,$$

são coeficientes de regressão parciais.

O erro da regressão linear é dado por:

$$(3.8) \quad (\beta_{0u.qu} X_u) = X_{0.12\dots k}$$

onde $E(X_{0.12\dots k}) = 0$ e $\text{var}(X_{0.12\dots k}) = \sigma_{0.12\dots k}^2$

Podemos escrever o coeficiente de regressão parcial na forma abaixo

$$(3.9) \quad \beta_{0i.12\dots(i-1)(i+1)\dots k} = \frac{\sigma_{01.12\dots(i-1)(i+1)\dots k}}{\sigma_{i.12\dots(i-1)(i+1)\dots k}^2}$$

(APÊNDICE 3)

Assim podemos escrever o coeficiente de trajetória por:

$$(3.10) \quad \rho_{0i} = \frac{\sigma_{0i.12\dots(i-1)(i+1)\dots k}}{\sigma_{i.12\dots(i-1)(i+1)\dots k}^2} \frac{\sigma_i}{\sigma_0}$$

Desde que,

$$\begin{aligned} & \sigma_{0i.12\dots(i-1)(i+1)\dots k} = \\ & = \sigma_{0.12\dots(i-1)(i+1)\dots k} \sigma_{i.12\dots(i-1)(i+1)\dots k} \rho_{0i.12\dots(i-1)(i+1)\dots k} \end{aligned}$$

(apêndice 3), então:

$$(3.11) \quad \rho_{0i} = \frac{\sigma_{0.12\dots(i-1)(i+1)\dots k}}{\sigma_{i.12\dots(i-1)(i+1)\dots k}} \frac{\sigma_i}{\sigma_0} \rho_{0i.12\dots(i-1)(i+1)\dots k}$$

ou,

$$(3.12) \quad \rho_{0i} = \frac{\frac{\sigma_{0.12\dots(i-1)(i+1)\dots k}}{\sigma_0}}{\frac{\sigma_{i.12\dots(i-1)(i+1)\dots k}}{\sigma_i}} \rho_{0i.12\dots(i-1)(i+1)\dots k}$$

Sendo $R_0^2(12\dots k)$ o coeficiente de correlação múltipla entre a variável X_0 e as variáveis X_1, X_2, \dots, X_k , temos:

$$(3.13) \quad 1 - R_0^2(12\dots k) = \frac{\sigma_{0.12\dots k}^2}{\sigma_0^2} \quad (\text{Apêndice 4})$$

Assim,

$$(3.14) \quad p_{0i} = \sqrt{\frac{1-R_0^2(12\dots(i-1)(i+1)\dots k)}{1-R_1^2(12\dots(i-1)(i+1)\dots k)}} \rho_{0i.12\dots(i-1)(i+1)\dots k}$$

Em particular para $i=1$, encontramos:

$$(3.15) \quad p_{01} = \sqrt{\frac{1-R_0^2(23\dots k)}{1-R_1^2(23\dots k)}} \rho_{01.23\dots k}$$

Ou, relacionando o coeficiente de trajetória p_{01} com os coeficientes de correlação parciais:

$$(3.16) \quad p_{01} = \sqrt{\frac{(1-\rho_{02}^2)(1-\rho_{03.2}^2)\dots(1-\rho_{0k.23\dots(k-1)}^2)}{(1-\rho_{12}^2)(1-\rho_{13.2}^2)\dots(1-\rho_{1k.23\dots(k-1)}^2)}} \rho_{01.23\dots k}$$

OBSERVAÇÃO: As expressões (3.15) e (3.16) foram obtidas apenas para comprovação, desde que elas se acham no artigo de Wright [17].

3.4 - MODELOS DE TRAJETÓRIA QUANDO AS VARIÁVEIS INDEPENDENTES SÃO NÃO CORRELACIONADAS

Seja agora um modelo de trajetória associado ao diagrama dado na figura (3.1), também com as variáveis X_0, X_1, \dots, X_k consideradas como variáveis aleatórias com uma distribuição normal multivariada, mas com as variáveis independentes não-correlacionadas entre si.

Neste caso os coeficientes de trajetória p_{0i} , coincidem com os coeficientes de correlação ρ_{0i} (Apêndice 5). Isto significa que toda influência de cada uma das variáveis causa sobre a variável efeito é medida pelo coeficiente de

trajetória. Portanto os estimadores dos coeficientes de trajetória p_{0i} são dados por r_{0i} , estimadores dos coeficientes de correlação.

Resta agora encontrar um estimador para o coeficiente de trajetória residual p_{0u} :

Para isto, nós vamos desenvolver certas relações que serão importantes para a obtenção do estimador do coeficiente de trajetória residual.

A partir da equação de trajetória (3.2), podemos escrever:

$$(3.17) \quad \text{var}(Z_0) = \text{var}\left(\sum_{i=1}^k p_{0i} Z_i + p_{0u} Z_u\right),$$

ou

$$p_{0u}^2 = 1 - \sum_{i=1}^k p_{0i} \rho_{0i}.$$

Sendo $C = (\rho_{ij})$, $(i, j=0, 1, \dots, k)$, a matriz de correlação e C_{ij} $(i, j=0, 1, \dots, k)$ seus cofatores, podemos escrever:

$$(3.18) \quad |C| = \det(C) = \sum_{i=0}^k \rho_{0i} C_{0i}$$

Desde que $\beta_{0i, q_i} = -\sigma_0 C_{0i} / \sigma_i C_{00}$ (apêndice 3), podemos escrever:

$$(3.19) \quad p_{0u}^2 = 1 + \sum_{i=1}^k \frac{\sigma_i \sigma_0}{\sigma_0 \sigma_i} \frac{C_{0i}}{C_{00}} \rho_{0i}$$

Ou,

$$(3.20) \quad p_{0u}^2 = \frac{C_{00} + \sum_{i=1}^k C_{0i} \rho_{0i}}{C_{00}} = \frac{1}{C_{00}} \sum_{i=0}^k C_{0i} \rho_{0i},$$

pois $\rho_{00} = 1$.

Assim, temos a partir de (3.18):

$$(3.21) \quad p_{0u}^2 = \frac{\det(C)}{C_{00}} = \frac{|C|}{C_{00}}$$

Como podemos escrever o coeficiente de correlação múltipla $R_0^2(12\dots k)$ pela expressão $1 - R_0^2(12\dots k) = |C|/C_{00}$, então podemos escrever:

$$(3.22) \quad p_{0u}^2 = 1 - R_0^2(12\dots k)$$

Como estimador do coeficiente de trajetória residual p_{0u} , temos:

$$(3.23) \quad \hat{p}_{0u} = \frac{\sqrt{\sum_{j=1}^n (X_{0j} - \sum_{i=1}^k \hat{\beta}_{0i} \cdot q_i X_{ij})^2}}{\sqrt{\sum_{j=1}^n (X_{0j} - \bar{X}_0)^2}},$$

pois

$$1 - \hat{R}_0^2(12\dots k) = S_{0.12\dots k} / S_0^2$$

e

$$nS_{0.12\dots k}^2 = \sum_{j=1}^n (X_{0j} - \sum_{i=1}^k \hat{\beta}_{0i \cdot qi} X_{ij})^2$$

Os estimadores dos coeficientes de regressão $\hat{\beta}_{0i \cdot qi}$ são obtidos pelo método de mínimos quadrados.

3.5 - DISTRIBUIÇÃO AMOSTRAL DE \hat{p}_{0i}

Temos dois casos a considerar: o primeiro é quando as variáveis independentes são não correlacionadas entre si e o segundo é quando as variáveis independentes são correlacionadas entre si.

Para o primeiro caso vimos que o coeficiente de trajetória p_{0i} coincide com o coeficiente de correlação ρ_{0i} da distribuição marginal de X_0 e X_i .

Por outro lado, se

$$\underline{X} = \begin{pmatrix} X_1 \\ \sim 1 \\ X_2 \\ \sim 2 \end{pmatrix}$$

tem distribuição $N(\underline{u}, \Sigma)$ com

$$\underline{u} = \begin{pmatrix} u_1 \\ \sim 1 \\ u_2 \\ \sim 2 \end{pmatrix} \quad \text{e} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

tem distribuição marginal de X_1 é $N(u_1, \Sigma_{11})$ (Anderson [1]).

Portanto dentro da suposição de normalidade, a distribuição amostral de \hat{p}_{0i} coincide com a distribuição amostral do estimador do coeficiente de correlação de uma dis-

tribuição bivariada normal. Esta distribuição foi apresentada no capítulo 2 desta monografia.

Para o segundo caso, vimos que o coeficiente de trajetória está relacionado com os coeficientes de correlação parciais por meio da expressão (3.16).

Sabemos também que os coeficientes de correlação parciais $\rho_{0i,12\dots(i-1)(i+1)\dots k}$ são coeficientes de correlação de distribuições condicionais de X_0 e X_i dados $X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_k$.

Por outro lado se

$$\underline{X} = \begin{pmatrix} X_1 \\ \underline{x}_1 \\ X_2 \\ \underline{x}_2 \end{pmatrix}$$

tem distribuição $N(\underline{u}, \Sigma)$ com

$$\underline{u} = \begin{pmatrix} \underline{u}_1 \\ \underline{u}_2 \end{pmatrix} \quad \text{e} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

então a distribuição condicional de X_1 dado $X_2 = \underline{x}_2$ é normal com média

$$\underline{u}_1 + \Sigma_{22}^{-1} \Sigma_{21} (\underline{x}_2 - \underline{u}_2)$$

e matriz de covariância

$$\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

(Anderson [1]).

Portanto dentro da suposição de normalidade, a distribuição amostral do estimador de cada coeficiente de cor-

relação parcial está completamente determinada.

No entanto a distribuição do estimador do coeficiente de trajetória para este caso, não é apresentada na literatura citada neste trabalho.

CAPÍTULO 4

APLICAÇÕES

4.1 - APLICAÇÕES DA ANÁLISE DE TRAJETÓRIAS EM GENÉTICA

A análise de trajetórias é utilizada em genética na determinação das várias relações existentes entre os pais e seus descendentes, e na determinação das relações existentes entre cruzamentos de indivíduos correlacionados.

De uma maneira sucinta, encontramos todos esses resultados em Li [12], capítulos 13 e 14.

Apenas citando um exemplo, sem entrar em detalhes, das relações existentes entre os pais e seus descendentes, que estão apresentados no livro acima citado, tomamos a trajetória do zigoto determinado pela união de dois gametas.

O zigoto é considerado como linearmente, completamente e igualmente determinado pela união de dois gametas. Esta relação pode ser representada pelo esquema causal simples (figura 4.1):

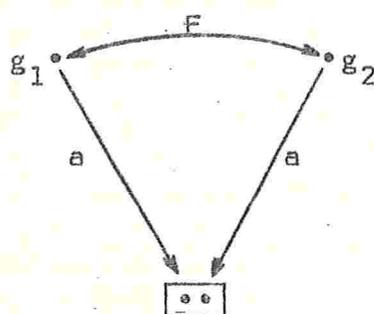


FIGURA (4.1)

$$Z \text{ (zigoto)} = g_1 \text{ (esperma)} + g_2 \text{ (ovo)}$$

Desde que Z é igualmente determinado por g_1 e g_2 para genes autossomais, os dois coeficientes, um do ovo para o zigoto e o outro do esperma para o zigoto devem ser iguais. Usando a notação de Wright, seja a o coeficiente de trajetória do gameta (ovo ou esperma) para o zigoto e seja F o coeficiente de correlação entre os gametas. Pela fórmula da de terminação completa (1.18), temos:

$$(4.1) \quad 2a^2 + 2a^2F = 1$$

Assim,

$$(4.2) \quad a^2 = \frac{1}{2(1+F)},$$

e

$$a = \sqrt{\frac{1}{2(1+F)}}$$

Assim encontramos por meio de uma aplicação da análise de trajetória uma geração genética importante.

4.2 - APLICAÇÃO DA ANÁLISE DE TRAJETÓRIAS NO ESTUDO DE INFLUÊNCIA DE ALGUNS FATORES SOBRE A TAXA SALARIAL DE PESSOAL ESPECIALIZADO DAS INDÚSTRIAS DE SÃO PAULO

A pesquisa foi baseada em uma amostra de 22.587 trabalhadores especializados em 688 indústrias tomadas em 11 setores mais produtivos de São Paulo.

Foram estudados os seguintes fatores:

1. Educação (E), medida por meio de uma escala proporcional ao número de anos de escolaridade.
2. Nível da influência ocupacional (I) medida por uma escala proporcional a uma hierarquia estabelecida na empresa.
3. Idade (A), medida em anos.
4. Tempo de serviço (S).
5. Anos no trabalho atual (J).
6. Setor Industrial: uma variável controle incluída para determinar os efeitos das variáveis pelo setor considerado.

Foi utilizada a análise de trajetórias para medir as influências dos fatores acima descritos, na taxa salarial (W) do pessoal especializado, medida pela soma de todos os salários e benefícios dividido pelo número de horas que o funcionário trabalhou para a indústria durante o ano.

O modelo causal considerado é dado pelo diagrama de trajetórias abaixo:

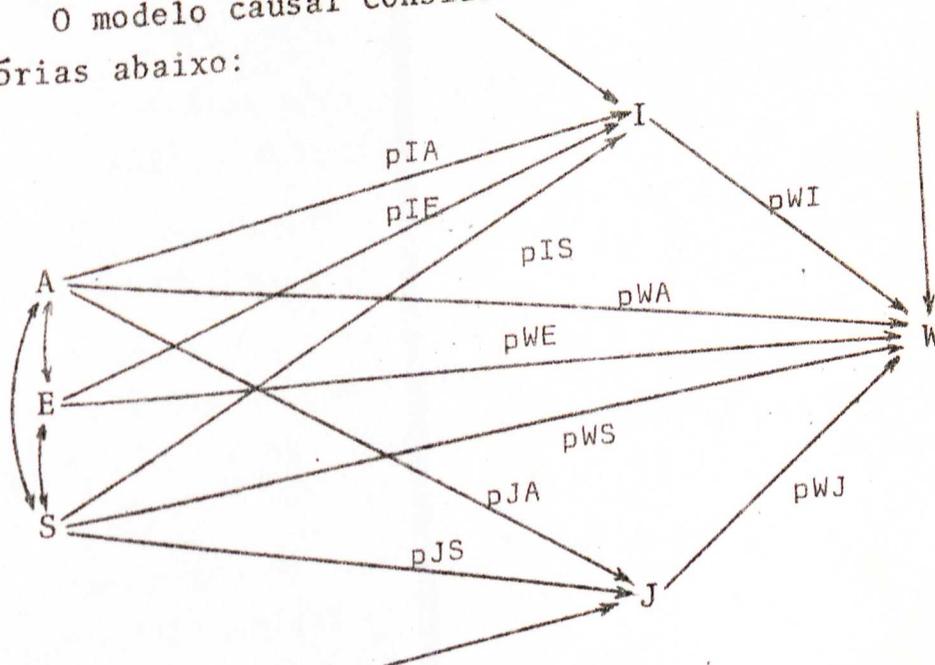


FIGURA (4.2)

W: taxa salarial
I: nível de influência ocupacional
J: anos no trabalho atual
A: idade
E: educação
S: tempo de serviço

Este modelo foi construído a partir de fatos conhecidos ou supostos pelo pesquisador. Assim, o nível de influência ocupacional (I) foi considerado como tendo um efeito direto na variável dependente. A educação (E) foi considerada como uma variável independente tomada como medida de competência treinada, afetando diretamente o salário e também indiretamente em razão de sua posição na empresa. A idade (A) foi considerada como uma variável indicando uma experiência geral tendo efeito direto e indiretos através da posição na empresa e da experiência no trabalho.

O tempo de serviço (S) também foi considerado como uma variável independente tendo mesmo tipo de efeitos que a idade.

A variável anos no trabalho atual (J) foi considerada como tendo apenas efeito direto sobre o salário.

Um sumário detalhado de toda análise de trajetórias está apresentado na tabela (4.1) às páginas 41 e 42.

Notamos que para a amostra total, os efeitos diretos maiores sobre a taxa salarial (W), são devido à educação ($p_{WE} = 0,529$) e ao nível de influência ocupacional ($p_{WI} = 0,323$).

Em termos de maiores efeitos totais, temos a educação ($\pi_{WE} = 0,459$), nível de influência ocupacional ($\pi_{WI} = 0,323$) e a idade ($\pi_{WA} = 0,314$). Notamos também, que o tempo de serviço e os anos de trabalho atual tem efeitos totais

SETOR	VARIÁVEL i	EFEITO TOTAL DA i-ésima VARIÁVEL ($\pi_{\omega i}$)	EFEITO DIRETO DA i-ésima VARIÁVEL ($p_{\omega i}$)	EFEITO INDIRETO	TAMANHO DA AMOSTRA
TODOS OS SETORES	I	0,323	0,323	-	22.587
	J	-0,059	-0,059	-	
	A	0,314	0,258	0,056	
	E	0,459	0,329	0,130	
	S	0,047	0,069	0,022	
PRODUÇÃO DE ALIMENTOS	I	0,316	0,316	-	1.341
	J	-0,103	-0,103	-	
	A	0,258	0,249	0,027	
	E	0,482	0,347	0,135	
	S	-0,033	0,007	-0,040	
VESTUÁRIO e SAPATOS	I	0,253	0,253	-	183
	J	-0,107	-0,107	-	
	A	0,167	0,158	0,009	
	E	0,541	0,433	0,108	
	S	0,168	0,240	-0,072	
TÉXTEIS	I	0,343	0,343	-	1.578
	J	0,014	0,014	-	
	A	0,310	0,272	0,038	
	E	0,426	0,279	0,147	
	S	-0,056	-0,062	0,006	
QUÍMICOS	I	0,311	0,311	-	1.418
	J	0,094	0,094	-	
	A	0,356	0,221	0,065	
	E	0,529	0,369	0,161	
	S	0,104	0,048	0,056	
FARMACÊU- TICOS	I	0,457	0,457	-	1.354
	J	-0,014	-0,014	-	
	A	0,314	0,197	0,117	
	E	0,386	0,220	0,166	
	S	-0,019	-0,042	0,023	
PRODUTOS DE PAPEL	I	0,307	0,307	-	530
	J	-0,015	-0,015	-	
	A	0,363	0,296	0,067	
	E	0,431	0,284	0,147	
	S	-0,004	0,023	0,027	

continua

SETOR	VARIÁVEL i	EFEITO TOTAL DA i-ésima VARIÁVEL (π_{wi})	EFEITO DIRETO DA i-ésima VARIÁVEL (ρ_{wi})	EFEITO INDIRETO	TAMANHO DA AMOSTRA
VIDRO E CIMENTO	I	0,381	0,381	-	335
	J	-0,024	-0,024	-	
	A	0,256	0,196	0,060	
	E	0,510	0,363	0,142	
	S	-0,035	-0,003	-0,032	
METAIS	I	0,321	0,321	-	2.561
	J	-0,047	-0,047	-	
	A	0,305	0,260	0,045	
	E	0,537	0,386	0,151	
	S	0,094	0,113	-0,019	
EQUIPAMEN- TO MECÂNICO	I	0,422	0,422	-	2.282
	J	0,026	0,026	-	
	A	0,347	0,253	0,094	
	E	0,463	0,309	0,154	
	S	0,058	0,065	-0,007	
EQUIPAMEN- TO ELÉTRICO	I	0,360	0,360	-	3.620
	J	-0,044	-0,044	-	
	A	0,345	0,274	-0,071	
	E	0,509	0,351	0,158	
	S	0,119	0,126	-0,007	
VEÍCULOS MOTORIZA- DOS	I	0,301	0,301	-	7.853
	J	-0,011	-0,011	-	
	A	0,336	0,274	0,026	
	E	0,393	0,297	0,096	
	S	0,102	0,078	0,024	

onde i = I, J, A, E, S.

TABELA (4.1)

muito pequenos ($\pi_{WS} = 0,047$ e $\pi_{WJ} = -0,059$).

Portanto, a influência ocupacional, a educação e a idade são determinantes poderosos dos níveis salariais das indústrias brasileiras.

A educação tem a maior influência sobre o salário porque além do efeito direto, há um efeito indireto através da variável I (0,130).

A idade também tem um substancial efeito total nos salários com a amostra total. Verificamos que seu efeito in direto é de 0,056, sendo positivo através do nível de influ ência ocupacional e negativo através da experiência de trabalho atual.

As análises feitos com os vários setores industriais apresentam resultados análogos. A educação é a variá vel mais influente nos seus efeitos totais. Depois seguem o nível de influência ocupacional, a idade, o tempo de serviço e a experiência no trabalho atual.

Só um setor industrial apresenta resultados diferentes: a indústria de vestuários e sapatos.

Neste setor a influência ocupacional e a idade tem um efeito menor que nos demais setores, mas o tempo de serviço possui um efeito maior.

4.3 - APLICAÇÃO DA ANÁLISE DE TRAJETÓRIAS NO ESTUDO DA SEGREGAÇÃO

RESIDENCIAL NAS CIDADES NORTE-AMERICANAS

Aqui a análise foi feita utilizando como unidades amostrais

149 cidades norte-americanas com populações maiores que 50.000 habitantes em 1960 e com um mínimo de 2.500 não-brancos, sendo 95% ou mais, negros.

Foram estudados os seguintes fatores:

1. Diferenciação em rendas
2. Diferenciação em empregos
3. Porcentagem de não-brancos
4. Número de não-brancos
5. Diferenciação no crescimento entre as populações de brancos e não-brancos.

Os fatores diferenciação em rendas e por diferenciação em empregos foram medidos por índices adequados e conhecidos na sociologia.

A diferenciação no crescimento das populações brancas e não-brancas, foi medida pelas diferenças entre as taxas de crescimento dos brancos e dos negros entre 1950 e 1960. Essas taxas foram consideradas como razões da população não-branca (branca) em 1960 para a população não-branca (branca) em 1950. A população não-branca foi considerada com a transformação logarítmo e o tamanho relativo foi considerado pela porcentagem de não-brancos na população total de cada cidade.

A análise de trajetórias foi realizada para medir as influências desses fatores na segregação residencial (6).

O modelo causal considerado é dado na página seguinte.

O modelo causal foi proposto pelos sociólogos a partir de uma série de considerações, e três hipóteses foram estipuladas:

- 1º - As diferentes posições sociais entre os brancos e os negros são fatores que afetam a segregação resi

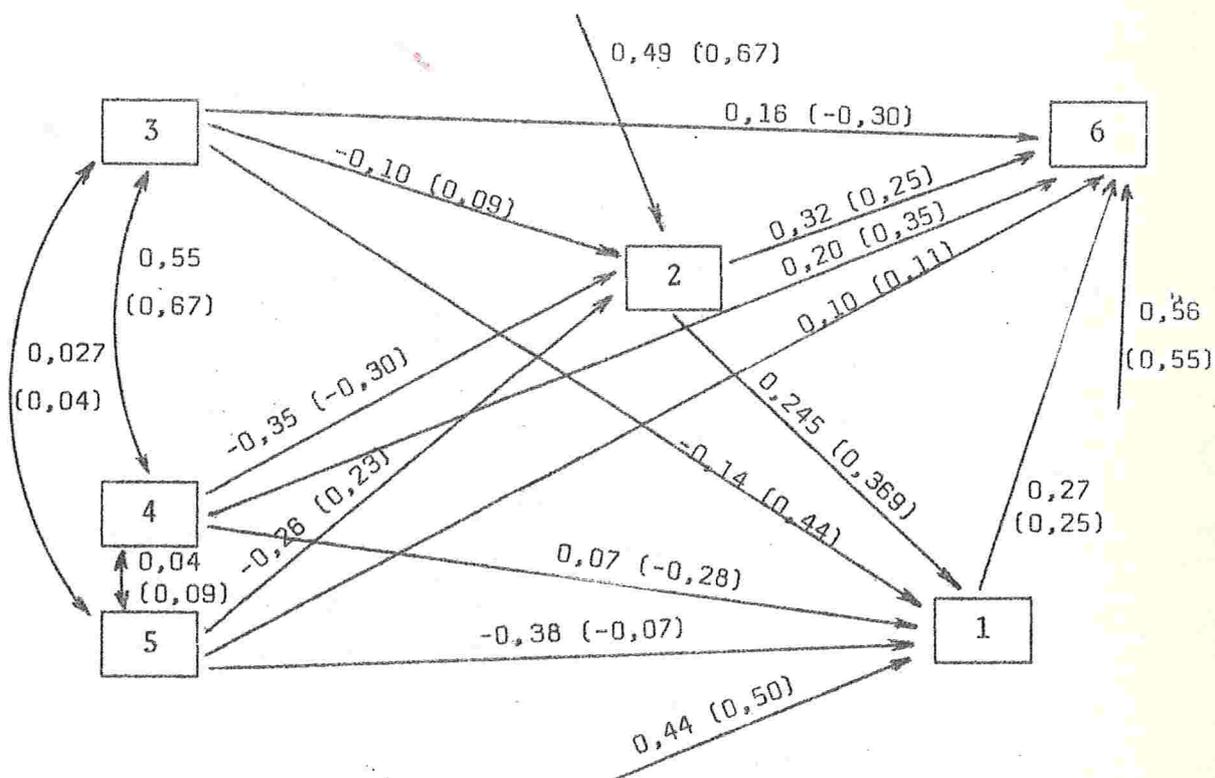


FIGURA 4.3 - Os coeficientes entre parênteses são relativos às cidades sulistas.

dencial.

- 2º - Os tamanhos absolutos e relativos das populações negras têm efeitos importantes na segregação.
- 3º - A taxa de crescimento da população negra relativamente à taxa de crescimento da população branca afeta a segregação.

Os resultados são apresentados na tabela (4.2) apresentada na página seguinte.

De uma análise geral dos resultados concluímos que a segregação residencial é grandemente influenciada pelas diferentes posições sociais entre brancos e negros.

VARIÁVEL DEPENDENTE	EFEITO TOTAL	EFEITO DIRETO	EFEITO INDIRETO
Diferença em rendas	0,27 (0,25)	0,27 (0,25)	
Diferença em empregos	0,25 (0,32)	0,32 (0,25)	-0,07 (0,09)
Porcentagem não-branca. . . .	0,10 (-0,04)	0,16 (-0,30)	-0,08 (0,14)
Número de não-brancos	0,13 (0,08)	0,20 (0,35)	-0,12 (-0,17)
Diferença de crescimento brancos/não-brancos	-0,08 (0,18)	0,10 (0,11)	-0,21 (0,06)

TABELA (4.2) - EFEITOS DE VÁRIOS FATORES NA SEGREGAÇÃO RESIDENCIAL.

(a) região não-sul: 78 cidades; região sul: 71 cidades.

(b) Os resultados em parênteses nas tabelas dadas são para a região sul.

Assim como nos outros exemplos, pelo fato de se trtarem de modelos complexos, de acordo com os resultados do capítulo 3 não podemos realizar testes de hipóteses para verificar se alguns coeficientes são significativamente diferentes de zero ou não.

APÊNDICE 1

1.1 - PROPOSIÇÃO 1 - "Se \underline{X}_2 tem distribuição $N(u_2 \mathbf{1}, I_n \sigma^2)$, então $\underline{X}_2 - \bar{X}_2 \mathbf{1}$ tem uma distribuição $N(\underline{0}, \Sigma)$, onde $\Sigma = (\sigma_{ij})$, $i, j = 1, 2, \dots, n$ e

$$\sigma_{ij} = \frac{\sigma^2(n-1)}{n}$$

para $i=j$ e $\sigma_{ij} = -\frac{\sigma^2}{n}$ para $i \neq j$ ".

Como $\underline{X}_2 - \bar{X}_2 \mathbf{1}$ é uma combinação linear de variáveis aleatórias distribuídas normalmente, então também terá uma distribuição normal. Notamos também que sua média é igual ao vetor nulo.

Agora verificaremos que a matriz de covariância é dada na forma acima.

Seja $\sigma_{ii} = \text{var}(X_{2i} - \bar{X}_2)$. Desenvolvendo, temos:

$$(A1.1) \quad \text{var}(X_{2i} - \bar{X}_2) = \text{var}(X_{2i}) + \text{var}(\bar{X}_2) - 2 \text{cov}(X_{2i}, \bar{X}_2)$$

Sabemos que $\text{cov}(X_{2i}, \bar{X}_2) = E(X_{2i} \bar{X}_2) - E(X_{2i})E(\bar{X}_2)$ e como (X_{21}, \dots, X_{2n}) é uma amostra aleatória, podemos escrever:

$$(A1.2) \quad E(X_{2i} \bar{X}_2) = \frac{1}{n} \left\{ E(X_{2i} X_{21} + \dots + X_{2i}^2 + \dots + X_{2i} X_{2n}) \right\} = \\ = \frac{1}{n} \left\{ u_2^2 + u_2^2 + \dots + E X_{2i}^2 + \dots + u_2^2 \right\}$$

ou

$$(A1.3) \quad E(X_{2i} \bar{X}_2) = u_2^2 + \frac{\sigma^2}{n}$$

Assim,

$$(A1.4) \quad \text{cov}(X_{2i} - \bar{X}_2, \bar{X}_2) = \frac{\sigma^2}{n}$$

Substituindo-se em (A1.1):

$$(A1.5) \quad \text{var}(X_{2i} - \bar{X}_2) = \sigma^2 + \frac{\sigma^2}{n} - 2 \frac{\sigma^2}{n} = \frac{\sigma^2(n-1)}{n}$$

Também,

$$(A1.6) \quad \begin{aligned} \sigma_{ij} = \text{cov}(X_{2i} - \bar{X}_2, X_{2j} - \bar{X}_2) &= E(X_{2i} - \bar{X}_2)(X_{2j} - \bar{X}_2) - \\ &- E(X_{2i} - \bar{X}_2)E(X_{2j} - \bar{X}_2) \end{aligned}$$

Então podemos escrever:

$$(A1.7) \quad \sigma_{ij} = -\frac{\sigma^2}{n}, \quad i \neq j,$$

pois

$$E(X_{2i} - \bar{X}_2)(X_{2j} - \bar{X}_2) = -\frac{\sigma^2}{n}$$

e

$$E(X_{2i} - \bar{X}_2) = E(X_{2j} - \bar{X}_2) = 0$$

1.2 - PROPOSIÇÃO 2 - "Considerando \underline{X}_2 com distribuição

$$N(\underline{X}\beta, \sigma^2 I_n) \text{ e}$$

\underline{X}_1 como vetor não aleatório, as estatísticas U e V dadas por

$$U = \underline{X}_1 \cdot (\underline{X}_2 - \bar{X}_2 \underline{1}) / \sqrt{n} \sigma S_1 \text{ e}$$

$$V = \text{SSE}/\sigma^2 = \underline{X}'_2 [I_n - X(X'X)^{-1}X'] X_2 / \sigma^2$$

são independentes, onde:

$$X = \begin{pmatrix} 1 & X_{11} \\ 1 & X_{12} \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_{1n} \end{pmatrix}; \quad \underline{X}_1 = \begin{pmatrix} X_{11} \\ X_{12} \\ \cdot \\ \cdot \\ X_{1n} \end{pmatrix}; \quad \underline{X}_2 = \begin{pmatrix} X_{21} \\ X_{22} \\ \cdot \\ \cdot \\ X_{2n} \end{pmatrix}$$

Para demonstrarmos essa proposição, vamos utilizar o seguinte fato: se \underline{X} tem distribuição $N(\underline{u}, V)$, então $\underline{X}'\underline{A}\underline{X}$ e $\underline{B}\underline{X}$ são distribuídas independentemente, se e somente se $\underline{B}\underline{A} = 0$ (Searle [15]).

Então escrevemos:

$$(A1.8) \quad U = \frac{1}{\sqrt{n}\sigma S_1} (-\bar{X}_1, 1) \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{12} & \dots & X_{1n} \end{pmatrix} \begin{pmatrix} X_{21} \\ \cdot \\ \cdot \\ X_{2n} \end{pmatrix}$$

ou

$$(A1.9) \quad U = \frac{1}{\sqrt{n}\sigma S_1} (-\bar{X}_1, 1) X' X_2$$

Fazemos então:

$$(A1.10) \quad \underline{B} = \frac{1}{\sqrt{n}\sigma S_1} (-\bar{X}_1, 1) X'$$

Logo:

$$(A1.11) \quad U = \underline{B}\underline{X}_2$$

Como \underline{X}_2 tem uma distribuição $N(\underline{X}\underline{\beta}, \sigma^2 \underline{I}_n)$ e podemos escrever $SSE = \underline{X}_2' \underline{A} \underline{X}_2$, onde

$$A = \frac{1}{\sigma^2} [\underline{I}_n - X(X'X)^{-1}X'],$$

temos:

$$(A1.12) \quad BVA = \frac{1}{\sqrt{nc}S_1} (-\bar{X}_1, 1) X' (\underline{I}_n - X(X'X)^{-1}X') = \underline{0}$$

Portanto, U e V são independentes.

APÊNDICE 2

PROPRIEDADE DE INVARIÂNCIA DOS ESTIMADORES DE
MÍNIMOS QUADRADOS DOS COEFICIENTES DE TRAJETÓRIA

Em correspondência ao modelo de regressão (3.1) da do no capítulo 3, determinamos os estimadores de mínimos quadrados dos coeficientes de regressão a partir das equações normais abaixo

$$(A2.1) \quad \begin{pmatrix} n & \sum_1^n X_{1j} & \sum_1^n X_{2j} & \dots & \sum_1^n X_{kj} \\ \sum_1^n X_{1j} & \sum_1^n X_{1j}^2 & \sum_1^n X_{1j}X_{2j} & \dots & \sum_1^n X_{1j}X_{kj} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_1^n X_{kj} & \sum_1^n X_{kj}X_{1j} & \dots & \dots & \sum_1^n X_{kj}^2 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\beta}_{01} \\ \dots \\ \hat{\beta}_{0k} \end{pmatrix} = \begin{pmatrix} \sum_1^n X_{0j} \\ \sum_1^n X_{1j}X_{0j} \\ \dots \\ \sum_1^n X_{kj}X_{0j} \end{pmatrix}$$

Seja W_i o determinante da matriz obtida substituindo-se a i -ésima coluna da matriz $X'X$ pelo vetor coluna $X'X_0$ e seja $W = \det(X'X)$. Pela regra de Cramer, temos:

$$(A2.2) \quad \hat{\alpha}_1 = \frac{W_1}{W}, \quad \hat{\beta}_{0i} = \frac{W_{i+1}}{W}; \quad i=1, \dots, k.$$

A partir da expressão (1.7), teremos estimadores dos

$$(A2.6) \quad \begin{pmatrix} \sum Z_{1j}^2 & \sum Z_{1j} Z_{2j} & \dots & \sum Z_{1j} Z_{kj} \\ \sum Z_{2j} Z_{1j} & \sum Z_{2j}^2 & \dots & \sum Z_{2j} Z_{kj} \\ \dots & \dots & \dots & \dots \\ \sum Z_{kj} Z_{1j} & \sum Z_{kj} Z_{2j} & \dots & \sum Z_{kj}^2 \end{pmatrix} \begin{pmatrix} \hat{p}_{01} \\ \hat{p}_{02} \\ \dots \\ \hat{p}_{0k} \end{pmatrix} = \begin{pmatrix} \sum Z_{1j} Z_{0j} \\ \sum Z_{2j} Z_{0j} \\ \dots \\ \sum Z_{kj} Z_{0j} \end{pmatrix}$$

Pela regra de Cramer, fazendo $V = \det(Z'Z)$ e V_i o determinante da matriz obtida substituindo-se a i -ésima coluna da matriz $Z'Z$ pelo vetor $Z'Z_0$, encontramos:

$$\hat{p}_{0i} = \frac{V_i}{V}, \quad i=1, 2, \dots, k, \quad \text{ou,}$$

$$\hat{p}_{0i} = \frac{\begin{vmatrix} nS_1^2/S_1^2 & nS_{12}/S_1 S_2 & \dots & nS_{10}/S_1 S_0 & \dots & nS_{1k}/S_1 S_k \\ nS_{12}/S_1 S_2 & nS_2^2/S_2^2 & \dots & nS_{20}/S_2 S_0 & \dots & nS_{2k}/S_2 S_k \\ \dots & \dots & \dots & \dots & \dots & \dots \\ nS_{1k}/S_1 S_k & nS_{2k}/S_2 S_k & \dots & nS_{k0}/S_k S_0 & \dots & nS_k^2/S_k^2 \end{vmatrix}}{\begin{vmatrix} nS_1^2/S_1^2 & nS_{12}/S_1 S_2 & \dots & nS_{1j}/S_1 S_j & \dots & nS_{1k}/S_1 S_k \\ nS_{12}/S_1 S_2 & nS_2^2/S_2^2 & \dots & nS_{2j}/S_2 S_j & \dots & nS_{2k}/S_2 S_k \\ \dots & \dots & \dots & \dots & \dots & \dots \\ nS_{1k}/S_1 S_k & nS_{2k}/S_2 S_k & \dots & nS_{kj}/S_k S_j & \dots & nS_k^2/S_k^2 \end{vmatrix}}$$

onde

$$Z_{fi} = \frac{X_{fi} - \bar{X}_f}{S_f}, \quad f=1, \dots, k$$

e

$$nS_{gh} = \sum_{i=1}^n (X_{gi} - \bar{X}_g)(X_{hi} - \bar{X}_h)$$

Então,

$$(A2.8) \quad \hat{p}_{0i} = \frac{\frac{1}{S_0 S_1^2 \dots S_{i-1}^2 S_i S_{i+1}^2 \dots S_k^2} \begin{vmatrix} nS_i^2 & nS_{12} \dots nS_{10} \dots nS_{1k} \\ nS_{12} & nS_2^2 \dots nS_{20} \dots nS_{2k} \\ \dots & \dots & \dots & \dots \\ nS_{1k} & nS_{2k} \dots nS_{k0} \dots nS_k^2 \end{vmatrix}}{\frac{1}{S_1^2 S_2^2 \dots S_k^2} \begin{vmatrix} nS_1^2 & nS_{12} \dots nS_{1i} \dots nS_{1k} \\ nS_{12} & nS_2^2 \dots nS_{21} \dots nS_{2k} \\ \dots & \dots & \dots & \dots \\ nS_{1k} & nS_{2k} \dots nS_{ki} \dots nS_k^2 \end{vmatrix}}$$

Observamos que as expressões (A2.8) e (A2.5) são i d \hat{e} nticas, mostrando assim que os estimadores de m \hat{i} nimos qua drados dos coeficientes de trajet \hat{o} rias gozam de proprieda de de invari \hat{a} ncia.

APÊNDICE 3

A3.1 - RELAÇÃO ENTRE COEFICIENTES DE CORRELAÇÃO PARCIAIS E COEFICIENTES DE CORRELAÇÃO DE ORDEM ZERO

Sejam X_0, X_1, \dots, X_k variáveis aleatórias com uma distribuição normal multivariada.

$$(A3.1) \quad f(x_0, x_1, \dots, x_k) = \frac{|\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}} e^{-\frac{1}{2}(\underline{X}-\underline{u})' \Sigma^{-1} (\underline{X}-\underline{u})}$$

onde

$$\underline{X} = \begin{pmatrix} X_0 \\ X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_k \end{pmatrix}; \quad \underline{u} = \begin{pmatrix} u_0 \\ u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_k \end{pmatrix}; \quad \Sigma = \begin{pmatrix} \sigma_{00} & \sigma_{01} & \dots & \sigma_{0k} \\ \sigma_{10} & \sigma_{11} & & \sigma_{1k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma_{k0} & \sigma_{k1} & & \sigma_{kk} \end{pmatrix}$$

ou poderíamos escrever:

$$\underline{X} \sim N(\underline{u}, \Sigma)$$

Seja C a matriz dos coeficientes de correlação populacionais:

$$(A3.2) \quad C = \begin{pmatrix} 1 & \rho_{01} & \rho_{02} & \dots & \rho_{0k} \\ \rho_{10} & 1 & \rho_{12} & \dots & \rho_{1k} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k0} & \rho_{k1} & \rho_{k2} & \dots & 1 \end{pmatrix}$$

Consideramos C_{ij} como o cofator do (i,j) -ésimo elemento da matriz simétrica C , sendo $|C|$ o determinante da matriz C .

O cofator C_{ij} pode ser escrito por:

$$(A3.3) \quad C_{ij} = (-1)^{i+j} |M_{ij}|,$$

onde $|M_{ij}|$ é o menor da matriz C .

Definimos $\rho_{01.23\dots k}$ como o coeficiente de correlação parcial de X_0 e X_1 dados X_2, X_3, \dots, X_k fixas, na forma:

$$(A3.4) \quad \rho_{01.23\dots k} = - \frac{C_{01}}{(C_{00} C_{11})^{1/2}}$$

Assim, por exemplo,

$$\rho_{12.3} = - \frac{C_{12}}{(C_{11} C_{22})^{1/2}} = \frac{\rho_{12} - \rho_{13} \rho_{23}}{\{(1 - \rho_{13}^2)(1 - \rho_{23}^2)\}^{1/2}}$$

$\rho_{12.3}$ é o coeficiente de correlação parcial de X_1 e X_2 com X_3 fixado.

A3.2 - REGRESSÃO PARCIAL LINEAR

Seja o modelo dado na forma:

$$(A3.5) \quad E(X_0 | X_1, X_2, \dots, X_k) = \beta_{01.q1} X_1 + \beta_{02.q2} X_2 + \dots + \beta_{0k.qk} X_k,$$

onde X_0, X_1, \dots, X_k são variáveis aleatórias com uma distribuição normal multivariada dada em (A3.1).

Em (A3.5), consideramos:

$$(A3.6) \quad \beta_{0j.qj} = \beta_{0j.12\dots(j-1).(j+1)\dots k}$$

Podemos escrever (A3.5) na forma:

$$(A3.7) \quad X_0 = \beta_{01.q1} X_1 + \dots + \beta_{0k.qk} X_k$$

Consideramos o resíduo por:

$$(A3.8) \quad X_{0.12\dots k} = X_0 - E(X_0 | X_1, \dots, X_k)$$

ou

$$E(X_0 | X_1, \dots, X_k) + X_{0.12\dots k} = X_0$$

onde

$$E(X_{0.12\dots k}) = 0$$

e

$$\text{var}(X_{0.12\dots k}) = \sigma_{0.12\dots k}^2 = E(X_{0.12\dots k}^2)$$

Supondo agora as $(k+1)$ variáveis aleatórias X_0, X_1, \dots, X_k com uma distribuição normal multivariada, temos:

$$(A3.9) \quad \frac{E(X_0 | X_1, \dots, X_k)}{\sigma_0} = - \sum_{j=1}^k \frac{C_{0j}}{C_{00}} \frac{X_j}{\sigma_j}$$

Podemos escrever os coeficientes de regressão parciais quando regredimos X_0 em X_j com as outras $(k-2)$ variáveis fixas, na forma

$$(A3.10) \quad \beta_{0j \cdot qj} = - \frac{\sigma_0}{\sigma_j} \frac{C_{0j}}{C_{00}}$$

A3.3 - RELAÇÃO ENTRE A VARIÂNCIA RESIDUAL COM O COEFICIENTE DE REGRESSÃO PARCIAL

Vimos que

$$\text{var}(X_{0.12\dots k}) = \sigma_{0.12\dots k}^2$$

ou

$$\sigma_{0.12\dots k}^2 = E[\{X_0 - E(X_0 | X_1, \dots, X_k)\}^2]$$

(variância residual em torno da regressão).

Daí,

$$(A3.11) \quad \sigma_{0.12\dots k}^2 = E[\{X_0 - \sum_{j=1}^k \beta_{0j \cdot qj} X_j\}^2]$$

Desenvolvendo:

$$(A3.12) \quad \begin{aligned} \sigma_{0.12\dots k}^2 &= E[X_0(X_0 - \sum_{j=1}^k \beta_{0j \cdot qj} X_j) - \\ &\quad - \sum_{j=1}^k \beta_{0j \cdot qj} X_j (X_0 - \sum_{j=1}^k \beta_{0j \cdot qj} X_j)] \end{aligned}$$

Como o segundo produto é igual a zero, temos:

$$(A3.13) \quad \sigma_{0.12\dots k}^2 = \sigma_0^2 - \sum_{j=1}^k \beta_{0j} \cdot q_j \sigma_{0j}$$

A3.4 - RELAÇÕES ENTRE AS VARIÂNCIAS, AS REGRESSÕES E AS CORRELAÇÕES DE DIFERENTES ORDENS

De (A3.13) e (A3.10), podemos escrever:

$$(A3.14) \quad \sigma_{0.12\dots k}^2 = \sigma_0^2 + \sum_{j=1}^k \frac{\sigma_0}{\sigma_j} \frac{C_{0j}}{C_{00}} \sigma_{0j}$$

Daí,

$$\frac{\sigma_{0.12\dots k}^2}{\sigma_0^2} = 1 + \frac{1}{C_{00}} \sum_{j=1}^k C_{0j} \rho_{0j} = 1 + \frac{1}{C_{00}} (|C| - C_{00}) = \frac{|C|}{C_{00}}$$

Assim,

$$(A3.15) \quad \sigma_{0.12\dots k}^2 = \sigma_0^2 \frac{|C|}{C_{00}}$$

Do mesmo modo,

$$(A3.16) \quad \text{cov}(X_{\ell.q\ell}, X_{m.qm}) = \sigma_\ell \sigma_m \frac{|C|}{C_{\ell m}}$$

Quando $\ell = m$, temos

$$\sigma_{m.q}^2 = \sigma_m^2 \frac{|C|}{C_{mm}}$$

como foi visto em (A3.15).

Também podemos escrever

$$(A3.17) \quad \text{cov}(X_{\ell.r}, X_{m.r}) = \beta_{\ell m.r} \sigma_{m.r}^2$$

$$(A3.18) \quad \text{cov}(X_{\ell.r}, X_{m.r}) = \beta_{m \ell.r} \sigma_{\ell.r}^2$$

$$(A3.19) \quad \text{cov}(X_{\ell.r}, X_{m.r}) = \rho_{\ell m.r} \sigma_{\ell.r} \sigma_{m.r}$$

A3.5 - OUTRAS RELAÇÕES

Podemos escrever, algumas outras relações:

$$(A3.20) \quad \sigma_{0.q}^2 = \sigma_{0.q-1}^2 (1 - \rho_{01.q}^2)$$

Em particular:

$$(A3.21) \quad \sigma_{1.0}^2 = \sigma_1^2 (1 - \rho_{01}^2)$$

Observamos que em (A3.20), o índice q-1 significa um conjunto de q variáveis excluindo X_1 .

Da expressão (A3.20), temos:

$$(A3.22) \quad \sigma_{0.q-1}^2 = \sigma_{0.q-1-2}^2 (1 - \rho_{02-q-1}^2)$$

Aplicando-se sucessivamente (A3.20), temos:

$$(A3.23) \quad \sigma_{0.12\dots k}^2 = \sigma_0^2 (1 - \rho_{0k}^2) (1 - \rho_{0(k-1).k}^2) (1 - \rho_{0(k-2).(k-1)k}^2) \dots (1 - \rho_{01.2\dots k}^2)$$

Portanto:

$$(A3.24) \quad \frac{\sigma_{0.12\dots k}^2}{\sigma_0^2} = (1 - \rho_{01}^2) (1 - \rho_{02.1}^2) \dots (1 - \rho_{0k.12\dots(k-1)}^2)$$

ou

$$(A3.25) \quad \frac{\sigma_{0.12\dots k}^2}{\sigma_0^2} = \frac{|C|}{C_{00}}$$

Uma relação para a covariância entre X_0 e X_1 para X_k fixado, pode ser escrita por:

$$(A3.26) \quad \sigma_{01.k} = \sigma_{01} - \frac{\sigma_{0k} \sigma_{k1}}{\sigma_k^2}$$

Também, pela definição de coeficiente de regressão parcial, temos

$$(A3.27) \quad \beta_{ij.k} = \sigma_{ij.k} / \sigma_{j.k}^2$$

Para outras relações, ver Kendall [9].

APÊNDICE 4

4.1 - COEFICIENTE DE CORRELAÇÃO MÚLTIPLA

Considerando as mesmas variáveis aleatórias X_0, X_1, \dots, X_k vistas no apêndice 3, juntamente com suas respectivas suposições, definimos por coeficiente de correlação múltiplo $R_{0(1\dots k)}$ entre X_0 e X_1, \dots, X_k a:

$$(A4.1) \quad 1 - R_{0(1\dots k)}^2 = \sigma_{0.1\dots k}^2 / \sigma_0^2,$$

onde $\sigma_{0.1\dots k}^2$ foi definida no apêndice 3.

Também:

$$(A4.2) \quad 0 \leq R^2 \leq 1.$$

Vimos também que

$$\sigma_{0.12\dots k}^2 = E(X_{0.12\dots k}^2)$$

e que

$$E(X_{0.12\dots k}^2) = E(X_0 X_{0.12\dots k}).$$

Desde que $E(X_{0.12\dots k}) = 0$, temos:

$$(A4.3) \quad \sigma_{0.12\dots k}^2 = \text{var}(X_{0.12\dots k}) = \text{cov}(X_0, X_{0.12\dots k})$$

Considerando agora a correlação entre X_0 e sua esperança condicional

$$E(X_0 | X_1, \dots, X_k) = X_0 - X_{0.12\dots k},$$

nós encontramos:

$$(A4.4) \quad \frac{\text{cov}(X_0, X_0 - X_{0.12\dots k})}{\{\text{var } X_0 \text{ var}(X_0 - X_{0.12\dots k})\}^{1/2}} =$$
$$= \frac{\text{var } X_0 - \text{cov}(X_0, X_{0.12\dots k})}{[\text{var } X_0 \{\text{var } X_0 + \text{var } X_{0.12\dots k} - 2 \text{cov}(X_0, X_{0.12\dots k})\}]^{1/2}}$$

Usando A4.3, podemos escrever:

$$(A4.5) \quad \frac{\sigma_0^2 - \sigma_{0.1\dots k}^2}{\{\sigma_0^2(\sigma_0^2 - \sigma_{0.1\dots k}^2)\}^{1/2}} = \left\{ \frac{\sigma_0^2 - \sigma_{0.1\dots k}^2}{\sigma_0^2} \right\}^{1/2} = R_{0(12\dots k)}$$

Também podemos escrever de (A4.1) e (A3.24) que:

$$(A4.6) \quad 1 - R_{0(1\dots k)}^2 = \frac{|C|}{C_{00}} = (1 - \rho_{01}^2)(1 - \rho_{02.1}^2) \dots (1 - \rho_{0k.1\dots(k-1)}^2)$$

APÊNDICE 5

5.1 - O COEFICIENTE DE TRAJETÓRIAS E O COEFICIENTE DE CORRELAÇÃO DE ORDEM ZERO, QUANDO AS VARIÁVEIS INDEPENDENTES NÃO SÃO INTER-CORRELACIONADAS

Vamos considerar agora uma regressão linear da variável X_0 nas variáveis aleatórias X_1, X_2, \dots, X_k . Considerando as variáveis independentes como não correlacionadas entre si.

A matriz de correlação aparece na forma:

$$(A5.1) \quad C = \begin{pmatrix} 1 & \rho_{01} & \rho_{02} & \dots & \rho_{0k} \\ \rho_{10} & 1 & 0 & 0 & \dots & 0 \\ \rho_{20} & 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{k0} & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Notamos que

$$(A5.2) \quad C_{00} = (-1)^0 \begin{vmatrix} 1 & & & \\ & 1 & 0 & \\ & & \ddots & \\ & 0 & & 1 \end{vmatrix} = 1$$

$$C_{01} = (-1)^1 \begin{vmatrix} \rho_{10} & 0 & 0 & \dots & 0 \\ \rho_{20} & 1 & 0 & \dots & 0 \\ \rho_{30} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k0} & 0 & 0 & \dots & 1 \end{vmatrix} = -\rho_{10}$$

$$C_{02} = (-1)^2 \begin{vmatrix} \rho_{10} & 1 & 0 & 0 & \dots & 0 \\ \rho_{20} & 0 & 0 & 0 & \dots & 0 \\ \rho_{30} & 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{k0} & 0 & 0 & 0 & \dots & 1 \end{vmatrix}$$

Logo:

$$C_{02} = (-1)^5 \rho_{20} \begin{vmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & \cdot & & \\ & 0 & & & \cdot & \\ & & & & & \cdot & \\ & & & & & & 1 \end{vmatrix} = -\rho_{20}$$

$$C_{03} = (-1)^3 \begin{vmatrix} \rho_{10} & 1 & 0 & \dots & \dots & 0 \\ \rho_{20} & 0 & 1 & 0 & 0 & \dots & 0 \\ \rho_{30} & 0 & 0 & 0 & 0 & \dots & 0 \\ \rho_{40} & 0 & 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{k0} & 0 & 0 & 0 & 0 & \dots & 1 \end{vmatrix} = (-1)^7 \rho_{30} \det(I) = -\rho_{30}$$

Dai, teremos:

$$C_{0k} = (-1)^k \begin{vmatrix} \rho_{10} & 1 & 0 & 0 & \dots & 0 \\ \rho_{20} & 0 & 1 & 0 & \dots & 0 \\ \rho_{30} & 0 & 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{k0} & 0 & 0 & 0 & 0 & \dots & 0 \end{vmatrix} = (-1)^k (-1)^{k+1} \rho_{k0} \det(I) = -\rho_{k0}$$

Assim, concluímos que:

$$(A5.3) \quad C_{0j} = -\rho_{0j}, \text{ para } j=1, 2, \dots, k$$

Como o coeficiente de trajetória é dado por:

$$p_{0j} = \frac{\sigma_j}{\sigma_0} \beta_{0j.qj} \text{ onde } \beta_{0j.qj} = -\frac{\sigma_0}{\sigma_j} \frac{C_{0j}}{C_{00}}$$

temos:

$$p_{0j} = \frac{-C_{0j}}{C_{00}}$$

Por (A5.2) e (A5.3), teremos:

$$p_{0j} = -(-\rho_{0j}), \quad j=1, 2, \dots, k.$$

Ou:

$$(A5.4) \quad p_{0j} = \rho_{0j}, \quad j=1, \dots, k.$$

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] - ANDERSON, T.W. - An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc., (1964), 4a. edição.
- [2] - ALWIN, D.F. e Hauser, R.M. "The decomposition of effects in path analysis", Amer. Soc. Review, 40 (1975), pp.37-47.
- [3] - BLAU, P.M. e DUCAN, O.D. - "The process of Stratification", The American Occupational Structure, ed. John Wiley & Sons, Inc., (1967), pp. 163-205.
- [4] - BOSE, R.C. - Statistical Models, notas mimeografadas da Universidade de Carolina do Norte, U.S.A.
- [5] - DUNCAN, O.D. - "Path Analysis: sociological examples", Amer. Journal of Sociology, 72 (July 1966), pp.1-16.
- [6] - DRAPER, N.R. e SMITH, H. - Applied Regression Analysis, New York, John Wiley and Sons Inc. (1966).
- [7] - HALLER, A.O. e PASTORE, J. - "Training, Position and Experience in the wage-rates of Specialized Personnel in São Paulo's manufacturing Firms", notas mimeografadas do Instituto de Pesquisas Econômicas da Universidade de São Paulo.
- [8] - KEMPTHORNE, O. - An Introduction to Genetic Statistics, John Wiley and Sons, Inc. (1968).
- [9] - KENDALL, M.G. and STUART, A. - The Advanced Theory of Statistics, Charles Griffin and Co. Ltd., London (1958), vol.2.

- [10] - KLEMMACK, D.L. e LEGGETTE, T.A. e MAYER, L.S. - "Non-random exogenous variables in path analysis", American Sociological Review, vol. 38, (December 1973), pp. 778-784.
- [11] - LAND, K.C. - "Principles of Path Analysis", Sociological Methodology, (1969), pp. 1-37.
- [12] - LI, C.C. - "Population Genetics", (1955), pp. 174.
- [13] - LI, C.C. - "The concept of path coefficients and its impact on population genetics", Biometrics, (1956), 12, pp. 190-210.
- [14] - SEARLE, S.R. - Linear Models, John Wiley and Sons, Inc.
- [15] - TUKEY, J.W. - "Causation, Regression and path analysis", in Statistics and Mathematics in Biology, Ed. Kempthorne et al. Iowa State College Press. Ames., Iowa, (1954), pp. 35-66.
- [16] - TURNER, R. and STEVENS, C.D. - "The Regression Analysis of Causal Paths", Biometrics, 15 (1959), pp. 236-258.
- [17] - WRIGHT, S. "The method of path coefficients", Ann.Math.Stat., 5, (1934), pp. 161-215.
- [18] - WRIGHT, S. "The Interpretation of Multivariate Systems", in Statistics and Mathematics in Biology, Ed. Kempthorne, et al. Iowa State College Press, Iowa, (1954), pp. 11-33.
- [19] - WRIGHT, S. - "Path Coefficients and path regression: alternative or complementary concepts?", Biometrics, 16, (junho 1960), pp. 189-202.