Ministério do Planejamento e Coordenação Geral
Fundação IBGE
INSTITUTO BRASILEIRO DE GEOGRAFIA

GEOGRAFIA QUANTITATIVA

J. P. Cole

GEOGRAFIA QUANTITATIVA

1972

Cole, J P

Geografia quantitativa. Rio de Janeiro, Instituto Brasileiro de Geografia, 1972.

120 p. 23cm.

Acima do título: Ministério do Planejamento e Coordenação Geral Fundação IBGE...

1. Geografia quantitativa. I. Brasil. Instituto Brasileiro de Geografia. II. Título.

Biblioteca
do
IBG

100b
SWB
C689

MINISTÉRIO DO PLANEJAMENTO E COORDENAÇÃO GERAL
Fundação IBGE
INSTITUTO BRASILEIRO DE GEOGRAFIA

GEOGRAFIA QUANTITATIVA

J. P. Cole

RIO DE JANEIRO
1972

Apresentação

Com a revolução quantitativa ocorrida na Geografia, principalmente na Suécia, Inglaterra e Estados Unidos, o Departamento de Geografia do IBG não poderia ficar à margem do acontecimento e, aproveitando a vinda ao Brasil do eminente geógrafo inglês Prof. Dr. J. P. Cole da Universidade de Nottingham, promoveu um curso de Iniciação à Geografia Quantitativa, no qual temas como "Aplicação de Matrizes e Gráficos na Geografia"; "Análise Fatorial e de Grupamento"; "Trend Surface", etc. foram sucessivamente apresentados.

O intuito de tal curso, que teve a duração de 3 meses, foi o de promover uma primeira aproximação dos Geógrafos do IBG às novas técnicas matemático-estatísticas que vêm sendo utilizadas, com grande êxito, em pesquisas geográficas.

Esta obra, lançada agora pelo IBG, é tão sòmente a coletânea de apostilas preparadas pelo Dr. J. P. Cole para o Curso e sua divulgação tem por objetivo difundir êstes conhecimentos iniciais a todos aquêles que se interessam pela realização de pesquisas geográficas com utilização de técnicas de maior precisão e dentro da mais moderna orientação metodológica, no sentido da obtenção de resultados cada vez mais objetivos nos estudos geográficos.

MIGUEL ALVES DE LIMA

Diretor Superintendente do
Instituto Brasileiro de Geografia

sumário

I — Números, Matrizes e Gráficos 1

II — Aplicações de Matrizes e de Gráficos na Geografia 7

III — Três testes Estatísticos 15

IV — Amostragem 27

V — Conjuntos 33

VI — Notas sôbre Distância e Localização em Geografia Humana 39

VII — O Problema de Difusão 49

VIII — A Curva de Lorenz 63

IX — Análise Fatorial (factor analysis) e Análise de Agrupamento (Cluster Analysis) no Brasil e no Peru 69

X — "Trend Surface Analysis" aplicado ao Estudo da Distribuição da População Rural no Norte de Portugal 87

XI — Projeções Demográficas com computador 103

1

Números, Matrizes e Gráficos

SISTEMAS DE NÚMEROS

Na matemática se reconhece a existência de vários tipos ou sistemas de números. É importante distinguir entre os números inteiros (por exemplo 3; 112; 0; — 19) e os números reais (por exemplo: 2.61; 0.333; π; — 3.96; 3.000). Os números inteiros positivos e 0 (zero) se usam para contar. Os números inteiros positivos chamam-se números naturais. Cada número inteiro tem um número real equivalente: 1, 1.0; 15, 15.00. Os números inteiros são um *subconjunto* dos números reais. Em inglês os números inteiros chamam-se "*integers*" e os números reais "*real numbers*".

BASES

Para expressar números, usamos um método no qual a posição de um símbolo indica o seu valor. Por exemplo: no número 1 969, o primeiro 9 vale 100 vêzes mais que o segundo 9, devido à sua posição. Ao contrário, no número romano XXX, cada símbolo X tem o mesmo valor, apesar de ter cada símbolo uma posição diferente. Em qualquer método que usa a posição do símbolo no número para indicar seu valor, é indispensável um símbolo 0 (zero), ou algum símbolo equivalente.

É possível, porém, expressar os mesmos números com menor ou maior quantidade de símbolos. O número de símbolos determina a base.

O sistema de números que usa só dois símbolos, 1 (um) e 0 (zero), chama-se base 2 ou binário (em inglês *binary*). Êste sistema tem a desvantagem de necessitar mais símbolos (aproximadamente 3,2 vêzes) que o sistema base 10 (*decimal* ou *denary*) para escrever o mesmo número. Por exemplo, 37 se escreve em binário 100101. Porém o sistema "binary" tem certas vantagens importantes. Em primeiro lugar, é a base mais conveniente para aplicação nos computadores electrônicos, os quais "traduzem" qualquer informação recebida em forma "binary" para processá-la. Em segundo lugar, oferece um método claro para manipular dados que expressam presença ou ausência de um atributo. Em terceiro lugar, é interessante notar, por exemplo, que o número aparentemente tão expressivo como 100, se escreve em "binary" 1100100, um número qualquer.

Para entender como funciona um método de expressar números com outra base, é preciso considerar primeiro o método que usamos normalmente. O número 305 significa da direita para a esquerda $5\times1 + 0\times10 + 3\times100$. Cada posição no número tem um valor 10 vêzes maior que a posição imediatamente à sua direita. No sistema base 2, cada posição no número tem um valor só duas vêzes maior que a posição à sua direita. O número em "binary" 101 significa, lendo da direita para esquerda, $1\times1 + 0\times2 + 1\times4$ e representa o número 5 em base 10.

MATRIZES E GRÁFICOS

A matriz e o gráfico se empregam agora com muita freqüência na matemática e também na geografia. Uma matriz ("matrix" ou "array" em inglês) é um conjunto de números, apresentados na forma de linhas e colunas. Tem um número determinado de elementos, cada elemento sendo a intersecção de uma linha com uma coluna. Indica-se cada linha e cada coluna com um número *inteiro* positivo. Por exemplo, é possível ter linha 3, linha 4, linha 5, mas não linha 3.5 ou linha 0.72.

Um gráfico é um plano ou uma superfície de área infinita, com um número infinito de pontos (no sentido da geometria de Euclides). Na prática, atualmente só se usa uma parte do gráfico. A parte usada, geralmente, é a parte ao redor de um ponto central. O Centro é aquêle ponto onde se cruzam (ou se entrecortam) dois eixos ortogonais (isto é um ângulo reto). Usando os dois eixos como escalas de distância do ponto de origem (zero) é possível descrever a posição de qualquer ponto no gráfico, empregando um par de números reais, com tanta precisão quanto se deseja, usando casas depois do ponto decimal. Geralmente o eixo "horizontal" (da esquerda para a direita) se chama eixo x, o eixo "vertical", eixo y.

Vale a pena notar quatro diferenças fundamentais entre uma matriz e um gráfico.

1) A matriz tem um número finito (limitado) de elementos *distintos*, enquanto o gráfico tem um número infinito de pontos numa superfície *contínua*. No caso do gráfico, não é possível usar mais que uma parte muito pequena do número total dos pontos disponíveis, enquanto na matriz é necessário preencher cada elemento (ou entrada) da matriz com um número.

2) A posição de um elemento na matriz se estabelece com dois números inteiros, primeiro o número da linha, depois o número da coluna. A posição de um ponto no gráfico se estabelece com dois números reais, os quais se chamam um par ordenado. O primeiro número dá a posição do ponto com referência ao eixo x, o segundo sua posição com referência ao eixo y.

3) Convencionalmente, a matriz começa na parte superior, à esquerda, enquanto que o gráfico começa no ponto de origem, zero. Um gráfico que só usa posições positivas começa na parte inferior, à esquerda.

4) Na matriz, todos os elementos têm posições positivas. No gráfico é possível representar posições negativas, que ficam ou abaixo do eixo x, ou à esquerda do eixo y, ou em ambas posições.

Concluindo, deve-se notar que matemàticamente é possível ter uma matriz com mais de duas dimensões (a linha e a coluna sendo consideradas, cada uma como uma dimensão). Do mesmo modo, é possível ter um gráfico com mais de duas dimensões (cada eixo sendo considerado como uma dimensão). É difícil apresentar uma matriz ou um gráfico de mais de três dimensões na superfície de uma página, e impossível apresenta-los quando têm mais de três dimensões. É possível, porém, apresentar e manipular matrizes e gráficos com qualquer número de dimensões com a ajuda da álgebra. Tais matrizes e gráficos multidimensionais têm várias aplicações na matemática e nas ciências em geral.

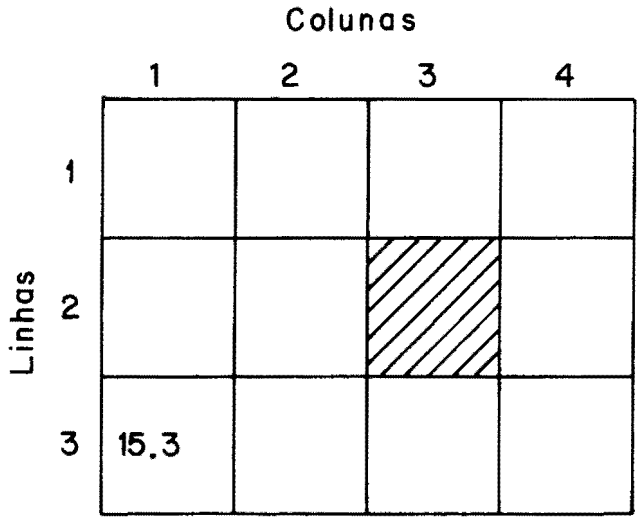


Fig.1 DivEd/D- J.A.C

Matriz com 3 linhas e 4 colunas.

Em geral uma matriz tem M linhas e N colunas.

A célula com hachuras no diagrama onde cruzam a linha 2 e a coluna 3, chama-se 2 3 (diz-se duas, três).

Na célula 3 1 (diz-se três, um) encontra-se o elemento 15,3

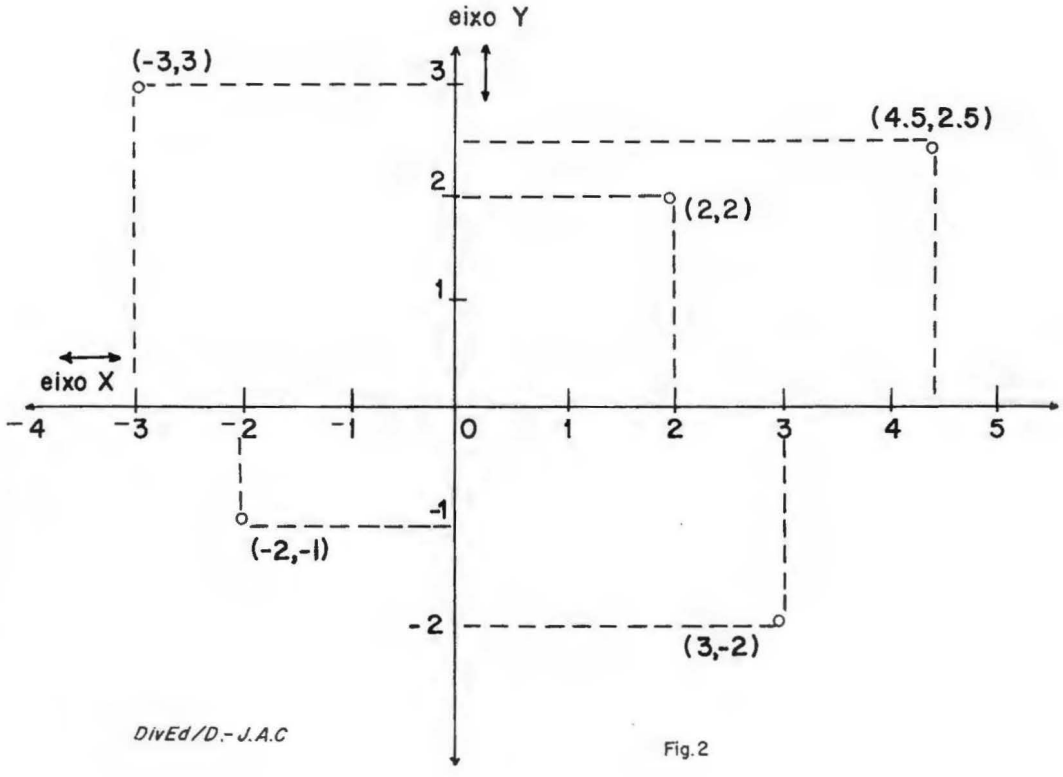


Fig. 2

2

Aplicações de Matrizes e de Gráficos na Geografia

MATRIZES

A matriz tem a vantagem de expressar numa forma numérica, simultâneamente suscinta e precisa, uma quantidade grande de informações. É fácil encontrar numa matriz qualquer elemento, usando os números da linha e da coluna para indicar ou procurar a posição.

Seguem vários exemplos da aplicação de matrizes em situações de interêsse potencial para o geógrafo. Nos exemplos foram usadas matrizes de tamanho reduzido e com valôres apenas de um ou dois números para economizar espaço.

Exemplo 1 — Numa matriz dêste tipo pode-se considerar as relações entre setores da economia de um país. Nesse exemplo se apresentam os valôres (em unidades monetárias) das transações entre cada par de setores ou ramos da economia de um país. Em inglês, esta aplicação de uma matriz chama-se "input — output matrix". É empregada, por exemplo, no estudo da economia dos Estados Unidos, da URSS, de Israel. A matriz para a economia da URSS no ano de 1966 tem 82 linhas e colunas. Uma matriz dêste tipo é quadrada (tem o mesmo número de linhas e colunas) mas não é simétrica devido ao fato de que o valor das transações entre um par de setores raramente é igual em ambas as direções. Os valôres na diagonal principal representam as transações dentro de cada setor.

Exemplo 2 — Esta matriz apresenta a relação entre cada estado e cada setor da economia de um país. Nesse exemplo, usam-se só os números 1 e 0 para expressar presença ou ausência de um cultivo no estado. Seria

possível usar quantidades em vez de sòmente 1 ou 0. Esta matriz não tem que ser quadrada, embora fôsse quadrada no caso de ter igual número de estados e de setores.

Exemplo 3 — Esta matriz apresenta as relações entre cada estado do país. Como no caso da matriz do primeiro exemplo, esta matriz é quadrada, devido ao fato de que tem os mesmos estados (poderiam ser cidades, portos, pessoas) nas linhas e nas colunas. Nesse exemplo usa-se só 1 ou 0 para expressar a presença ou ausência de migração, entre cada par de estados. Os valôres de 1 na diagonal principal representam a migração dentro de cada estado. Seria possível incluir nesta matriz a quantidade de pessoas migrantes entre cada par de estados em vez de usar só o valor de 1. Outro tipo de informação nesta matriz seria a distância (por exemplo, em quilômetros) ou o preço da passagem de ônibus entre cada par de lugares.

Exemplo 1

SETORES	Carvão	Aço	Madeira	Móveis	Ferrovia
Carvão	15	20	0	0	2
Aço	2	25	1	1	3
Madeira	1	0	8	12	0
Móveis	0	0	0	6	0
Ferrovia	10	7	3	1	10

Exemplo 2

	Algodão	Açúcar	Café	Cacau
Ceará	1	1	0	0
Bahia	1	1	0	1
São Paulo	1	1	1	0
Rio Grande do Sul	0	0	0	0
Goiás	0	0	0	0

Exemplo 3

De / Para	Ceará	Bahia	S. Paulo	R. G. S.	Goiás
Ceará	1	0	0	0	0
Bahia	0	1	0	0	0
São Paulo	1	1	1	0	1
Rio Grande do Sul	0	0	0	1	0
Goiás	1	1	0	1	1

Exemplo 4 — Empregam-se duas matrizes, a primeira (4a) como a matriz do exemplo 2, a segunda (4b) como a matriz do exemplo 3. A matriz 4a contém informação sôbre 11 cidades grandes do Brasil. Cada coluna representa um atributo. Os atributos são: tamanho (mais de um milhão de habitantes), diversificação (número de serviços superior a um mínimo determinado) e extensão territorial da zona de influência da cidade (mais de dois milhões de quilômetros quadrados). É interessante notar que as características de cada cidade se expressam na matriz por 3 números, usando 1 ou 0. A linha de 1 ou 0 de cada cidade representa um número atual em "*binary*" (base 2). O número equivalente em base 10 pode servir como cifra ou indicação de tipo da cidade. Por exemplo, São Paulo e Rio são igual e têm o mesmo número em base 10, isto é, 7, que poderia servir como referência ao tipo. É possível calcular outro índice da matriz: o total ou a soma dos valôres 1 de cada linha na matriz (isto é, cada cidade). O resultado ou marcação neste caso representa o número de pontos ganhos pela cidade por ter tal ou qual atributo.

Os dados na matriz 4a podem ser usados para calcular índices de semelhança entre cada par de cidades. Para incluir todos os resultados é necessário uma matriz com o número das linhas e das colunas igual ao número das cidades. Calcula-se o número de vêzes que um par de cidades tem o mesmo valor, seja 1 ou 0. Na matriz 4a, Rio tem o mesmo valor que São Paulo em todos os três atributos. Assim, são idênticos e têm um índice de semelhança de 3 (o máximo possível neste exemplo). Salvador e Fortaleza também têm um índice de semelhança de 3, porque onde o primeiro tem 0, o outro tem 0, e onde o primeiro tem 1, o outro também tem 1. Ao contrário, São Paulo e Santos têm um índice de semelhança de 0, o mínimo possível neste exemplo, porque sempre têm valôres contrários.

Os resultados dos cálculos de todos os índices de semelhança se encontram na matriz 4b. A diagonal principal tem um índice de 3, o índice inevitável de semelhança completa entre cada cidade e ela mesma. Neste caso não é necessária preencher a diagonal principal da matriz, mas é útil salientar o fato.

Vale a pena salientar várias extensões dos conceitos no exemplo 4 e outras maneiras de expressar a informação.

a) É possível somar os valôres em cada coluna (ou linha) da matriz 4b. Desta maneira se obtém um índice total de semelhança de cada cidade com tôdas as outras (incluindo ela mesma se somar-se o valor da diagonal principal). Neste exemplo, São Paulo e Rio são menos parecidos com o grupo inteiro, enquanto Pôrto Alegre, Salvador, Fortaleza e Curitiba são os mais parecidos com o grupo inteiro.

4a	Tamanho	Diversificação	Influência territorial	EQUIVALÊNCIA	
				Código	Escore
São Paulo	1	1	1	7	3
Rio de Janeiro	1	1	1	7	3
Recife	1	1	0	6	2
Belo Horizonte	1	1	0	6	2
Pôrto Alegre	0	1	0	2	1
Salvador	0	1	0	2	1
Fortaleza	0	1	0	2	1
Curitiba	0	1	0	2	1
Belém	0	1	1	3	2
Santos	0	0	0	0	0
Campinas	0	0	0	0	0

4b	São Paulo	Rio de Janeiro	Recife	Belo Horizonte	Pôrto Alegre	Salvador	Fortaleza	Curitiba	Belém	Santos	Campinas
São Paulo	3	3	2	2	1	1	1	1	2	0	0
Rio de Janeiro	3	3	2	2	1	1	1	1	2	0	0
Recife	2	2	3	3	2	2	2	2	1	1	1
Belo Horizonte	2	2	3	3	2	2	2	2	1	1	1
Pôrto Alegre	1	1	2	2	3	3	3	3	2	2	2
Salvador	1	1	2	2	3	3	3	3	2	2	2
Fortaleza	1	1	2	2	3	3	3	3	2	2	2
Curitiba	1	1	2	2	3	3	3	3	2	2	2
Belém	2	2	1	1	2	2	2	2	3	1	1
Santos	0	0	1	1	2	2	2	2	1	3	3
Campinas	0	0	1	1	2	2	2	2	1	3	3
TOTAL	16	16	21	21	24	24	24	24	19	17	17

b) Se considerarmos que cada atributo na matriz 4a representa uma dimensão, é possível elaborar um gráfico em três dimensões (fig. 4c) em forma de um cubo com 8 vértices. Cada vértice representa uma combinação de atributos. Das oito posições possíveis, só 5 estão representadas pelas 11 cidades consideradas neste exemplo.

c) O exemplo serve também para ilustrar alguns conceitos do estudo de *conjuntos* (em inglês "sets") e *diagramas* Venn. Por exemplo São Paulo é grande *e* diversificado *e* tem área de influência grande. Assim pertence a três conjuntos. No diagrama, cada círculo representa um conjunto.

Cada cidade foi colocada no círculo ou círculos que representa(m) o(s) conjunto(s) aos quais pertence. São Paulo coloca-se na intersecção dos 3 conjuntos, porque pertence a todos os três. Seria possível usar êste método para considerar as qualificações de cada cidade e definir a área metropolitana.

O critério poderia ser, por exemplo, que a cidade deveria ser grande (população) *ou* ter serviços diversificados *ou* ter zona de influência muito grande. A palavra *ou* é menos exigente que a palavra *e* porque para ser admitida no conjunto de áreas metropolitanas, seria suficiente a cidade pertencer pelo menos a um dos três *conjuntos.*

GRÁFICO

Também tem várias aplicações na geografia: além de sua função como meio de representar dados estatísticos, serve ainda para indicar relações entre variáveis e transferir dados geográficos do mapa para o computador.

a) é possível representar gràficamente equações e inequações algébricas. $x = y$ ou $x - y = 0$ é uma equação e no gráfico representa a linha na qual o valor no eixo x é igual ao valor no eixo y. A equação $x - 2y = 0$ é indicada pela linha na qual o valor de y é duas vêzes o valor de x. $x \leq 1$ é uma inequação e em vez de ser representada no gráfico por uma linha, é representada por uma área. A área que satisfaz as duas inequações, $x \leq 1$ e $y \leq 1$ é marcada no gráfico com hachuras.

b) Se considerarmos o mapa como um tipo de gráfico podemos aproveitar o sistema de eixos como coordenadas e usar o gráfico para precisar a posição de cada lugar representado por um ponto. Recorrendo a um cálculo relativamente fácil e bem conhecido, pode-se preparar um programa para um computador capaz de calcular a distância entre muitos pares de lugares no mapa. Ao mesmo tempo, usando o conceito de um "banco" de dados, é possível colocar-se na "memória" de um computador vários dados numéricos a respeito de cada lugar.

O método para calcular a distância entre qualquer par de pontos num gráfico é o seguinte (usando como exemplo A e B no gráfico):

Estabelecer a posição de A e de B em têrmos de pares ordenados de números: A (6,4) B (2,1). A distância entre A e B é a hipotenusa de um triângulo retângulo. O comprimento dos outros dois lados

do triângulo são a diferença entre A e B no eixo x, e a diferença entre A e B no eixo y, isto é, 4 e 3. É suficiente saber a posição de cada lugar no mapa para que o computador faça os cálculos necessários para saber a distância entre cada par de pontos. A fórmula é a seguinte:

$$AB = \sqrt{(Ax - Bx)^2 + (Ay - By)^2}$$

Ax é a posição do ponto A no eixo x, isto é, 6.

Bx é a posição do ponto A no eixo x, isto é, 2. Etc.

Assim: $AB = \sqrt{(6 - 2)^2 + (4 - 1)^2}$

$$AB = \sqrt{4^2 + 3^2}$$

$$AB = \sqrt{16 + 9}$$

$$AB = \sqrt{25} = \underline{5}$$

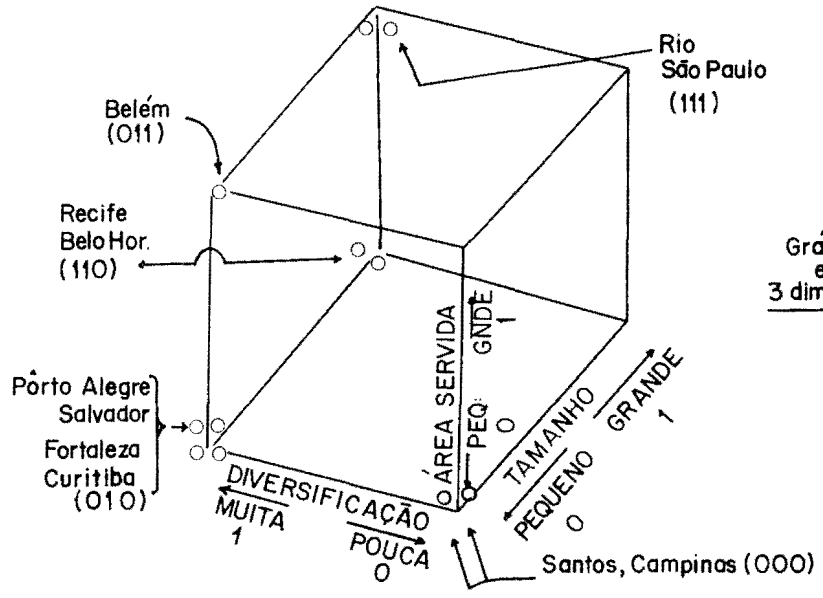


Gráfico em 3 dimensões

U (UNIVERSO OU CONJUNTO UNIVERSAL)

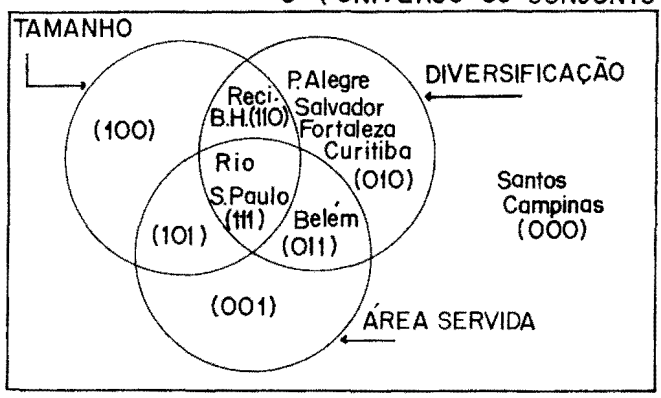


Diagrama Venn

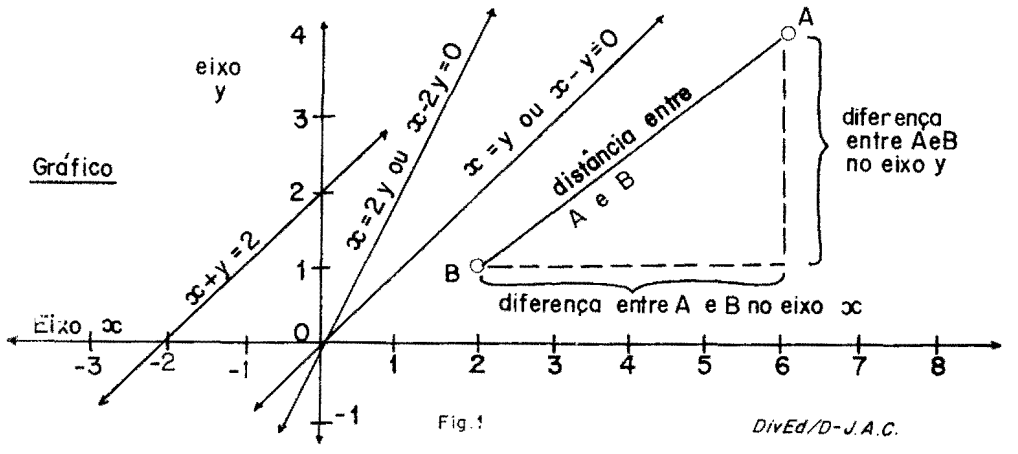


Fig. 1

DivEd/D-J.A.C.

3

Três Testes Estatísticos

INTRODUÇÃO

Existem na estatística muitas técnicas ou testes para investigar e medir o grau de correlação entre grupos de dados de vários tipos. Os exemplos neste trabalho mostram três testes que foram adaptados para serem aplicados a três tipos diferentes de dados.

Para entender mais profundamente a estrutura e a aplicação dos testes é necessário consultar um livro sôbre Estatística. Recomenda-se o seguinte livro em português: *Estatística Elementar* por Paul G. Hoel (Editôra Fundo de Cultura, Rio de Janeiro, 1963).

Na estatística se reconhecem três tipos fundamentalmente diferentes de informações numéricas: nominal, ordinal e com intervalos (em inglês "interval"). Outro tipo, o de proporção (inglês "ratio"), se distingue do tipo "interval" por certas características, mas os mesmos testes servem igualmente para dados "interval" e dados "ratio".

Se considerarmos 4 cidades brasileiras — Rio, São Paulo, Belo Horizonte e Brasília — podemos expressar ou descrever a distância do litoral, de cada cidade com três tipos diferentes de informação:

1) À pergunta: a cidade está no litoral? Pode-se responder: *Sim* ou *não*.
2) À pergunta: em que ordem estão distantes do litoral? Pode-se ordenar as 4 cidades de 1 a 4.
3) À pergunta: qual é a distância das 4 cidades do litoral? Pode-se medir a distância do litoral (em linha reta ou por rodovia) e obter uma variável.

	NOMINAL Está no litoral?	ORDINAL Ordem de distância do litoral	COM INTERVALOS Distância (km) do litoral
Rio de Janeiro.....	Sim (1)	1	0
São Paulo.....	Não (0)	2	50
Belo Horizonte.....	Não (0)	3	350
Brasília.....	Não (0)	4	900

É evidente na tabela que em geral a informação torna-se mais precisa indo-se do nominal até o "interval".

Note-se:

1) Muitas vêzes os dados em forma nominal podem ser expressos ou em forma verbal ou em forma numérica. Aqui encontramos a intersecção da língua com a matemática. Na tabela podemos responder à pergunta *está* no litoral com a resposta *sim* ou *não* ou com *1* ou *0*, como já foi indicado em outra ocasião.

2) Geralmente é melhor usar dados do tipo "interval" sempre que fôr possível. Muitas vêzes, porém, os dados disponíveis têm a forma ordinal ou nominal. Por exemplo, no caso dos dados ordinais, nos exames ou nas competições cada juiz dá uma posição numa escala ordinal a cada um dos candidatos, em vez de dar-lhes uma nota específica. Sabe-se a ordem dos candidatos, mas não se sabe a distância entre cada par.

   No caso dos dados nominais, sabe-se apenas quantos elementos estão contidos em cada classe. Não é possível pôr em ordem as classes. Por exemplo, há 20 mulheres, 12 com vestido vermelho, 6 com vestido amarelo e 2 com vestido branco. A classe de mulheres com vestido vermelho é a classe modal ou a moda porque tem o maior número de elementos. Mas não se pode dizer que a classe de mulheres com vestido vermelho seja melhor ou pior que a classe de mulheres com vestido branco ou amarelo. Em contraste, no caso, por exemplo, de 20 crianças colocadas em 3 grupos segundo os resultados de um exame, o grupos com notas entre 8 e 10 é superior ao grupo com notas entre 5 e 7 ou 2 e 4.

   As notas podem ser consideradas em uma escala com intervalos, as côres de vestidos não o podem.

3) Os dados nominais e ordinais chamam-se *não-paramétricos*. Os dados com intervalos são *paramétricos*. Nesse sentido, paramétrico indica a existência de uma *média aritmética* ou só *média*

(em inglês "*mean*"). Cada tipo de escala tem sua própria *medida* descritiva (em inglês "measure of central tendency").

Nos *dados nominais,* a classe modal ou moda é equivalente ao valor médio. A classe modal é a classe que tem o maior número de observações. Num histograma é a coluna mais alta. No exemplo da tabela das cidades a classe *não no litoral* constitui a classe modal porque é composta de 3 elementos, São Paulo, Belo Horizonte e Brasília. A outra classe tem só Rio de Janeiro.

Nos *dados ordinais,* a mediana (*median* em inglês) é equivalente ao valor médio. A mediana de um conjunto de valôres é a medida do meio, após os valôres terem sido colocados por ordem de grandeza. Por exemplo, quando a quantidade de valôres é ímpar:

1, 2, 3, [4], 5, 6, 7, o valor 4 é a mediana. Quando a quantidade de valôres é par:

1, 2, 3, [4,5], 6, 7, 8, a mediana é o valor que fica entre os dois valôres mais centrais na lista de números, 4 e 5.

O valor médio ou a *média* dos dados em "interval" é a soma dos valôres dividida pelo número de observações. No exemplo das distâncias das cidades do Brasil em relação ao litoral seria: $(0 + 50 + 350 + 900) \div 4 = 325$.

COEFICIENTE DE CORRELAÇÃO DE POSTOS

(em inglês: Spearman Rank Correlation coefficient).

Êste teste usa dados em forma ordinal. Na fórmula encontrada na tabela 1, os valôres 1 e 6 ficam constantes. Σ (sigma) significa a soma de todos os valôres que seguem (ne ste caso a soma dos valôres d^2). O valor *d* é a diferença entre os valôres da primeira coluna e da segunda relativas à classificação, para cada estado, *n* é o número de estados considerados. O índice de correlação *rho* sempre fica entre + 1 e — 1. Uma valor perto de zero indica uma correlação fraca que se obteria fàcilmente por casualidade. Um valor perto de mais um (+ 1.0) indica uma correlação forte positiva, enquanto um valor perto de menos um (— 1.0) indica uma correlação forte negativa.

A interpretação do índice *rho* depende do número de estados considerados. Tabelas encontradas nos livros de estatística dão os valôres convencionalmente usados para indicar os índices de correlação suficientemente altos que só raramente podem ser obtidos por casualidade.

O conceito de nível de confiança implica que o índice atingido teria sido obtido por casualidade menos de uma vez em vinte (1/20 ou com um nível de confiança de 95%) ou menos de uma vez em cem (1/100 ou com um nível de confiança de 99%).

Tabela 1 — *Coeficiente de correlação de postos*

ESTADOS	DADOS INICIAIS		CLASSIFICAÇÃO		d	d^2
	X	Y				
Maranhão	34	18	9	10	- 1	1
Piauí	30	35	10	9	+ 1	1
Ceará	45	73	8	3	+ 5	25
Rio Grande do Norte	57	72	3	4	- 1	1
Paraíba	54	59	7	7	0	0
Pernambuco	60	94	2	2	0	0
Alagoas	51	58	6	8	- 2	4
Sergipe	55	65	5	6	1	1
Bahia	57	71	4	5	- 1	1
Minas Gerais	71	139	1	1	0	0

$\Sigma\ d^2 = 34$

$$rho = 1 - \frac{6 \Sigma d^2}{n^3 - n} = 1 - \frac{6 \times 34}{1.000 - 10} = 1 - \frac{204}{990}$$

$$= 1 - 0.206 = + 0.794$$

Tabela 2 — *Notamos os níveis seguintes:*

N (número de casos)	NÍVEL DE 95%	NÍVEL DE 99%
10	0,56	0,75
15	0,44	9,60
20	0,38	0,53
25	0,34	0,47
30	0,31	0,43

Obtendo-se dois casos com os mesmos valôres (por exemplo ambos situados em 3.º lugar), é necessário dar o valor médio do 3.º e do 4.º (que não existe), isto é 3.5, ficando com 1, 2, 3.5, 3.5, 5, 6. Tendo-se três ou qualquer outro número ímpar de casos na mesma posição, é necessário usar o valor médio das posições que ocupariam se não fôssem iguais, 1, 2, 3, 3, 3, 6, 7 seria 1, 2, 4, 4, 4, 6, 7 etc....

"PEARSON PRODUCT MOMENT TEST"

Êste teste é parecido com o teste anterior "Spearman Rank Correlation" no sentido que o índice de correlação r, como o valor *rho* de Spearman, fica sempre entre + 1 e — 1, incluindo êstes valôres extremos, mas usa dados em forma "interval". Além disso, o cálculo é mais complicado, fato porém que não tem importância quando se emprega o computador para fazer os cálculos. No exemplo se empregam os dados iniciais que foram usados para calcular os índices Spearman. Os valôres foram arredondados para minimizar os cálculos neste exemplo. É importante, porém, observar que não se usam dados ordinais. Os dados e os cálculos estão na tabela 3. As duas variáveis, renda por pessoa e veículos por mil pessoas aparecem como variáveis X e Y na tabela 3. Para usar a fórmula no exemplo (há várias fórmulas para calcular o índice) é preciso saber o valor de N (neste caso é 10, os 10 estados) e a soma de cada coluna. O índice de correlação r, calculado aproximadamente neste exemplo, é mais alto (+ 0.875 em vez de + 0.794) que no caso do teste Spearman aplicado aos mesmos dados expressos em forma ordinal. Em geral, nota-se uma diferença dêste tipo devido ao fato de que o teste Pearson usa mais eficientemente os dados que o teste Spearman.

Os valôres estatìsticamente significativos do índice r para diferentes valôres de N se encontram na tabela 4.

Tabela 3

ESTADOS	X(renda)	Y(veículos)	XY	X^2	Y^3
Maranhão	3	2	6	9	4
Piauí	3	4	12	9	16
Ceará	5	7	35	25	49
Rio Grande do Norte	6	7	42	36	49
Paraíba	5	6	30	25	36
Pernambuco	6	9	54	36	81
Alagoas	5	6	30	25	36
Sergipe	6	7	42	36	49
Bahia	6	7	42	36	49
Minas Gerais	7	14	98	49	196
TOTAL	52	69	391	286	565

A fórmula é:

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{(N\Sigma X^2 - (\Sigma X)^2)(N\Sigma Y^2 - (\Sigma Y)^2)}}$$

É preciso calcular os cinco valôres seguintes, que são as somas dos valôres em cada coluna da tabela

$\Sigma X = 52 \qquad \Sigma X^2 = 286$

$\Sigma Y = 69 \qquad \Sigma Y^2 = 565$

$\Sigma XY = 391 \qquad N = 10$ Neste caso

$$r = \frac{10 \times 391 - 52 \times 69}{\sqrt{(10 \times 286 - 2.700)\,(10 \times 565 - 4.700)}}$$

$$r = \frac{3.910 - 3.580}{\sqrt{(2.860 - 2.700)\,(5.650 - 4.760)}} = \frac{330}{\sqrt{160 \times 890}}$$

$$r = \frac{330}{\sqrt{142.500}} = \frac{330}{377} = + 0.875$$

Tabela 4 — *Valôres para r* (Correlação Pearson Product Moment)

N-2	95%	99%	N-2	95%	99%
1.....	0,997	1,000	24.....	0,388	0,496
2.....	0,950	0,990	25.....	0,381	0,487
3.....	0,878	0,959	26.....	0,374	0,478
4.....	0,811	0,917	27.....	0,367	0,470
5.....	0,754	0,874	28.....	0,361	0,463
6.....	0,707	0,834	29.....	0,355	0,456
7.....	0,666	0,798	30.....	0,349	0,449
8.....	0,632	0,765	35.....	0,325	0,418
9.....	0,602	0,735	40.....	0,304	0,393
10.....	0,576	0,708	45.....	0,288	0,372
11.....	0,553	0,684	50.....	0,273	0,354
12.....	0,532	0,661	60.....	0,250	0,325
13.....	0,514	0,641	70.....	0,232	0,302
14.....	0,497	0,623	80.....	0,217	0,283
15.....	0,482	0,606	90.....	0,205	0,267
16.....	0,468	0,590	100.....	0,195	0,254
17.....	0,456	0,575	125.....	0,174	0,228
18.....	0,444	0,561	150.....	0,159	0,208
19.....	0,433	0,549	200.....	0,138	0,181
20.....	0,423	0,537	300.....	0,113	0,148
21.....	0,413	0,526	400.....	0,098	0,128
22.....	0,404	0,515	500.....	0,088	0,115
23.....	0,396	0,505	1.000.....	0,062	0,081

1) **Os níveis de confiança de 95% e 99% são equivalentes às probabilidades de $\frac{1}{20}$ e $\frac{1}{100}$ ou de 0.05 e 0.01. Existem outros níveis de confiança (por exemplo 999 por mil).**

2) Na prática, em geografia, não vale a pena fazer uma correlação com menos de 10 casos.
3) N, o número de casos (por exemplo pessoas, estados), é equivalente na tabela a N-2. Por exemplo, com N-10, procura-se 8 na tabela.

O TESTE "QUI-QUADRADO"

O teste "qui-quadrado" tem a mesma função que os dois outros testes porque também dá um índice de correlação. Usa-se uma tabela de probabilidades, (ver tabela) para averiguar se tal índice poderia ser derivado fàcilmente por casualidade. O índice de "qui-quadrado" não se encontra porém entre + 1 e — 1, mas entre 0 (zero) e um número muito grande. Além disso, a maneira de apresentar os dados para o teste "qui-quadrado" é diferente. Êste teste é aplicado exclusivamente nas situações em que os dados representam quantidades de entidades distintas (por exemplo, pessoas, cidades, fazendas). As quantidades são expressas em números inteiros.

Exemplo 1 — Há 40 fazendas hipotéticas, 20 na zona plana, 20 na zona acidentada. Ao todo há 20 tratores. Potencialmente há 4 tipos de fazenda: fazenda na zona plana com trator; fazenda na zona acidentada com trator; fazenda na zona plana que não tem trator e fazenda na zona acidentada que não tem trator.

Os números observados de fazendas de cada tipo podem ser apresentados na forma de um matriz com duas linhas e duas colunas. O número observado se escreve na célula apropriada e representa a freqüência de vêzes que foi contada ou observada cada fazenda do tipo definido na linha e na coluna.

O teste "qui-quadrado" é empregado para se saber se há uma correlação entre duas variáveis. Neste caso, queremos verificar se a forma do terreno influi no uso de tratores. Na tabela 5, no primeiro caso é evidente intuitivamente que as condições físicas não influem na presença dos tratores, enquanto no segundo caso é evidente que quase tôdas as fazendas com trator ficam na zona plana. No terceiro caso, porém, é mais difícil assegurar se as fazendas da zona plana têm mais tratores que as fazendas da zona acidentada só por casualidade ou pela influência das condições físicas.

A fórmula para calcular o índice "qui-quadrado" é a seguinte:

$$\text{Qui-quadrado} = \Sigma \frac{(E - O)^2}{E}$$

O é cada valor observado e E é o correspondente ao valor esperado. *Esperado* não no sentido de ser valor que o experimentador esperaria,

mas no sentido de ser o valor que seria obtido se a distribuição dos tratores nas fazendas não tivesse sido influenciada pelas condições físicas. Neste exemplo todos os valôres esperados são 10. As três situações hipotéticas e o cálculo do índice "qui-quadrado" para o terceiro caso se encontram na tabela 5.

Exemplo 1:

Tabela 5

Primeiro caso

		ZONA FÍSICA		
		Plana	Acidentada	Total
Fazenda	Com trator	10	10	20
	Sem trator	10	10	20
	TOTAL	20	20	40

Total geral

Segundo caso

		ZONA FÍSICA		
		Plana	Acidentada	Total
Fazenda	Com trator	19	1	20
	Sem trator	1	19	20
	TOTAL	20	20	40

Total geral

Terceiro caso

		ZONA FÍSICA		
		Plana	Acidentada	Total
Fazenda	Com trator	13	7	20
	Sem trator	7	13	20
	TOTAL	20	20	40

Total geral

Índice de "qui-quadrado" (X^2): "qui-quadrado" $= \frac{\Sigma(O - E)^2}{E}$

O é o valor *observado*
E é o valor *esperado*
Consideramos a matriz de valôres no terceiro caso (as linhas, coluna por coluna)

Linha	Coluna	O	E	(O—E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
1	1	13	10	3	9	.9
1	2	7	10	—3	9	.9
2	1	7	10	—3	9	.9
2	2	13	10	3	9	.9

$$\frac{\Sigma(O - E)^2}{E} = 3.6$$

Muitas vêzes o valor esperado não é o mesmo para cada célula na matriz. O método geral de calcular os valôres esperados é sempre o seguinte para qualquer célula na matriz:

A soma dos valôres observados na linha da célula, multiplicada pelo número total de valôres observados na coluna da célula, dividida pelo total geral.

$$\frac{\text{total da sua linha} \times \text{total da sua coluna}}{\text{Total geral}}$$

para qualquer célula, assim que linha I, coluna J, seria:

$$\frac{\text{total I} \times \text{total J}}{\text{total geral}}$$

No exemplo que se segue, o valor esperado na célula 1 1 é 60 (total da primeira linha) × 40 (total da primeira coluna) ÷ 100 (total geral), isto é, 24. Na tabela 6 se encontra o método de calcular os valôres esperados e o índice de "qui-quadrado".

Tabela 6

Valôres observados	Zona baixa	Zona média	Zona alta	Total
Granja cultiva milho.....	10	35	15	60
Granja não cultiva milho.....	30	5	5	40
TOTAL.....	40	40	20	100

Total geral

60 × 40/100

Valôres esperados			
	24	24	12
	16	16	8

Linha	Coluna	O	E	(O−E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
1	1	10	24	−14	196	8.2
1	2	35	24	11	121	5.0
1	3	15	12	3	9	0.8
2	1	30	16	14	196	13.1
2	2	5	16	−11	121	7.6
2	3	5	8	− 3	9	1.1

35.8

Depois de obter o valor de "qui-quadrado" é necessário estabelecer o número de *graus* de liberdade na matriz. O número de graus de liberdade é igual ao número de linhas na matriz menos 1, multiplicado pelo

número de colunas na matriz menos um: (L — 1) × (C — 1). No exemplo da tabela 6, é (2 — 1) × (3 — 1, isto é, 2.

A tabela 7 dá os valôres de "qui-quadrado" para os níveis de confiança de 95% e 99% para selecionados graus de liberdade. Tabelas mais amplas são encontradas em muitos livros de estatística. O valor de "qui-quadrado" calculado na tabela 6, 35.8 (com 2 graus de liberdade) indica uma correlação muito forte entre altitude e cultivo de milho nas granjas. Ao contrário, o índice de "qui-quadrado" de 3,6 (com 1 grau de liberdade) na tabela 5 fica abaixo do nível de confiança de 95% e assim indica uma desproporção entre os valôres que poderiam ser obtidos por casualidade mais de uma vez em vinte.

Tabela 7 — *Índices de "qui-quadrado"*

GRAU DE LIBERDADE	Nível de confiança 95% (1/20 ou 0,05)	99% (1/100 ou 0,01)
1	3,84	6,63
2	6,00	9,21
3	7,81	11,34
4	9,49	13,28
5	11,07	15,09
6	12,59	16,81
7	14,07	18,48
8	15,51	20,09
9	16,92	21,67
10	18,31	23,21
11	19,68	24,73
12	21,03	26,22
13	22,36	27,69
14	23,68	29,14
15	25,00	30,58
16	26,30	32,00
17	27,59	33,41
18	28,87	34,81
19	30,14	36,19
20	31,41	37,57
25	37,65	44,31
30	43,77	50,89
40	55,76	63,69
50	67,50	76,15

4

Amostragem

COMO PREPARAR UMA AMOSTRAGEM

Preparado apenas para esclarecer os conceitos sôbre amostragem e fornecer material para discussão, êste exercício trata de uma *população* total de cem fazendas. Em estatística, a palavra população se aplica não só a pessoas, mas a qualquer grupo definido de objetos ou de observações. Na realidade, seria desejável obter dados para tôdas as cem fazendas, visto o número tão pequeno da população.

O tamanho de uma amostragem pouco depende do tamanho da população total.

A idéia de ser necessário usar uma amostragem de 5%, de 10% ou de outra proporção da população total é quase inválida e irrelevante. O tamanho da amostragem depende sobretudo da *natureza* das perguntas do inquérito.

O número de perguntas e o tamanho da população pouco influem. Algumas vêzes, inclusive, o tamanho da população total é completamente desconhecido, de modo que não seria possível determinar o tamanho da amostragem tendo como base a população total.

Neste exercício, uma Companhia que vende equipamentos agrícolas vai distribuir um inquérito a certo número de fazendas. O número total das fazendas é 100. A Companhia tem uma lista numérica das fazendas e um mapa que indica a posição de cada fazenda. Para economizar tempo e dinheiro vai utilizar uma amostragem. O inquérito contém apenas 3 perguntas:

(1) A fazenda produz batata-inglêsa?

(2) A fazenda tem trator?
(3) A fazenda tem eletricidade?

Deve-se determinar:

(1) O tamanho da amostragem
(2) As fazendas às quais vai ser distribuído o inquérito.

Note-se que há fazendas em zona plana, em zona acidentada, em zona de floresta, junto à estrada asfaltada e perto das vilas. As fazendas situam-se em três municípios: A, B e C.

É evidente que a amostragem deve ser representativa da população total de fazendas. Embora fôsse mais fácil visitar só as fazendas perto da estrada o resultado assim obtido correria o risco de ser pouco exato, dado que provàvelmente elas seriam mais modernizadas (dispondo de eletricidade e trator) que as fazendas mais afastadas da estrada.

Há vários métodos que ajudam a assegurar a elaboração de uma amostragem representativa e imparcial.

AMOSTRAGEM ESTRATIFICADA

Incluir fazendas de cada tipo (plano, estrada, floresta) em números proporcionais à sua porcentagem dentro da população total. É relativamente difícil, porém, calcular quantas fazendas há de cada tipo.

Recomenda-se o método seguinte (a tabela tem apenas fazendas selecionadas para mostrar o método).

POR EXEMPLO	Na estrada	Perto da vila	Na zona acidentada	Na floresta	Tipo "Binary Class"	
Fazenda:						
n.º 1.....	0	0	0	0	0000	0
n.º 2.....	0	0	0	0	0000	0
n.º 22.....	0	0	1	0	0010	2
n.º 46.....	1	1	0	0	1100	12
n.º 64.....	0	0	0	1	0001	1
n.º 97.....	1	0	0	0	1000	8

a) Sem considerar a questão dos 3 municípios, há 4 critérios: estrada, vila, zona acidentada e floresta. Existem 16 combinações possíveis dos 4

critérios (calculado pela fórmula 2^n, neste caso 2^4). Na tabela cada fazenda recebe um número em "binary", o qual convertido em "denary" dá uma classe. Neste exemplo, não existem tôdas as classes possíveis. Não há a classe 1 1 1 1 ou 15 porque não há fazendas situadas, simultâneamente, na estrada, perto da vila, em zona acidentada e na floresta.

b) Sem considerar as várias condições na região em questão, é possível escolher a amostragem usando só os números das fazendas na lista. Pode-se, por exemplo, visitar cada quinta fazenda, 1, 6, 11, 16, 21..., para incluir na amostragem 20 fazendas. Às vêzes, porém, existe o perigo, usando-se números em ordem regular, de encontrar uma regularidade na própria lista. Por exemplo, preparando uma amostragem de casas num subúrbio, seria possível tomando-se cada décima casa, encontrar tôdas as vêzes uma casa de esquina.

Recomenda-se usar números aleatórios.

EXERCÍCIO: ETAPAS

(1) Determinar o tamanho da amostragem (ou seja de quantas fazendas constará a amostragem).

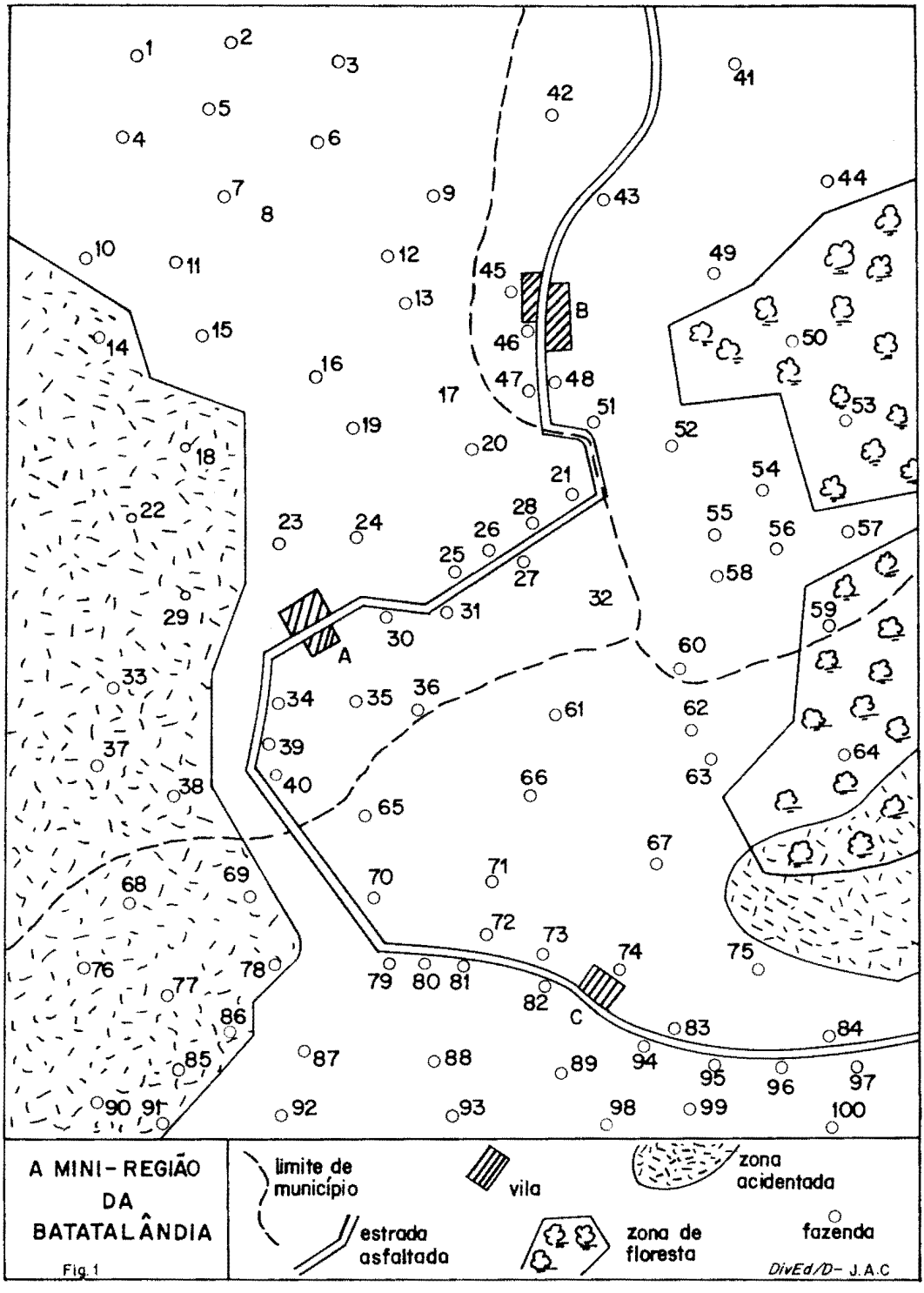
(2) Decidir que fazendas serão as escolhidas, usando o método que lhe parecer melhor.

(3) Referir-se à tabela de dados. A tabela representa a realidade, isto é, as características da população total, o que não seria disponível na prática (de outro modo não seria preciso preparar os inquéritos).

(4) Comparar os totais obtidos na amostragem com os totais verdadeiros. Se, por exemplo, com a amostragem de 20 obtém-se 14 fazendas com batata-inglêsa, 3 com trato e 3 com eletricidade, poder-se-ia predizer 5 vêzes mais, de cada tipo da população inteira, à proporção da amostragem em relação à população total sendo 1 $\times$ 5 isto é: 70, 15 e 15. Qual é o grau de divergência entre a sua previsão e a verdade? Consultar a tabela "a realidade de Batatalândia" para saber a verdade. Que relação existe entre a(s) pergunta(s) feita(s) no inquérito e o tamanho da amostragem? Como ilustração, consideremos o caso de Batatalândia.

Na realidade, 60 fazendas produzem batata-inglêsa. Amostragem de vinte revelariam muitas vêzes quantidades de 9, 10, 11, 12, 13, 14, etc. fazendas cultivando batata-inglêsa, as quais, multiplicadas por 5, dariam estimativas de 45, 50, 55, 60, 65, 70. O grau de divergência com a população total geralmente variaria entre 0 e 30%. No caso de outro atributo (por exem-

plo, piscina), se na realidade apenas 10 fazendas possuíssem piscina, numa amostragem de 20, encontrar-se-iam valôres de 0, 1, 2, 3, 4, raramente valôres superiores. Multiplicados por 5, êstes valôres dariam estimativas de 0, 5, 10, 15, 20, 25, os quais com exceção do 10 divergiriam fortemente do número verdadeiro de fazendas com piscina.



A MINI-REGIÃO DA BATATALÂNDIA

Fig. 1

DivEd/D- J.A.C

A realidade de Batatalândia — 1 = presente
B — Batata-inglêsa, T — Trator, E — Eletricidade — 0 = ausente

N.º DA FAZENDA	B	T	E	N.º DA FAZENDA	B	T	E
1	1	1	0	52	1	1	0
2	1	1	0	53	0	0	0
3	1	1	0	54	0	1	0
4	1	1	0	55	0	1	0
5	1	0	0	56	0	0	0
6	1	1	0	57	0	0	0
7	1	0	0	58	0	0	0
8	1	0	0	59	0	0	0
9	0	1	0	60	0	0	0
10	1	0	0	61	0	1	0
11	1	0	0	62	0	1	0
12	0	1	0	63	0	1	0
13	0	1	0	64	0	0	0
14	1	0	0	65	1	0	0
15	1	0	0	66	1	0	0
16	1	1	0	67	0	0	0
17	0	0	0	68	1	0	0
18	1	0	0	69	1	0	0
19	1	1	0	70	1	1	0
20	0	1	0	71	1	1	0
21	1	0	1	72	1	1	0
22	1	0	0	73	1	0	0
23	1	1	0	74	1	0	1
24	0	0	0	75	0	0	0
25	1	0	1	76	1	0	0
26	1	0	1	77	1	0	0
27	1	0	1	78	1	0	0
28	1	0	1	79	1	0	1
29	1	0	0	80	1	0	1
30	1	0	0	81	1	0	1
31	1	0	0	82	1	0	1
32	0	0	0	83	1	0	1
33	1	0	0	84	0	0	1
34	1	0	0	85	1	0	0
35	1	0	0	86	1	0	0
36	1	0	0	87	0	1	0
37	1	0	0	88	0	1	0
38	1	0	0	89	0	1	0
39	1	0	0	90	0	0	0
40	1	0	0	91	1	0	0
41	0	1	0	92	0	1	0
42	0	1	0	93	0	0	0
43	0	1	0	94	1	0	1
44	0	1	0	95	0	0	1
45	1	0	1	96	0	0	1
46	1	0	1	97	0	0	1
47	0	0	1	98	1	0	0
48	1	1	1	99	0	0	0
49	0	0	0	100	0	0	0
50	0	0	0				
51	1	0	0	TOTAL	60	30	20

5

Conjuntos

INTRODUÇÃO

O conceito de conjunto (em inglês "set") desempenha importante papel na vida cotidiana e na matemática moderna. À medida que o vocabulário da nova matemática vai penetrando na geografia, torna-se cada vez mais freqüente o uso de vários têrmos da linguagem dos *conjuntos*. Além disso, certos geógrafos nos Estados Unidos, na Inglaterra e na URSS já mostraram as vantagens de se empregar o conceito de *conjuntos* diretamente no vocabulário e nos métodos geográficos.

Da mesma maneira que se empregam os números em muitas situações (por exemplo, usa-se 1, 2, 3, 4, para contar pessoas, árvores, navios, dias), os conjuntos também têm aplicações gerais. Por exemplo, as letras da palavra gato, G, A, T, *e* O, são quatro elementos de um *conjunto*. Os números 2, 4, 6, 8, formam o *conjunto* de números inteiros positivos entre 1 e 9. Na geometria, uma linha pode ser considerada como um conjunto de pontos, como também uma superfície. Um conjunto pode ser finito ou infinito. O conjunto das letras que formam a palavra GATO é finito. O conjunto de números inteiros é infinito.

Entre as vantagens do uso do conceito e do vocabulário dos conjuntos na geografia, notamos os pontos seguintes:

1) Os conjuntos, aliados à lógica, ajudam a pensar claramente.
2) Os conjuntos dão uma forma de expressão comum a muitas disciplinas e facilitam o intercâmbio de idéias e de métodos entre elas.
3) Os conjuntos facilitam a organização e classificação e oferecem uma linguagem muito concisa.

VOCABULÁRIO

A teoria dos *conjuntos* é um tipo de álgebra. Usa certas palavras e certos conceitos. Muitas vêzes a álgebra pode ser acompanhada por diagramas Venn (nome de um matemático, John Venn). Êstes diagramas permitem representar gràficamente as situações expressas algèbricamente.

O conjunto das oito cidades na tabela foi escolhido como um exemplo quase geográfico para ilustrar o vocabulário e os símbolos empregados na teoria dos conjuntos. O grupo de oito cidades chama-se conjunto universal, ou universo de discurso, porque é o conjunto que está sendo considerado. Devemos definir os elementos que compõem o conjunto, seja escrevendo os elementos entre *chaves* seja (no caso de um conjunto que contém muitos elementos ou um número infinito de elementos) com alguma definição em forma de palavras ou de têrmos matemáticos. Neste exemplo, podemos definir o conjunto universal do seguinte modo (usando formas abreviadas dos nomes para economizar espaço): { SPO, RDJ, REC, PAL, SAL, BHO, BRS, MAN }. Os elementos do conjunto podem ser escritos em qualquer ordem. Êste conjunto tem 8 elementos. Podemos nos referir ao conjunto com uma letra, por exemplo A. Podemos dizer "São Paulo" *pertence* ao conjunto A, ou *SPO* ε *A*. Um conjunto com só um elemento chama-se *conjunto unitário.* (Neste exemplo, o conjunto de cidades capitais federais tem um só elemento, Brasília). Um conjunto que não tem nenhum elemento chama-se conjunto vazio. É representado pelo símbolo { } ou ϕ. (Neste exemplo, o conjunto de cidades brasileiras com estação elétrica nuclear é ainda um conjunto vazio).

		Na costa	Pôrto marítimo	Pôrto fluvial	Cidade antiga	Capital federal	Refinaria de petróleo	Estação elétrica nuclear
1. São Paulo.....	SPO	0	0	0	1	0	1	0
2. Rio de Janeiro	RDJ	1	1	0	1	0	1	0
3. Recife.....	REC	1	1	0	1	0	0	0
4. Pôrto Alegre..	PAL	1	1	0	1	0	1	0
5. Salvador.....	SAL	1	1	0	1	0	1	0
6. Belo Horizonte	BHO	0	0	0	0	0	1	0
7. Brasília.....	BRS	0	0	0	0	1	0	0
8. Manaus.....	MAN	0	0	1	0	0	1	0
Subconjunto...	—	A	B	C	D	E	F	G

Igualdade — Dois conjuntos são iguais se tiverem os mesmos elementos. Assim, no exemplo citado, o conjunto de cidades na costa é igual ao conjunto de cidades com pôrto marítimo.

Subconjunto — Se todos os elementos de um conjunto forem também elementos de outro conjunto, diz-se que o primeiro conjunto é *subconjunto* do segundo. No exemplo, o conjunto de cidades antigas é um *subconjunto* do conjunto universal. Note-se que êste subconjunto pode também ser considerado como um conjunto nôvo. Qualquer conjunto pode ser considerado subconjunto dêle mesmo, mas neste caso não é um subconjunto *próprio*. O símbolo C é usado para indicar que um conjunto é subconjunto do outro.

Reunião — A reunião de dois (ou mais de dois) conjuntos contém elementos que pertencem *ou* ao primeiro conjunto, *ou* ao segundo, *ou* aos dois conjuntos. Na linguagem o mesmo conceito se expressa pela palavra *ou* quando *ou* significa *ou* uma coisa *ou* outra coisa, *ou* as duas coisas. Ilustra-se, êste conceito no diagrama Venn, na página com figuras. No exemplo, Brasília é a única cidade que não pertence à reunião dos dois conjuntos. O símbolo U é usado para indicar a reunião de conjunto. *Interseção* — A interseção de dois (ou mais de dois) conjuntos contém elementos que pertencem ao primeiro conjunto, *e* ao segundo. Na linguagem é como a palavra *e*. É mais exigente do que a palavra *ou*. O símbolo $\cap$ indica a interseção dos conjuntos.

Conjuntos disjuntos * — Dois conjuntos são disjuntos se não têm nenhum elemento em comum. Por exemplo: o conjunto de portos marítimos e o conjunto de portos fluviais. Se definirmos êstes dois conjuntos como B e C, podemos expressar a idéia de conjuntos disjuntos com os símbolos $B \cap C = \{\ \}$, ou dizer que a interseção dos conjunto B e C é um conjunto vazio.

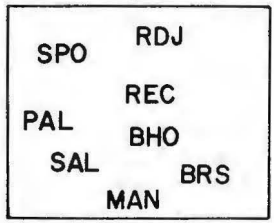
Complemento de um conjunto — Um exemplo na matemática: o complemento do conjunto dos números inteiros pares é o conjunto dos números inteiros ímpares. No caso do conjunto de 8 cidades do exemplo, o complemento do conjunto de cidades capitais federais (Brasília, um conjunto único) são tôdas as outras cidades. O complemento do conjunto é indicado pela mesma letra do conjunto com um traço em cima: D e $\bar{D}$.

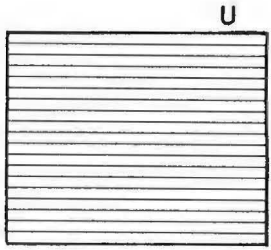
Correspondência — Dois conjuntos são correspondentes ou equivalentes ($\longleftrightarrow$), se a cada elemento de um conjunto é possível pôr exatamente um elemento correspondente do outro conjunto. Por exemplo, numa biblioteca cada livro possue uma ficha de identificação que ficará ar-

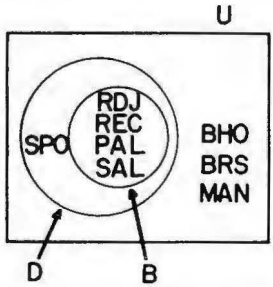
* "disjoint"

quivada no caso de empréstimo. Cada ônibus tem *uma* posição para o motorista. Quando estamos contando, usamos os números inteiros positivos (e zero), colocando em uma correspondência de um a um, um número e um objeto.

U (conjunto universal)

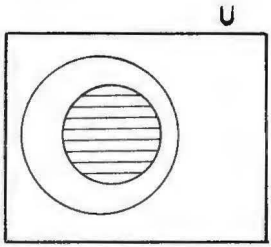


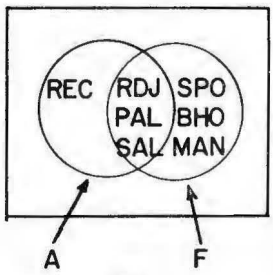




Subconjunto

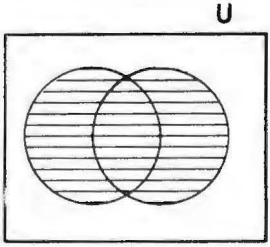
$B \subset D$





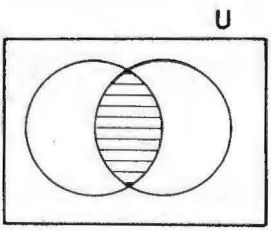
Reunião

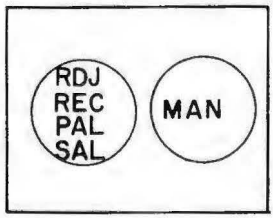
$A \cup F$



Intersecção

$A \cap F$

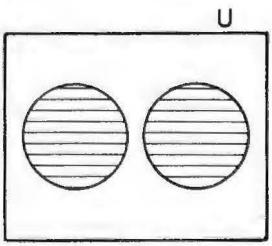


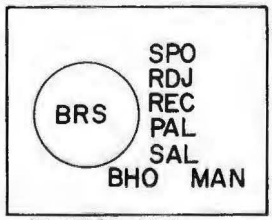


Conjuntos disjuntos

$B \cap C = \{\}$

(A intersecção de A e B é um conjunto vazio)





Complemento

$D \text{ e } \overline{D}$

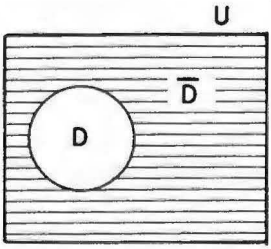


Fig.1

DivEd/D-J.A.C

Notas Sôbre Distância e Localização em Geografia Humana

DIFERENTES ESPÉCIES DE DISTÂNCIA EM GEOGRAFIA

Névora é uma pequena cidade em um país em desenvolvimento. É a residência, entre outras coisas, de um médico-assistente que viaja uma vez por semana, para cada uma das cinco vilas mais próximas, a fim de prestar serviços, por curto período, nas clínicas locais. A situação de viajante do médico-assistente de Névora é apresentada para ilustrar, da maneira mais simples possível, uma situação geral que ocorre no mundo, numa escala variável de regiões e diversos graus de complexidade: a relação (ou a falta de relação) entre a distância espacial direta, a distância real (p. ex., a estrada mais curta), o tempo-distância e o custo-distância.

O quadro n.º 1 mostra as informações básicas sôbre a viagem média, utilizada pelo médico de Névora.

Quadro 1

DE NÉVORA	DISTÂNCIA EM km		Velocidade km p/Hora	Tempo minutos	Custo centavos
	Direta	Real			
A.....	6	8	8	60	80
B.....	4	6	6	60	30
C.....	5	6	9	40	60
D.....	3	6	3	120	0
E.....	2	4	2	120	0
TOTAL....	20	30	(28/5)	400	170
MÉDIA....	4	6	(28/5)	80	34

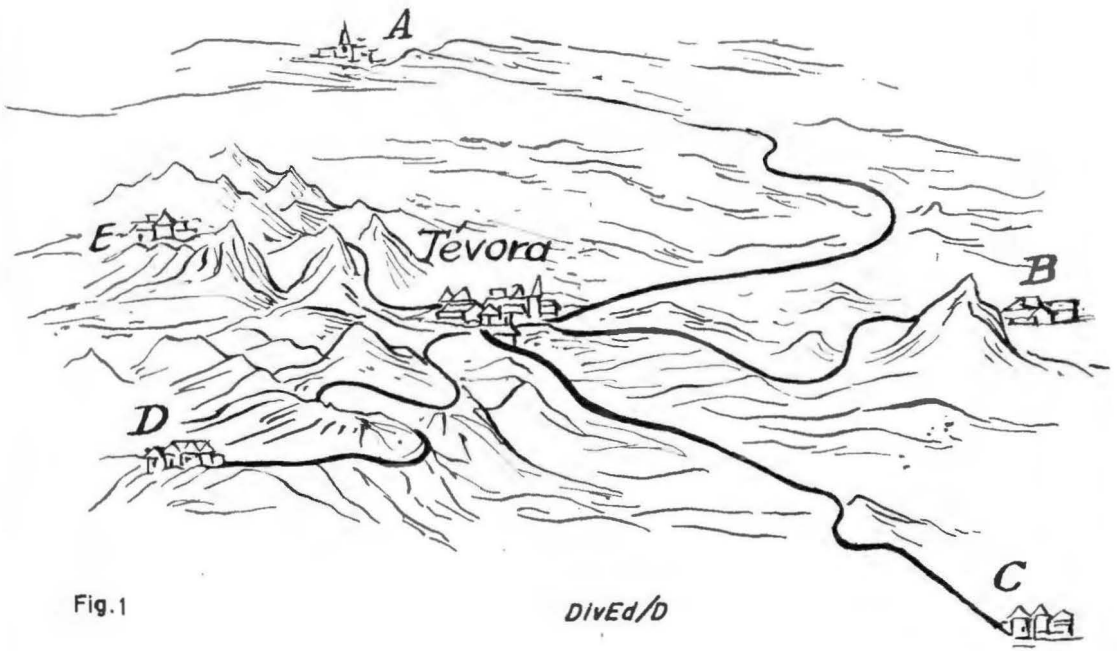


Fig.1 DivEd/D

As localidades A e C são alcançadas por estrada, utilizando-se uma carroça que viaja à velocidade de 8 e 9 km/h, respectivamente. A localidade B é alcançada a lombo de cavalo à velocidade de 6 km/h. D e E são alcançadas sòmente por trilhas difíceis, a pé e à velocidade de 3km/h e 2 km/h. A viagem de carroça (até A e C) custa 10 centavos por km, a viagem a cavalo (até B) 5 centavos por km.

Os dados podem ser processados de diversas maneiras, para mostrar como as voltas ou desvios, tempos e custos modificam as distâncias diretas iniciais (euclidianas).

1. Distância *direta* versus distância *real*, sendo a distância real o meio de viagem viável mais curto (num dado contexto). Neste exemplo, os seguintes índices podem ser elaborados:

a() Distância *Direta* dividida pela distância *Real* (opção adicional: multiplicar por 100 para eliminar a casa decimal).

(b) Distância *Real* dividida pela distância *Direta* (opção adicional: multiplicar por 100 para eliminar a casa decimal).

(c) Distância *Real* menos distância *Direta* para obter o excesso da distância Real (em têrmos absolutos) sôbre a distância Direta.

Quadro 2 — *Índices de Tévora às Localidade A-E*

PARA	1(a)	1(b)	1(c)	1(d)
A.....	75	133	2	— 17
B.....	67	150	2	0
C.....	83	120	1	— 30
D.....	50	200	3	+ 50
E.....	50	200	2	+ 50

O efeito comparativo das voltas ou desvios nas cinco viagens pode ser demonstrado sob a forma de mapa. Um dos métodos para fazê-lo é o de encontrar o grau médio de voltas, expressando a distância total real viajada (30 km) em percentagem da distância total direta (20 km), isto é, um índice de volta média de 150. Isto pode, então, ser comparado com os índices da tabela 2, coluna 1(b). Os resultados da subtração dos índices médios de 150 de cada índice individual são apresentados na coluna 1(d). Digressão: As coisas não são tão simples como parecem à primeira vista. O índice de percurso médio obtido da soma dos valôres na coluna 1(b) é 803,5 ou 160,6, e não 150. Por que assim? Uma é média ponderada a outra não.

É possível executar um reajustamento da distância direta para a distância tempo, ou da distância real da estrada para a distância tempo. Um meio (tentar achar outros meios) de relacionar a distância direta com a distância tempo é comparar a parte que cada viagem tem, do total da distância direta e do total do tempo de viagem e expressar um dêsses valôres como parte proporcional do outro. Por exemplo, a distância direta de Tévora para C, 5 km, é 25% de tôda a extensão do percurso (20 km), mas leva apenas 40 minutos, 10% de todo o tempo de viagem.

Quadro 3

	km	%	Minutos	%	Índice
A.....	6	30	60	15	50
B.....	4	20	60	15	75
C.....	5	25	40	10	40
D.....	3	15	120	30	200
E.....	2	10	120	30	300
TOTAL....	20	100	400	100	

A tabela 3 mostra os índices de favorabilidade de tempo para as cinco localidades: percentagem de minutos dividida pela percentagem de km, multiplicada por 100.

OS MAPAS

O mapa I mostra a situação básica, com as distâncias diretas e reais, dadas em km.

O mapa II mostra as cinco localidades A-E situadas de acôrdo com suas distâncias reais de Tévora.

O mapa III mostra as cinco localidades A-E relocadas de acôrdo com o seu tempo de viagem de Tévora.

O mapa IV mostra as cinco localidades A-E relocadas de acôrdo com o seu custo-distância de Tévora. Naturalmente seria possível caminhar para A, B e C como para D e E.

O mapa V mostra a divergência de cada lugar da média global causada por seu índice de voltas ou desvios. Os valôres da coluna 1(d) da tabela 2 são usados.

O mapa VI mostra a divergência de cada localidade da média global da distância real, motivada pela diferença na velocidade de viagem; os valôres da tabela III, coluna dos Índices são usados.

Se o que foi anteriormente dito foi fàcilmente compreendido por todos, desenvolvam as idéias abaixo:

1. O médico-assistente é pago a razão de 100 centavos a hora. O tempo de viagem até os cinco vilarejos está incluído em seu tempo de trabalho. Presumindo-se que êle gaste igual período de tempo em cada vila, classifique as cinco vilas de acôrdo com o que sua viagem custa ao govêrno para cada uma delas.

2. Suponha que V. está considerando as várias espécies de distância não entre um determinado lugar e alguns outros, mas entre cada local e cada um dos outros locais. Suas informações serão ordenadas sob a forma de matriz. Formule um exemplo simples e considere as distâncias direta, real e a distância-custo entre cada par de lugares. No mapa III foi possível relocar cada vila em uma nova posição de acôrdo com o "tempo corrigido" em relação à Tévora. A pergunta que vale 1.000 centavos é a seguinte: Pode V. fazer um ordenamento semelhante para *relocar* cada localidade na sua nova posição de "tempo corrigido" em relação a cada uma das outras localidades?

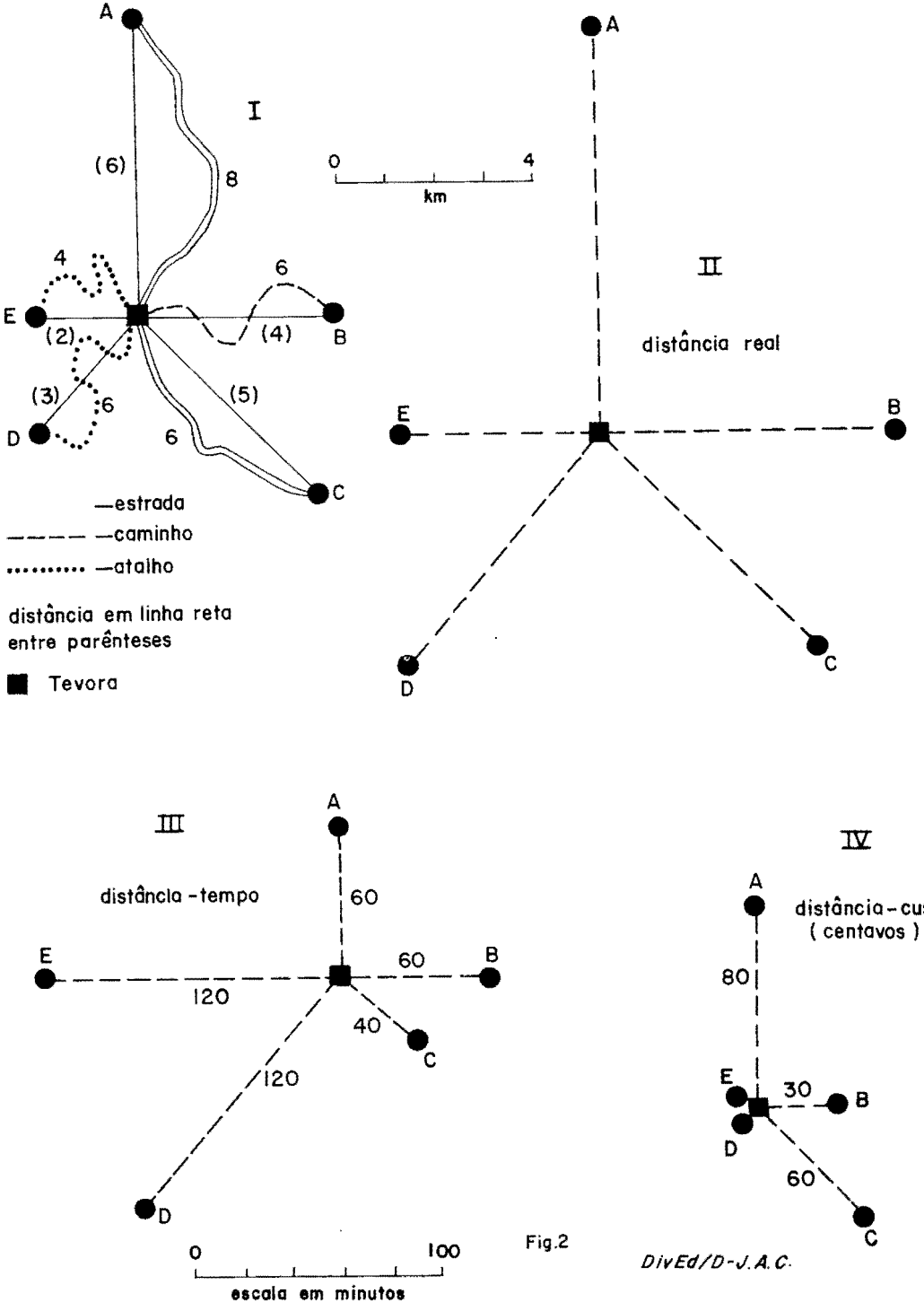


Fig.2

DivEd/D-J.A.C.

V

divergência do índice global de voltas de 150

VI

divergência da distância real causada pelas diferentes velocidades de viagens.

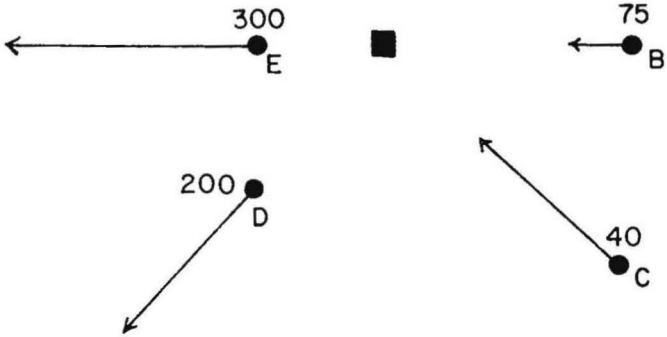


Fig.3

DivEd/D-J.A.C.

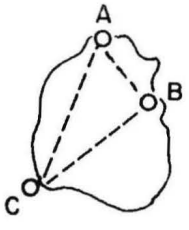
DISTÂNCIA DIRETA E REAL ENTRE MUITOS PARES DE LOCALIDADES

O transporte por terra é caracterizado pelo uso de canais de comunicação que formam uma rêde. Sôbre essas rêdes é raramente possível fazer-se um percurso direto (linha reta numa superfície plana) entre duas localidades. Os desvios do caminho direto podem ser medidos de tal modo que índices comparáveis de desvios podem ser produzidos, não obstante a dimensão da área sob consideração. Tal índice pode ser derivado como se segue:

$$\frac{\text{Distância atual mais curta} \times 100}{\text{Distância direta}}$$

quando vários pares de lugares estão sob consideração a informação pode ser tabulada sob a forma de matriz: veja diagrama abaixo.

As somas das colunas nas duas primeiras matrizes podem ser usadas para dar os índices globais para cada localidade em relação a tôdas as outras na rêde. O total global, ou a soma das somas das colunas pode, de igual modo, ser usada para a caracterização da rêde total. Notar que as três matrizes são quadradas e simétricas.



	A	B	C
A	0	96	200
B	96	0	165
C	200	165	0

	A	B	C
A	0	80	100
B	80	0	110
C	100	110	0

	A	B	C
A	0	120	200
B	120	0	150
C	200	150	0

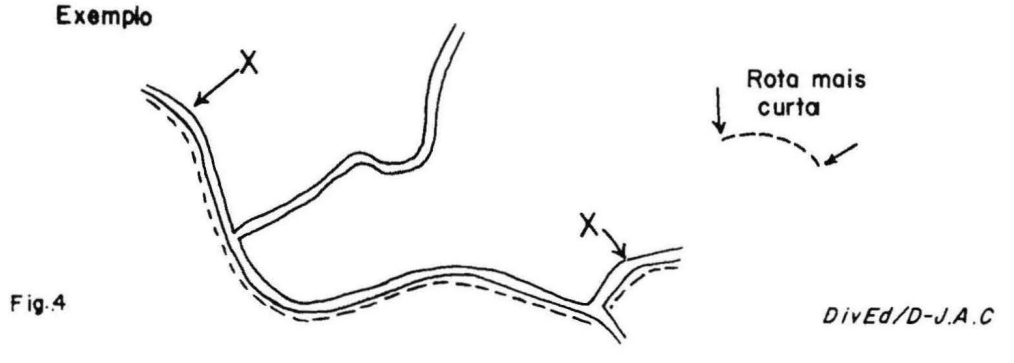


Fig.4

DivEd/D-J.A.C

De um mapa 1"O.S., de área conhecida, tomar uma parcela de 10 × 10 km. Usando números aleatórios, situar oito localidades. Calcular os índices acima descritos e comparar seu resultado global com o dos outros estudantes.

Notar que a maioria das localidades escolhidas ocasionalmente estará situada fora das estradas. Traçar a rota mais curta e tomar a distância mais direta para a estrada mais próxima.

LOCALIZANDO UMA USINA SIDERÚRGICA

Até agora Fishport tem importado todo o aço de que necessita. Para evitar isto decidiu construir uma usina siderúrgica. Os principais materiais de que necessita são o coque metalúrgico e o minério de ferro. O calcário, o minério de manganês, o suprimento de água e, naturalmente, operários, são também necessários. Neste exercício apenas o carvão e o minério de ferro serão considerados. Quando o aço fôr produzido será enviado aos mercados de Fishport. Para a produção de uma tonelada de aço são necessárias uma tonelada de carvão e duas toneladas de minério. Algumas sucata de aço pode também ser usada, mas isto não será levado em conta. No sistema ferroviário apresentado no mapa, o frete para o transporte de uma tonelada de carvão, minério ou aço é representado por uma unidade de dinheiro para cada trecho da linha férrea entre qualquer localidade vizinha, marcada com pontos prêtos.

Oito localidades foram propostas como possíveis sítios para a siderúrgica. Estão designados no mapa. O carvão é minerado sòmente em Basset e o minério de ferro é extraído apenas em Ullen. Está prevista a construção de apenas uma siderúrgica.

EXEMPLO DE EXERCÍCIO

Se todo o aço é vendido em Prescott, qual é o melhor lugar para a instalação da usina. Procure qual será o frete ferroviário se a construção fôr efetuada em Prescott. Resolva a questão para uma tonelada de aço.

Duas toneladas de minério (Ullen para Prescott)	2 unidades de distância custam	4
Uma tonelada de carvão (Basset para Prescott)	2 unidades de distância custam	2
Uma tonelada de aço vendida no local	0 unidade de distância custa	0
	Custo total	6

Quanto seria o frete se a construção fôsse em Basset?

Duas toneladas de minério (Ullen para Basset)	4 unidades de distância custam	8
Uma tonelada de carvão no local	0 unidade de distância custa	0
Uma tonelada de aço (Basset para Prescott)	2 unidades de distância custam	2
	Custo total	10

Exercício 1. Calcular os fretes ferroviários para a entrega de uma tonelada de carvão e duas toneladas de minério para a construção em Gruver, Camden, Page, Ullen.

Exercício 2. Se o aço é vendido em Prescott e em Lucas, pela mesma quantia, qual é a localidade com o frete ferroviário mais baixo agora? E qual é o mais alto?

Sugestão: Calcular fretes ferroviários para duas toneladas de carvão e aço e quatro toneladas de minério.

Exercício 3. Se o aço é vendido pelo mesmo preço em tôdas as 8 localidades, qual é a localidade com o frete ferroviário mais baixo?

Sugestão: Calcular os fretes ferroviários para oito toneladas de carvão e aço e dezesseis toneladas de minério.

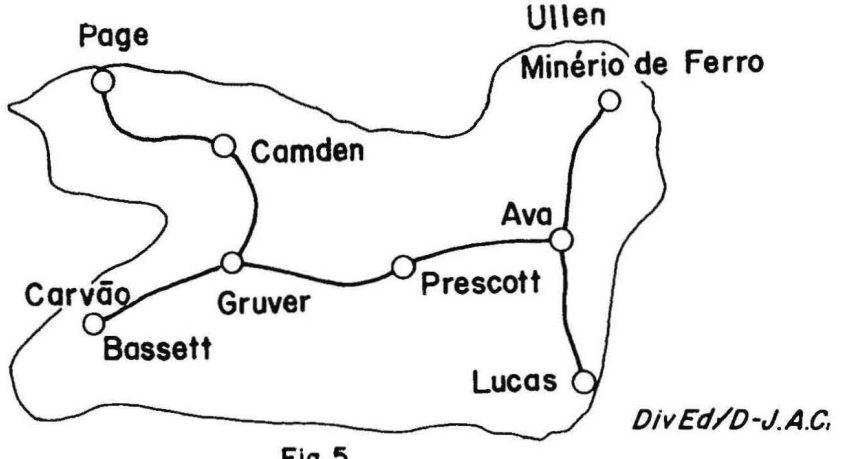


Fig.5

7

O Problema de Difusão

INTRODUÇÃO

A consciência da maneira pela qual muitas coisas, incluindo pessoas, espécies de plantas, invenções, doenças e idéias, espalham-se ou difundem-se sôbre um espaço, através do tempo, já de há muito faz parte da experiência do Homem. O estudo consciente, formal (e poder-se-ia dizer científico?) das formas de difusão por geógrafos é, entretanto, relativamente recente. Na verdade, para muitos de nós, êle começa em 1952 com o trabalho de T. Hagerstrand, "A propagação das Ondas de Inovação" (Lund Studies in Geography, Ser. B., Human Geography, n.º 4, Lund, Suécia).

Talvez, o conceito mais básico é o de que muitas coisas se espalham pelo espaço, durante o tempo (elas podem na verdade contrair-se de maneira semelhante). Muita distribuição espacial em um determinado momento pode ser uma distribuição em curso de alteração; ela pode ter sido menos extensa no passado e poderá tornar-se mais extensa no futuro.

Numa escala razoàvelmente pequena, a difusão da doença da bôca e dos pés (foot and mouth disease) a centenas de fazendas nos Midlands (Inglaterra) no período de 1967/1968, a partir de Oswestry, fornece um bom exemplo de um processo de difusão, cujo início, dispersão e consumação podem ser estudados com precisão. Na escala mundial, a difusão da alfabetização tanto "para baixo", no sentido social, quanto horizontal, particularmente fora das áreas urbanas, é um exemplo de um processo mais complexo ainda em desenvolvimento e de difícil visualização.

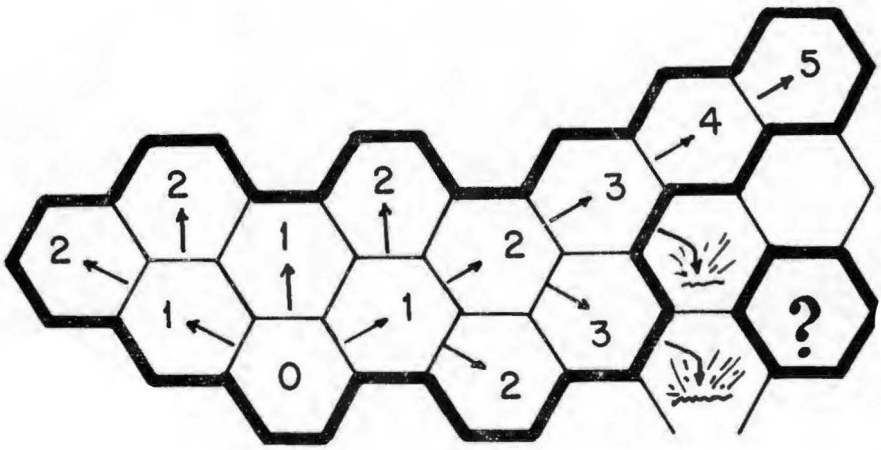


Fig.1 DIvEd/D-J.A.C.

O objetivo dêste artigo é apenas o de fornecer exemplos práticos, todos fictícios, de três diferentes tipos de difusão pelo espaço. Por conveniência serão êstes denominados de mecanicista, probabilístico e baseado em decisões; êstes têrmos poderão ser considerados por alguns como super-simplificados e inadequados.

O USUAL

A figura 2 mostra os têrmos usados no manuseio de um sistema de hexágonos. Os hexágonos são usados, de preferência, aos quadrados (ou aos triângulos) porque possuem seis lados, cada um dos quais confina com um hexágono adjacente. Em uma rêde de quadrados, há quatro lados e quatro cantos, de maneira que a proximidade dos quadrados na primeira camada é de dois tipos. Qualquer hexágono da rêde pode ser localizado com um par apenas de números ordenados, mas se as "linhas" forem horizontais (isto é, da esquerda para a direita, através da página) as colunas não estarão em ângulo reto, mas em ângulo de sessenta graus. A notação é mostrada na figura 2. Para jogos e modelos do tipo a ser descrito neste artigo, é conveniente dividir a área (espaço) em compartimentos numeráveis (hexágonos ou quadrados) ao invés de considerá-la contínua. Da mesma maneira, o tempo é dividido em compartimentos (turnos), tais como dias ou semanas.

A noção do que seja uma camada de hexágonos é também dada na figura 2. Qualquer camada dada, consiste em uma única combinação

de hexágonos, como camada de um dado lugar de origem (0). Note-se que um lugar de origem pode ter parte de sua primeira camada e ou outras camadas no mar.

A DIFUSÃO MECANICISTA

A figura 3 poderá ser usada como base para a experimentação da difusão mecânica. Suponhamos que exista uma fazenda leiteira em cada hexágono, excetuando-se aquêles ocupados por montanhas ou pântanos. Existem, ao todo, cento e cinqüenta hexágonos livres e, portanto, cento e cinqüenta fazendas.

Uma doença surge em determinada fazenda (daqui por diante denominada fazenda de origem). Na primeira semana, tôdas as fazendas dos hexágonos contíguos à fazenda de origem são contaminadas — isto é, tôdas as fazendas da primeira camada. Na semana seguinte, tôdas as fazendas dos hexágonos contíguos às fazendas contaminadas são afetadas, e assim por diante.

Calcular e tabular o número de fazendas afetadas cada semana, e o total de fazendas contaminadas ao cabo de 10 semanas, se a doença surge (1) em 01 10 (isto é na península Wayward), (2) em 10 10 (uma fazenda bastante central), (3) em 13 14, (4) em 17 14.

Exemplo de tabulação do local (1), 01 10. Seria da seguinte forma:

	Em uma única semana	Cumulativamente
Início	1	1
Fim de semana I	3	4
Fim de semana II	1	5
Fim de semana III	2	7
Fim de semana IV	2	9

e assim por diante.

Como não há fazendas nas montanhas, nos pântanos ou no mar, êstes não podem ser contaminados. Agem como barreiras. Se possível, use uma côr diferente de *crayon* ou de caneta hidrográfica para a difusão e parta de cada uma das quatro fazendas, de cada vez.

DIFUSÃO PROBABILÍSTICA

A figura 4 poderá ser usada como base para êste exercício. Ao contrário da difusão prèviamente estudada, esta não é mecânica. A idéia

de que uma difusão se espalha, partindo de um ponto, em direção externa é a mesma. Há porém um nôvo elemento de sorte (acaso, incerteza, imprevisibilidade) acrescentado.

Suponha que alguma nova idéia (inovação) tal como uma nova espécie de semente ou um nôvo tipo de fertilizante ou implemento agrícola, seja introduzida em uma fazenda ou vila por acaso, ou deliberadamente (por intervenção, por exemplo, de um perito-agrônomo das Nações Unidas). A inovação é gradualmente ouvida e aceita ou rejeitada por mais e mais fazendas ou vilas. Não é possível dizer-se em dado momento quais dêles aceitarão a inovação. O exercício seguinte é um modêlo *estocástico* de tipo muito simples. Mostra a espécie de resultado que poderia ser obtido da tal difusão fortuita. Muitos resultados diferentes são possíveis.

Para êste exercício será admitido que cada um dos cento e cinqüenta hexágonos livres do mapa da figura 4 contenha uma vila. Introduz-se uma nova técnica na vila 10 10 e esta a adota, como início da difusão. Esta vila marcada com 0 (origem) na figura 4, tem condições de informar uma nova vila durante o período de tempo (ou turno) um. Que nova vila será informada é coisa que dependerá da sorte. Mas a probabilidade é bastante grande de que esta vila esteja na primeira camada e é certo que ela esteja ou na primeira ou na segunda camada. (Esta regra foi estabelecida por mim para *êste jôgo em particular* e poderá ser alterada). Como então determinar qual a vila informada?

Use a tabela de "números aleatórios" anexa. Inicie em qualquer lugar distante do início. Utilize pares sucessivos de dígitos (ex. 17,53) mas não use os números de 91 a 99 ou 00. Na figura 1 (últimos diagramas), a idéia da "rêde de informação" está ilustrada. Imagine o 0 (centro) da rêde da esquerda localizado sôbre o nosso vilarejo inicial inovador, na posição 10 10. Há uma probabilidade 1/10 para cada um dos seis vilarejos da camada interna de ser informado da novidade e uma probabilidade de 1/30 para cada um dos 12 vilarejos da segunda camada. Note-se que o total das probabilidades é de UM. Os números "aleatórios" de 01 a 90 são atribuídos em 3s ou 9s de acôrdo com a camada, interna ou externa, para reproduzir as proporções das probabilidades. Os números são mostrados na rêde da direita, na figura 1. Isto será usado para a simulação da difusão. O "0" será imaginàriamente colocado em cada vila à medida que esta tenha condições de informar.

Exemplo efetuado, com resultados tabulados. O mapa da figura 4 poderá ser reaproveitado por V. Observe que uma informação caindo sôbre uma montanha, pântano ou um hexágono de mar, é desperdiçada. Os números abaixo são os obtidos. Utilize os seus próprios números.

O primeiro número aleatório, sorteado, foi o número 29. Isto atinge a vila imediatamente a NE de O. Coloque um 1 nesta vila. Fim do período de tempo 1. Anote os resultados na tabela abaixo.

Tanto a vila em 10 10 (que já sabia) e a nova vila em 09 10, tem condições de informar (passar adiante). É de grande conveniência, mas não é obrigatório, prosseguir sucessivamente ao longo das linhas através do mapa, do Oeste para o Leste, tomando de cada vez uma vila que sabe. Da tabela de "números aleatórios" *tire um nôvo número para cada vila.* Anote-o. O número sorteado para a vila em 09 10 foi de 81. Isto atinge a vila da segunda camada de 09 10, ou seja, a vila em 10 12. Registre um 2 ali. Para a vila em 10 10 o número sorteado é 88. Isto atinge uma montanha em 12 12 e é perdido, pois não há vila ali. Anote os resultados na tabela e prossiga no período de tempo n.º 3. Os três números seguintes foram tirados: 22 para a vila situada em 09 10 (anote um 3 em 08 10), 39 para a vila situada em 10 10 (coloque um três na posição 10 09) e 56 para a vila situada em 10 12 (coloque um três na posição 10 14). Anote os resultados na tabela. Prossiga durante 8 a 10 períodos de tempo. Compare os seus resultados com os de outros estudantes. Se V. tiver tempo, reinicie a inovação partindo de outra vila de origem. Verifique a proporção de informações aproveitadas e perdidas. Que modificações pode V. sugerir para o processo? Que situação real mundial poderia ser testada desta maneira?

Exemplo de tabela

	Período de Tempo		Cumulativo	
	informado	*perdido*	*informado*	*perdido*
Início da operação	1	0	1	0
Fim do período de tempo 1	1	0	2	0
Fim do período de tempo 2	1	1	3	1
Fim do período de tempo 3	3	0	6	1

(segue)

DIFUSÃO BASEADA EM DECISÕES

A mesma base, reproduzida novamente na figura n.º 5, será usada para êste exercício. Neste caso, a colonização de uma área por dois grupos rivais é simulada. O jôgo exige dois "jogadores" que tomarão, cada um, as decisões para cada grupo de fazendeiros. Existem 5 fazendeiros em cada "time": o "time" A (vermelho e o "time" B (azul). Os colonos são designados por A1, A2, A3, A4, A5 e B1, B2, B3, B4, B5. O sucesso é medido pelo número de pontos obtidos por cada "time". Os pontos são obtidos da maneira seguinte:

(1) O hexágono de base sôbre o qual é colocado o ponto inicial de colonização não é contado.

(2) Quatro pontos são contados para cada hexágono vazio adquirido se êste se encontra na primeira camada em volta do estabelecimento original.

(3) Dois pontos são contados para cada hexágono vazio adquirido se êste se encontra na segunda camada em volta do estabelecimento original.

(4) Dois e um pontos são contados (ao invés de quatro e dois, respectivamente) se o hexágono adquirido está na montanha ou no pântano.

O sistema de recompensa leva em consideração, embora de maneira imperfeita, a) a influência da distância na utilidade da terra; b) a influência das condições físicas.

Os estágios do jôgo (marco 1) são os seguintes.

Jogam de cada vez A1, B1 A2, B2 e assim por diante.

Um hexágono é tomado para cada fazenda, cada ano. Sòmente um hexágono contíguo a outro já possuído, poderá ser adquirido. O jôgo para o primeiro ano foi marcado na figura n.º 5. Cada colono obteve um hexágono valendo 4 pontos.

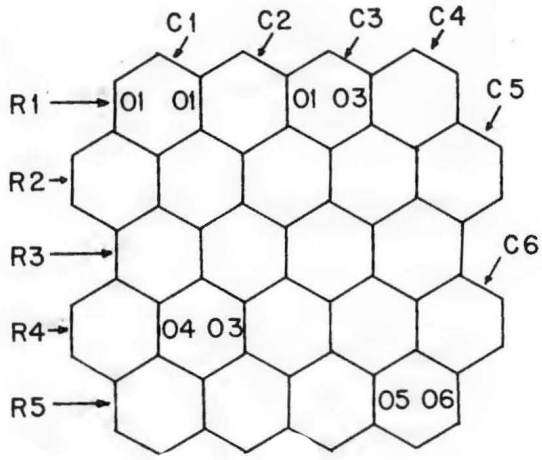
Prossiga o jôgo até quinze anos.

Conte os pontos obtidos por "time", separadamente, quando todos os hexágonos dentro de duas camadas de cada fazenda tiverem sido adquiridos. Algumas fazendas esgotarão os hexágonos antes do fim dos quinze anos (não há "score" para terra além de duas camadas). Contar um ponto para cada, anualmente. (Êstes poderão ser investidos em um barco de pesca ou um hotel de turistas).

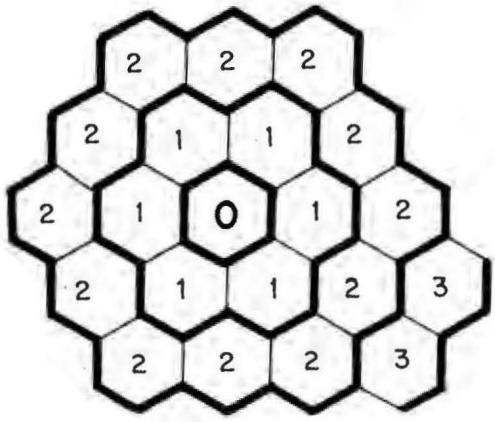
Êste jôgo possui vários dos aspectos dos chamados jogos operacionais, usados atualmente em geografia.

Se V. tiver tempo, comece tudo novamente, escolhendo de cada vez onde colocar os colonos A1, B1, A2, B2, etc.

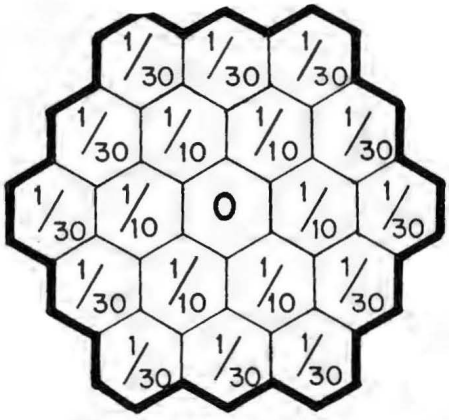
Um exercício adicional (não tão fácil quanto parece) é a otimização da distribuição dos dez colonos de tal maneira que o número de pontos obtidos em conjunto (jôgo da soma de não-zero) — (a perda de A não é necessàriamente um ganho para B e vice-versa) é maximizado.



Sistema sugerido para a referência aos quadros específicos no sistema hexagonal. R é equivalente à linha numa matriz e C à coluna. Os números iniciados por zeros, são usados com algarismos simples (isto é, abaixo de 10), quando dez ou mais colunas e/ou linhas são usadas.

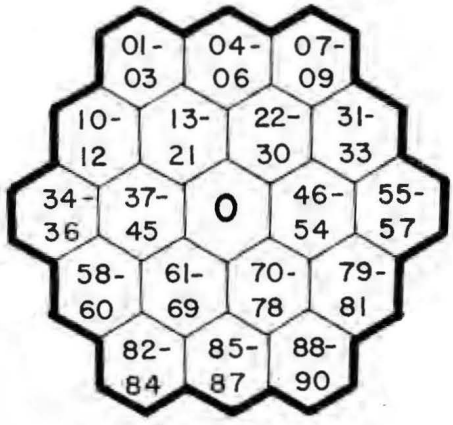


Qualquer hexágono dado tem seis outros adjacentes a êle. Para a finalidade dêste artigo êstes serão denominados de primeira camada. As segunda e terceira camadas são também mostradas. A numeração de uma camada é feita naturalmente, em relação a determinado hexágono original (0)



Probabilidades de informação

DivEd/D J.A.C



Probabilidades de informação expressas através de números aleatórios (pares de algarismos) 90 99 devem ser ignorados

Fig.2

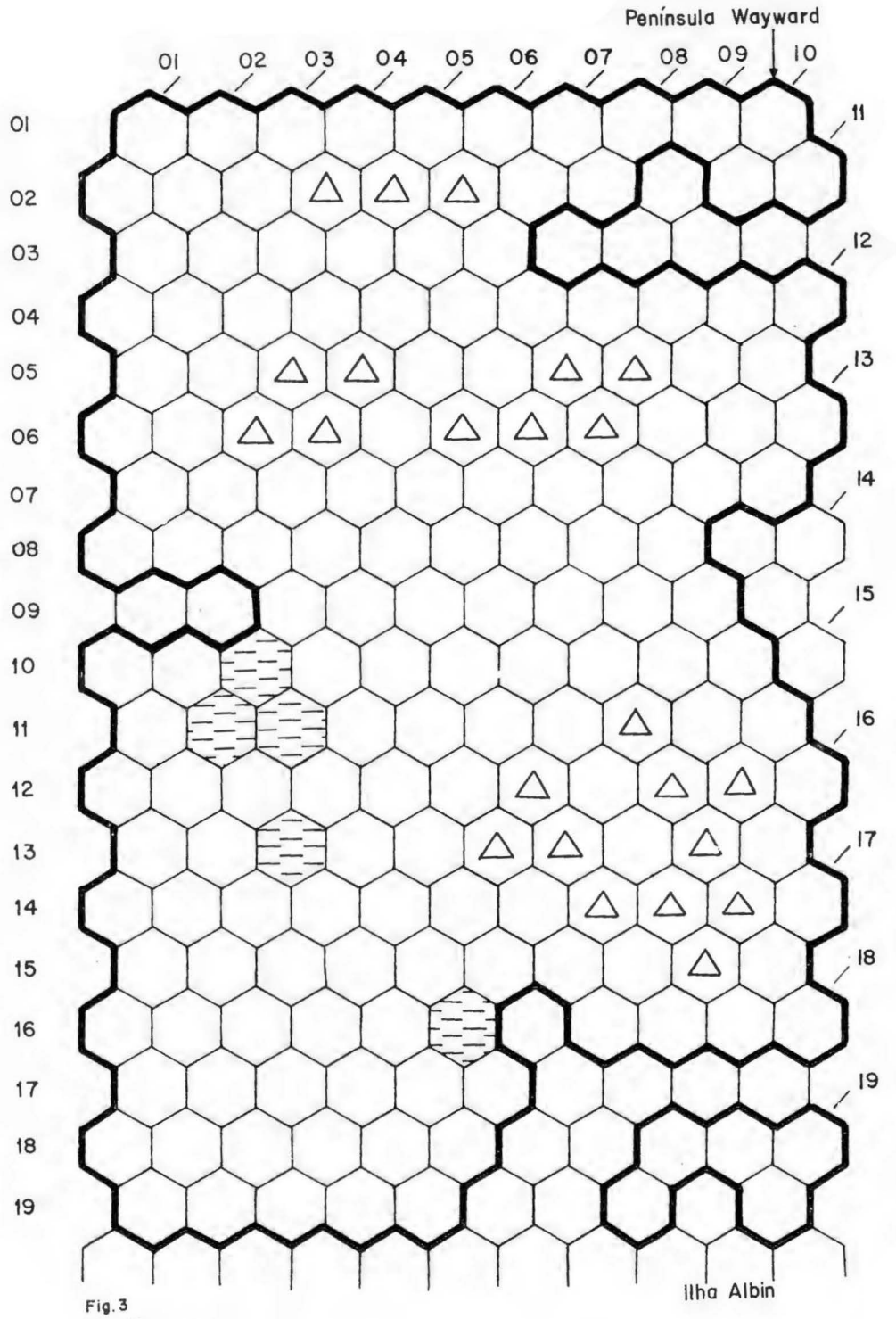


Fig. 3
DIvEd/D-J.A.C.

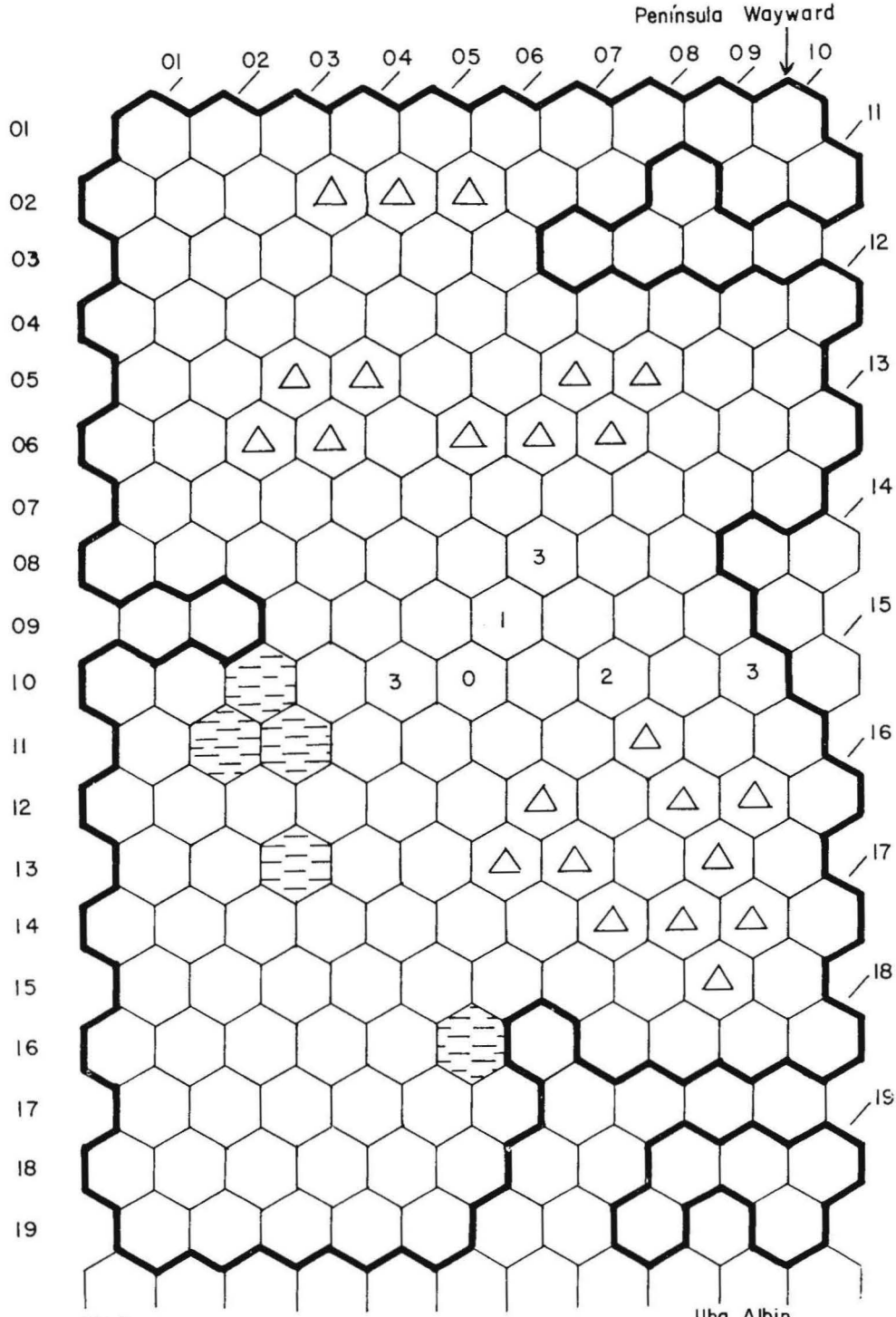


Fig 4
DivEd/D-J.A.C.

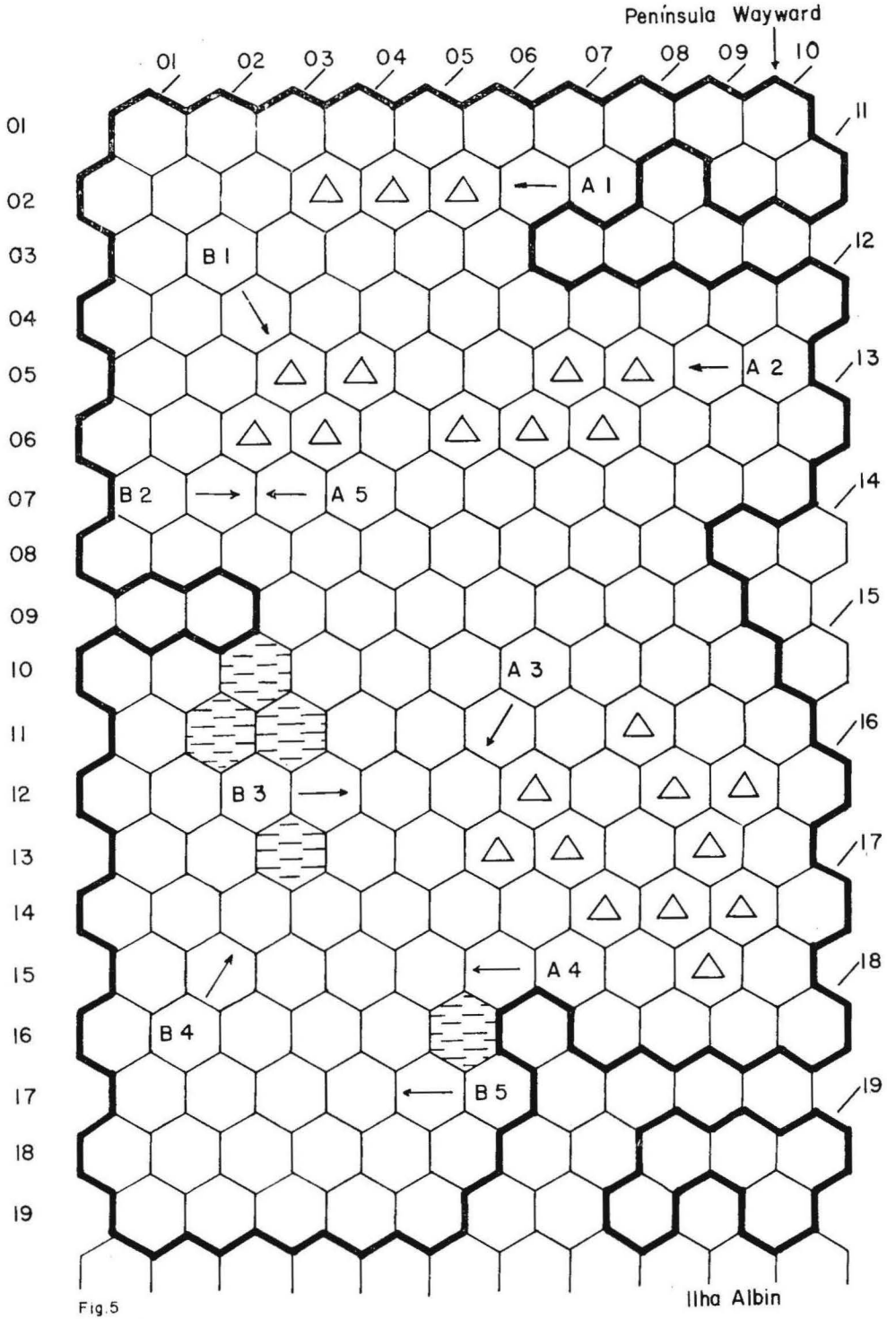


Fig.5
DivEd/D-J.A.C

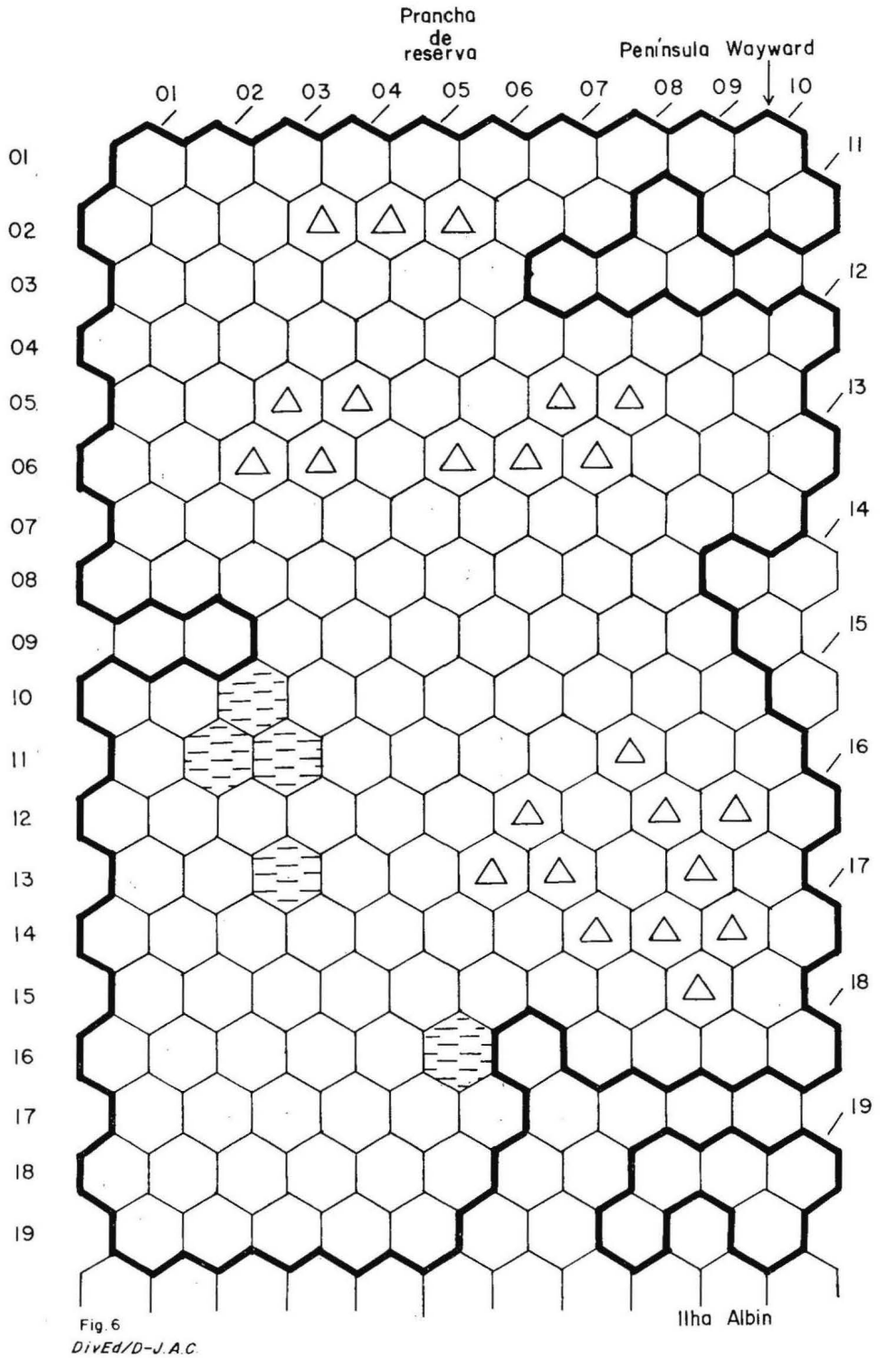


Fig. 6
DivEd/D-J.A.C.

Tabela de números aleatórios

5	7	0	7	9	0	3	1	3	4	6	3	9	8	0	0	0	3	6	7	7	0	5	8	3	6
8	7	9	0	3	9	3	2	4	5	4	4	1	4	0	0	2	8	9	4	9	0	7	9	8	0
3	8	9	6	5	6	3	8	0	9	7	5	8	9	3	9	9	3	7	5	1	9	7	7	0	8
0	0	0	3	8	7	8	7	9	9	8	8	8	5	8	2	8	0	9	0	3	8	7	9	0	4
1	9	8	0	9	6	9	6	0	2	7	4	9	3	4	9	0	9	7	4	1	2	6	0	7	0
4	3	8	6	9	3	6	7	8	3	7	9	5	4	6	4	1	4	3	5	3	7	3	0	6	9
4	2	3	8	6	8	5	7	0	1	9	5	3	7	0	5	2	4	1	1	1	5	0	5	3	1
1	5	3	3	3	0	8	2	2	7	1	5	5	1	7	8	7	0	7	5	1	0	5	7	3	8
9	3	6	2	5	1	2	1	6	9	0	7	6	5	8	6	5	5	1	7	2	7	1	4	2	6
0	9	9	8	1	9	9	3	1	6	7	8	7	9	3	4	1	9	0	2	0	6	8	9	4	4
9	0	5	2	6	3	1	4	9	6	3	1	3	2	4	5	6	3	4	2	2	5	8	6	1	2
4	8	5	7	8	0	1	8	7	6	5	0	4	1	5	6	0	5	3	5	5	2	6	3	8	2
6	1	2	3	6	2	0	6	1	5	6	2	5	6	4	9	6	0	0	9	6	8	6	2	7	6
0	8	5	6	8	1	9	2	3	8	6	1	5	1	1	1	2	7	9	5	1	9	6	8	7	4
6	1	1	7	3	1	2	3	5	2	2	8	3	9	2	3	3	8	0	0	5	7	7	6	2	4
7	8	2	7	0	1	0	9	0	5	7	7	0	4	2	4	1	4	5	9	3	3	2	6	8	6
4	6	8	1	3	7	0	2	1	4	6	2	6	7	4	8	0	8	7	0	5	2	0	4	0	5
7	8	2	3	3	0	3	5	5	7	9	3	8	1	6	4	9	0	5	5	2	6	9	9	5	9
9	2	4	3	8	5	4	6	9	6	9	2	6	2	2	8	6	9	6	3	7	8	7	8	9	5
9	0	2	9	6	7	5	0	1	2	7	3	1	6	5	2	0	2	1	2	5	1	7	5	5	8
6	8	3	1	7	5	3	1	9	2	4	1	8	3	8	7	6	4	0	8	8	7	9	4	9	1
5	7	5	3	7	4	9	4	9	3	4	7	7	5	3	2	5	2	9	7	0	8	9	0	9	5
0	8	4	8	0	4	3	7	6	2	3	7	6	1	6	2	2	1	1	2	9	4	6	4	7	0
7	2	5	0	2	7	0	3	7	6	7	1	0	6	8	2	3	3	2	3	5	7	0	3	0	6
6	5	4	7	7	9	6	7	0	7	3	0	3	9	2	6	9	3	7	1	6	4	9	7	7	7
5	9	4	6	0	2	1	2	3	1	2	0	6	0	9	1	5	3	2	2	1	5	4	9	8	5
0	4	2	9	1	6	2	2	1	7	3	2	9	4	3	6	7	5	2	6	6	8	9	0	9	3
1	5	8	3	3	6	3	2	3	6	7	9	7	5	5	0	7	5	5	0	3	5	0	4	7	7
2	7	1	0	7	3	3	6	2	2	2	0	0	7	3	8	1	7	9	8	9	6	7	3	1	6
5	6	0	8	3	9	4	6	1	2	4	0	4	8	4	9	4	4	8	8	9	5	5	7	4	9
3	0	9	6	6	7	3	8	5	3	0	6	7	4	0	5	0	5	5	2	4	4	7	3	8	7
4	8	3	5	6	0	6	1	7	5	4	9	0	3	9	4	5	2	5	6	1	7	2	6	5	9
7	2	7	3	6	3	3	0	2	5	9	1	7	2	6	2	2	3	3	1	4	4	2	0	5	2
7	0	0	2	2	5	6	2	1	3	1	1	9	2	4	7	0	8	3	3	7	8	7	2	1	1
8	0	7	1	2	4	3	4	4	1	4	4	2	0	0	2	6	0	2	8	5	9	5	0	9	4

1 6 6 3 1 0 6 1 2 3 7 1 7 9 4 6 2 2 4 5 7 8 1 1 4 2

9 1 0 3 9 1 7 5 1 1 4 4 2 4 7 4 2 1 1 8 7 1 2 3 7 6

2 2 8 6 6 4 9 7 7 8 8 8 1 4 1 2 4 6 4 6 7 3 4 9 8 9

7 9 9 5 2 1 3 7 9 5 0 2 0 9 6 1 0 3 4 1 2 8 3 0 6 3

6 7 2 7 8 0 4 5 9 9 9 7 0 4 2 3 6 4 6 2 1 6 6 2 6 8

0 0 4 8 0 9 1 0 9 6 2 5 1 4 0 5 0 6 6 3 3 4 2 8 3 1

4 7 2 5 9 5 1 6 2 1 0 9 6 4 4 7 9 4 5 8 3 1 3 6 4 8

8 6 0 3 3 7 5 2 0 2 7 1 8 9 5 4 5 0 5 5 3 3 2 1 0 9

7 1 4 2 1 1 1 6 3 1 9 3 2 7 1 3 5 7 2 7 2 0 6 7 7 7

4 6 5 0 0 9 0 6 2 3 4 4 2 3 5 0 1 9 9 8 3 7 7 0 6 1

5 7 6 6 1 9 1 0 7 2 2 5 2 1 1 8 6 5 7 2 2 1 0 6 0 1

1 5 5 0 3 1 6 6 5 3 6 1 1 5 4 5 4 6 9 7 3 9 9 0 2 3

9 5 1 2 0 2 5 2 9 9 9 7 8 0 7 8 4 0 6 3 4 7 7 2 5 1

2 9 7 4 3 0 8 2 5 9 0 7 3 3 9 0 8 1 9 3 3 3 3 8 8 4

0 6 8 1 2 9 6 5 0 2 5 4 1 5 2 2 1 5 5 8 7 0 8 8 9 3

9 6 6 0 8 2 8 8 7 0 5 4 4 6 0 6 0 7 7 0 9 6 7 0 4 5

1 7 5 4 8 7 9 9 8 2 6 8 8 0 1 7 3 3 2 9 6 8 3 3 3 4

2 2 6 7 7 9 1 4 8 5 3 0 5 6 3 4 3 9 6 5 2 7 3 6 1 7

6 6 6 5 3 7 6 8 4 5 2 1 1 3 9 9 5 2 1 5 5 7 5 6 6 1

2 0 3 1 2 5 4 2 4 6 3 6 2 1 3 4 9 2 0 2 9 1 2 2 7 2

4 8 3 0 9 3 2 8 9 2 8 7 5 5 1 7 8 2 4 6 3 8 9 5 7 5

6 4 8 1 5 9 1 8 5 4 0 7 6 3 1 6 6 3 4 1 3 6 3 5 8 2

0 6 3 0 4 3 4 6 3 0 8 0 0 7 8 3 3 1 7 3 5 4 2 2 5 6

4 2 6 4 8 7 4 1 1 1 9 6 8 6 1 6 1 3 2 1 9 5 3 7 8 4

0 1 0 8 9 7 4 4 4 9 1 0 0 1 5 8 0 5 1 4 3 1 3 0 4 6

9 2 4 6 4 4 5 4 4 3 9 6 2 9 3 6 1 1 9 5 6 5 7 9 6 4

6 6 5 0 8 1 3 5 2 6 9 5 1 1 7 1 7 9 4 8 4 5 4 7 8 4

8 3 7 6 9 1 5 9 4 3 8 6 9 7 7 4 4 1 9 4 5 7 1 4 4 2

3 5 2 0 7 6 2 6 5 7 8 0 6 9 2 5 0 2 2 1 6 7 0 7 2 6

9 2 6 4 4 0 9 7 3 0 0 6 8 7 6 2 5 1 0 1 6 2 6 6 1 9

5 5 6 0 7 8 0 8 6 6 1 8 6 9 7 3 6 2 9 0 8 0 6 8 6 1

5 4 5 7 2 5 2 4 1 9 2 7 6 1 8 1 6 6 5 3 2 7 1 3 8 2

6 5 9 5 0 6 3 8 6 8 9 0 0 0 5 1 0 8 1 2 4 6 3 6 7 0

6 5 0 5 3 6 5 1 2 9 9 8 2 0 1 8 3 3 5 5 7 4 2 8 2 3

3 9 9 8 6 0 6 3 6 1 3 7 2 2 5 8 8 5 7 2 5 4 1 7 3 9

1 9 5 9 6 1 9 1 8 9 4 3 9 5 1 9 4 9 8 5 8 0 4 9 5 9

8

A Curva de Lorenz

INTRODUÇÃO

A curva de Lorenz é uma representação diagramática, relativamente simples, de distribuições de certas espécies. Será ilustrada por meio de dois exemplos, muito simples, de natureza geográfica. Ser-lhe-á solicitado, então, que a aplique aos conjuntos de dados para o País de Gales e para a Itália. É de se notar que a quantidade de detalhes e a precisão da curva crescem na proporção da quantidade de informações disponível.

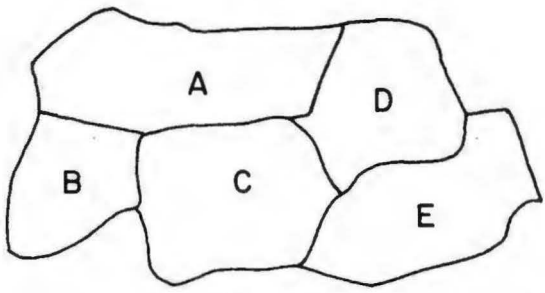


Fig.1 *DivEd/D-J.A.C.*

A ilha de Nébula é dividida em cinco condados. Acêrca dêles, possuímos as seguintes informações:

	Área do Condado (milhares de km²)	*População do Condado* (milhares)	*Pessoas na Indústria* (milhares)
A.....	23	280	7
B.....	12	690	55
C.....	18	110	3
D.....	16	180	18
E.....	19	440	23

DISTRIBUIÇÃO DA POPULAÇÃO PELA ÁREA

1.º passo: **Calcular a densidade da população (pessoas por km²) de cada unidade de área: A:12, B:57, C:6, D:11, E:23.**

2.º passo: **Classificar os condados de acôrdo com a densidade da população, como abaixo:**

	Área	*População*	*Densidade*
B.....	12	690	57
E.....	19	440	23
A.....	23	280	12
D.....	16	180	11
C.....	18	110	6
	88	1 700	

Nota: A densidade foi calculada apenas para permitir a classificação dos condados para a curva de Lorenz.

3.º passo: **Somar as colunas de área e de população.**

4.º passo: **Calcular a área de cada unidade como percentagem da área total. Registrar êsses valôres.**

5.º passo: **Calcular a população de cada unidade como percentagem da população total. Registrar êsses valôres.**

6.º passo: **Adicionar os novos valôres percentuais, cumulativamente, por área e, em seguida, por população. Êsses novos valôres determinarão a construção da curva de Lorenz. A tabela terminada deverá se apresentar como abaixo:**

	Área	*População*	% *Área*	% *População*	*CUMULATIVA* % *Área*	% *População*
B.....	12	690	14	41	14	41
E.....	19	440	22	26	36	67
A.....	23	280	26	16	62	83
D.....	16	180	18	11	80	94
C.....	18	110	20	6	100	100
	88	1 700	100	100		

Os valôres (pares ordenados) para cada condado na percentagem cumulativa de área e de população formam as posições coordenadas nos eixos dos y (vertical) e dos x (horizontal), respectivamente. Um pouco de reflexão mostrará que se todos os condados de Nébula tivessem a mesma densidade demográfica, a distribuição da população na curva de Lorenz se apresentaria sob a forma de uma reta, atravessando diagonalmente o gráfico de (0 0) a (100 100).

PESSOAS NA INDÚSTRIA DISTRIBUÍDAS PELA POPULAÇÃO TOTAL, POR CONDADO EM NÉBULA

1.º passo: Calcular a proporção das pessoas na indústria em relação à população total, expressando-a em percentagem: A:2,5%, B:8%, C:2,7%, D:10%, E:5 %, isto é (ind ÷ total) × 100.

2.º passo: Classificar os condados de acôrdo com as suas proporções na indústria.

	POPULAÇÃO		*Percentagem*
	Total	*Na Indústria*	*Na Indústria*
D.....	180	18	10
B.....	690	55	8
E.....	440	23	5
C.....	110	3	2,7
A.....	280	7	2,5
	1 700	106	

3.º passo: Somar as colunas da população total e da população na indústria.

4.º passo: Calcular a população total de cada unidade como percentagem da população total de Nébula. Registrar êsses valôres.

5.º passo: Calcular a população na indústria de cada condado como percentagem da população total de Nébula. Registrar êsses valôres.

6.º passo: Adicionar os novos valôres percentuais cumulativamente, primeiro os da percentagem sôbre a população total e, em seguida, os da percentagem de população na indústria. Êsses

novos valôres determinarão a construção da curva de Lorenz. A tabela completa deverá se apresentar como abaixo:

	POPULAÇÃO		*POPULAÇÃO*		*CUMULATIVA*	
	Total	*Indústria*	% *Total*	% *Indústria*	% *Total*	% *Indústria*
D.....	180	18	11	17	11	17
B.....	690	55	41	52	52	69
E.....	440	23	26	21	78	90
C.....	110	3	6	3	84	93
A.....	280	7	16	7	100	100
	1 700	106	100	100		

7.º passo: Construir a curva de Lorenz para a distribuição na indústria em Nébula.

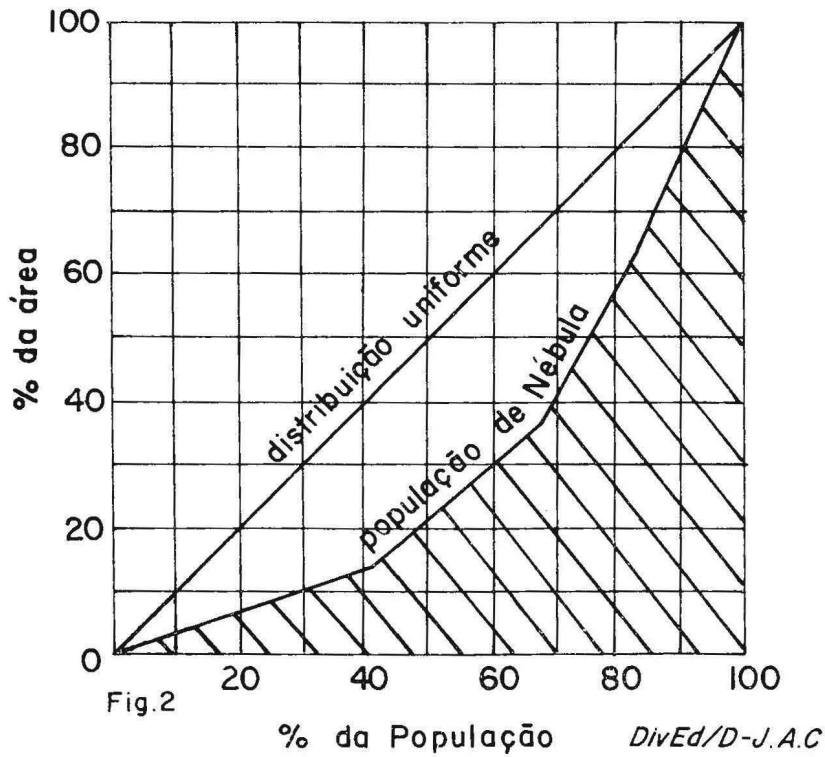


Fig.2

Condados do País de Gales (e Monmouthshire), 1961.

	Área (milhares de acres)	*População* (aproximada até centena)
Anglesey	177	51,700
Brecon	469	55,200
Caernarvon	364	121,800
Cardigan	443	53,600
Carmarthen	588	168,000
Denbigh	428	174,200
Flint	164	150,100
Glamorgan	523	1,229,700
Merioneth	422	38,300
Monmouth	347	444,700
Montgomery	510	44,200
Pembroke	393	94,100
Radnor	301	18,500

A densidade pode ser calculada em pessoas por 1.000 acres dividindo-se a população pela área. Esta será, então, a base para a reclassificação dos condados nos cálculos adicionais.

Siga a ordem de operações do exemplo de Nébula e construa a curva de Lorenz. Como os dados são padronizados, os condados galeses podem ser colocados no gráfico de Nébula para fins de comparação.

Manufatura na Itália em 1961

Região	*Emprêgo na Indústria*	*População Total* (em milhares)	*Total de Empregados por Mil*
Piemonte	673,100	3,914	172
Valle d'Aosta	2,600	101	25
Liguria	202,700	1,735	117
Lombardia	1,646,700	7,406	222
Trentino-Alto A.	43,400	786	55
Veneto	342,800	3,847	90
Friuli-Venezia G.	91,700	1,204	77
Emilia-Romagna	319,500	3,667	87
Marche	78,300	1,347	58
Toscana	320,800	3,286	98
Umbria	36,100	795	45
Lazio	225,500	3,959	57
Campania	179,500	4,761	38
Abruzzi	39,600	1,206	33
Molise	8,800	358	25
Puglia	104,800	3,421	31
Basilicata	13,700	644	21
Calabria	44,700	2,045	22
Sicilia	126,700	4,721	27
Sardegna	33,000	1,419	23
ITÁLIA	4,533,800	50,624	90

Nota: Um índice de indústria um pouco melhor seria o das pessoas na indústria como percentagem da população total *empregada* e não da população total.

9

Análise Fatorial (fator analysis) *e Análise de Agrupamento* (Cluster analysis) *no Brasil e no Peru*

INTRODUÇÃO

O método matemático denominado análise fatorial (*factor analysis*) já foi aplicado várias vêzes em estudos de psicologia e de educação. Nos últimos anos foi empregado também por geógrafos, sobretudo nos Estados Unidos, em trabalhos que exigem estudo de diversas variáveis ao mesmo tempo, em contraste com estudos nos quais entram sòmente uma ou duas variáveis. Não se pode descrever aqui a matemática complicada do método da análise fatorial, nem se pode considerar a preparação de um programa para computadores eletrônicos. É importante saber, entretanto, que a aplicação do método exige o uso de um computador, dada a imensidade dos cálculos quando se considera mais de quatro ou cinco variáveis.

Do ponto de vista geográfico, a análise fatorial pode ser considerada como um método que compara vários mapas de uma região em estudo. Cada mapa só contém uma distribuição, (por exemplo, da temperatura média anual, ou da percentagem de população que vive em centros urbanos). Pode-se considerar o grau de "covariação" (*covariance*) entre duas distribuições (ou variáveis). O nível de correlação pode ser descrito em têrmos numéricos, usando um índice de correlação *product-moment*. O índice r se encontra sempre entre 1 e — 1 ($-1 \leq r \leq 1$). Quando o grau de correlação é elevado, registra-se com um valor aproximado a 1 (ou a — 1 se a correlação é negativa). Uma correlação fraca se situa em tôrno de 0.

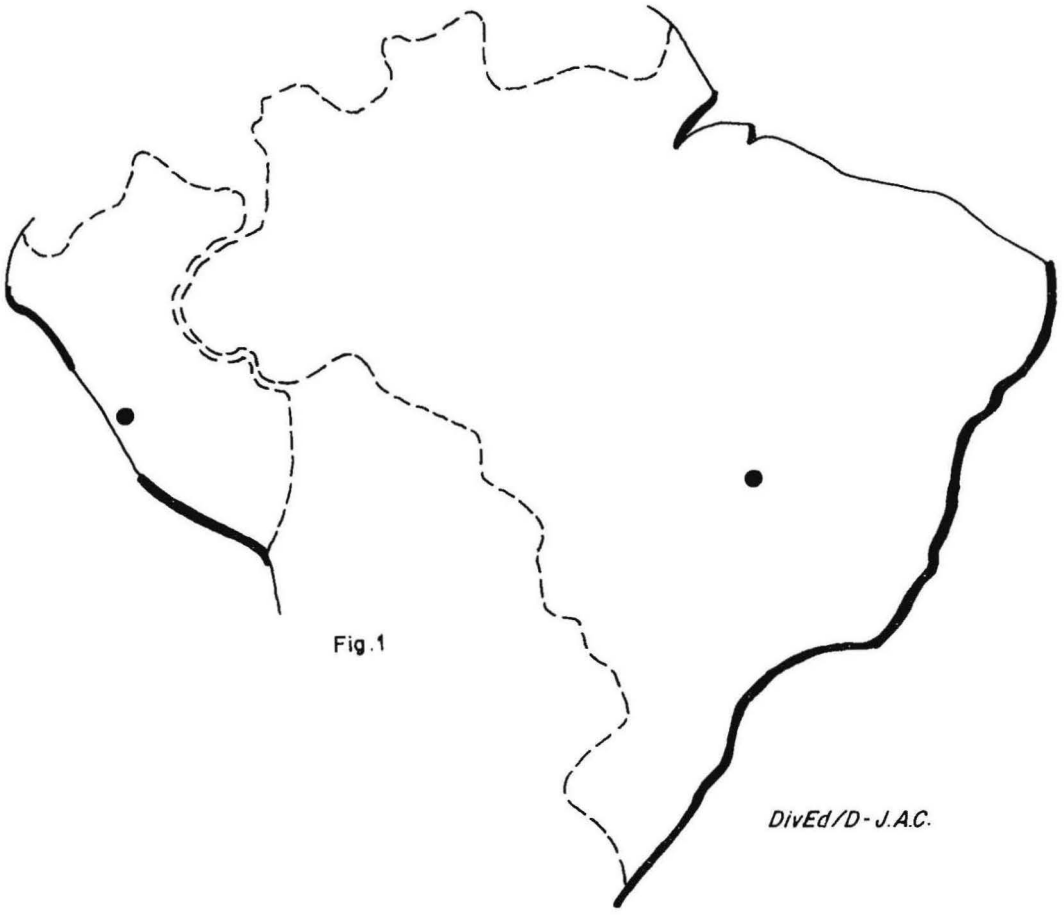


Fig. 1

Calcula-se inicialmente o índice de correlação entre cada par de variáveis. A segunda etapa é mais interessante porque procura relações entre várias variáveis, não sòmente duas (por exemplo, entre urbanização, nível de vida, nível de alfabetização, etc.). Os grupos ou "famílias" de variáveis que se destacam são chamados "fatôres" ou "componentes principais". Costuma-se considerar os fatôres em ordem de importância, segundo a proporção "explicada" por cada fator da variação total entre as variáveis. Freqüentemente, 15 a 25 variáveis originais se agrupam em dois ou três fatôres importantes, os quais, possìvelmente, contêm cêrca de 50% ou 60% da variação total.

A etapa seguinte do processo é um exame dos casos (por exemplo, indivíduos, objetos, ou províncias). Neste artigo estamos considerando

os estados do Brasil e os departamentos do Peru. Êstes compartimentos são os "casos" neste artigo. Esta etapa dá a "posição" relativa de cada caso (por exemplo, estado) em cada fator.

A última etapa considerada neste artigo é o problema da classificação dos casos com base, não em uma variável mas em um ou em vários fatôres. Isto é, pode-se derivar uma classificação e regionalização dos casos em base multivariável.

Não se podem considerar aqui nem as vantagens nem os perigos da análise fatorial. Nem tampouco há espaço para interpretar mais do que uma pequena parte dos resultados. Espera-se, entretanto, que os leitores que não conheçam a análise fatorial tenham uma idéia de suas possibilidades, e que os dados apresentados neste trabalho sejam úteis e interessantes.

Os pontos seguintes devem ser considerados:

1. Necessita-se de um número suficiente de áreas (províncias, estados). Pode-se trabalhar com 10 a 20 áreas. É melhor entretanto considerar de 50 a 100. Assim, o estudo do Peru baseado nos dados para as 144 províncias é superior ao estudo baseado nos dados para os 24 departamentos. Num estudo dêste tipo é desejável (mas nem sempre possível) que os compartimentos sejam mais ou menos do mesmo tamanho. Além disso, para eliminar as variações de área e de população entre os compartimentos do sistema, é essencial expressar os dados sob a forma de unidades por área especificada ou por pessoa (ou por cem pessoas).

2. Os resultados do trabalho dependem, naturalmente, das variáveis incluídas. Neste estudo incluiu-se, sobretudo, variáveis econômicas e sociais. Infelizmente, não é sempre possível obter todos os dados que se deseja para tal estudo, de modo que, às vêzes, tem-se que prescindir dêles ou buscar outros dados semelhantes.

PERU

A aplicação da análise fatorial no caso do Peru é explicada passo a passo:

Quadro 1B — Êste quadro contém uma descrição de cada variável usada no estudo dos departamentos. A maior parte das variáveis foi retirada dos vários tomos do censo do Peru de 1961. Para eliminar as diferenças

de número de habitantes entre os departamentos, converteu-se os dados originais em percentagens (por exemplo, pessoas que vivem em centros urbanos como percentagem da população total) ou em outros valôres apropriados. Esta operação é necessária mas, infelizmente, faz com que os departamentos venham a ter igual "pêso" no estudo. Isto tem que ser considerado na interpretação dos dados. É menos importante no estudo de 144 províncias do que no estudo dos 24 departamentos.

A seleção das 23 distribuições ou variáveis depende de várias circunstâncias. O leitor tem que observar que para qualquer estudo dêste tipo, há um número enorme de possíveis "misturas" de variáveis. No caso dêste estudo, levou-se em consideração o seguinte:

a) É desejável, mas não essencial, que o número de casos (departamentos) supere, pelo menos em mais um, o número de variáveis.

b) A "mistura" das variáveis depende em primeiro lugar da disponibilidade dos dados. Felizmente, graças à extensão do censo de 1961, existem muitas informações para o Peru, ou por departamento ou por província. Entretanto, não foram publicadas algumas informações interessantes como, por exemplo, dados sôbre os impostos, o poder de aquisição, os salários, os veículos motorizados. Além disso, os dados quanto aos aspectos físicos do país (por exemplo, precipitação de chuva, qualidade dos solos), são em geral escassos.

c) A "mistura" de variáveis pode representar todos os aspectos do país (altitude, clima, vegetação, população, atividades produtivas, aspectos sociais, opinião política), ou pode considerar sòmente um aspecto como por exemplo, o aspecto social ou econômico.

Êste estudo considera sobretudo os aspectos demográfico, econômico e social. Naturalmente, como a análise fatorial dá pêso igual a cada variável considerada, êste estudo dá mais ênfase às relações entre as variáveis demográficas, econômicas e sociais do que às físicas. Pode-se dizer que o elemento pessoal entra na seleção das variáveis.

Uma vez decidido quais as variáveis a serem empregadas para tal estudo, é importante dar-se conta da ordem na qual elas vão ser consideradas. Convém agrupar as variáveis segundo o aspecto do país ao qual elas se referem. (pg. 76).

Quadro 2B — Neste quadro se apresentam os dados utilizados no estudo. Apresentam-se êles em uma matriz 24×23. Cada *entry* (entrada) na matriz tem que ser ocupada por *um* número (valor). Em geral os dados são percentagens, mas em certos casos são valôres absolutos (por exemplo, altitude), ou por mil (por exemplo, 472, 468, pessoas com mais de 19 anos por mil pessoas da população total). Seria o mesmo escrever esta última variável como percentagem: 47.2%, 46.8%, etc.

Deve-se fixar o que significa cada variável. No presente estudo pôs-se a situação "boa" em posição alta para as variáveis econômicas e sociais. Assim a percentagem da população alfabetizada (variável 15) dá um valor alto (por exemplo, 74% em Arequipa) onde há mais alfabetização e um valor baixo (por exemplo, 23% em Apurimac) onde há mais analfabetos. No caso do consumo de coca, (variável 18) como é um aspecto social não desejável, calculou-se a percentagem da população que *não* masca coca como percentagem da população total.

O primeiro passo do processo *factor analysis* é calcular o índice de correlação "*Pearson product moment*" entre cada par de variáveis. O índice sempre se encontra entre + 1 e — 1 e é referido como r. Assim, pode-se dizer: uma correlação de 1 indica uma correlação *positiva* perfeita entre duas variáveis. Uma correlação de — 1 indica uma correlação *negativa* perfeita entre duas variáveis. Uma correlação em tôrno de 0 indica nenhuma correlação ou correlação fraca. Uma correlação positiva indica que onde há um valor alto em uma variável, o há também na outra, e onde há um valor baixo em uma variável, o há também na outra. Com uma comparação das variáveis 15 (população alfabetizada) e 16 (população eleitoral) pode-se notar que existe uma correlação positiva.

Não há espaço aqui para explicar o mecanismo do "*product moment correlation*". Basta dizer que ainda que os valôres não pareçam comparáveis na matriz 2.B, não há dificuldade matemática em normalizá-los. Assim os dados na matriz 2.B são os dados iniciais para a análise fatorial.

O primeiro processo é o da produção de uma matriz de correlações. Tal matriz está reproduzida na tabela 3.B. (pg. 78).

Quadro 3B — Nesta matriz se reproduz o índice de correlação entre cada par de variáveis. Para assegurar que a matriz completa seja reproduzida em uma só página, modificou-se a forma de expressão do índice de correlação da seguinte maneira: cada índice foi multiplicado por 100 (isto é, a vírgula decimal foi movida duas casas para a direita). Um número sublinhado indica uma correlação negativa. Assim por exemplo, o índice de correlação — 0.18 entre altitude e precipitação se reproduz como *18* na matriz. A correlação positiva perfeita + 1.00, indica-se como *100*. É necessário uma matriz 23×23 porque existem 23 variáveis. Naturalmente, cada variável tem uma correlação de exatamente + 1.00 com ela mesma, de modo que a diagonal principal da matriz, a qual vai do canto superior esquerdo até o canto inferior direito consiste de centenas (na realidade + 1.00).

Na interpretação da matriz, o mais importante é levar-se em conta que cada índice de correlação dá uma indicação da fôrça da relação

(covariação) entre um par de variáveis, para o Peru inteiro. A correlação baseia-se em dados para 24 departamentos, mas o índice de correlação os *considera em seu total.*

Examinando brevemente a matriz, nota-se que os índices de correlação sugerem a existência de vários pares de variáveis entre os quais as correlações são relativamente altas. É importante notar que a "fôrça" da correlação é uma questão relativa. Por conveniência, podemos dizer que uma correlação entre + 0.50 e 1.00 ou entre — 0.50 e — 1.00 é relativamente forte. (Isto é, entre 50 e 100 ou entre — 50 e — 100).

Notar-se-á na matriz várias correlações "fortes". Além disso, não é difícil encontrar não sòmente pares de variáveis "intercorrelacionadas" mas também grupos. Por exemplo, há uma correlação forte entre 15 e 16 de 96 (isto é + 0.96). Mas há também uma correlação forte entre 11 e 15 de 81 (isto é + 0.81) e entre 11 e 16 de 83 (+ 0.83). Assim, existem "famílias" de variáveis que "covariam" entre si. Entretanto, a distribuição e as relações entre variáveis são muito complexas, e não é fácil "tirar" as famílias de variáveis da matriz. Felizmente, a etapa seguinte da análise fatorial ajuda nisto. Matemàticamente, pode-se considerar a intercorrelação de 23 variáveis como se feita em 23 dimensões. Não houvesse nenhuma intercorrelação entre as 23 variáveis, e cada variável teria sua própria "direção" neste espaço de 23 dimensões. Se houvesse correlação total entre as 23, tôdas iriam na mesma direção ou no mesmo eixo. A realidade, em geral, fica entre êstes dois extremos. Para o exame da matriz que se segue, deve-se ter em conta que a variação total entre as 23 variáveis é expressa sob a forma de *23 unidades de variação.* (pg. 79)

Quadro 4B — A primeira parte do quadro mostra a proporção da variação total entre as 23 variáveis que se acumula em cada *fator.* Cada fator pode representar várias variáveis. Os fatôres diminuem em importância ou pêso de modo que o número 1 é o mais importante, 11 o seguinte, etc. O primeiro *fator,* 1, leva aproximadamente 45% da variação total entre tôdas as 23 variáveis, o segundo fator, 11, leva 18%. Pode-se dizer que os quatro primeiros fatôres levam 78% da variação total. Isto é, só 4 fatôres fazem 78% do "trabalho" de 23 variáveis; os outros fatôres (V-XXIII) diminuem em importância e não se consideram. Entretanto, nenhum dos fatôres reflete exatamente uma variável. Melhor dizendo, representam grupos de variáveis, os quais tem uma relação com a variável proporcional ao valor indicado no quadro *factor loadings.* O grau de relação de cada variável com o fator 1 depende de sua distância de 0. Assim, por exemplo, a variável 15 (alfabetizados) com — 0.93, bem como a variável 9 (agricultores) com + 0.95, têm uma relação forte com o fator 1. As duas variáveis têm uma correlação negativa forte entre elas

de — 0.82 (veja a matriz de correlações). A variável 16 (população eleitoral) com — 0.94 vai com a variável 15, com a correlação forte positiva de + 0.96 já mencionada. (pg. 79)

Quadro 5B — No fator 1 é fácil distinguir a presença de várias variáveis econômicas e sociais. Ao mesmo tempo nota-se a importância da altitude e a ausência da variável eleitoral (23). Infelizmente, os departamentos do Peru são muito grandes, não são compactos, e muitos dêles contêm partes de duas regiões físicas (por exemplo Ancash, costa e serra). Felizmente êste defeito é remediado em grande parte no estudo das 144 províncias.

No quadro 5B pode-se ter uma idéia das "famílias" de variáveis que existem no Peru, mas o quadro não dá nenhuma indicação dos departamentos. É a matriz seguinte que introduz de nôvo os departamentos. (pg. 79)

Quadro 6B — É uma matriz de departamentos por fatôres. Foram reproduzidos sòmente dados para os seis primeiros fatôres (I-VI). Já foi notado que o primeiro fator, número 1, é o fator que agrupa mais "as variáveis" do setor econômico e social. Vem a ser como um eixo de desenvolvimento. Cada departamento tem um valor que indica sua posição em uma escala que vai desde o departamento menos "desenvolvido" até o departamento mais "desenvolvido". Assim, em ordem, têm-se: Huancavelica 15.34, Apurimac 13.30, Ayacucho 13.08, até Tacna — 12.40, Lima — 19.31 e Callao — 24.61. A interpretação do segundo fator é mais difícil. Parece que há distinção entre a parte mais antiga do Peru onde se fala quichúa, num extremo e a parte mais moderna que inclui as partes mais longínquas da capital, no norte.

Infelizmente, não há mais espaço para discutir os defeitos e as vantagens da análise fatorial. Basta anotar que é um método matemático capaz de usar e processar muitas informações e que já está sendo usado em vários estudos. A realidade geográfica do Peru é, entretanto, muito mais complexa. Com a aplicação de dados para as províncias em vez dos departamentos, é possível aumentar a informação disponível; na opinião dos autores, a análise fatorial é capaz de processar os dados de maneira imparcial e de produzir resultados relativamente precisos e de forma compacta. A interpretação dos resultados depende, naturalmente, da pessoa que faz o estudo. (pg. 80)

Quadro 1B — Peru — Descrição de variáveis

FONTE	LOCALIZAÇÃO NA FONTE	ITEM	DATA	DESCRIÇÃO COMPLETA
1 CP	vários mapas	Altitude		Altitude média em centenas de metros acima do nível do mar de tôdas as capitais de províncias.
2 RO	P. 185	Precipitação		Média anual em centímetros
3 LO		Distância por terra de Lima		Tempo em horas de Lima por rodovia (ou por via fluvial)
4 TI	N.º 3	Densidade demográfica	1961	Pessoas por km 2
5 TI	N.º 14	Urbana	1961	População urbana como percentagem da população total.
6 TI	N.º 10	Sexo		Homens por 100 mulheres
7 TI	N.º 16	População acima de 19 anos	1961	Pessoas acima de 19 anos, por mil da população total
8 TI	N.º 24	Migração		Pessoas nascidas em outros departamentos como percentagem da população total.
9 TIV	N.º 92	Trabalhadores agrícolas	1961	Pessoas empregadas na agricultura, como percentagem da população empregada total.
10 TIV	N.º 92	Minas e indústrias manufatureiras	1961	Pessoas empregadas em minas e indústrias manufatureiras como percentagem da população total empregada.
11 TIV	N.º 92	Comércio e serviços	1961	Pessoas empregadas no comércio e prestação de serviços como percentagem da população total empregada
12 TIV	N.º 92	População empregada	1961	População empregada (econômicamente ativa) como percentagem da população total
13 TIV	N.º 95 (N.º 92)	Empregos domésticos	1961	Pessoas na categoria de ocupação de empregados domésticos por mil da população total empregada
14 TV	N.º 141	Uso de calçado	1961	Pessoas que usam normalmente algum tipo de calçado, como percentagem da população total
15 TIII	N.º 60	Alfabetização	1961	População alfabetizada como percentagem da população total acima de 16 anos de idade
16 ELTI	(N.º 16)	Índice de eleitores	1961	Eleitores, como percentagem da população acima de 19 anos de idade
17 TII	N.º 33	"Status" marital	1961	Número de pessoas (homens mais mulheres) casados por dez pessoas vivendo maritalmente
18 TV	N.º 147	Mascar coca	1961	Pessoas que *Não* mascam coca como percentagem da população total acima de 19 anos de idade
19 TIII	N.º 49	Falando espanhol	1961	Pessoas que falam espanhol como percentagem da população total acima de 4 anos
20 TIV	N.º 116	Empregados públicos	1961	Pessoas empregadas pelo Estado, (funcionários públicos) como percentagem da população total empregada.
21 TIV	N.º 95	"Trabalhadores"	1961	"Trabalhadores" masculinos como percentagem de todos os empregados masculinos.
22 TIV	N.º 95	Pessoal de escritório	1961	Pessoas que trabalham em escritórios, homens e mulheres, como percentagem da população total empregada.
23 EL		Opinião eleitoral	1966	Eleições locais de 1966, número de pessoas que votaram no APRA e no UNO por 100 pessoas que votaram no AP e no DC.

FONTES PERUANAS.
CP Centros Poblados (4 tomos), Dirección Nacional de Estadística y Censos.
RO Romero, E., Geografia Economica del Peru, Lima 1963 p 185.
LO Peru, Mapa de Carreteras, Compania Petrolera Lobitos, 1963.
TI–TV Tomos del Censo Nacional de Población, Dirección Nacional de Estadística y Censos.
I Quadros comparativos, distribuição geográfica, idade e sexo, lugar de nascimento.
II Migração. Estado conjugal.
III Idioma, alfabetização, assistência escolar, nível de educação.
IV Características econômicas.
V Usos e costumes locais.

Quadro 2B — *Dados usados*

	PERU	1 Altitude	2 Precipitação	3 Dist. Lima	4 Dens. Pop.	5 Pop. Urb.	6 Homens	7 Acim. 19	8 Migração	9 Trab. Agric.	10 Minas Ind.	11 Serviços. Com.
D1	Amazonas	20	100	60	3	39	104	472	18	74	5	14
D2	Ancash	25	30	14	16	33	93	468	9	63	14	14
D3	Apurimac	30	60	35	14	20	91	480	3	77	8	11
D4	Arequipa	20	20	20	6	65	105	482	22	35	18	31
D5	Ayacucho	30	70	29	9	25	89	472	4	77	7	12
D6	Cajamarca	23	70	23	21	15	95	486	4	78	11	8
D7	Callao	0	0	0	2 893	96	110	506	51	8	25	42
D8	Cuzco	32	80	45	8	32	100	500	7	66	13	20
D9	Huancavelica	34	50	23	14	19	91	430	4	78	9	9
D10	Huanuco	23	80	21	9	21	98	446	7	76	6	10
D11	Ica	3	5	6	12	54	109	446	22	50	16	23
D12	Junin	31	80	15	12	49	98	468	14	53	19	19
D13	La Libertad	23	70	11	25	42	98	454	12	57	15	19
D14	Lambayeque	0	10	15	21	62	104	466	17	48	17	22
D15	Lima	19	20	0	60	86	101	500	40	15	22	45
D16	Loreto	1	300	183	1	39	106	400	8	58	8	26
D17	Madre de Dios	5	320	94	0	25	157	462	40	53	20	25
D18	Moquegua	18	20	24	3	48	108	486	25	58	13	20
D19	Pasco	32	80	16	6	35	106	454	18	48	30	14
D20	Piura	7	15	21	20	45	101	432	4	56	15	20
D21	Puno	36	60	35	10	18	93	482	2	72	11	13
D22	San Martin	6	150	88	3	59	99	394	6	76	7	13
D23	Tacna	18	15	27	5	70	119	492	39	39	19	31
D24	Tumbes	0	40	26	12	61	116	436	22	43	10	40

	PERU	12 Pop. Emp.	13 Emp. Dom.	14 Sapatos	15 Alfabet.	16 Eleit.	17 Casam.	18 Coca (não)	19 Espanhol	20 Func. Pub.	21 Trabalh.	22 Pes. Esc.	23 Op. El.
D1	Amazonas	31	30	52	58	44	30	79	88	24	31	6	114
D2	Ancash	29	41	86	48	37	29	80	33	10	31	6	116
D3	Apurimac	28	38	71	23	15	34	68	4	13	16	4	87
D4	Arequipa	36	106	95	74	64	70	93	66	11	46	12	113
D5	Ayacucho	28	36	89	27	19	107	54	5	14	15	4	81
D6	Cajamarca	28	24	61	45	29	15	89	97	11	18	3	130
D7	Callao	35	51	100	94	89	46	100	90	24	58	23	99
D8	Cuzco	34	60	74	33	21	26	56	10	11	27	6	75
D9	Huancavelica	28	28	79	29	20	87	27	6	10	18	3	99
D10	Huanuco	29	36	65	43	28	20	56	29	12	23	5	114
D11	Ica	34	48	98	84	74	37	99	83	9	64	11	79
D12	Junin	31	42	92	61	50	63	68	39	9	37	8	120
D13	La Libertad	29	45	95	63	50	16	84	99	9	39	9	236
D14	Lambayeque	29	38	91	70	52	15	99	96	9	58	11	0
D15	Lima	37	97	99	90	90	46	98	81	14	46	23	85
D16	Loreto	27	46	57	67	46	11	100	81	28	30	9	77
D17	Madre de Dios	41	33	83	72	45	14	95	68	13	61	9	63
D18	Moquegua	36	60	90	64	57	42	89	59	17	35	8	107
D19	Pasco	30	32	91	52	36	29	65	35	8	47	9	177
D20	Piura	29	33	78	55	41	40	100	99	12	45	7	111
D21	Puno	33	32	64	34	20	44	54	51	12	21	4	44
D22	San Martin	26	33	47	68	53	59	97	88	32	13	7	92
D23	Tacna	41	89	96	74	65	45	90	60	17	47	13	163
D24	Tumbes	31	23	94	77	59	36	100	99	45	53	10	98

Quadro 3B — *Matriz de índices de correlação*

	1	2	3	4
1	1.00			
2	—0.18	1.00		
3	—0.29	0.88	1.00	
4	—0.31	—0.20	—0.20	1.00

Os índices de correlação foram multiplicados por 100 depois de arredondados até a 2.ª casa decimal. A barra sob o número significa que êle é negativo.

Matriz de Correlação para o Peru

Altitude	1	100																							
Precipitação	2	*18*	100																						
Dist. Lima	3	*29*	88	100																					
Dens Pop.	4	*31*	*20*	*20*	100																				
Pop, Urb.	5	*56*	*36*	*23*	51	100																			
Homens	6	*53*	48	28	09	22	100																		
Acima 19	7	34	*45*	*54*	32	19	01	100																	
Migração	8	*45*	*06*	*17*	53	73	67	42	100																
Trab. Agric.	9	45	26	28	*55*	*87*	*39*	*36*	*84*	100															
Minas Ind.	10	*10*	*19*	*38*	37	51	38	38	65	*77*	100														
Serviços Com.	11	*66*	*13*	*08*	44	84	45	25	80	*91*	50	100													
Pop. Emp.	12	*13*	*04*	*15*	17	41	08	57	79	*59*	52	59	100												
Emp, Dom.	13	01	*27*	*17*	05	58	10	46	50	*60*	39	58	62	100											
Sapatos	14	*10*	*49*	*61*	26	53	23	42	59	*71*	69	58	49	46	100										
Alfabetos	15	*75*	*04*	*02*	38	86	54	01	79	*82*	51	81	49	43	45	100									
Eleit.	16	*62*	*22*	*18*	44	92	38	15	81	*86*	53	83	51	56	53	96	100								
Casam.	17	33	*30*	*25*	05	09	*34*	03	*08*	04	*10*	*04*	*04*	17	20	*20*	*05*	100							
Coca (não)	18	*84*	06	16	20	65	47	*08*	51	*57*	28	62	27	27	19	82	74	*40*	100						
Espanhol	19	*74*	02	09	19	52	33	*18*	35	*42*	16	46	09	03	03	74	63	*42*	83	100					
Func. Pub.	20	*54*	22	41	20	34	21	*37*	18	*12*	*32*	40	*08*	*16*	*24*	35	28	*01*	39	39	100				
Trabalho	21	*61*	*12*	*22*	29	61	67	15	74	*80*	70	70	58	28	68	78	71	*29*	61	51	01	100			
Pes. Esc.	22	*48*	*22*	*22*	59	90	33	33	85	*95*	66	89	53	60	58	84	90	*04*	58	44	18	67	100		
Op. Eleitoral	23	28	*13*	*23*	*02*	00	*09*	*00*	04	*05*	21	*08*	*03*	10	18	03	07	*06*	*04*	06	*10*	*04*	00	100	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	

Variáveis

Matriz de Correlação

Mostra correlações de pares entre cada uma das 23 variáveis.

Matriz de correlação é matriz de variável por variável.

Quadro 4B

Variação nos fatôres	Fatôres (sòmente seis dos 23)					
de 23	I	II	III	IV	V	VI
como Percentagem	10.3035	4.1434	2.0075	1.4858	1.1030	0.9713
Percentagem	45%	18%	9%	6%		
acumulada	45%	63%	72%	78%		

Variáveis Quadro 5B "PONDERAÇÕES"

		I	II	III	IV	V	VI
1.	Altitude.....	0.61	— 0.68	0.17	0.07	0.18	0.14
2.	Precipitação..	0.20	0.69	0.54	0.20	— 0.08	0.26
3.	Dist. de Lima	0.21	0.79	0.27	0.34	— 0.17	0.11
4.	Dens. Pop....	— 0.50	— 0.10	— 0.27	0.16	0.54	0.46
5.	Pop. Urb.....	— 0.89	— 0.03	— 0.36	0.14	— 0.09	— 0.01
6.	Homens.....	— 0.54	0.39	0.65	0.03	0.08	0.07
7.	Acima 19....	— 0.30	— 0.68	0.23	0.11	0.21	— 0.15
8.	Migração.....	— 0.91	— 0.07	0.23	0.15	0.08	0.16
9.	Trab. Agric..	0.95	0.16	0.04	— 0.04	— 0.01	— 0.08
10.	Minas ind....	— 0.68	— 0,37	0.29	— 0.26	0.07	0.22
11.	Serviços.....	— 0.91	0.05	— 0.07	0.24	— 0.05	— 0.01
12.	Pop. Emp....	— 0.66	— 0.24	0.56	0.24	— 0.07	— 0.12
13.	Emp. Dom...	— 0.56	— 0.39	0.09	0.13	— 0.52	— 0.23
14.	Sapatos.....	— 0.65	— 0.55	0.05	— 0.14	0.01	0.01
15.	Alfabet.....	— 0.93	0.27	— 0.09	— 0.07	— 0.11	— 0.01
16.	Eleit.....	— 0.94	0.07	— 0.20	0.01	— 0.15	— 0.02
17.	Casam.....	0.13	— 0.44	— 0.35	0.55	— 0.12	0.10
18.	Coca (não)...	— 0.72	0.50	— 0.12	— 0.21	— 0.10	— 0.22
19.	Espanhol.....	— 0.57	0.53	— 0.25	— 0.39	— 0.06	— 0.13
20.	Func. Pub....	— 0.22	0.63	— 0.38	0.35	— 0.02	0.20
21.	Trabalh.....	— 0.84	0.05	0.27	— 0.28	0.16	— 0.10
22.	Pes. Esc.....	— 0.93	— 0.10	— 0.12	0.13	0.08	0.09
23.	Op. El.....	— 0.02	— 0.23	— 0.04	— 0.47	— 0.55	0.59

Quadro 6B

Peru: pesos individuais para os primeiros seis fatôres

Departamentos	Fatôres					
	I	II	III	IV	V	VI
Amazonas	4.68	3.62	— 0.36	0.18	— 0.08	0.02
Ancash	5.19	— 2.31	— 0.28	— 1.09	0.13	— 0.14
Apurimac	13.30	— 2.02	0.11	0.59	0.62	— 0.24
Arequipa	— 9.63	— 3.81	— 0.05	1.19	— 1.90	— 0.90
Ayacucho	13.08	— 4.02	— 1.24	2.42	0.29	0.21
Cajamarca	9.50	1.02	— 0.29	— 2.27	0.35	— 0.28
Callao	—24.61	— 1.92	— 2.51	1.13	2.87	2.19
Cuzco	7.25	— 3.19	2.12	1.35	0.34	— 0.51
Huancavelica	15.34	— 3.72	— 0.79	1.32	0.27	0.87
Huanuco	10.90	0.02	0.16	— 0.44	0.21	0.13
Ica	— 8.92	0.31	— 0.31	— 1.52	0.28	— 1.33
Junin	0.98	— 3.54	0.08	— 0.07	— 0.33	0.58
La Libertad	— 0.50	— 1.00	— 0.53	— 3.48	— 1.70	1.29
Lambayeque	— 6.13	1.49	— 0.76	— 1.50	1.93	— 2.35
Lima	—19.31	— 3.94	— 0.64	1.37	— 1.29	— 0.92
Loreto	2.39	11.60	0.43	1.46	— 1.07	0.53
Madre de Dios	— 7.61	6.09	7.44	0.47	1.03	0.34
Moquegua	— 3.19	— 1.51	0.46	0.35	— 0.44	— 0.63
Pasco	0.95	— 3.49	1.90	— 2.40	— 0.16	1.92
Piura	0.36	1.82	— 1.61	— 2.06	0.12	— 0.62
Puno	11.32	— 1.93	0.69	0.93	1.06	— 0.77
San Martin	5.55	8.16	— 3.31	1.22	— 1.05	0.42
Tacna	—12.40	— 3.12	1.70	0.73	— 1.80	0.12
Tumbes	— 8.69	5.39	— 2.40	0.12	0.32	0.08

BRASIL

A análise fatorial foi aplicada neste artigo, a variáveis para 21 Estados do Brasil, com o intuito de ilustrar simples e òbviamente o que a análise fatorial pode fazer.

***Quadro 1A* — Doze variáveis foram usadas para vinte e um Estados do Brasil. Os quatro territórios e o Estado do Acre foram excluídos por razões técnicas. As variáveis estão definidas no quadro 1A e dadas as suas fontes. A figura dois mostra os 21 estados.** (pg. 82)

Quadro 2A — Para eliminar a influência (direta) da dimensão (em população ou em área) dos estados, os dados foram convertidos apropriadamente em têrmos de *por área* (isto é densidade demográfica) ou *por população* (isto é população urbana como percentagem da população total). Os dados são mostrados no quadro 2A. (pg. 83)

Quadro 3A — Êste quadro mostra os valôres r da correlação momento-produto de Pearson (*Pearson product moment*) para cada par de variáveis. Os valôres de 1.00 na diagonal principal representam a correlação de cada variável consigo mesma. A matriz é quadrada e simétrica, daí não haver necessidade de reproduzir a metade superior direita. (pg. 15)
Quadro 4A — Os valôres Eigen dos primeiros dois fatôres são mostrados após o Quadro 3A. Neste exemplo, os primeiros dois fatôres (I e II) "explicam" o valor inusitadamente elevado de 80% de tôda a variação. (pg. 83)

Quadro 5A — Mostra as ponderações (*factor loadings*) de cada uma das variáveis individuais nos primeiros dois fatôres. São mostradas as seguintes:

1 — Valôres não rotacionados

a) como índice de correlação

b) como percentagem da variável *computada* por fator. Êste valor é o assinalado em *a* elevado ao quadrado e multiplicado por 100. Note que quando um valor mais perto de 0 do que 1 ou — 1 é elevado ao quadrado, torna-se menor; isto é: $0{,}7 \times 0{,}7 = 0{,}49$.

2 — Valôres rotacionados

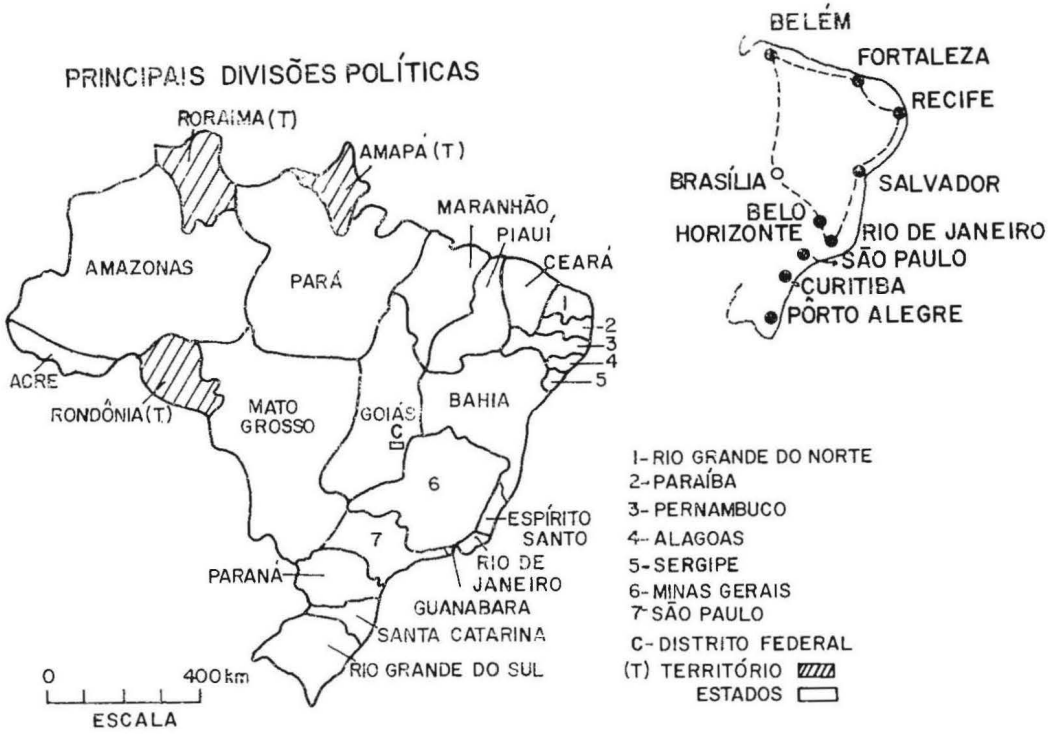
a) como índice de correlação

b) como percentagem da variável computada por fator.

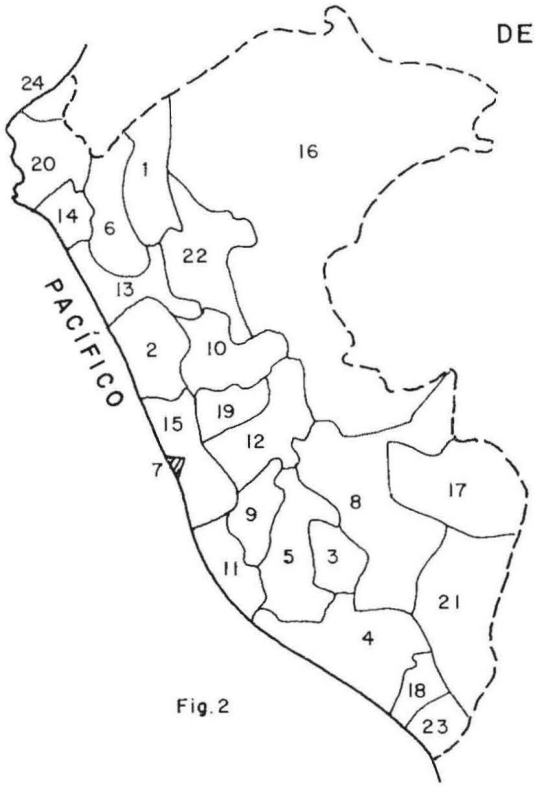
Note-se que a *rotação* neste exemplo reduz a concentração de variáveis de desenvolvimento suportadas pelo primeiro fator n/rotacionado (*unrotated*). (pg. 84)

Quadro 6A — Êste quadro mostra a classificação e os "*scores*" dos 21 estados no fator n/rotacionado, o "concenso do desenvolvimento" para o Brasil. Note-se que os valôres são relativos e não representam (embora reflitam) qualquer valor absoluto particular (tal como a renda *per capita*). O valor médio é o zero. A distribuição é estruturada em direção aos valôres mais baixos. O consenso é um, com a contribuição de muitas variáveis e é portanto uma visão "multivariável" do desenvolvimento no Brasil.

ESTADOS DO BRASIL



DEPARTAMENTOS DO PERU



1- AMAZONAS
2- ANCASH
3- APURIMAC
4- AREQUIPA
5- AYACUCHO
6- CAJAMARCA
7- CALLAO
8- CUZCO
9- HUANCAVELICA
10- HUANUCO
11- ICA
12- JUNÍN
13- LA LIBERTAD
14- LAMBAYEQUE
15- LIMA
16- LORETO
17- MADRE DE DIOS
18- MOQUEGUA
19- PASCO
20- PIURA
21- PUNO
22- SAN MARTIN
23- TACNA
24- TUMBES

Fig. 2

DivEd/D-J.A.C.

A figura 4 mostra o agrupamento dos 21 Estados brasileiros, com a utilização de todos os doze fatôres. Representa, portanto, uma classificação em bases multivariáveis. Muito pouca informação é perdida no emparelhamento dos primeiros indivíduos ou grupos. Por exemplo, a grande semelhança entre os Estados de Alagoas (9) e Sergipe (10) revelados pela árvore de ligação (*linkage tree*) é confirmada pela semelhança de seus valôres respectivos nas doze variáveis do quadro 2A. (pg. 83)

Quadro 1A — Definição das variáveis do Brasil:

1) Densidade da população, pessoas por quilômetro quadrado em 1970.
2) Média do crescimento anual natural da população em pessoas por milhar, durante o período de 1950 a 1960.
3) Pessoas nascidas em outros estados como percentagem da população total do estado em 1950.
4) Brancos, como percentagem da população total.
5) População urbana, como percentagem da população total.
6) Pessoas empregadas na indústria por milhar de população total.
7) Eletricidade consumida por habitante em milhares de kwh.
8) Telefones por milhares de habitantes.
9) Automóveis por milhares de habitantes.
10) Orçamentos estaduais em cruzeiros por habitantes por ano.
11) População alfabetizada como percentagem da população total.
12) Leitos de hospitais disponíveis para cada 10.000 habitantes.

FONTES PRINCIPAIS: *Anudrio Estatístico do Brasil* (principalmente na Edição de 1968) — IBGE.

Subsídios à regionalização, IBGE.

Contribuições para o estudo da demografia do Brasil IBGE, 1961.

Quadro 2A — Matriz de dados do Brasil:

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.	12.
1.	Amazonas	0.6	34	10	37	32	7	7	6	7	75	43	29
2.	Pará	1.6	29	6	29	41	8	6	4	7	53	49	23
3.	Maranhão	10.8	30	10	34	18	7	1	1	3	11	25	4
4.	Piauí	5.7	35	8	28	24	5	1	2	3	15	26	14
5.	Ceará	25.9	37	4	44	34	8	4	7	7	30	31	17
6.	Rio Grande do Norte	24.4	33	8	29	37	20	5	4	7	31	32	19
7.	Paraíba	40.0	33	6	67	35	21	9	4	3	30	29	18
8.	Pernambuco	48.1	30	6	50	45	21	17	5	9	63	32	25
9.	Alagoas	50.6	34	6	41	34	30	7	3	5	30	24	20
10.	Sergipe	38.7	34	6	50	39	20	9	2	6	30	34	20
11.	Bahia	12.4	28	3	30	35	9	9	4	7	40	32	13
12.	Minas Gerais	20.2	30	3	58	40	15	29	12	12	63	44	35
13.	Espírito Santo	42.1	34	11	59	32	11	9	9	10	46	47	22
14.	Rio de Janeiro	106.9	24	16	60	61	33	43	15	16	89	56	39
15.	Guanabara	3 592.9	17	43	70	98	53	75	99	43	395	85	89
16.	São Paulo	67.3	26	13	86	63	70	79	36	42	322	65	51
17.	Paraná	36.3	29	32	86	31	14	16	6	17	70	53	22
19.	Rio Grande do Sul	24.5	25	1	89	45	31	19	12	20	127	66	45
20.	Mato Grosso	1.1	32	18	53	40	9	6	6	13	37	51	24
21.	Goiás/DF	5.1	28	23	58	40	5	11	12	18	535	33	24

Quadro 3A — Matriz de Correlação do Brasil:

1.	Densidade da população	1.00											
2.	Acréscimo natural	—0.65	—1.00										
3.	Em migração	0.70	—0.57	1.00									
4.	Côr (branca)	0.18	—0.29	0.35	1.00								
5.	Percentagem urbana	0.80	—0.80	0.58	0.36	1.00							
6.	Empregos na indústria	0.46	—0.51	0.29	0.61	0.73	1.00						
7.	Consumo de eletricidade	0.61	—0.73	0.50	0.54	0.87	0.88	1.00					
8.	Telefones	0.94	—0.74	0.70	0.33	0.91	0.65	0.82	1.00				
9.	Automóveis	0.62	—0.70	0.63	0.66	0.82	0.81	0.92	0.83	1.00			
10.	Orçamentos estaduais	0.48	—0.60	0.60	0.37	0.61	0.42	0.60	0.64	0.74	1.00		
11.	Alfabetização	0.58	—0.60	0.54	0.70	0.73	0.66	0.74	0.71	0.85	0.44	1.00	
12.	L. Hospital	0.77	—0.60	0.55	0.59	0.88	0.77	0.85	0.88	0.88	0.59	0.89	1.00

Quadro 4A — Valôres de Eigen I 8.3, II 1.3,...

Percentagem de explicação total I 69, ...

Quadro 5A — Pesos das variáveis nos fatôres I e II:

		FATÔRES N/ROTACIONADOS				FATÔRES ROTACIONADOS			
		ÍNDICE		PERCENTAGEM		ÍNDICE		PERCENTAGEM	
		I	II	I	II	I	II	I	II
1.	Densidade	0.80	—0.47	63	22	0.89	0.02	79	0
2.	Acréscimo natural	—0.80	0.25	64	6	—0.35	—0.09	12	1
3.	Em emigração	0.70	—0.38	49	14	0.40	0.17	16	3
4.	Côr	0.59	0.67	34	44	0.02	0.93	0	86
5.	População urbana	0.93	—0.15	86	2	0.58	0.10	33	1
6.	Pess./indústria	0.79	0.41	62	17	0.24	0.33	6	11
7.	Eletricidade	0.92	0.16	85	3	0.33	0.24	11	6
8.	Telefones	0.93	—0.29	87	8	0.76	0.10	57	1
9.	Automóveis	0.95	0.18	91	3	0.32	0.37	11	14
10.	Orçamento do Estado	0.71	—0.17	50	3	0.23	0.15	5	2
11.	Alfabetização	0.85	0.27	73	7	0.35	0.49	12	24
12.	L. Hospital	0.95	0.10	89	1	0.59	0.36	35	13

Quadro 6A — "scores" das áreas no fator não rotacionado I "consenso do desenvolvimento":

Guanabara	351
São Paulo	180
Rio de Janeiro	70
Rio Grande do Sul	55
Santa Catarina	23
Goiás — DF	22
Paraná	17
Minas Gerais	— 5
Mato Grosso	— 20
Pernambuco	— 25
Espírito Santo	— 35
Pará	— 41
Amazonas	— 48
Sergipe	— 51
Paraíba	— 52
Rio Grande do Norte	— 58
Alagoas	— 60
Bahia	— 61
Ceará	— 74
Maranhão	— 92
Piauí	— 96

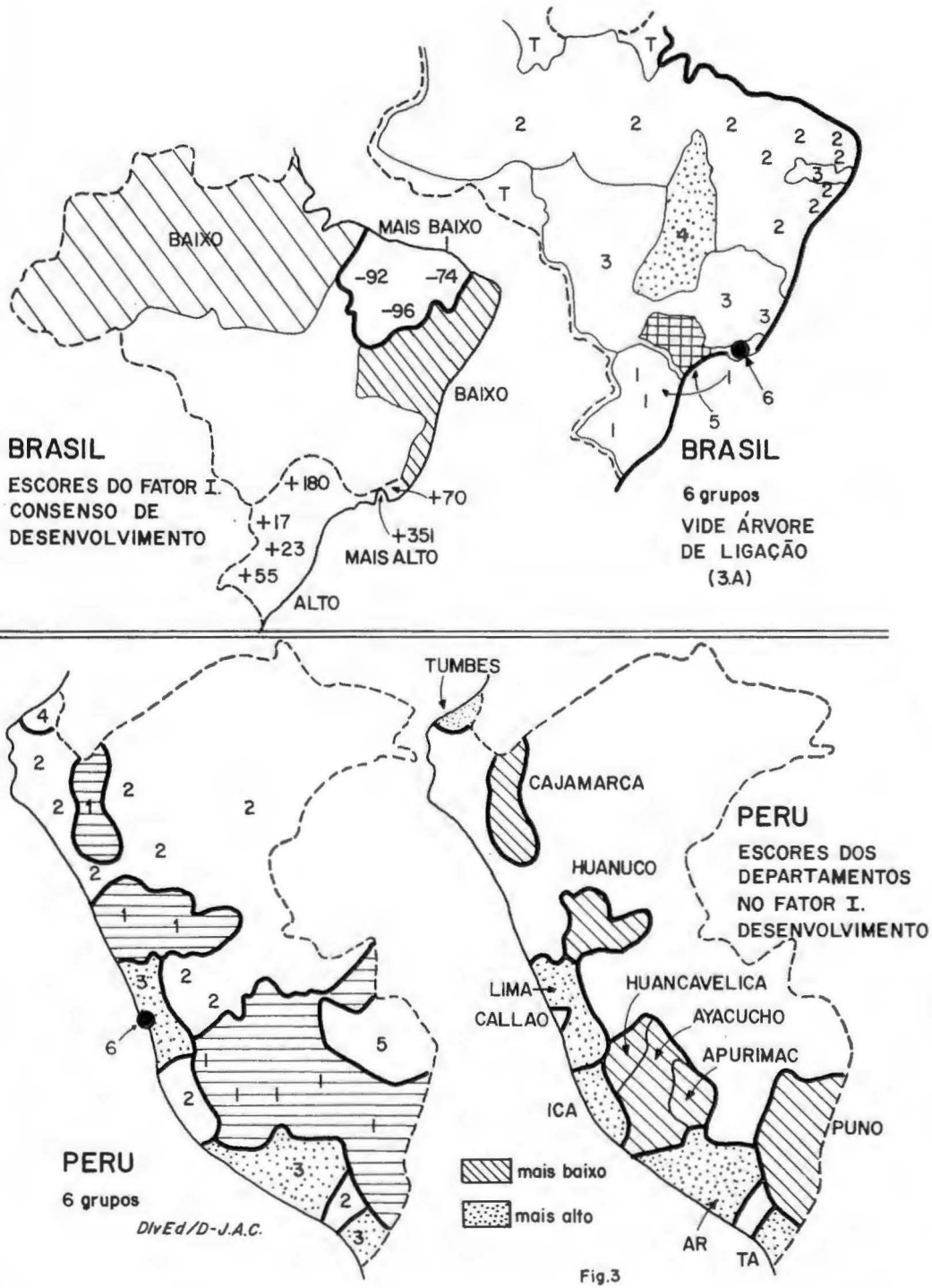
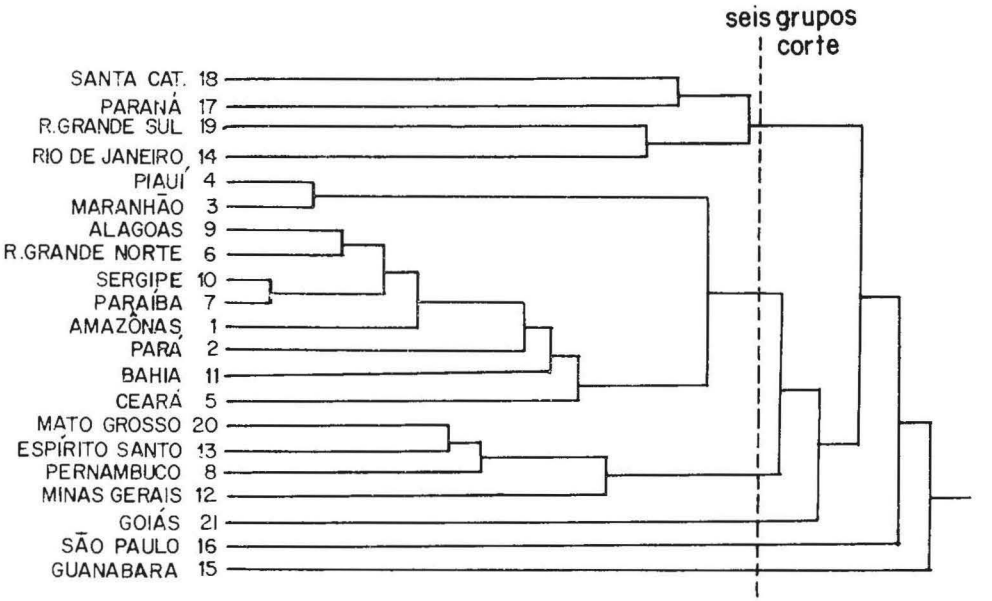

BAIXO
MAIS BAIXO
-92
-74
-96
BAIXO
BRASIL
ESCORES DO FATOR I.
CONSENSO DE
DESENVOLVIMENTO
+180
+70
+17
+23
+55
+351
MAIS ALTO
ALTO
BRASIL
6 grupos
VIDE ÁRVORE
DE LIGAÇÃO
(3.A)
TUMBES
CAJAMARCA
HUANUCO
PERU
ESCORES DOS
DEPARTAMENTOS
NO FATOR I.
DESENVOLVIMENTO
LIMA
CALLAO
HUANCAVELICA
AYACUCHO
APURIMAC
ICA
PUNO
AR
TA
mais baixo
mais alto
PERU
6 grupos
DivEd/D-J.A.C.

Fig.3

Árvore de ligação: 21 estados do Brasil.



Árvore de ligação: 24 departamentos do Peru.

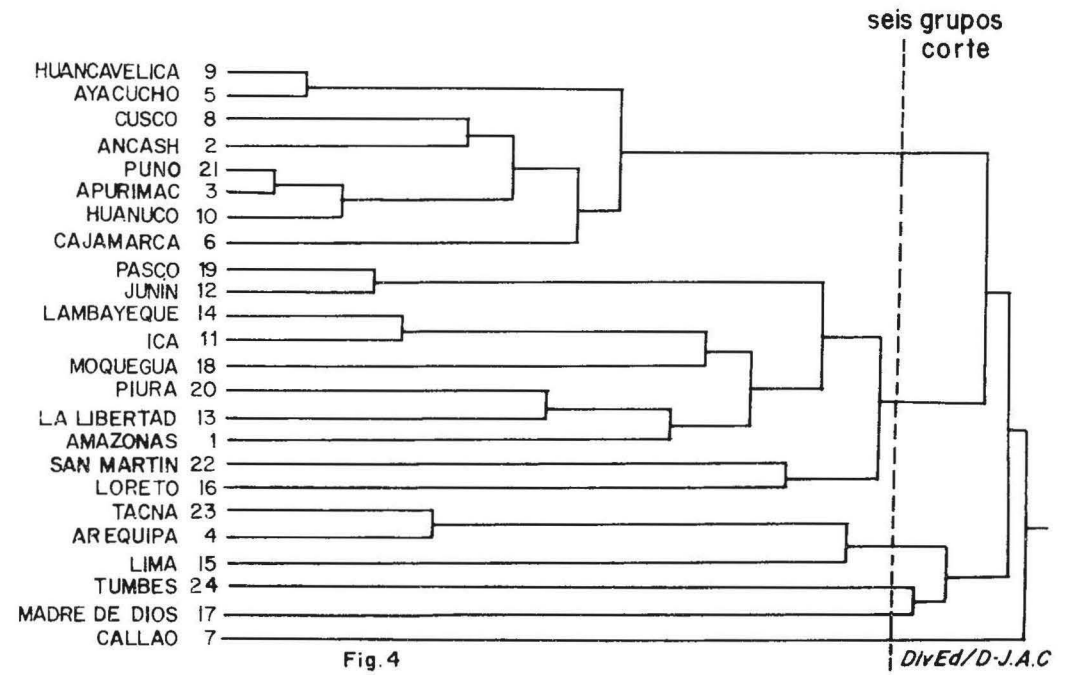


Fig. 4

DivEd/D·J.A.C

10

"Trend Surface Analysis" aplicado ao Estudo da Distribuição da População Rural no Norte de Portugal

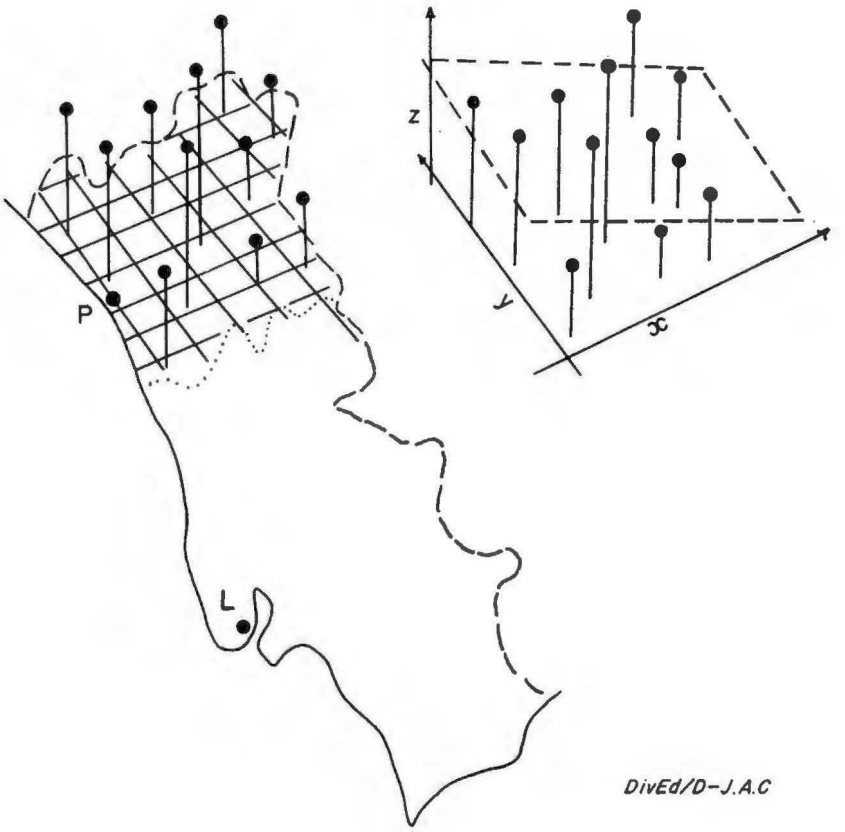
INTRODUÇÃO

O método "*trend surface analysis*" (análise de superfícies de tendência) se aplica para ajustar superfícies a distribuições de pontos em 3 dimensões. A "primeira" superfície é uma superfície plana, a "segunda" é curva e as outras têm formas mais complicadas. As 3 dimensões nas quais se encontram os pontos representam respectivamente:

1. a posição de um lugar (geográfico) numa direção oeste-este (eixo x)
2. a posição do mesmo lugar numa direção sul-norte (eixo y)
3. o valor do lugar numa variável (por exemplo, altitude, densidade de população).

Em muitas distribuições geográficas, os valôres observados numa variável para um número determinado de lugares revelam uma tendência para aumentar ou diminuir de um lado da região para outro. Por exemplo, no caso do norte de Portugal, a densidade da população rural diminui de uma maneira geral de oeste para leste. Observam-se, no entanto, exceções a esta tendência geral. Na faixa ocidental notam-se certas áreas com uma densidade relativamente baixa, enquanto que na parte oriental há concelhos que apresentam uma densidade relativamente alta.

O objetivo do processo de *trend surface analysis* é de produzir e de descrever matemàticamente a forma de uma série de superfícies que caracterizem a tendência geral dos valôres observados para diminuir de uma parte à outra da região. Ao mesmo tempo, o processo calcula um valor "residual" para cada lugar incluído no estudo.



Conceitualmente, a superfície em *trend surface analysis* corresponde a um valor médio em 3 dimensões. Esta idéia se representa na figura 1, (a), (b) e (c). No caso de uma série de valôres numa escala linear, o valor médio corresponde a um ponto nessa escala. Para valôres distribuídos numa gráfico a duas dimensões, a posição equivalente à da média é representada por uma linha. No caso de valôres distribuídos num espaço a 3 dimensões, os valôres médios encontram-se numa superfície a 2 dimensões. Como a superfície representa uma média, alguns valôres estarão acima dessa superfície e outros abaixo.

O valor residual indica a posição, positiva ou negativa, do lugar estudado, em relação à superfície que caracteriza a tendência geral (veja-se figura 1) (d). Quando o valor atual ou "observado" do lugar fica abaixo da superfície expectativa (ponto 3), o residual tem um valor positivo. Este valor é o resultado do cálculo seguinte: subtrai-se o valor observado do valor expectativo. O valor expectativo é o nível do ponto na superfície que se encontra verticalmente acima do lugar. Ao contrário,

quando o valor observado do lugar fica acima da superfície (ponto 1), há um residual negativo.

Os valôres dos resíduos representam anomalias em relação à tendência geral, e *trend surface analysis*, ao revelar estas anomalias, indica os lugares que merecem investigações mais detalhadas. A primeira superfície tem geralmente a forma de um plano inclinado do lado da área que tem os valôres mais altos para o lado que tem os valôres mais baixos. Como a primeira superfície não é capaz de descrever mais do que uma parte da distribuição, convencionalmente expressa-se o grau em que a superfície descreve a distribuição dos pontos em forma de percentagem da "explicação". Se, porém, a distribuição dos pontos fôsse tal que a superfície de tendência se ajustasse perfeitamente, essa superfície explicaria 100% da distribuição. Se, ao contrário, os pontos estivessem distribuídos aleatòriamente, seria difícil que a superfície caracterizasse mais do que 5% — 10% da variação.

É possível ajustar uma superfície não plana, por exemplo, curva (veja-se figura 1) (g) que caracterize com mais precisão a distribuição dos pontos em três dimensões. Da mesma maneira, no caso duma distribuição de pontos num gráfico, é possível que, em certas situações, uma linha curva caracterize melhor a distribuição do que uma linha reta (veja-se figura 1) (e). A aplicação duma segunda superfície em *trend surface analysis* diminui em geral o valor dos resíduos e aumenta a percentagem da explicação. As superfícies mais complicadas, isto é, a segunda (quadrática), a terceira (cúbica), a quarta, etc., ajustam-se cada vez, com maior precisão, à distribuição. Como explicam progressivamente uma maior percentagem da variação, a distância dos valôres anômalos (residuais) em relação à superfície tendencial é cada vez menor. Isto é, a superfície aproxima-se cada vez mais da distribuição mesma dos pontos. Mas as superfícies, de ordem mais elevada, exigem progressivamente mais cálculos. Os cálculos necessários até para a primeira superfície necessitam da ajuda de um computador eletrônico.

APLICAÇÃO DO MÉTODO AO NORTE DE PORTUGAL

A Figura 2 mostra a posição dos 123 concelhos estudados.

O Quadro 1 mostra a posição exata de cada concelho num gráfico de coordenadas usadas neste estudo, com ponto de origem ao sudoeste da parte de Portugal considerada:

colunas 1 e 2: a posição para leste e norte, respectivamente

coluna 3 : a densidade de população rural observada

coluna 4 : a densidade de população rural expectativa segundo a primeira superfície

coluna 5 : a diferença entre colunas 3 e 4

coluna 6 : o valor em coluna 5 como percentagem do valor em coluna 3.

Figura 3 — mostra o valor da densidade de população rural para cada concelho. Nota-se visualmente neste mapa uma diminuição geral dos valôres de oeste para leste. No centro da área considerada, porém, encontram-se os concelhos de mais elevada densidade de população rural (por exemplo concelhos números 44 — Mesão Frio, 51 Sta Marta de Penaguião, 48 Peso da Régua.

A *figura 4* mostra os valôres da 1.ª superfície de tendência. Neste exemplo a 1.ª superfície explica 28,5% da variação. Os valôres na fig. 4 chamam-se os valôres expectativos enquanto que os valôres da fig. 3 chamam-se os valôres observados. O valor expectativo é o valor teórico que cada concelho teria se coincidisse com a superfície mesma. A diferença entre o valor observado e o valor expectativo de cada concelho é o valor residual.

A *figura 5* indica a inclinação da primeira superfície e a *figura 6* mostra o valor residual de cada concelho. Estes valôres encontram-se no quadro 1, coluna 5. Veja-se a fig. 2 para a localização dos concelhos.

A diferença entre o valor observado e o valor espectativo, como já explicamos na primeira seção, calcula-se subtraindo o valor observado do valor expectativo. Por isso, os valôres residuais são positivos quando os valôres observados são inferiores aos valôres expectativos. Por exemplo, no caso do concelho n.º 2 (Caminha), o valor observado é 11, mas o valor expectativo é aproximadamente 35. O valor residual é aproximadamente 24. Revela uma forte anomalia da densidade de população rural do concelho de Caminha em relação aos concelhos vizinhos, um fato que se observa fàcilmente no mapa da figura 3.

No caso do concelho n.º 44 — Mesão Frio, o valor observado é 68, o valor expectativo é 23.4. Há um valor residual muito elevado de — 44.6. Isto indica que o valor expectativo é muito inferior ao valor da verdadeira densidade de população rural do concelho.

INTERPRETAÇÃO DOS RESULTADOS

A primeira superfície de tendência é uma superfície plana inclinada de oeste para leste ou, melhor, O.-N.O. para E.-S.E., sugerindo, portanto, que os valôres de densidade de população rural na área considerada apresentem uma tendência geral para decrescer nessa direção. A inclinação da superfície revela uma faixa ocidental de fortes densidades de população rural. Note-se, no entanto, que esta faixa estreita-se ao sul do Douro. A área baixa do norte do Minho com fortes densidades de popu-

lação rural estende-se desde o litoral até às montanhas do Alto Minho (concelho de Terras de Bouro (n.º 19) e área oriental dos concelhos de Arcos de Valdevez (n.º 1) e Ponte da Barca (n.º 6). Ao sul do Douro, as montanhas do norte da Beira estão mais próximas do Atlântico e a área baixa costeira é mais reduzida.

Na periferia oriental da área em estudo, observam-se os mais baixos valôres de densidade de população. São, com efeito, os planaltos transmontanos e, ao sul do Douro, os concelhos fronteiriços da Beira.

Esta primeira superfície explica 28,5% da distribuição. Na fig. 6 estão representados os resíduos em relação à superfície de tendência, ou sejam, os valôres que distam desta, numa direção positiva (abaixo da superfície) ou negativa (acima da superfície).

São os concelhos da seção média do vale do Douro que apresentam os valôres residuais negativos mais elevados: concelhos n.º 44 (Mesão Frio) com — 44, n.º 51 (Sta Marta de Penaguião) com — 37, n.º 48 (Peso da Régua) com — 34, n.º 88 (Lamego) com — 31, n.º 91 (Resende) com — 22. São os concelhos da região vinhateira do Douro com densidades muito elevadas de população rural. Tôda a área baixa da província do Minho apresenta valôres de densidade de população rural acima da superfície de tendência. Notem-se particularmente os concelhos de Felgueiras (n.º 26), Fafe (n.º 16), Amares (n.º 11) e Vila Verde (n.º 22) com valôres residuais de respectivamente — 19, — 13, — 12, — 9, Mesmo concelhos do litoral, como Póvoa do Varzim (n.º 36) e Esposende (n.º 15), apresentam valôres residuais negativos elevados. Pelo contrário, para os concelhos do litoral ao norte de Viana do Castelo (concelho n.º 9), os valôres residuais são positivos, isto é, as densidades de população rural são mais baixas do que os valôres expectativos da superfície de tendência. Este fato nota-se particularmente para a vila piscatória de Caminha.

A cidade do Pôrto, áreas urbanas vizinhas e concelhos, onde a proximidade da cidade se faz sentir mais diretamente, apresentam relativamente fracas densidades de população rural. Observam-se, também, elevados valôres residuais positivos nos concelhos da montanha do Minho. Os concelhos números 1 e 6, Arcos de Valdevez e Ponte da Barca, estendem-se muito na direção leste-oeste e incluem áreas da montanha com baixas densidades de população. Note-se como a área dêstes concelhos foi incluída na superfície de densidades mais elevadas (veja-se superfície de tendência, fig. 5). Contudo, os verdadeiros valôres de densidade de população na montanha do Minho são francamente anômalos em relação a esta tendência geral, isto é, encontram-se muito abaixo desta e os valôres residuais são positivos. Para os concelhos de Arcos de Valdevez (n.º 1) e de Ponte da Barca (n.º 6) são respectivamente 3 e 5. O concelho de Terras de Douro (n.º 19) engloba uma área montanhosa que continua pelas serras do Gerês e Larouco (concelho n.º 46 Montalegre). Note-se como os valôres residuais positivos para estes dois concelhos são elevados: 17 e 13 respectivamente.

Em todo o Trás-os-Montes e montanhas do norte da Beira, as densidades de população rural são relativamente baixas. Contudo, Viseu (concelho n.º 108), centro mais importante desta última área apresenta um valor residual negativo em relação à tendência geral. O mesmo acontece com os concelhos vizinhos de Penalva do Castelo (n.º 103), Nelas (n.º 101), Satão (n.º 106) e Mangualde (n. 99).

As superfícies de ordem mais elevada ajustam-se progressivamente melhor à distribuição. A 6.ª superfície contém 61,7% da explicação. Encontra-se representada na fig. 7. Na área do Baixo-Minho e seção média do vale do Douro (concelhos vizinhos de Peso da Régua) observam-se as mais elevadas densidades de população rural. Estas decrescem para o litoral, montanhas do Minho e, de uma maneira geral, ao sul do Douro e Trás-os-Montes. As densidades mais baixas de população observam-se na área leste e nordeste transmontana e montanhas do norte.

Observe-se o mapa de resíduos (fig. 8). O contraste entre as densidades de população rural do Baixo-Minho e a montanha são realmente mais acusados. A vila piscatória de Caminha constitui uma flagrante exceção em relação aos concelhos minhotos da orla marítima e das margens do rio Minho. As densidades de população rural são muito baixas nas montanhas do norte da Beira. Aumentam sensìvelmente em direção ao Douro, para a região de Viseu, ao sul, e para o litoral de Aveiro. Os planaltos de Trás-os-Montes apresentam densidades abaixo da superfície de tendência e no nordeste observam-se os valôres mais baixos. Chaves (conc. n.º 43) constitui uma flagrante exceção. Note-se também que os concelhos da Terra Quente do Douro, como Moncorvo e Vila Nova de Fozcoa, apresentam densidades de população rural relativamente elevadas.

CONCLUSÃO

Trend surface analysis tem as seguintes aplicações.

1. Dá uma descrição geral de uma distribuição complicada.
2. Revela anomalias numa distribuição, indicando a necessidade de investigar mais detalhadamente.
3. No caso de a população observada ser apenas uma amostragem, a superfície dá a possibilidade de estimar, com dado nível de confiança, valôres para lugares distribuídos entre os pontos escolhidos especìficamente para calcular a superfície. Êste aspecto não foi desenvolvido no exemplo usado neste estudo.

QUADRO 1

RESÍDUOS PARA SUPERFÍCIE DE ORDEM 1

N.º de ordem		X'	Y'	Z Valôres Observados	Z Valôres Expectativos	Diferença	Percentagem da Diferença
1	Arcos de Valdevez	6 300	18 400	27 000	30 862	3 862	14 303
2	Caminha	3 400	18 200	11 000	34 949	23 949	117 719
3	Melgaço	7 500	19 800	31 000	29 537	— 1 463	4 719
4	Monção	6 000	19 800	29 000	31 680	2 680	9 241
5	Paredes de Coura	4 900	18 700	35 000	32 945	— 2 055	5 870
6	Ponte da Barca	7 000	17 500	24 000	29 611	5 611	23 381
7	Ponte de Lima	4 600	17 000	42 000	32 901	— 9 099	21 665
8	Valença	4 500	19 500	30 000	33 739	3 739	12 465
9	Viana do Castelo	3 400	16 600	23 000	34 504	11 504	50 017
10	Vila Nova da Cerveira	3 900	18 800	22 000	34 402	12 402	56 372
11	Amares	6 600	16 000	42 000	29 765	—12 235	29 130
12	Barcelos	4 500	15 000	40 000	32 487	— 7 513	18 782
13	Braga	6 000	15 000	39 000	30 344	— 8 656	22 194
14	Cabeceiras de Basto	9 600	14 900	23 000	25 174	2 174	9 450
15	Esposende	3 300	15 000	42 000	34 201	— 7 799	18 568
16	Fafe	8 000	14 200	41 000	27 265	—13 735	33 501
17	Guimarães	6 800	14 000	35 000	28 923	— 6 077	17 362
18	Póvoa do Lanhoso	7 300	15 400	35 000	28 598	— 6 402	18 290
19	Terras de Bouro	7 700	16 800	11 000	26 417	17 417	58 333
20	Vieira do Minho	8 400	15 700	20 000	27 111	7 111	35 553
21	Vila Nova de Famalição	5 200	13 600	30 000	31 098	1 098	3 659
22	Vila Verde	5 900	16 500	42 000	30 905	—11 095	26 418
23	Celorico de Basto	8 900	13 600	23 000	25 812	2 812	12 225
24	Amarante	8 800	12 300	29 000	25 593	— 3 407	11 748
25	Baião	9 100	11 100	37 000	24 830	—12 170	32 891
26	Felgueiras	7 600	13 000	47 000	27 502	—19 498	41 485
27	Gondomar	5 300	10 700	22 000	30 148	8 148	37 035
28	Lousada	7 000	12 300	41 000	28 164	—12 836	31 306
29	Maia	4 400	12 000	27 000	31 795	4 795	17 760
30	Marco de Canaveses	7 900	11 100	36 000	26 545	— 9 455	26 265
31	Matosinhos	3 800	11 700	29 000	32 569	3 569	12 307
32	Paços de Ferreira	6 200	12 400	30 000	29 335	— 0 665	2 216
33	Paredes	6 100	11 600	33 000	29 255	— 3 745	11 347
34	Penafiel	6 800	11 000	30 000	28 088	— 1 912	6 372
35	Pôrto	4 200	11 200	26 000	31 858	5 858	22 532
36	Póvoa de Varzim	3 500	13 800	51 000	33 582	—17 418	34 153
37	Santo Tirso	5 500	12 600	27 000	30 391	3 391	12 558
38	Valongo	5 300	11 600	20 000	30 398	10 398	51 991
39	Vila do Conde	3 900	13 000	28 000	32 788	4 788	17 099
40	Vila Nova de Gaia	4 500	10 400	29 000	31 207	2 207	7 611
41	Alijó	13 100	12 400	23 000	19 478	— 3 522	15 314
42	Boticas	11 200	16 200	15 000	23 250	8 250	54 997
43	Chaves	13 300	16 900	21 000	20 444	— 0 556	2 646
44	Mesão Frio	10 100	11 200	68 000	23 430	—44 570	65 545
45	Mondim de Basto	10 000	13 400	14 000	24 185	10 185	72 748
46	Montalegre	10 300	17 300	11 000	24 841	13 841	25 831
47	Murça	13 400	13 600	19 000	19 383	0 383	2 017
48	Peso da Régua	10 700	11 200	57 000	22 573	—34 427	60 399
49	Ribeira de Pena	10 700	14 600	20 000	23 519	3 519	17 593
50	Sabrosa	12 200	12 000	21 000	20 652	— 0 348	1 656
51	Sta. Marta de Penaguião	10 800	11 800	60 000	22 597	—37 403	62 339
52	Valpaços	14 200	15 500	19 000	18 769	— 0 231	1 215
53	Vila Pouca de Aguiar	12 000	14 500	16 000	21 634	5 634	35 211
54	Vila Real	11 000	12 500	30 000	22 506	— 7 494	24 981
55	Alfândega da Fé	17 200	13 100	9 000	13 815	4 815	53 505
56	Bragança	19 000	17 200	9 000	12 385	3 385	37 610
57	Carrazeda de Anciães	14 500	11 800	13 000	17 311	4 311	33 161
58	Freixo de E. à C.	17 800	10 500	8 000	12 235	4 235	52 936
59	Macedo de Cav.	17 500	15 000	11 000	13 916	2 916	26 506
60	Miranda do Douro	21 700	14 400	8 000	7 749	— 0 251	3 142
61	Mirandela	15 400	14 500	11 000	16 777	5 777	52 514
62	Mogadouro	19 400	12 800	7 000	10 589	3 589	51 273
63	Tôrre de Moncorvo	16 700	11 100	11 000	13 973	2 973	27 030
64	Vila Flor	15 600	12 600	13 000	15 962	2 962	22 785
65	Vimioso	20 400	15 200	7 000	9 828	2 828	40 405

QUADRO 1

RESÍDUOS PARA SUPERFÍCIE DE ORDEM 1

N.º de ordem		X'	Y'	Z Valôres Observados	Z Valôres Expectativos	Diferença	Porcentagem da Diferença
66	Vinhais	16 500	18 000	11 000	16 179	5 179	47 082
67	Agueda	5 700	5 400	16 000	28 102	12 102	75 635
68	Albergaria A. V.	5 200	6 500	18 000	29 122	11 122	61 788
69	Anadia	5 500	4 200	27 000	28 053	1 053	3 901
70	Arouca	7 000	8 800	19 000	27 190	8 190	43 108
71	Aveiro	4 400	5 900	27 000	30 098	3 098	11 473
72	Castelo de Paiva	6 700	9 700	20 000	27 870	7 870	39 348
73	Espinho	4 200	9 500	13 000	31 385	18 385	41 425
74	Estarreja	4 600	7 200	27 000	30 174	3 174	11 755
75	Feira	5 000	9 200	24 000	30 159	6 159	25 662
76	Ilhavo	3 800	5 600	27 000	30 872	3 872	14 339
77	Mealhada	5 300	3 200	41 000	28 061	—12 939	31 559
78	Murtosa	4 000	7 200	27 000	31 031	4 031	14 930
79	Oliveira de Azemeis	5 500	8 000	26 000	29 111	3 111	11 964
80	Oliveira do Bairro	7 600	4 700	47 000	25 192	—21 808	46 399
81	Ovar	4 200	8 300	19 000	31 051	12 051	63 428
82	São João da Mad	5 300	8 600	30 000	29 563	— 0 437	1 455
83	Sever do Vouga	6 200	6 800	27 000	27 777	0 777	2 877
84	Vagos	3 500	4 800	32 000	31 077	— 0 923	2 883
85	Vale de Cambra	6 400	7 900	28 000	27 797	— 0 203	0 724
86	Armamar	11 500	10 500	29 000	21 235	— 7 765	26 776
87	Cinfães	8 300	9 700	29 000	25 584	— 3 416	11 780
88	Lamego	10 500	10 300	54 000	22 608	—31 392	58 134
89	Moimenta da Beira	11 800	9 100	17 000	20 417	3 417	20 099
90	Penedono	13 700	9 400	13 000	17 786	4 786	36 815
91	Resende	9 600	10 300	46 000	23 894	—22 106	48 057
92	S. João da Pesqu	13 500	10 700	17 000	18 433	1 433	8 432
93	Sernancelhe	12 900	8 700	12 000	18 734	6 734	56 117
94	Tabuaço	12 400	10 500	19 000	19 949	0 949	4 996
95	Tarouca	10 800	9 600	32 000	21 984	—10 016	31 298
96	Vla. Nova de Paiva	10 900	8 200	15 000	21 452	6 452	43 014
97	Carregal do Sal	9 100	4 000	23 000	22 855	— 0 145	0 631
98	Castro d'Aire	9 700	8 700	21 000	23 305	2 305	10 979
99	Mangualde	11 200	5 500	22 000	20 272	— 1 728	7 854
100	Mortagua	7 000	3 700	11 000	25 771	14 771	34 285
101	Nelas	10 100	4 800	24 000	21 649	— 2 351	9 797
102	Oliv. de Frades	7 200	6 700	20 000	26 320	6 320	31 602
103	Penalva do Castelo	11 700	6 300	25 000	19 780	— 5 220	20 878
104	Sta. Comba Dão	8 000	3 500	24 000	24 287	0 287	1 196
105	S. Pedro do Sul	8 200	7 600	17 000	25 142	8 142	47 896
106	Sátão	11 500	7 300	22 000	20 344	— 1 656	7 525
107	Tondela	8 000	4 900	26 000	24 677	— 1 323	5 090
108	Viseu	9 800	6 400	31 000	22 523	— 8 477	27 346
109	Vouzela	8 000	6 300	23 000	25 066	2 066	8 984
110	Aguiar da Beira	12 700	7 400	15 000	18 658	3 658	24 387
111	Almeida	17 300	5 800	8 000	11 641	3 641	45 516
112	Celorico da Beira	13 700	5 700	14 000	16 756	2 756	19 688
113	Figueira de C. R	17 000	8 300	7 000	12 766	5 766	82 365
114	Fornos de Alg	12 900	6 000	17 000	17 983	0 983	5 781
115	Gouveia	12 200	7 500	16 000	19 400	3 400	21 251
116	Guarda	14 900	7 700	14 000	15 599	1 599	11 419
117	Manteigas	12 700	3 500	7 000	17 573	10 573	51 039
118	Meda	14 800	8 800	12 000	16 048	4 048	33 730
119	Pinhel	15 800	6 800	12 000	14 062	2 062	17 187
120	Sabugal	16 600	3 400	12 000	11 973	— 0 027	0 221
121	Seia	11 200	3 200	13 000	19 632	6 632	51 017
122	Trancoso	14 100	7 500	13 000	16 686	3 686	28 352
123	V. Nova de Foz Côa	15 600	9 900	11 000	15 211	4 211	38 280

Observados SS = $0.1801787^{10} + 05$

Expectativas SS = $0.5142717^{10} + 04$

Percentagem Explicação = 28.5423144%

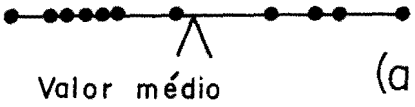
Aumento = 28.5423144

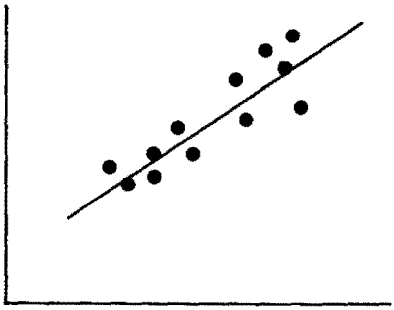
Os coeficientes são:

$0.3474183^{10} + 02 = 0.1428583^{10} + 01$

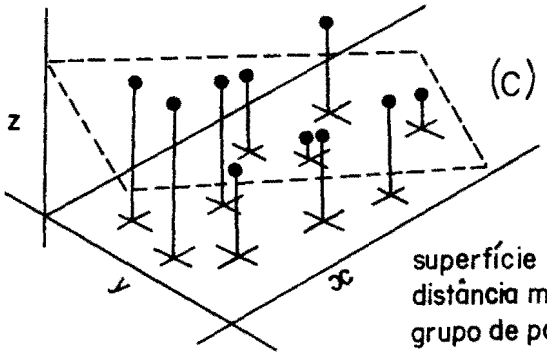
$0.2782654^{10} + 00$

Para a 1.º superfície

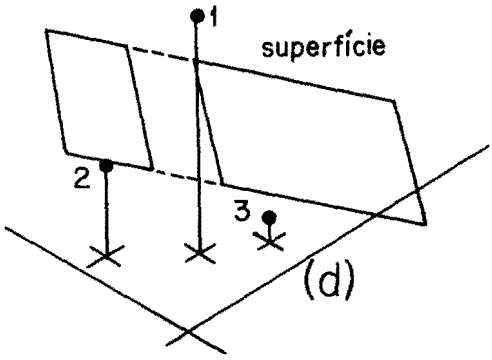




linha a uma distância mínima de um grupo de pontos numa superfície



superfície plana a uma distância mínima de um grupo de pontos em 3 dimensões



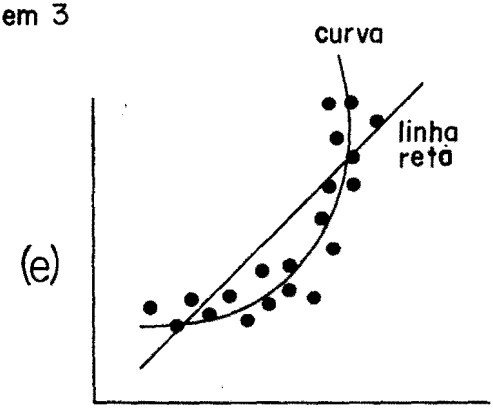
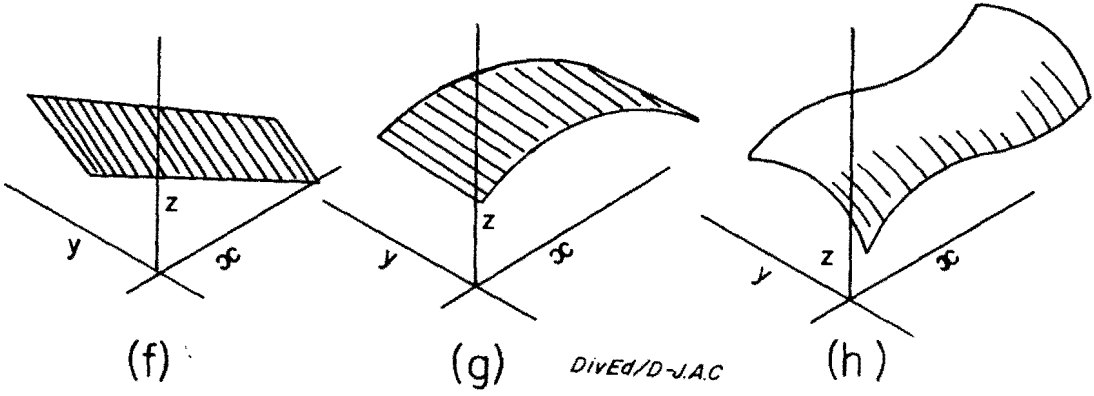


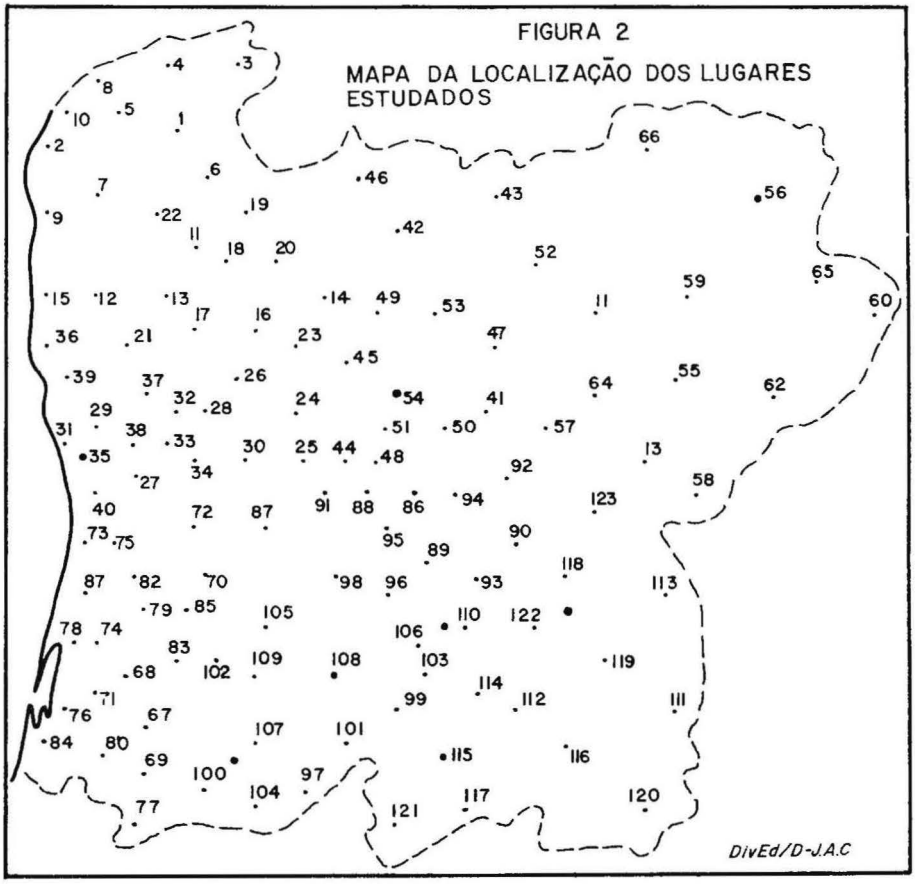
FIGURA 1



DivEd/D-J.A.C

FIGURA 2

MAPA DA LOCALIZAÇÃO DOS LUGARES ESTUDADOS



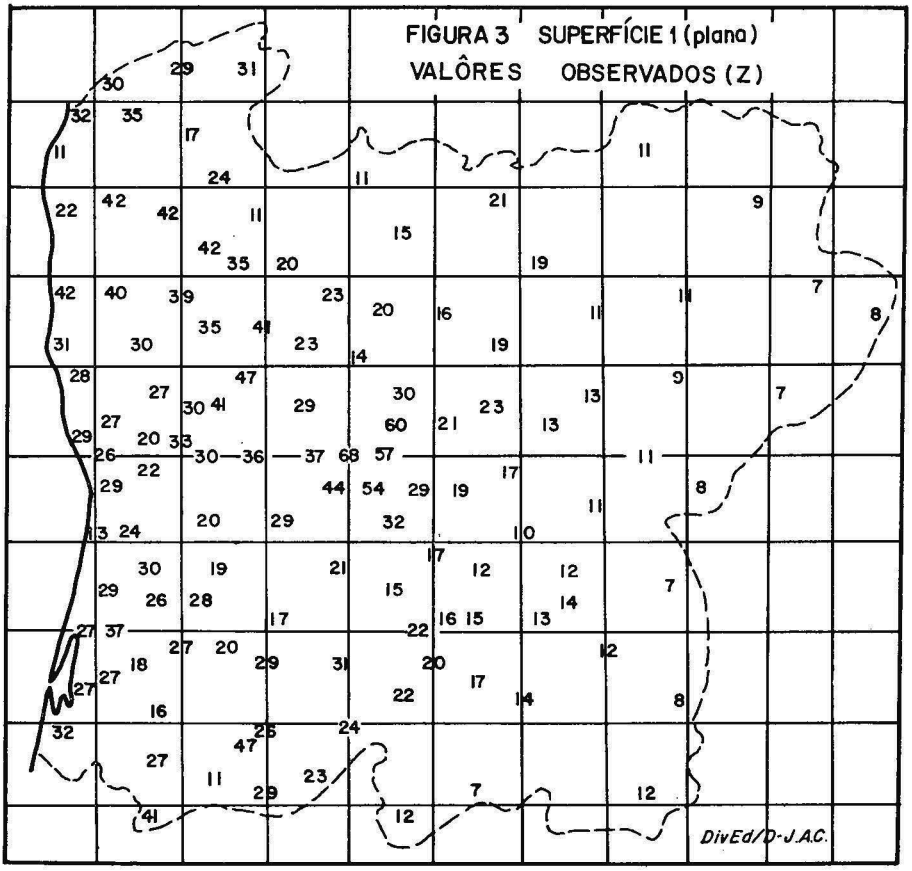


FIGURA 3 SUPERFÍCIE 1 (plano)
VALÔRES OBSERVADOS (Z)
DivEd/D-J.A.C.

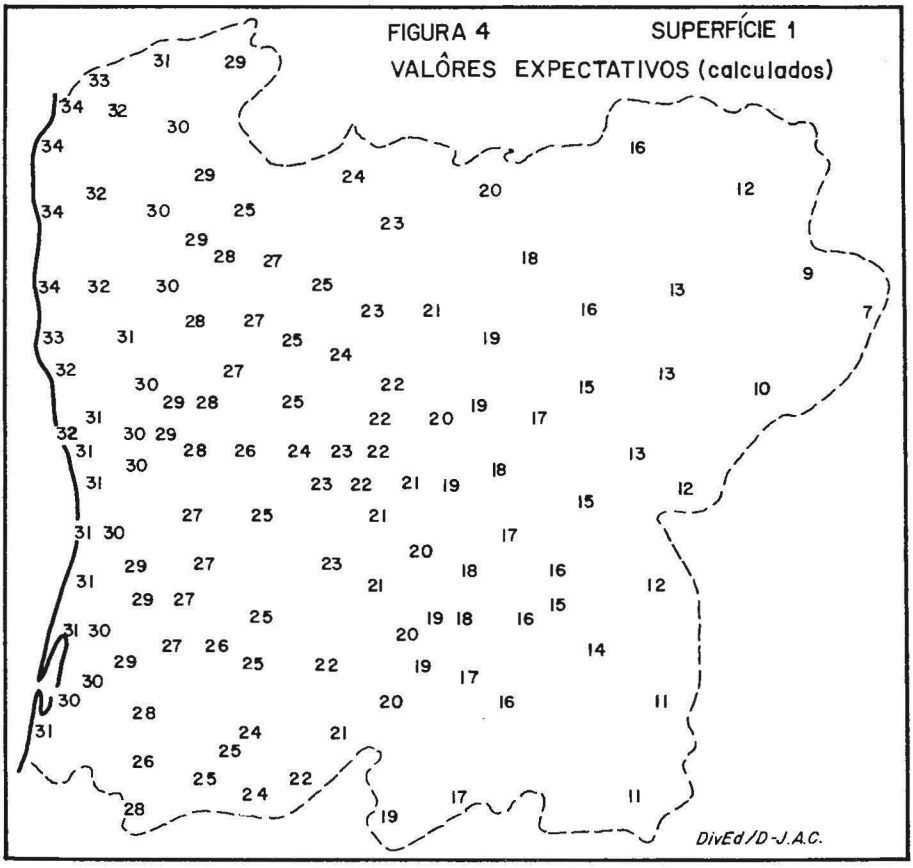
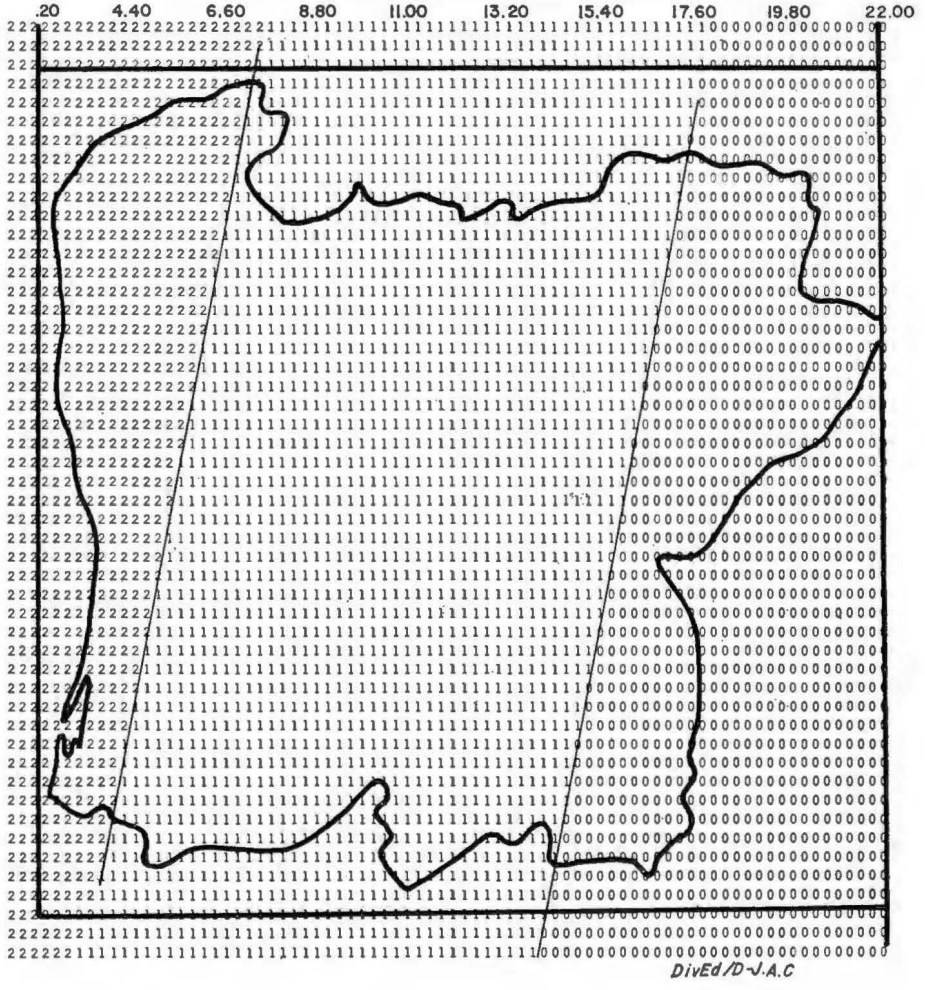


FIGURA 4 SUPERFÍCIE 1
VALÔRES EXPECTATIVOS (calculados)
DivEd/D-J.A.C.

FIGURA 5 - INCLINAÇÃO DA PRIMEIRA SUPERFÍCIE



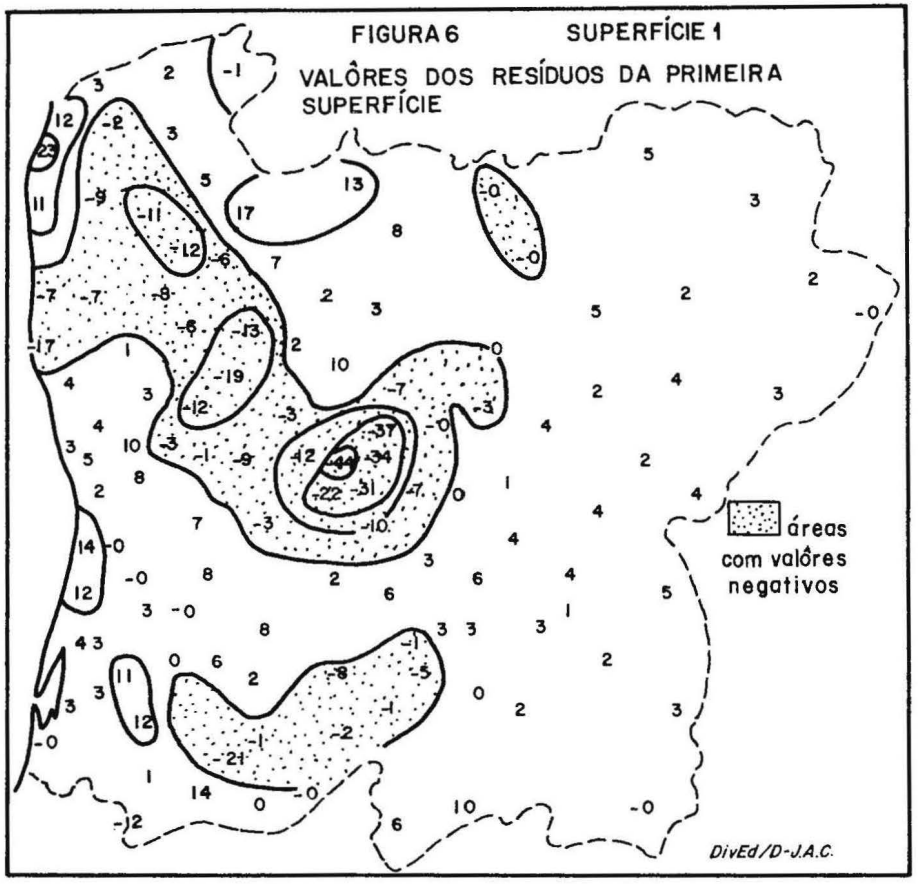
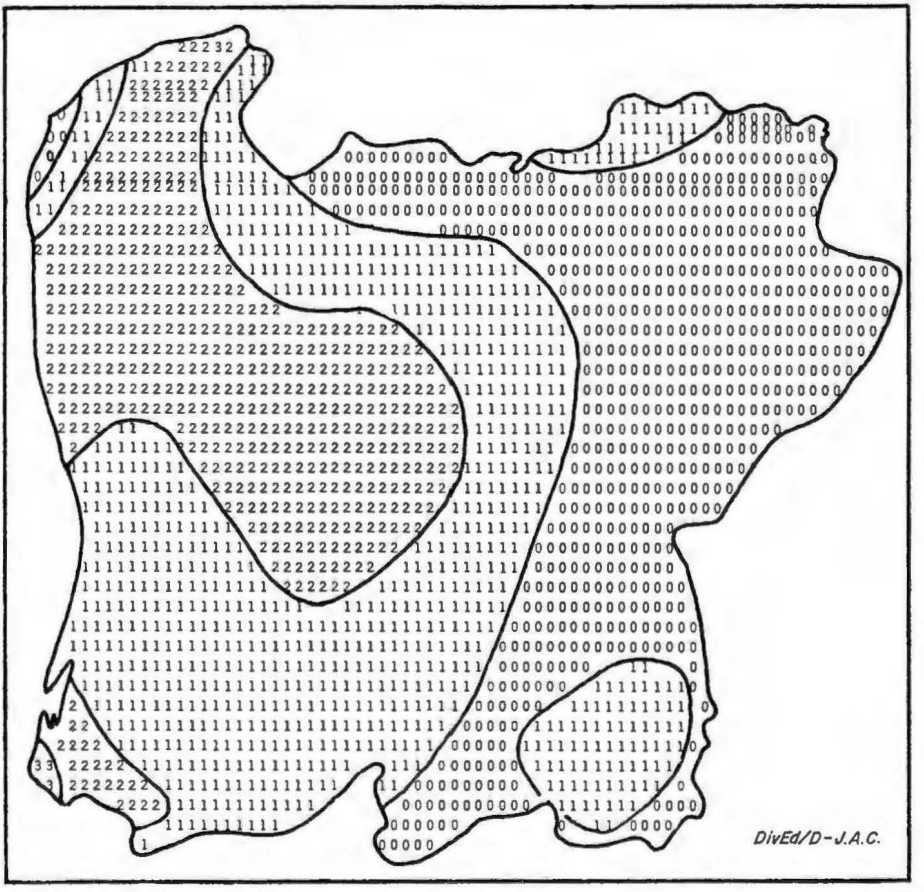


FIGURA 6 SUPERFÍCIE 1
VALÔRES DOS RESÍDUOS DA PRIMEIRA SUPERFÍCIE
áreas com valôres negativos
DivEd/D-J.A.C.

Figura-7 FORMA DA SEXTA SUPERFÍCIE



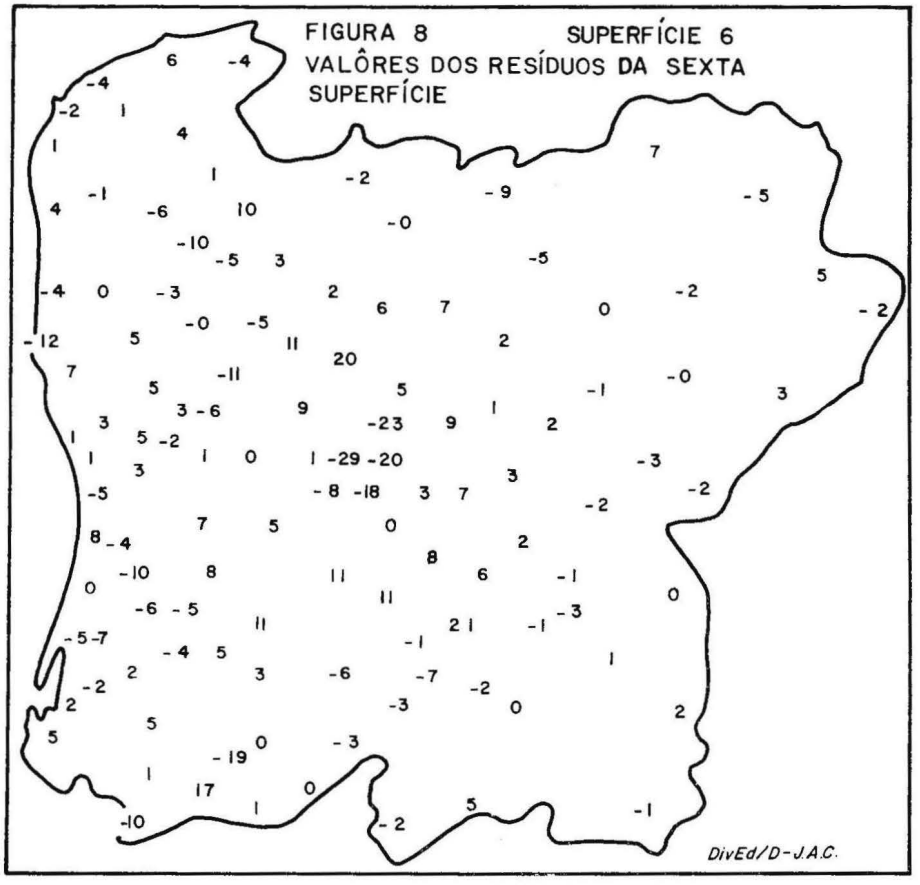


FIGURA 8 SUPERFÍCIE 6
VALÔRES DOS RESÍDUOS DA SEXTA SUPERFÍCIE
DivEd/D-J.A.C.

11

Projeções Demográficas com Computador

INTRODUÇÃO

O programa MALTHUS constitui um tipo relativamente simples da cadeia de Markov projetado para efetuar as operações abaixo, durante determinado número escolhido de anos (NYR) ou períodos de outra duração.

1. Alterar a população (DATA (1,I) de um dado número (NPL) de lugares segundo um índice (DATA (2,I)) de crescimento vegetativo (aumento, redução ou inalterabilidade).

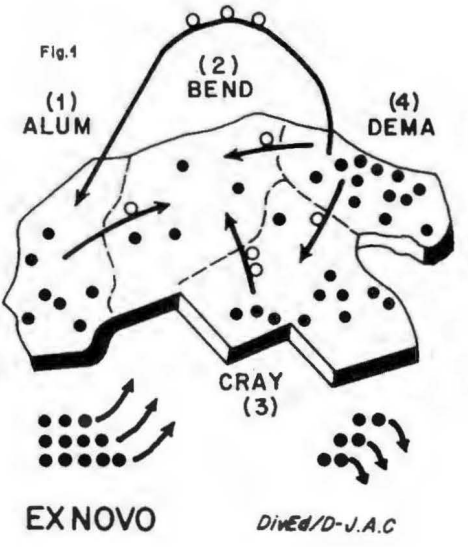
2. Modificar o índice de crescimento vegetativo segundo outro índice (DATA (3,I)) que produz, quando se deseja, uma diminuição, um aumento ou nenhuma alteração no ritmo de crescimento vegetativo.

3. Decidir com o auxílio de um índice (DATA (4,I)) quantos anos deveriam passar, desde o princípio do período estudado, antes que o redutor deva começar a operar. 0.0 significa que êste opera desde o início do primeiro ano.

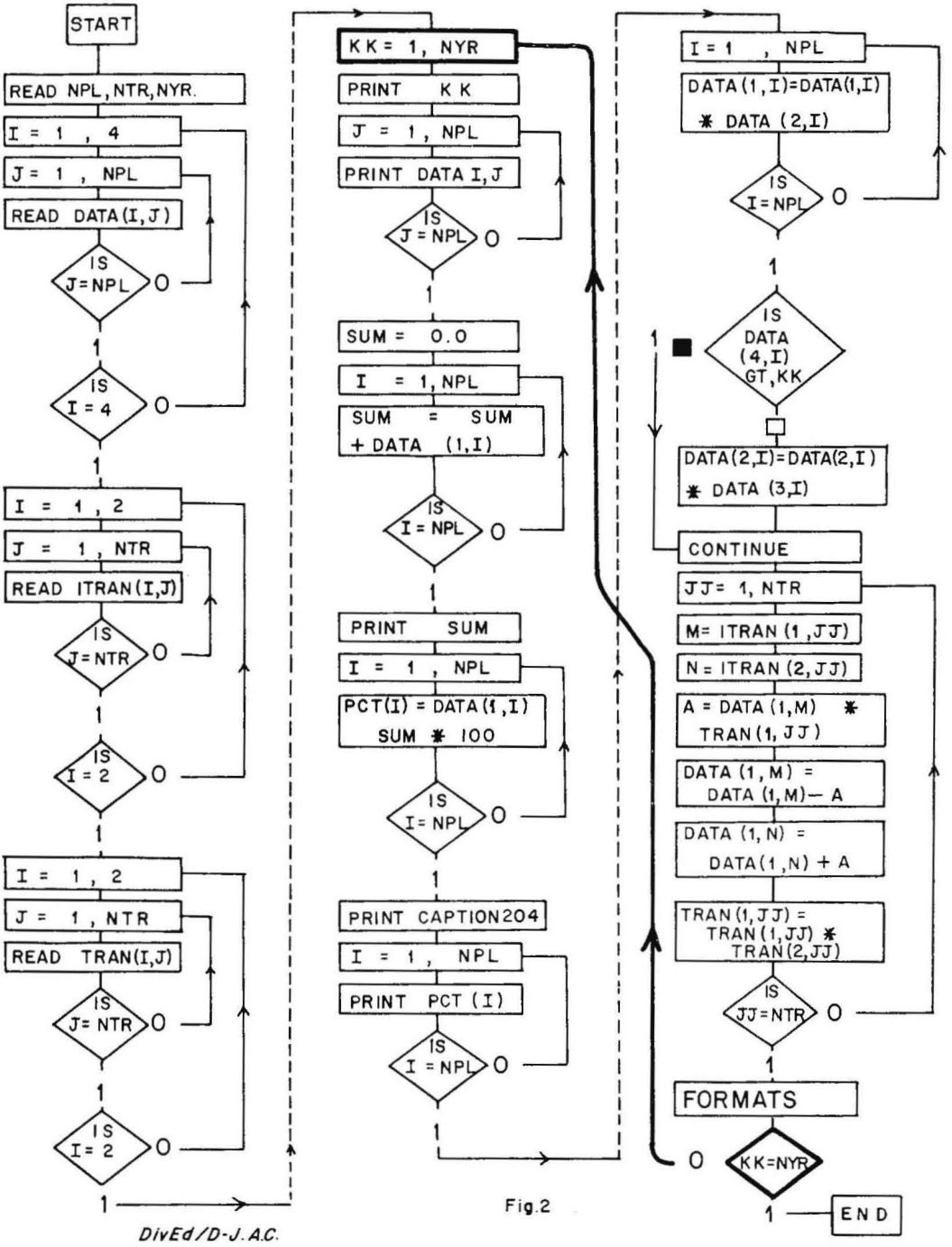
4. Subtrair uma determinada proporção (TRAN (1,I) da população de um lugar (ITRAN 1) e acrescentá-la à população de outro local, (ITRAN 2), representando assim, a migração entre pares de lugares. O número de pares de lugares escolhidos (NTR) é optativo.

5. Alterar o índice de migração segundo um índice (TRAN (2,I)) que produz, quando se deseja, uma redução, um aumento ou nenhuma modificação na taxa de migração populacional.

A *figura 1* mostra a situação em um exemplo de pequena grandeza, a ilha de EXNOVO, além dos *inputs* e etapas escolhidas no programa.



A figura 2 mostra o fluxograma do programa.



DivEd/D-J.A.C.

Fig.2

A *figura 3* mostra uma cópia (printout) do programa em FORTRAN IV (EGDON).

Figura 3

```
          SOMA        "DATA"  FILEIRA 1
Etapa 1   | 2000.0 | —— | 600.0 + 400.0 + 400.0 + 600.0 |

          PCT
Etapa 2   | 30.0    20.0    20.0    30.0 | —— DATA (1,I)/2000.0*100.0

          DATA 1 NÔVO                      DATA 1 ANTERIORES
Etapa 3   | 606.0   404.0   408.0   618.0 | —— | 600.0  400.0  400.0  600.0 |

Etapa 4   | 1.009899 1.009899 1.02  1.03 | —— | 1.01   1.01   1.02   1.03 |
          DATA 2 NÔVO                      DATA 2 ANTERIORES

Etapa 5   a) M | 1 | —— 1 b) N | 2 | —— 2 c) A | .6 | —— 600.0*.001
          b), e) | 605.4 | —— 606.0 | 404.6 | —— 404.0                    etc.
```

A *figura 4* mostra parte dos resultados do programa (Output) para a ilha de EXNOVO.

Figura 4

	POPULAÇÃO NO FIM DO ANO				PERCENTAGEM			
	5	10	15	19	5	10	15	19
(1) Alum....	633	668	702	731	29	28	27	26
(2) Bend....	432	466	501	530	20	19	19	19
(3) Cray....	443	490	541	584	20	21	21	21
(4) Dema...	678	767	863	944	31	32	33	34
TOTAL	2187	2390	2608	2789	100	100	100	100

NOTA 1) O tempo é considerado da seguinte maneira: O programa inicia-se em um determinado instante do tempo, tempo zero, ou o início do ano 1 (ou algum outro período). As mudanças ocorrem durante o ano. Uma nova situação foi criada e os resultados imprimem-se para o *fim* do ano 1, que é o mesmo que o início do ano 2. Tal processo é repetido NYR vêzes.

NOTA 2) Para que não se obtenham resultados expressos em números excessivamente grandes, recomenda-se empregar o milhar de pessoas como unidade de grandeza, a menos que se trate de números muito pequenos.

ETAPAS DOS CÁLCULOS

1.ª Etapa: Somar os valôres da população na linha 1 da matriz DATA para produzir em SOMA a população total de EXNOVO, isto é 2.000. Imprimir êste resultado.

2.ª Etapa: Calcular os valôres da população de cada região, como percentagem da população total. Colocar os resultados em PCT e imprimir PCT.

3.ª Etapa: Alterar a população de cada lugar de acôrdo com o crescimento vegetativo, isto é DATA linha 1 multiplicado por DATA linha 2. Os novos valôres substituem os anteriores em DATA linha 1. Note-se que (por exemplo) um aumento de 1% de 600 (a população total) é obtido multiplicando-se 600 por 1.01.

4.ª Etapa: Se o número estipulado em DATA linha 4 é maior do que o número de anos (KK) que passaram, o índice de crescimento vegetativo em DATA linha 2 não é modificado pelas alterações do índice em DATA linha 3. Em caso contrário o referido índice é modificado.

O resultado da operação de multiplicar um aumento vegetativo de 1.01 muitas vêzes por 0,9999 seria o de reduzir 1.01 a 1.00 (ou seja, chegar a uma situação de estabilidade da população) em 80 anos aproximadamente. Um índice de 0.99 seria demasiadamente alto. Note-se que no exemplo de EXNOVO, no primeiro ano, o índice modificador em DATA linha 3 funciona apenas em ALUM e BEND, por que em DATA linha 4, CRAY e DEMA possuem índices de 10.0 impedindo assim a operação dos índices em DATA linha 3 até o fim de um período de dez anos.

5.ª Etapa: As migrações são efetuadas em diversos estágios:

(a) M toma o valor apropriado em ITRAN (1,JJ)

(b) N toma o valor apropriado em ITRAN (2,JJ)

M e N contêm agora as posições na matriz DATA de um dado (JJ) par de lugares entre os quais (de M a N) está ocorrendo migração.

(c) A toma o valor do número de pessoas que devem migrar i.e. a população (DATA 1,M) do lugar que perde população multiplicado pelo valor que determina a proporção que sai (TRAN 1,M). No primeiro caso, é 600 (a população de ALUM) multiplicado por 0.001 (em TRAN 1, lugar M).

(d) DATA (1, ALUM) torna-se DATA (1, ALUM) menos 0.6, o valor em A.

(e) DATA (1, BEND) torna-se DATA (1, BEND) mais 0.6, o valor em A.

(f) TRAN (1) é multiplicado por TRAN (2) (nesse caso 0.001 não muda porque o valor de TRAN 2 é 1.0).

A etapa 5 repete-se para cada transação, isto é, continua NTR vêzes (no presente caso, 5 vêzes).

Quando a 5.ª Etapa estiver concluída, voltar à 1.ª Etapa e repetir NYR vêzes. (Neste exemplo NYR = 20).

PARTE 2 (i) BRASIL

1. *Divisões territoriais usadas*: 23 estados e territórios além de 9 áreas metropolitanas definidas pelo Instituto Brasileiro de Geografia e Estatística. Consideram-se como uma só unidade Acre e Roraima. O Distrito Federal (Brasília) está incluído no Estado de Goiás. [1]

Tabela 1 — *Matriz. Dados iniciais*

I — População em 1960. (Em milhares)
II — Aumento vegetativo anual inicial
III — Índice para reduzir o aumento vegetativo anual a 1.00 dentro de aproximadamente 80 anos
IV — Período de anos após o qual funcionará o índice de redução.

*ÁREAS METROPOLITANAS	População I	Mudança vegetativa II	Redutor de II III	Anos quando redutor inicia IV
1. Rondônia	71	1.030	0.9997	30
2. Acre/Roraima	189	1.030	0.9997	30
3. Amazonas	721	1.034	0.99965	30
4. Pará	1.114	1.030	0.9997	30
5. *Belém**	437	1.025	0.99975	10
6. Amapá	69	1.030	0.9997	30
7. Maranhão	2.492	1.030	0.9997	30
8. Piauí	1.263	1.035	0.99965	30
9. Ceará	2.725	1.040	0.9996	30
10. *Fortaleza**	613	1.025	0.99975	10
11. Rio Grande do Norte	1.157	1.033	0.99965	20
12. Paraíba	2.018	1.033	0.99965	20
13. Pernambuco	2.976	1.032	0.9997	20
14. *Recife**	1.161	1.025	0.99975	10
15. Alagoas	1.271	1.034	0.99965	20
16. Segipe	760	1.034	0.99965	20
17. Bahia	5.316	1.030	0.9997	20
18. *Salvador**	674	1.020	0.9998	10
19. Minas Gerais	8.902	1.032	0.9997	20
20. *Belo Horizonte**	897	1.020	0.9998	10

[1] Vide figura 2 da pág. 82.

*ÁREAS METROPOLITANAS	População I	Mudança vegetativa II	Redutor de II III	Anos quando redutor inicia IV
21. Espírito Santo	1.189	1.034	0.99965	20
22. Rio de Janeiro	1.678	1.030	0.9997	20
23. *Guanabara + Rio de Janeiro**	5.032	1.015	0.99985	0
24. São Paulo — Estado	8.242	1.030	0.9997	10
25. *São Paulo — Área Metropolitana**	4.773	1.015	0.99985	0
26. Paraná	3.782	1.030	0.9997	10
27. *Curitiba**	496	1.015	0.99985	0
28. Santa Catarina	2.147	1.073	0.9996	10
29. Rio Grande do Sul	4.385	1.025	0.99975	10
30. *Porto Alegre**	1.064	1.015	0.99985	0
31. Mato Grosso	910	1.032	0.9997	20
32. Goiás e D.F.	2.097	1.028	0.9997	20

2. ***População Inicial***: Censo do Brasil, 1960.

3. ***Premissas***:

(a) O aumento vegetativo da população (coluna II) varia como indicado. Fontes dos dados para os estudos *Subsídios*...

É de se notar o aumento vegetativo relativamente baixo (1,5% ao ano, isto é 1,015) nas quatro áreas metropolitanas da parte sul e sudeste do país e os índices geralmente mais baixos de tôdas as aglomerações urbanas. Neste modêlo, a população, ao movimentar-se de várias partes do país para as áreas urbanas, adota imediatamente hábitos de família menor.

(b) Um redutor lento (coluna III) é utilizado para reduzir o índice de crescimento vegetativo em tôdas as áreas aproximadamente a 1.00 (isto é sem alteração) em cêrca de 90 a 100 anos.

(c) Mas os dados na coluna IV asseguram que o redutor na coluna III principia a operar em alguns lugares mais tarde do que em outros. Começa imediatamente nas quatro áreas urbanas mais ao sul, mas sòmente após trinta anos, nas áreas "menos desenvolvidas" do Norte e Nordeste.

(d) A migração ocorre cntre pares selecionados de lugares (vide ITRAN) e representa sòmente uma pequena parte proporcional da população total da maior parte das áreas onde se originam os migrantes.

(e) A taxa de migração é presumida, neste exemplo, a reduzir-se sucessivamente a partir do primeiro ano, sendo multiplicada cada ano por 0,99.

Tabela 2 — Brasil, matriz de migração. Índices de migração

	5 Belém	7 Maranhão	10 Fortaleza	14 Recife	18 Salvador	20 Belo Horizonte	22 Rio de Janeiro	23 Guanabara
4. Pará	.003							
8. Piauí		.0005	.003					
9. Ceará		.0005	.001	.001				.0005
10. *Fortaleza*				.0005				.0005
11. Rio Grande do Norte			.0025	.0025				.0005
12. Paraíba			.002	.002				.001
13. Pernambuco				.0025				.0015
14. *Recife*								.0025
15. Alagoas				.0025	.0025			
16. Sergipe				.002	.003			
17. Bahia		.001			.0025			
18. *Salvador*								.0025
19. Minas Gerais							.001	.0015
20. *Belo Horizonte*						.002		.0005
22. Rio de Janeiro								.003
24. São Paulo (Estado)								
26. Paraná								
29. Rio Grande do Sul								

	24 São Paulo (Estado)	25 São Paulo (Cidade)	26 Paraná	27 Curitiba	30 Pôrto Alegre	31 Mato Grosso	32 Goiás
4. Pará							
8. Piauí							
9. Ceará	.0005	.0005	.0005				
10. *Fortaleza*		.0005					
11. Rio Grande do Norte	.001	.0005	.001				
12. Paraíba	.001	.001	.0005				
13. Pernambuco	.001	.0015	.001				
14. *Recife*		.0025					
15. Alagoas	.0015		.0015			.0005	.0015
16. Sergipe	.001		.001				.0005
17. Bahia	.0005		.0015			.0005	.001
18. *Salvador*		.0025					
19. Minas Gerais	.0005	.001	.001			.0005	.001
20. *Belo Horizonte*							
22. Rio de Janeiro	.001	.0015	.001				.001
24. São Paulo (Estado)		.001	.0025			.001	.0015
26. Paraná				.0015			
29. Rio Grande do Sul	.001		.003		.003	.0005	.0005

4. *Resultados*:

(a) A população *total* do Brasil aumenta de 3 a 4 vêzes em 50 anos, e de 6 a 7 vêzes, em 100 anos.

(b) As duas áreas metropolitanas, Guanabara e São Paulo, registram um aumento de cêrca de quatro vêzes, em 100 anos. (Em outro exemplo, com uma migração duas vêzes maior e sem redução do índice de migração registraram aumento de 10 vêzes).

(c) Guanabara, São Paulo e Pôrto Alegre têm menor proporção da população total do Brasil em 2060 do que agora. Curitiba mantém a mesma proporção, enquanto que outras áreas metropolitanas aumentaram suas proporções. A população em milhões de habitantes, em 2060 seria:

Guanabara	18	Belo Horizonte	7	Pôrto Alegre	4
São Paulo	17	Fortaleza	7	Curitiba	3
Recife	10	Salvador	6	Belém	3

(d) Os aumentos mais dramáticos observam-se no Ceará que em 2060 passa a ter mais de 8 por cento da população do Brasil em comparação com 4% em 1960, e no Maranhão, Piauí e Amazonas.

(e) Há uma drástica redução de proporção no Rio Grande do Sul.

Tabela 3 —*Resultados para o Brasil*

	POPULAÇÃO ESTIMADA EM MILHARES					
	1960	1980	2000	2020	2040	2060
1. Rondônia	70	126	225	359	508	632
2. Acre/Roraima	189	341	606	969	1 372	1 709
3. Amazonas	721	1 407	2 694	4 455	6 695	8 478
4. Pará	1 114	1 905	3 236	4 983	6 850	8 334
5. *Belém*	437	807	1 329	1 949	2 553	2 976
6. Amapá	69	125	221	354	501	624
7. Maranhão	2 492	4 734	8 729	14 332	20 731	26 221
8. Piauí	1 263	2 358	4 368	7 215	10 441	13 106
9. Ceará	2 725	5 501	11 206	19 411	29 410	37 918
10. *Fortaleza*	613	1 375	2 564	4 117	5 773	7 068
11. Rio Grande do Norte	1 157	1 914	3 022	4 239	5 261	5 751
12. Paraíba	2 018	3 370	5 359	7 563	9 434	10 352
13. Pernambuco	2 976	4 874	7 683	10 964	14 108	16 246
14. *Recife*	1 161	2 350	4 073	6 166	8 237	9 709
15. Alagoas	1 271	2 067	3 230	4 507	5 591	6 129
16. Sergipe	760	1 294	2 098	3 018	3 838	4 290
17. Bahia	5 316	8 451	12 912	17 833	22 184	24 717
18. *Salvador*	674	1 406	2 421	3 614	4 775	5 615
19. Minas Gerais	8 902	14 316	22 235	31 345	39 935	45 629
20. *Belo Horizonte*	897	1 804	3 076	4 590	6 093	7 216
21. Espírito Santo	1 189	2 321	4 208	6 634	9 091	10 772
22. Rio de Janeiro	1 678	2 952	4 874	7 151	9 334	10 791
23. *GB*/Rio	5 032	7 474	10 406	13 541	16 408	18 390
24. São Paulo (Estado)	8 242	14 177	22 033	30 536	37 715	41 394
25. *São Paulo* (Metrópole)	4 773	7 159	10 048	13 151	15 993	17 960
26. Paraná	3 782	8 406	15 415	24 005	32 306	37 736
27. *Curitiba*	496	830	1 312	1 908	2 526	3 017
28. Santa Catarina	2 147	4 344	7 624	11 403	14 533	15 778
29. Rio Grande do Sul	4 385	6 181	8 175	9 984	11 218	11 568
30. *Pôrto Alegre*	1 064	1 719	2 473	3 228	3 862	4 247
31. Mato Grosso	910	2 289	4 760	8 252	12 256	15 658
32. Goiás DF	2 097	4 144	7 386	11 396	15 350	18 042
TOTAL (milhões)	71	122	200	293	384	448

	POPULAÇÃO POR MIL DA POPULAÇÃO TOTAL					
	1960	1980	2000	2020	2040	2060
1. Rondônia	1	1	1	1	1	1
2. Acre/Roraima	3	3	3	3	4	4
3. Amazonas	10	12	14	16	17	19
4. Pará	16	16	16	17	18	19
5. *Belém*	6	7	7	7	7	7
6. Amapá	1	1	1	1	1	1
7. Maranhão	35	39	43	49	54	59
8. Piauí	18	19	22	25	27	29
9. Ceará	39	45	55	66	77	85
10. *Fortaleza*	9	11	13	14	15	16
11. Rio Grande do Norte	16	16	15	14	14	13
12. Paraíba	20	28	27	26	25	23
13. Pernambuco	42	40	39	37	37	36
14. *Recife*	16	19	20	21	21	22
15. Alagoas	18	17	16	15	15	14
16. Sergipe	11	11	11	10	10	10
17. Bahia	75	69	65	61	58	55
18. *Salvador*	10	11	12	12	12	13
19. Minas Gerais	126	117	112	107	104	102
20. *Belo Horizonte*	13	15	15	16	16	16
21. Espírito Santo	17	19	21	23	24	24
22. Rio de Janeiro	24	24	24	24	24	24
23. *GB/Rio*	71	61	53	46	43	41
24. São Paulo (Estado)	117	116	111	104	98	93
25. *São Paulo* (Metrópole)	68	58	51	45	42	40
26. Paraná	54	69	77	82	84	84
27. *Curitiba*	7	7	7	7	7	7
28. Santa Catarina	30	35	38	39	38	35
29. Rio Grande do Sul	62	51	41	34	29	26
30. *Pôrto Alegre*	15	14	12	11	10	9
31. Mato Grosso	13	19	24	28	32	35
32. Goiás DF	30	34	37	39	40	40
TOTAL (milhões)	1 000	1 000	1 000	1 000	1 000	1 000

PARTE 2 (ii) PERU

1. *Divisão Territorial*: 24 departamentos.

2. *População Inicial*: Censo de 1961. (Vide Col. I do quadro 4)

3. *Suposições*:

(a) Crescimento vegetativo da população (vide col. II do quadro 4) varia como indicado. O aumento anual médio para o país foi considerado ser de 2,5% por ano (isto é 1.025). O índice de cada departamento foi modificado, por falta de dados mais relevantes, segundo a "população de mães" por cem pessoas (*VI Censo Nacional de Población*, Tomo II, e Cuadro n.º 45).

(b) Foi usado um índice de 0.9975 (Coluna III) que reduz pouco a pouco o valor do aumento vegetativo, de modo que depois de 90-100 anos chega a ser de 1.00 (posição de não-alteração).

Tabela 4 — *Peru, entrada*

	I	II	III RedutorII	IV Quando começa	1 Amazonas	2 Ancash	3 Apurimac	4 Arequipa	5 Ayacucho
1. Amazonas	118	1.026	0.9975	0.0					
2. Ancash	583	1.026		0.30					
3. Apurimac	288	1.028		0.30				.0002	.0002
4. Arequipa	389	1.026		0.0					
5. Ayacucho	411	1.028		0.30				.0002	
6. Cajamarca	747	1.026		0.30	.0004	.0001			
7. Callao	214	1.024		0.0					
8. Cuzco	612	1.026		0.30			.0001	.0006	
9. Huancavelica	303	1.028		0.30					.0001
10. Huanuco	329	1.026		0.30					
11. Ica	256	1.024		0.0				.0002	
12. Junin	521	1.026		0.30					
13. La Libertad	582	1.026		0.0		.0008			
14. Lambayeque	342	1.024		0.0	.0001	.0001			
15. Lima	2.031	1.024		0.0		.0001			
16. Loreto	337	1.024		0.0					
17. Madre de Dios	15	1.022		0.0					
18. Moquegua	52	1.026		0.0				.002	
19. Pasco	138	1.026		0.30					
20. Piura	669	1.024		0.0	.0001	.0001			
21. Puno	686	1.028		0.30				.0012	
22. San Martin	162	1.024		0.0	.0001				
23. Tacna	66	1.024		0.0				.0004	
24. Tumbes	56	1.022		0.0					

(c) Os dados da coluna IV asseguram, contudo, que o índice que produz a redução não funciona antes de trinta anos nos departamentos da Serra. Assim, a idéia de reduzir o número de filhos não alcança os Andes antes de 1990.

(d) Há migração entre pares selecionados de departamentos (ver colunas 1 a 24). A proporção de pessoas que saem de cada departamento para outro departamento foi determinada na base dos dados do quadro n.º 27 (*VI Censo Nacional de População,* Tomo II, Población migrante... por lugar de nascimiento). Representa, assim, o padrão de migração no Peru aproximadamente entre 1910 e 1960.

(e) O índice de migração permanece inalterado neste exemplo.

4. *Resultados*:

(a) A população total do Peru aumenta aproximadamente 3 vêzes entre 1960 e 2010 e cinco vêzes entre 1960 e 2060.

(b) Lima e Callao teriam 12 1/2 milhões de habitantes em 2060, com 25% da população total do país.

(c) Feitas neste exemplo duas suposições (i) os departamentos da Serra têm um aumento vegetativo mais alto do que o dos departamentos da Costa e (ii) seu índice de aumento não começa a diminuir senão depois dos primeiros trinta anos; é evidente que a proporção da população do país que vive na serra aumentou, enquanto que a proporção que vive na região costeira fora de Lima, diminui.

Tabela 5 — *Continuação da tabela 4 — matriz de migração*

	6 Cajamarca	7 Callao	8 Cuzco	9 Huancavelica	10 Huanuco	11 Ica	12 Junin	13 La Libertad	14 Lambayeque
1. Amazonas	.0002							.0002	
2. Ancash		.0003			.0001	.0001		.0002	
3. Apurimac		.0002	.0008			.0004	.0003		
4. Arequipa		.0004	.0003			.0002	.0001		
5. Ayacucho		.0002		.0002		.0008	.0002		
6. Cajamarca								.001	.0007
7. Callao									
8. Cuzco									
9. Huancavelica						.0006	.0016		
10. Huanuco							.0002		
11. Ica		.0004							
12. Junin		.0002		.0002					
13. La Libertad	.0003	.0002							.0001
14. Lambayeque	.0003	.0002						.0003	
15. Lima		.0004				.0001	.0002		
16. Loreto					.0002				
17. Madre de Dios									
18. Moquegua									
19. Pasco					.0004		.0001		
20. Piura	.0003	.0003						.0001	.0003
21. Puno			.0004						
22. San Martin					.0005			.0001	
23. Tacna		.0004							
24. Tumbes		.0004							

	15 Lima	16 Loreto	17 Madre de Dios	18 Moquegua	19 Pasco	20 Piura	21 Puno	22 San Martin	23 Tacna	24 Tumbes
1. Amazonas	.001	.0002						.0004		
2. Ancash	.0035									
3. Apurimac	.003				.0001					
4. Arequipa	.003			.0002			.0002		.0002	
5. Ayacucho	.0035									
6. Cajamarca	.001					.0001				
7. Callao	.004									
8. Cuzco	.0009		.0001				.0001			
9. Huancavelica	.0018									
10. Huanuco	.0018	.0002			.0004					
11. Ica	.0035									
12. Junin	.0025				.0004					
13. La Libertad	.002					.0001				
14. Lambayeque	.002					.0002				
15. Lima										
16. Loreto	.001							.0002		
17. Madre de Dios	.001	.001								
18. Moquegua	.0015								.0008	
19. Pasco	.002									
20. Piura	.001									.0002
21. Puno	.0004			.0002					.0004	
22. San Martin	.001	.002								
23. Tacna	.002			.0008						
24. Tumbes	.0012					.0025				

Tabela 6 — *Resultados para o Peru*

	POPULAÇÃO EM MILHARES					
	1961	1981	2001	2021	2041	2060
Amazonas	118	193	287	385	469	516
Ancash	583	922	1 437	2 045	2 628	3 035
Apurimac	288	453	702	997	1 281	1 482
Arequipa	389	612	877	1 149	1 377	1 507
Ayacucho	411	647	1 006	1 431	1 841	2 132
Cajamarca	747	1 184	1 849	2 637	3 400	3 941
Callao	214	361	549	755	939	1 053
Cuzco	612	1 006	1 632	2 419	3 243	3 904
Huancavelica	303	491	784	1 147	1 518	1 804
Huanuco	329	530	841	1 221	1 601	1 887
Ica	256	390	543	690	801	851
Junin	521	857	1 389	2 058	2 754	3 307
La Libertad	582	897	1 254	1 592	1 837	1 926
Lambayeque	342	515	703	874	988	1 017
Lima	2 031	3 524	5 490	7 716	9 798	11 197
Loreto	337	515	711	889	1 006	1 032
Madre de Dios	15	23	32	42	50	54
Moquegua	52	84	124	166	202	224
Pasco	138	227	369	547	734	884
Piura	669	987	1 318	1 594	1 745	1 734
Puno	686	1 136	1 855	2 770	3 743	4 541
San Martin	162	234	306	363	390	381
Tacna	66	106	155	208	254	283
Tumbes	56	80	103	121	128	124
Peru						
TOTAL	9 907	15 976	24 318	33 815	42 726	48 815

	PERCENTAGEM					
	1961	1981	2001	2021	2041	2060
Amazonas	1.2	1.2	1.2	1.1	1.1	1.1
Ancash	5.9	5.8	5.9	6.0	6.2	6.2
Apurimac	2.9	2.8	2.9	2.9	3.0	3.0
Arequipa	3.9	3.8	3.6	3.4	3.2	3.1
Ayacucho	4.1	4.1	4.1	4.2	4.3	4.4
Cajamarca	7.5	7.4	7.7	7.8	8.0	8.1
Callao	2.2	2.3	2.2	2.2	2.2	2.2
Cuzco	6.2	6.3	6.8	7.2	7.6	8.0
Huancavelica	3.1	3.1	3.2	3.4	3.6	3.7
Huanuco	3.3	3.3	3.4	3.6	3.7	3.9
Ica	2.6	2.4	2.2	2.0	1.9	1.7
Junin	5.3	5.4	5.7	6.1	6.4	6.8
La Libertad	5.9	5.6	5.2	4.7	4.3	3.9
Lambayeque	3.5	3.2	2.9	2.6	2.3	2.1
Lima	20.5	22.1	22.6	22.8	22.9	22.9
Loreto	3.4	3.2	2.9	2.6	2.4	2.1
Madre de Dios	0.2	0.1	0.1	0.1	0.1	0.1
Moquegua	0.5	0.5	0.5	0.5	0.5	0.5
Pasco	1.4	1.4	1.5	1.6	1.7	1.8
Piura	6.8	6.2	5.4	4.7	4.1	3.6
Puno	6.9	7.1	7.6	8.2	8.8	9.3
San Martin	1.6	1.5	1.3	1.1	0.9	0.8
Tacna	0.7	0.7	0.6	0.6	0.6	0.6
Tumbes	0.6	0.5	0.4	0.4	0.3	0.3
Peru						
TOTAL	100.0	100.0	100.0	100.0	100.0	100.0

PARTE 2 (iii) PORTUGAL

1. Divisões territoriais usadas: dezoito distritos com a modificação seguinte:

Lisboa C. contém os concelhos seguintes:

Lisboa, Loures, Vila Franca de Xira, Oeiras, Cascais; Almada, Barreiro, Moita, Seixal.

Pôrto C. contém os concelhos seguintes:

Pôrto, Maia, Vila Nova de Gaia, Matozinhos.

As áreas 12, 16 e 20 contêm os restantes concelhos dos respectivos distritos.

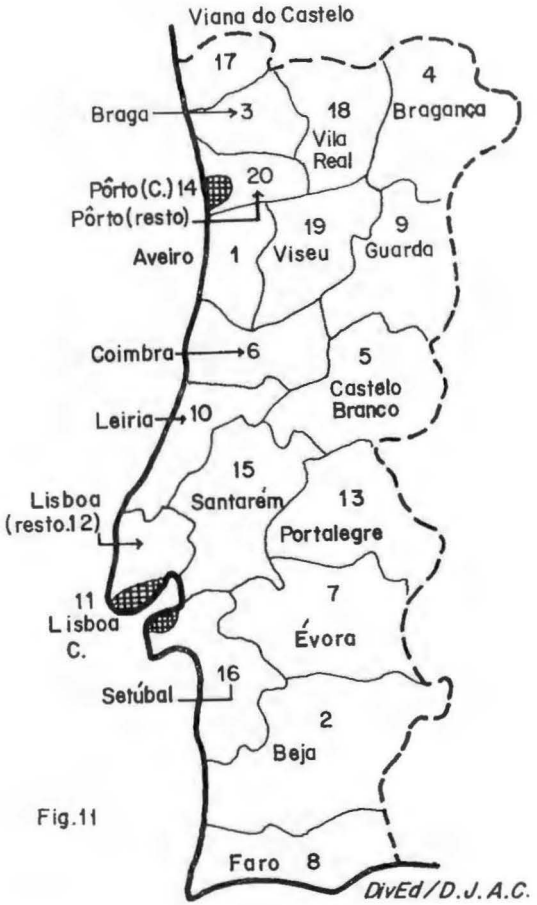


Fig.11

2. População inicial: Censo de 1960.

3. Suposições:

(a) Os dados sôbre o índice de aumento vegetativo (coluna II da matriz) baseiam-se no índice para meados do decênio 1960-1970. O índice para os distritos Lisboa e Pôrto foi ligeiramente modificado de forma que os concelhos menos próximos dessas cidades tivessem índices relativamente mais elevados do que as próprias cidades.

(b) Usou-se um índice (coluna III da matriz) para reduzir muito lentamente o aumento vegetativo de 4 distritos (1, 3, 14, 20) a metade dentro de cêrca de 100 anos. Assim, por exemplo, 0.9999 reduz 1.020 (um aumento vegetativo de 2% por ano) a 1.010.

(c) Neste exemplo o redutor começa a atuar imediatamente porque os valôres da coluna IV são 0.0.

(d) As migrações entre distritos encontram-se no quadro 8. Dispunha-se apenas de dados sôbre a população que emigrou para o distrito em 1960. Admitiram-se, a partir dêstes dados, as direções e índices de migração usados neste exemplo. Os principais centros de imigração são Lisboa e Pôrto.

(e) Empregou-se um redutor lento de migração de 0.99 em tôdas as transações.

Resultados:

(a) A população quase triplicou durante o período de cem anos.

(b) Dentro dos cem anos Lisboa e Pôrto terão respectivamente 4.5 e 2.5 milhões de habitantes. Quer dizer, as duas áreas em conjunto representam 23% da população total em 1960, e 30% em 2060.

(c) Em áreas rurais de forte emigração como Bragança, Beja, Évora, Portalegre a percentagem da população dêstes distritos em relação à população total diminui progressivamente desde 1960 — 2060.

Tabela 7 — *Matriz de dados iniciais*

	I População inicial em milhares	II Aumento vegetativo	III Índice que modifica II	IV Ano no qual começa a atuar III
1. Aveiro	522.2	1.015	0.99993	0.0
2. Beja	268.9	1.005	1.0	0.0
3. Braga	593.6	1.021	0.9999	0.0
4. Bragança	230.3	1.009	1.0	0.0
5. Castelo Branco	310.7	1.006	1.0	0.0
6. Coimbra	433.6	1.008	1.0	0.0
7. Évora	215.2	1.006	1.0	0.0
8. Faro	312.5	1.005	1.0	0.0
9. Guarda	276.5	1.006	1.0	0.0
10. Leiria	400.3	1.009	1.0	0.0
11. Lisboa (C)	1 271.0	1.008	1.0	0.0
12. Lisboa (D–C)[1]	285.9	1.010	1.0	0.0
13. Portalegre	183,8	1.004	1.0	0.0
14. Pôrto (C)	607.0	1.016	0.99993	0.0
15. Santarém	462.1	1.007	1.0	0.0
16. Setúbal (D–C)[2]	221.8	1.008	1.0	0.0
17. Viana do Castelo	275.3	1.011	1.0	0.0
18. Vila Real	322.6	1.013	1.0	0.0
19. Viseu	477.5	1.011	1.0	0.0
20. Pôrto (D–C)[2]	584.5	1.020	0.9999	0.0

(D–C)[1] significa população do distrito menos população de concelhos escolhidos.

(D–C)[2] significa população do distrito menos população dos concelhos incluídos em 11 (Lisboa C).

(C) significa concelho da cidade mais certos conselhos vizinhos.

Tabela 8 — Portugal. Índices de migração entre pares de distritos

PORTUGAL ÍNDICES DE MIGRAÇÃO ENTRE PARES DE DISTRITOS

	1 Aveiro	3 Braga	6 Coimbra	8 Faro	9 Guarda	10 Leiria	11 Lisboa (C)
1. Aveiro			0.0009			0.0004	
2. Beja				0.0002			0.002
3. Braga			0.0004				0.001
4. Bragança		0.001	0.0004		0.0013		0.001
5. Castelo Branco	0.001						0.002
6. Coimbra						0.0004	0.001
7. Évora							0.0024
8. Faro							0.0013
9. Guarda	0.0011					0.0003	0.001
10. Leiria			0.0005				0.0014
11. Lisboa (C)							
12. Lisboa (D–C)				0.00018		0.0006	0.0018
13. Portalegre							0.0023
14. Pôrto (C)							0.0016
15. Santarém (D–C)			0.0004			0.0006	0.005
16. Setúbal							0.005
17. Viana do Castelo		0.001					0.001
18. Vila Real	0.0009	0.0009			0.0007		0.0016
19. Viseu	0.002	0.0007	0.0009		0.001	0.0006	0.0013
20. Pôrto (D–C)							0.0016

PORTUGAL ÍNDICES DE MIGRAÇÃO ENTRE PARES DE DISTRITOS

	12 Portalegre	14 Pôrto (C)	15 Santarém	16 Setúbal	19 Viseu	20 Pôrto
1. Aveiro		0.002				
2. Beja	0.0004			0.0015		
3. Braga	0.0003	0.001				0.001
4. Bragança	0.0004	0.0005			0.002	0.0004
5. Castelo Branco	0.0004	0.0005		0.0003		
6. Coimbra		0.0005				
7. Évora	0.0002			0.001		
8. Faro		0.0005		0.0003		
9. Guarda	0.0003	0.0005			0.002	
10. Leiria		0.0005	0.002			
11. Lisboa (C)						
12. Lisboa (D–C)			0.003			
13. Portalegre	0.0002			0.0006		
14. Pôrto (C)						
15. Santarém (D–C)		0.0005		0.001		
16. Setúbal						
17. Viana do Castelo	0.0003	0.0016				0.0004
18. Vila Real	0.0004	0.0008			0.0015	0.0002
19. Viseu		0.0006				0.0003
20. Pôrto (D–C)						

Tabela 9 — *Resultados para Portugal*

	POPULAÇÃO ESTIMADA EM MILHARES					
	1960	1980	2000	2020	2040	2060
1. Aveiro	522.2	692.5	886.0	1 099.7	1 327.7	1 561.1
2. Beja	268.9	275.7	286.6	301.2	319.5	341.4
3. Braga	593.6	915.9	1 316.0	1 793.5	2 344.4	2 939.6
4. Bragança	230.3	242.5	261.4	287.1	320.3	361.9
5. Castelo Branco	310.7	324.4	343.4	367.8	397.5	433.0
6. Coimbra	433.6	525.7	633.9	760.8	909.6	1 083.7
7. Évora	215.2	227.1	242.6	261.7	284.6	311.5
8. Faro	312.5	334.4	360.0	389.4	422.9	460.8
9. Guarda	276.5	302.5	332.4	367.1	407.3	453.6
10. Leiria	400.3	466.5	545.8	640.9	755.0	891.5
11. Lisboa (C)	1 271.0	1 744.5	2 292.2	2 928.6	3 668.9	4 530.2
12. Lisboa (D–C)	285.9	333.3	391.9	464.1	552.7	661.2
13. Portalegre	183.8	188.1	194.6	202.9	213.1	225.1
14. Pôrto (C)	607.0	888.2	1 232.0	1 638.8	2 094.5	2 589.6
15. Santarém	462.1	496.2	541.8	599.3	669.5	753.4
16. Setúbal	221.8	264.1	310.7	363.3	423.5	493.2
17. Viana do Castelo	275.3	316.8	369.9	436.8	520.8	625.9
18. Vila Real	322.6	367.7	428.9	509.9	615.7	753.0
19. Viseu	477.5	491.1	524.4	576.7	649.2	744.6
20. Pôrto (D–C)	584.5	853.0	1 194.0	1 604.9	2 072.4	2 571.5
TOTAL	8 255	10 250	12 688	15 594	18 969	22 785

	POPULAÇÃO POR CEM DA POPULAÇÃO TOTAL					
	1960	1980	2000	2020	2040	2060
1. Aveiro	6.3	6.8	7.0	7.1	7.0	6.9
2. Beja	3.3	2.7	2.3	1.9	1.7	1.5
3. Braga	7.2	8.9	10.4	11.5	12.4	12.9
4. Bragança	2.8	2.4	2.1	1.8	1.7	1.6
5. Castelo Branco	3.8	3.2	2.7	2.4	2.1	1.9
6. Coimbra	5.3	5.1	5.0	4.9	4.8	4.8
7. Évora	2.6	2.2	1.9	1.7	1.5	1.4
8. Faro	3.8	3.3	2.8	2.5	2.2	2.0
9. Guarda	3.3	3.0	2.6	2.4	2.1	2.0
10. Leiria	4.8	4.6	4.3	4.1	4.0	3.9
11. Lisboa (C)	15.4	17.0	18.1	18.8	19.3	19.9
12. Lisboa (D–C)	3.5	3.3	3.1	3.0	2.9	2.9
13. Portalegre	2.2	1.8	1.5	1.3	1.1	1.0
14. Pôrto (C)	7.4	8.7	9.7	10.5	11.0	11.4
15. Santarém	5.6	4.8	4.3	3.8	3.5	3.3
16. Setúbal	2.7	2.6	2.4	2.3	2.2	2.2
17. Viana do Castelo	3.3	3.1	2.9	2.8	2.7	2.7
18. Vila Real	3.9	3.6	3.4	3.3	3.2	3.3
19. Viseu	5.8	4.8	4.1	3.7	3.4	3.3
20. Pôrto (D–C)	7.1	8.3	9.4	10.3	10.9	11.3
TOTAL	100	100	100	100	100	100

FUNDAÇÃO IBGE
SERVIÇO GRÁFICO