

Tipologia socioeconómica dos municípios continentais pela análise das correspondências

RESULTADOS DA ANÁLISE FACTORIAL DAS CORRESPONDÊNCIAS

1) INTRODUÇÃO

O trabalho que aqui se apresenta é o estudo de Portugal continental descrito por 25 variáveis sociais e económicas, utilizando a técnica de análise de dados conhecida por *análise factorial das correspondências* (AFC). As unidades de observação são os 274 concelhos (anexo 1) e os indicadores¹ (anexo 2) descrevem as classes sociais, a organização económica e alguns aspectos sanitários e de saúde de cada um daqueles concelhos.

Apesar de os dados utilizados serem todos anteriores ao processo de transformação iniciado em 25 de Abril de 1974, a análise mostra claramente as principais estruturas económicas actuais e a sua arrumação no País, assim como a distribuição das diferentes classes e camadas sociais no território nacional.

As 25 variáveis dão origem a dois factores que, só por si, explicam 63 % da informação total: a acumulação de riqueza social e as relações de produção na agricultura. Em função destes factores é possível definir regiões homogéneas, resultantes do agrupamento dos concelhos e que se distinguem entre si pela importância relativa das estruturas económicas.

Não consideramos, no entanto, ter obtido com este trabalho o seu objectivo final, pois não dispusemos para tal nem dos meios de recolha de dados

¹ O grande número de concelhos obrigou a diminuir drasticamente o número de variáveis para os dados poderem ser tratados em computador.

Se tivéssemos de retomar este trabalho, aproveitando os resultados e a experiência aqui adquiridos, agruparíamos os concelhos — por semelhança socioeconómica e proximidade territorial —, eliminaríamos variáveis que se revelaram pouco significativas e desdobraríamos sistematicamente os indicadores retidos (por exemplo, ao lado de ISA — percentagem de isolados poríamos NISA — percentagem de não isolados; ao lado de CIC e CIA usaríamos CIB = 100 - CIC - CIA; etc.).

nem do tempo de que necessitávamos. Muito ficou ainda por fazer, tratando-se, como é evidente, de um estudo exploratório em que eram essenciais os aspectos metodológicos.

A importância do problema e os resultados já obtidos animam-nos por isso a prosseguir este estudo logo que isso nos seja possível.

2) O MÉTODO DE AFC

Fazemos aqui uma breve introdução destinada aos leitores (não necessariamente matemáticos) não familiarizados com a AFC. Recomendamos as referências bibliográficas [1], [2] e [3] a quem estiver interessado no desenvolvimento matemático ou em pormenores práticos mais completos deste método estatístico de análise de dados multidimensionais.

A AFC é um método de análise factorial inicialmente delineado para tirar partido da estrutura probabilística de certos problemas. De facto, ela surgiu para o estudo do cruzamento de duas partições sobre uma mesma amostra (por exemplo, o estudo simultâneo sobre uma amostra da variável «classificação socioprofissional» e da variável «região-distrito em que habita» conduz à construção de um quadro de contingência ao qual se pode aplicar a AFC). A AFC foi entretanto tendo sucessivas generalizações e o seu campo de acção foi ampliado ao tratamento de dados descritos por variáveis positivas, variáveis binárias, variáveis ordinais, etc.

	$J \rightarrow$	1	...	j	...	p	
$I \downarrow$	1	K_{11}	...	K_{1j}	...	K_{1p}	$K_{1.}$
:	:	⋮		⋮		⋮	
i	K_{i1}	...	K_{ij}	...	K_{ip}	$K_{i.}$	
:	:	⋮		⋮		⋮	
n	K_{n1}	...	K_{nj}	...	K_{np}	$K_{n.}$	
	$K_{.1}$...	$K_{.j}$...	$K_{.p}$		

A AFC é hoje aplicada com êxito a todo o quadro de dados K de números positivos estabelecendo uma correspondência entre dois conjuntos, I e J , que mutuamente se descrevem.

Note-se que não é, afinal, muito restritiva a exigência de K conter apenas números positivos. Essa condição, que se prende com o desenvolvimento teórico da AFC a partir dos quadros de contingência, pode ser cumprida recorrendo a operações simples. De facto, é sempre possível reduzir um quadro de dados com números positivos e negativos a outro só com números positivos mediante o uso de uma ou de todas as operações seguintes: produto por uma constante, soma de uma constante (a aplicação sucessiva das duas provoca uma transformação linear dos dados) e desdobramento das variáveis.

A AFC repousa sobre a codificação e a escolha de métrica seguintes:

Cada elemento é representado pelo seu perfil condicionado (obtido dividindo os elementos de uma fila pela soma total dos elementos dessa fila);

A cada elemento é associada uma massa que é a frequência marginal correspondente;

A métrica do espaço é a distância do χ^2 .

Estas três escolhas (perfis, massas, métrica do χ^2) formalizam afinal três condições importantes:

Dois elementos de I (resp. J) são tanto mais próximos quanto mais semelhantes são os seus perfis condicionais de se associarem sobre J (resp. I);

Dois elementos com iguais leis condicionais, isto é, associados a filas proporcionais, podem ser substituídos por um só elemento cuja massa é a soma das massas, isto é, por um elemento que corresponde a uma fila que é soma das filas (princípio da equivalência distribucional);

A distância só depende das leis condicionais e das leis marginais.

No nosso caso, I é o conjunto dos 274 concelhos do continente ($n = 274$) e J é o conjunto das 25 variáveis socioeconómicas retidas ($p = 25$). O quadro K de partida é formado de números positivos e, no cruzamento da linha i com a coluna j , o valor $K(i,j)$ é o valor assumido pela variável j sobre o concelho i . Do ponto de vista da AFC, cada concelho i passa a ser descrito por um vector que designaremos por

$$f_j^i = \left\{ f_j^i \mid f_j^i = \frac{K(i,j)}{K(i)}, K(i) = \sum_j K(i,j), j = 1, \dots, 25 \right\}$$

e a sua massa é

$$f_i = K(i)/K, K = \sum_{i,j} K_{ij}$$

e o conjunto dos 274 concelhos constitui uma nuvem $N(I)$ de pontos com certa configuração no espaço 25 — dimensional R^{25} .

Dualmente, cada variável j está associada a um vector

$$f_1^i = \left\{ f_1^i \mid f_1^i = \frac{K(i,j)}{K(j)}, K(j) = \sum_i K(i,j), i = 1, \dots, 274 \right\}$$

de massa

$$f_j = K(j)/K, K = \sum_{i,j} K(i,j)$$

e o conjunto das variáveis constituem uma nuvem $N(J)$ mergulhada em R^{274} (a 274 dimensões).

A distância entre os elementos i e i' de $N(I)$, por exemplo, é tomada em relação ao *centro* da nuvem $N(I)$:

$$f_j = \left\{ f_j \mid f_j = \frac{K(j)}{K}, j = 1, \dots, 25 \right\}$$

e a sua fórmula é:

$$d^2(i,i') = \|f_j^i - f_j^{i'}\|^2 f_j = \sum_{j=1}^{25} \frac{(f_j^i - f_j^{i'})^2}{f_j}$$

(deste modo, dois perfis iguais, ou seja duas linhas proporcionais, produzirão uma distância nula).

Um dos objectivos de uma análise factorial é reduzir a um número dígito de dimensões, guardando o máximo de informação (isto é, mantendo tanto quanto possível as distâncias originais entre pares de elementos), a representação da nuvem que é dada inicialmente num espaço de dimensão elevada.

Sem entrar em detalhes², a análise factorial tem como efeito o seguinte: substituir o quadro de perfis inicial por dois outros quadros, a que chamaremos F e G — os quadros de factores; se designarmos por A o conjunto de índices dos factores, cada elemento $i \in I$ é descrito por

$$F_A(i) = \{F_\alpha(i) \mid \alpha \in A\}$$

e cada $j \in J$ é descrito por

$$G_A(j) = \{G_\alpha(j) \mid \alpha \in A\}$$

Para cada factor $\alpha \in A$, o par de vectores (F_α, G_α) , o primeiro com n componentes e o segundo com p componentes, é afinal o par de vectores de coordenadas de I e de J nos eixos factoriais associados em cada um dos espaços de $N(I)$ e $N(J)$ ao factor α .

² É o estudo dos invariantes simultâneos (vectores e valores próprios das formas quadráticas associadas à distância e às massas) que resolve o problema da análise factorial. Os vectores próprios unitários vão ser os versores das direcções do espaço associados aos factores; os valores próprios, λ_α , vão medir a importância dos factores.

A cada factor α está associado o valor λ_α , que representa a inércia da nuvem projectada na direcção associada ao factor; isto é, λ_α mede a fidelidade com que o eixo factorial associado respeita as distâncias originais.

Os factores são extraídos na ordem decrescente de λ_α , ou seja, na ordem da sua importância, e são não correlacionados (quer dizer, são ortogonais os eixos correspondentes).

Os λ_α medem simultaneamente a importância absoluta e relativa do factor α : λ_α é tanto maior quanto mais violentos são os contrastes pelo factor postos em evidência; e, pela sua posição na sucessão dos valores próprios: $\lambda_1 > \lambda_2 > \lambda_3 > \dots$, posiciona o factor em relação aos outros.

Sublinharemos ainda como importantes os seguintes aspectos da AFC:

- É possível interpretar, ao nível de cada factor, cada elemento de uma nuvem como baricentro dos pontos da outra nuvem, afectados de certas massas (fórmula de transição de AFC) — o que dá sentido à representação simultânea;
- É possível introduzir elementos suplementares, isto é, elementos que não intervieram na análise, mas cuja importância como indicadores dos factores ou de certos grupos os torna imprescindíveis à interpretação dos resultados;
- É possível reconstituir o quadro inicial de dados a partir dos quadros de novas coordenadas, completos, F e G , e dos momentos de inércia λ_α .

A fidelidade de um subespaço formado por um pequeno número de factores, como resumo da informação contida no quadro K inicial de dados, pode assim ser medida pelo desvio entre K e o quadro reconstituído a partir dos factores escolhidos.

3) RESULTADOS DA AFC

É a seguinte a sucessão dos cinco primeiros valores próprios λ_α e respectivas taxas de inércia ε_α [nota: $\varepsilon_\alpha = \lambda_\alpha / (\sum \lambda_\alpha)$]:

α	1	2	3	4	5
λ_α	0,1249	0,0435	0,0204	0,0144	0,0135
ε_α	46,7 %	16,2 %	7,6 %	5,4 %	5,0 %

O valor λ_1 aqui obtido é um valor «médio» quando comparado com a gama de valores possíveis (qualquer valor próprio deve estar entre 0 e 1) e com os valores próprios obtidos noutros estudos (argumento da experiência adquirida): valores da ordem de 0,01 são comuns em dados de mensuração, como, por exemplo, em taxonomia animal ou vegetal; por outro lado, valores superiores a 0,5 ou 0,6 correspondem a dicotomias mais ou menos perfeitas ao nível dos dados: quadro inicial com dois grupos de

linhas e, correspondentemente, dois grupos de colunas, decomposição da matriz em blocos. Aqui, um valor $\lambda_1 = 0,1249$ é um valor médio, que garante não existir uma dicotomia brutal ao nível dos dados, mas antes sugere a existência de um contínuo entre posições extremas, uma deformação «contínua» quando se passa de uma forma representada no extremo do primeiro eixo à forma existente no outro extremo. A mesma conclusão é aplicável aos factores seguintes, como é evidente.

Com os dois primeiros factores obtemos 62,9 % da inércia totalizada pelo conjunto de dados, isto é, praticamente $\frac{2}{3}$ da variabilidade total do sistema é resumida pela representação no plano factorial (1,2). Dada a quebra das taxas do segundo para o terceiro factores, é de esperar que o primeiro e o segundo factores tenham interpretações claras. É o que veremos a seguir.

3.1 OS FACTORES

O primeiro factor

O primeiro factor representa 46,7 % da inércia total.

O gráfico (ver página 57) figura as variáveis socioeconómicas no primeiro eixo factorial.

As variáveis sociais e económicas que mais contribuem para a formação deste factor (maior contribuição absoluta)³ e deste recebem uma parte

³ Tentemos aqui uma explicação breve, e portanto incompleta, dos conceitos de contribuição absoluta e contribuição relativa, que constituem ajudas à interpretação muito importantes para uma AFC. Uma direcção factorial α é tanto mais importante para a descrição de uma nuvem de pontos com massas

$$\{ (P_i, m_i) \mid i = 1, \dots, n \}$$

quanto maior for a inércia em relação à origem, λ_α , da nuvem projectada sobre essa direcção. Se representarmos por

$$(F_1(i), \dots, F_\alpha(i), \dots)$$

o vector de coordenadas do ponto i sobre o referencial dos eixos factoriais $(1, \dots, \alpha_1, \dots)$, tem-se:

$$\lambda_\alpha = \frac{\sum_{i=1}^n m_i F_\alpha^2(i)}{\sum_{i=1}^n m_i} \longrightarrow 1 = \frac{\sum_{i=1}^n F_\alpha^2(i)}{\lambda_\alpha}$$

Daqui se deduz que um elemento i é tanto mais importante para a explicação de um factor quanto maior for o valor de

$$m_i \frac{F_\alpha^2(i)}{\lambda_\alpha} \times 100 = \text{percentagem de contribuição absoluta}$$

do elemento i para a formação do factor α .

[$m_i F_\alpha^2(i)$ = contribuição absoluta de i para α .]

Por outro lado, designada por ρ_1 a distância do ponto i à origem, 0, tem-se:

$$\rho_1^2 = \sum_{\alpha \in A} F_\alpha^2(i) \longrightarrow 1 = \sum_{\alpha \in A} \left(\frac{F_\alpha(i)}{\rho_1} \right)^2$$

importante da sua própria explicação (grande contribuição relativa do factor) são do lado positivo do primeiro eixo [$(G_1(j) > 0)$]:

ISA (6,4; 61,5), PAS (11,4; 52,4), CIC (6,9; 56,6), AGR (8,4; 84,7),
NWC (6,2; 47,0)

e do lado negativo [$(G_1(j) < 0)$]:

EMT (4,1; 48,3), SEC (4,9; 75,9), OTE (9,4; 74,6), CIA (14,5; 56,2)

(a sucessão de dois números a seguir à sigla de cada variável é formada pela percentagem de contribuição absoluta e contribuição relativa correspondente àquele elemento).

Assim, as cinco variáveis citadas que saem à direita da origem explicam 39,3 % do primeiro factor e as quatro do lado negativo contribuem com 32,9 %. As nove variáveis totalizam 72,2 % de λ_1 , o que já é suficiente para sobre elas basear a interpretação.

Notemos, entretanto, que à direita da origem são ainda colocadas por este factor algumas variáveis que, tendo embora uma contribuição absoluta mais fraca, recebem ainda uma contribuição relativa razoável:

FAM (2,5; 31,9), PAI (3,5; 51,2), PPM (1,8; 17,7)

enquanto à esquerda da origem se projectam, nas mesmas condições, os indicadores seguintes

ELE (2,7; 30,1), MIG (1,0; 36,0), TER (2,4; 47,0), ETE (1,1; 60,0)

Acrescentando estes aos primeiros mencionados, obtemos 47,1 % para o grupo de variáveis com factor positivo (coordenadas $G_1 > 0$), 41,3 % para o grupo da esquerda (coordenadas negativas) e um total de 88,4 % (!) de contribuição absoluta a λ_1 para o conjunto de todas as variáveis citadas.

Antes de passar à interpretação do primeiro factor salientaremos como bons *indicadores* deste factor, isto é, como variáveis altamente correlacionadas (contribuição relativa superior a 50 %, ou seja, co-seno do ângulo superior a 0,70, em valor absoluto) as seguintes:

ISA, PAS, CIC, AGR, SEC, OTE, CIA, ETE

O primeiro factor traduz a acumulação de riqueza social

Este factor mostra no seu extremo esquerdo [$F_1(i) > 0$] os concelhos com percentagens elevadas de contribuição industrial correspondente ao

onde se conclui que a explicação de um elemento i se deve procurar na direcção ou direcções α para as quais é maior o quociente $F_\alpha^2(i)/\rho_1^2 =$ contribuição relativa da direcção α para o elemento i .

Note-se, de passagem, que $F_\alpha(i)/\rho_1 =$ co-seno do ângulo de OP_i com o eixo α .

O que está aqui exemplificado para a nuvem I aplica-se, *mutatis mutandis*, à nuvem J .

grupo A — trata-se da contribuição industrial paga por sociedades anónimas e em comandita por acções, por sociedades comerciais com capital superior a 3000 contos, por instituições de crédito e seguros, enfim, por outros contribuintes que nos três anos anteriores tenham tido média de rendimentos colectáveis igual ou superior a 300 contos (ver posição do ponto CIA no extremo negativo do primeiro eixo factorial). Os concelhos que se situam neste extremo do eixo apresentam taxas elevadas de activos no conjunto de actividades abrangidas pela construção, transportes e comunicações e indústrias transformadoras e extractivas (SEC) e, nestas últimas em especial, quer sejam empregados (ETE) quer sejam operários (OTE). As regiões situadas à esquerda da origem são ainda caracterizadas por taxas importantes de empregados (EMT), de activos no sector chamado terciário (TER), enfim, de um modo geral, de trabalhadores assalariados (ASS) e, em particular, dos que são pagos sem ser ao mês (OPT), em geral operários. Apresentam normalmente boa cobertura de electrificação (ELE) e um balanço migratório (MIG) positivo, isto é, são geralmente regiões receptoras de mão-de-obra.

Em contrapartida, as regiões correspondentes ao lado direito do primeiro factor, e em particular as que se situam mais extremadas, têm percentagens elevadas de trabalhadores familiares não remunerados (FAM), de camponeses *sem* assalariados (ISA), a quem o censo chama «isolados agrícolas»; de um modo geral, têm taxas elevadas de activos na agricultura (AGR), quer isolados, quer familiares, quer ainda assalariados (OPA). Aqui é importante a contribuição (CIC) da classe C para o total pago da contribuição industrial ao nível concelhio; lembremos que os contribuintes da classe C devem preencher simultaneamente várias condições, dentre as quais destacamos a de trabalharem sozinhos ou serem apenas auxiliados por familiares ou estranhos em número não superior a 3, se se tratar de indústria, ou a 1, se se tratar de comércio; devem utilizar, quando muito, um veículo automóvel ou dois de outra espécie; não devem ter escrita, ou tê-la tão rudimentar que não permita controlo sobre o rendimento havido; isto é, os contribuintes da classe C são sobretudo os artesãos e os pequenos comerciantes.

Quanto à estrutura agrícola, que é dominante nas regiões que aqui estamos a descrever, toma valores importantes, sobretudo para os concelhos que se situarem mais longe da origem, a taxa de empresas agrícolas autónomas (EAA) (em relação ao total de empresas agrícolas), ou seja, de empresas em que toda a mão-de-obra é familiar — sem recurso a assalariados, portanto — e em que nenhum membro da família tem actividade remunerada exterior à empresa.

Deste lado, a assistência médica é má — altas percentagens de partos sem assistência (PAS), grande número de pessoas por médico (PPM), além de ser, em geral, pior que do lado esquerdo da origem a sanidade básica (NWC).

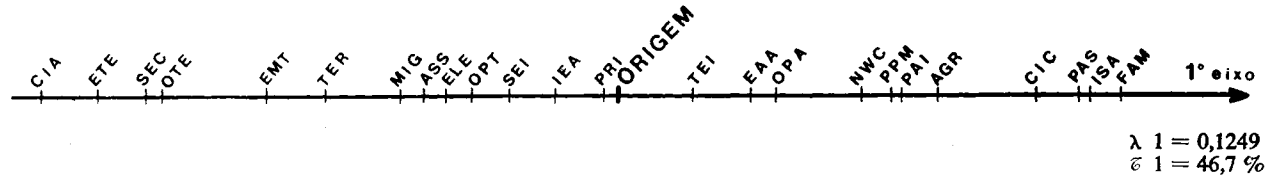
Se acrescentarmos que as regiões situadas à esquerda têm geralmente valores baixos para os indicadores socioeconómicos que se projectam à direita da origem e, inversamente, as variáveis que vão situar-se à esquerda vão assumir quase sempre valores baixos sobre os concelhos que se colocam no outro extremo do eixo, teremos uma imagem bastante nítida deste factor, que começamos por chamar *de acumulação de riqueza social*. Ele faz situar de um lado os concelhos com forte concentração de operários e de empregados e onde a importância das empresas mais desenvolvidas

Primeiro factor: acumulação de riqueza social

[GRÁFICO I]

• Variáveis que, pelos seus valores elevados, caracterizam o desenvolvimento industrial.

• Variáveis que, pelos seus valores elevados, tipificam o desenvolvimento centrado na agricultura.



Segundo factor: relações de produção na agricultura

[GRÁFICO II]

• Variáveis que, pelos seus valores elevados, caracterizam a agricultura pobre.

• Variáveis que tipificam uma agricultura mais desenvolvida.



se faz sentir (transformadoras, construção, transportes, comércio, bancos, seguros, serviços): é aqui que o desenvolvimento capitalista é maior, que é criada e acumulada a maior parte da mais-valia social; é onde temos a melhor cobertura médica e sanitária; aqui encontramos sem surpresa, entre outros, parte dos concelhos de Lisboa, Porto e Setúbal (os mais industrializados ou mais desenvolvidos) e ainda alguns de Leiria e Aveiro. Do outro lado da origem (lado \oplus) estão os concelhos essencialmente agrícolas, saindo mais estremados aqueles em que é predominante a pequena propriedade, em que se nota a ausência de grandes empresas, já que uma grande parte da contribuição industrial provém do artesão e dos pequenos comerciantes. Aqui vemos desfilar concelhos de Viseu, Vila Real, Viana do Castelo, Bragança, Guarda.

Para uma melhor visualização deste factor recomenda-se a observação atenta dos gráficos III e IV, que mostram, no primeiro plano factorial (sendo horizontal o primeiro eixo factorial), as variáveis e os distritos como elementos suplementares (gráfico III) e os concelhos (gráfico IV).

O segundo factor

Este factor concentra 16,2 % da inércia total.

Ele está sobretudo na dependência das variáveis seguintes: com coordenadas positivas [$G_2(j) > 0$],

OPA (24,3; 77,8), OPT (8,0; 46,2), NWC (8,6; 22,5), ASS (5,3; 29,5)

e com coordenadas negativas [$G_2(j) < 0$]:

PAI (4,5; 22,8), PAS (7,3; 11,6), ISA (6,9; 23,0), EAA (11,1; 36,7),
FAM (9,5; 42,2),

às quais juntaremos ainda PRI (1,5; 39,3), devido à sua contribuição relativa de quase 40 %.

Com estas dez variáveis, totalizando uma contribuição absoluta ao segundo factor igual a 81 % de λ_2 , iremos fazer a interpretação (gráficos II, pág. 57, e III, pág. 60).

O segundo factor estabelece, para a agricultura, a diferenciação entre as regiões onde predominam as relações de produção capitalistas (latifúndios) e as regiões onde predomina a pequena produção (minifúndios) ⁴.

Temos, por um lado, com a segunda coordenada positiva, as regiões caracterizadas por grandes percentagens de operários agrícolas (OPA) e com baixo nível de sanidade básica (NWC) e, por outro lado, as regiões

⁴ A variável OPA é um excelente indicador do segundo factor com a sua contribuição relativa de 77,8 %, ou seja, uma correlação com o factor de 0,88: concentração de assalariados agrícolas nos concelhos com coordenada $F_2 > 0$ e fraca taxa dos mesmos concelhos com coordenada $F_2 < 0$.

onde são muito elevadas as percentagens de trabalhadores familiares não remunerados (FAM) e de camponeses (ISA — isolados agrícolas), com percentagens relativamente importantes de empresas agrícolas familiares autónomas (EAA).

Assim, o segundo factor marca as diferenças, nas zonas onde a principal actividade é a agricultura, entre as regiões com relações de produção capitalistas, constituídas por concelhos de Beja, Évora, Portalegre, Santarém, Castelo Branco e Setúbal (zona de intervenção da Reforma Agrária) e as regiões onde a pequena economia mercantil é predominante, em que os produtores são também proprietários e por vezes trabalham a terra sem recurso a trabalho assalariado, isoladamente ou auxiliados por familiares não remunerados.

Podemos assim denominar este o *factor das relações de produção na agricultura*.

3.2 O PRIMEIRO PLANO FACTORIAL E OS TIPOS SOCIECONÓMICOS DE REGIÕES

Já sabemos que este plano dá conta de cerca de $\frac{2}{3}$ da inércia total. Como se viu no estudo dos primeiros factores, os concelhos distribuem-se: em primeiro lugar, em função do modo como neles é acumulada a riqueza social (maior ou menor industrialização, maior ou menor actividade agrícola) e, em segundo lugar, em função das formas de produção agrícolas para as regiões onde predomina esta actividade.

Renunciámos a representar simultaneamente os concelhos e as variáveis socioeconómicas, pois seria grande a acumulação de pontos. Em contrapartida, representamos graficamente, e em conjunto, a nuvem das variáveis e a nuvem dos distritos (introduzidos como elementos suplementares, a posição de cada distrito é o centro de gravidade dos respectivos concelhos). Evidentemente que nos escapa assim a dispersão ou heterogeneidade de cada distrito em particular, mas, mesmo assim, podemos tirar uma primeira ideia, a corrigir, das proximidades e afastamentos dos distritos entre si e da maior ou menor dependência das variáveis.

Representações das variáveis e dos distritos

Examinemos o gráfico III.

A configuração da nuvem de variáveis tem forma aproximadamente triangular: o primeiro vértice é consubstanciado no ponto CIA — percentagem de contribuição industrial da classe A; um segundo vértice situado ao alto no primeiro quadrante é ocupado por OPA — percentagem de assalariados agrícolas não pagos ao mês; o terceiro vértice está situado no quarto quadrante, ocupado por FAM — percentagem de activos familiares não remunerados.

Estes três vértices correspondem aos três tipos básicos de concelhos portugueses, encontrados atrás, na discussão do primeiro e segundo factores.

Assim, ao primeiro vértice, junto do qual se encontram os pontos OTE, SEC e ETE, OPNA (percentagem de assalariados não agrícolas, elemento suplementar, complementar de OPA) e CIA, estarão associadas

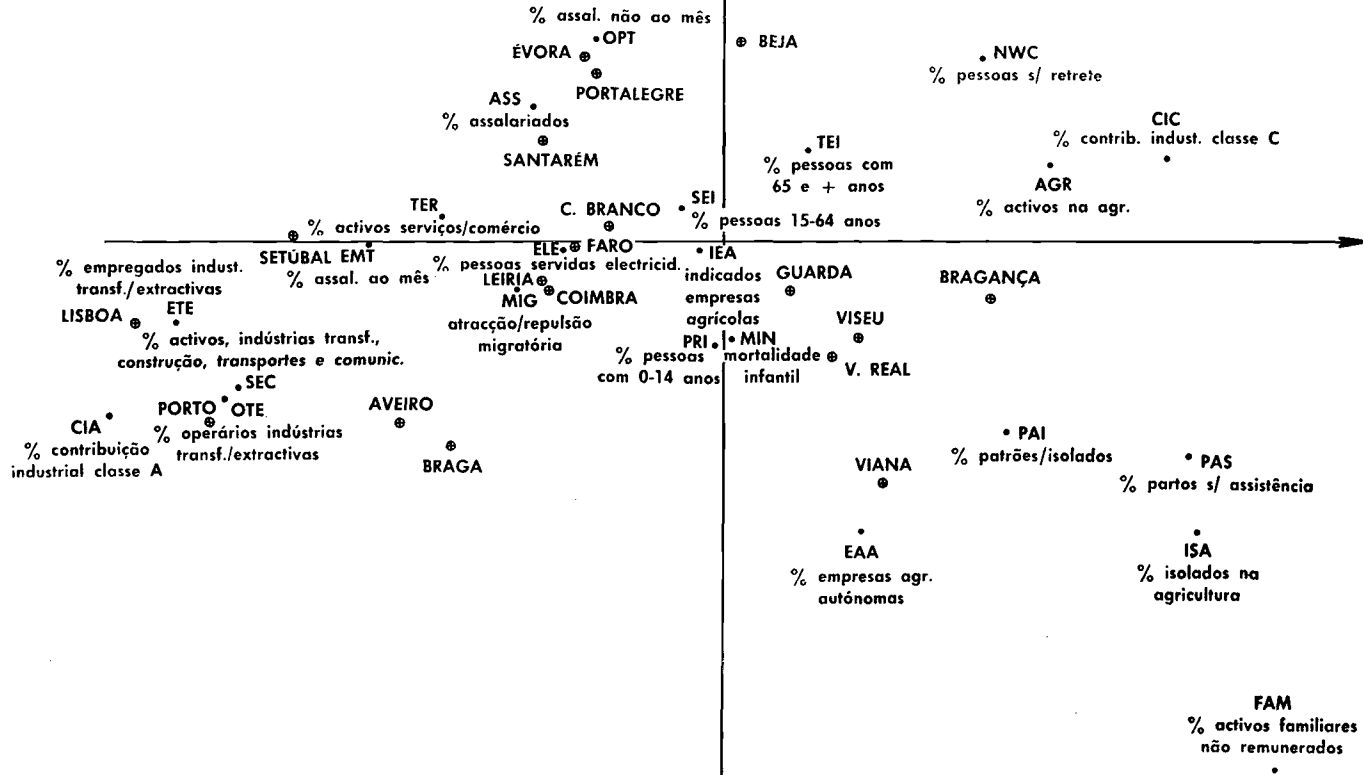
Segundo factor: relações de produção na agricultura

$\lambda 2 = 0,0435$
 $\epsilon 2 = 16,2\%$

Tipologia dos concelhos portugueses
 (Análise das correspondências)

Repartição das variáveis e dos distritos
 no plano factorial (1,2)
 Perc. de inércia explicada = 62,9 %

[GRÁFICO III]



Primeiro factor: acumulação da riqueza social
 $\lambda 1 = 0,1249$
 $\epsilon 1 = 46,7\%$

regiões com desenvolvimento quase exclusivamente industrial ou de serviços, com forte concentração de grandes empresas, onde é maior a percentagem de operários e empregados. Aqui se encontram concelhos de Lisboa e Porto e ainda alguns de Setúbal, Aveiro e Braga.

O segundo vértice — OPA — aparece relativamente isolado no gráfico. As variáveis mais próximas — NWC e TEI — ajudam, no entanto, a caracterizar a população das regiões em questão: tratar-se-á de regiões agrícolas com altas taxas de assalariados pagos «não ao mês», cuja população, em média mais idosa que a das outras regiões, vive, em geral, em fracas condições de sanidade básica. Encontramos aqui, entre outros, a maior parte dos concelhos pertencentes a Évora, Beja, Portalegre e Santarém.

Antes de passarmos às regiões associadas ao terceiro vértice notemos que as variáveis que se estendem na faixa compreendida entre o primeiro e o segundo vértices são as que caracterizam não só as regiões associadas a aqueles dois vértices, mas igualmente as zonas de transição entre eles.

Ao terceiro vértice, enfim, corresponderão sobretudo concelhos de Bragança, Vila Real, Viana, Viseu e Guarda: trata-se de concelhos essencialmente agrícolas, com fortes percentagens de empresas agrícolas autónomas, altas taxas de isolados e de familiares não remunerados e uma cobertura médica muito deficiente.

Representação dos concelhos: tipologia

Vejamos o gráfico iv.

A forma triangular da nuvem de variáveis reencontra-se aqui. Tal como fora previsto, não há separações brutais entre grupos de concelhos, com núcleos de pontos destacados uns dos outros. Há, no entanto, clareiras e corredores vazios que permitem isolar os grupos, tal como é mostrado pelo gráfico.

Reencontramos os três tipos puros de concelhos nos três vértices do «triângulo», correspondentes ao desenvolvimento industrial e/ou de serviços (tipo I), desenvolvimento agrícola com mão-de-obra assalariada (tipo II) e actividade agrícola (atrasada) com isolados e familiares não remunerados (tipo III).

Cada um destes tipos, por si só, tem uma certa coerência social e económica, mas não possui necessariamente homogeneidade ou proximidade geográfica. Desse ponto de vista, aquele em que é mais visível a proximidade geográfica é o grupo II, ao qual se segue o grupo III. A esse respeito consultar o gráfico v, que mostra Portugal continental e a tipologia dos concelhos.

As três zonas típicas, funcionando como pólos de atracção, definem entre si zonas de transição. Tal como os tipos «puros», também os tipos de transição têm certa unidade socioeconómica, mas, em geral, os concelhos estão afastados geograficamente.

Quando se progride de I para II, os concelhos vão perdendo o carácter industrial e tornando-se mais agrícolas, mantendo embora altas as percentagens totais de activos assalariados; a gradação aqui obtida permite diferenciar duas zonas nesta transição: chamaremos I-II à zona mais vizinha de I e II-I àquela que cola com II.

Procedendo da mesma forma com as transições de II para III (decremento da percentagem de assalariados agrícolas, subida da percentagem

de isolados e familiares) e de I para III (desaparecimento progressivo da indústria e serviços, substituídos por empresas agrícolas, em geral autónomas com isolados e familiares, e baixa radical dos assalariados de todas as categorias), surgem os grupos II-III, «terra-de-ninguém» II-III e III-II, por um lado, e I-III, «terra-de-ninguém» I-III e III-I, por outro.

Notar-se-á ainda a existência, em torno da origem, de um grupo de concelhos partilhando as características dos três tipos básicos e que foi «baptizado» I-II-III.

Para caracterizar completamente os diferentes tipos de concelhos portugueses teríamos de regressar aos indicadores socioeconómicos de que se partiu e estudar a distribuição de cada um dos indicadores sobre cada um dos grupos: destacaríamos assim o papel de cada uma das variáveis, isoladamente e em conjunto, na individualização dos tipos de concelhos, assim como nas gradações ou modificações sucessivas que se observem nos vários grupos quando se percorre o gráfico IV.

Deixaremos esse necessário estudo para a futura análise, já anunciada na introdução deste trabalho, e faremos aqui uma referência mais sintética sugerida pela comparação entre os gráficos III (dos indicadores) e IV (dos concelhos).

Assim, todos os concelhos de I, II e respectiva zona de transição têm, em geral, elevada a percentagem de assalariados, e em particular de operários, e baixa a taxa de activos isolados e activos familiares não remunerados. A passagem da zona industrial I para a zona de agricultura mais industrializada II é acompanhada pela diminuição global dos índices SEC (percentagem de activos nas indústrias transformadoras, construção, transporte e comunicações), OTE (percentagem de operários nas indústrias transformadoras e extractivas), ETE (percentagem de empregados nas indústrias transformadoras e extractivas) e CIA (contribuição industrial da classe A); haverá ainda tendência para diminuir, embora menos pronunciadamente, os valores de EMT (percentagem de assalariados pagos ao mês) e TER (percentagem de activos no comércio e serviços); e aumenta a percentagem de assalariados agrícolas não pagos ao mês — OPA.

Quanto à zona de transição das regiões agrícolas (passagem de II para III), ela é caracterizada por valores em regra baixos dos indicadores de desenvolvimento industrial (que caracterizam, como se viu, o pólo I) e valores altos de AGR — percentagem de activos na agricultura, de CIC — percentagem de contribuição industrial da classe C e de PPM — número de pessoas por médico. Quando se caminha de II para III, isto é, da agricultura desenvolvida para a agricultura atrasada, as regiões são caracterizadas por uma diminuição global progressiva de OPA — percentagem de assalariados agrícolas não pagos ao mês e simultaneamente por um aumento progressivo de ISA — percentagem de isolados na agricultura, de FAM — percentagem de efectivos familiares, de PAS — percentagem de partos e de EAA — percentagem de empresas agrícolas autónomas.

Enfim, quanto à terceira zona de transição, verifica-se, quando se passa de I — regiões industrializadas — para III — regiões agrícolas atrasadas —, que há uma quebra sucessiva dos valores de todos os indicadores característicos do desenvolvimento industrial e que saem do lado direito do gráfico III, em especial de CIA, ETE, OTE, SEC, ao mesmo tempo que aumentam os valores de EAA, ISA, FAM e PAS.

ANEXO 1

<i>Aveiro</i>	0100	<i>Braga</i>	0300
Águeda	0101	Amares	0301
Albergaria-a-Velha ..	0102	Barcelos	0302
Anadia	0103	Braga	0303
Arouca	0104	Cabeceiras de Basto ...	0304
Aveiro	0105	Celorico de Basto	0305
Castelo de Paiva	0106	Esposende	0306
Espinho	0107	Fafe	0307
Estarreja	0108	Guimarães	0308
Feira	0109	Póvoa de Lanhoso	0309
Ílhavo	0110	Terras de Bouro	0310
Mealhada	0111	Vieira do Minho	0311
Murtosa	0112	Vila Nova de Famalicão	0312
Oliveira de Azeméis ...	0113	Vila Verde	0313
Oliveira do Bairro ..	0114		
Ovar	0115	<i>Bragança</i>	0400
São João da Madeira ...	0116	Alfândega da Fé	0401
Sever do Vouga	0117	Bragança	0402
Vagos	0118	Carrazeda de Ansiães ...	0403
Vale de Cambra	0119	Freixo de Espada à Cinta	0404
		Macedo de Cavaleiros ...	0405
<i>Beja</i>	0200	Miranda do Douro	0406
Aljustrel	0201	Mirandela	0407
Almodôvar	0202	Mogadouro	0408
Alvito	0203	Torre de Moncorvo	0409
Barrancos	0204	Vila Flor	0410
Beja	0205	Vimioso	0411
Castro Verde	0206	Vinhais	0412
Cuba	0207		
Ferreira do Alentejo ...	0208	<i>Castelo Branco</i>	0500
Mértola	0209	Belmonte	0501
Moura	0210	Castelo Branco	0502
Odemira	0211	Covilhã	0503
Ourique	0212	Fundão	0504
Serpa	0213	Idanha-a-Nova	0505
Vidigueira	0214	Oleiros	0506

Penamacor 0507	Portimão 0811
Proença-a-Nova 0508	São Brás de Alportel 0812
Sertã 0509	Silves 0813
Vila de Rei 0510	Tavira 0814
Vila Velha de Ródão 0511	Vila do Bispo 0815
<i>Coimbra</i> 0600	Vila Real de Santo António ... 0816
Arganil 0601	<i>Guarda</i> 0900
Cantanhede 0602	Aguiar da Beira 0901
Coimbra 0603	Almeida 0902
Condeixa-a-Nova 0604	Celorico da Beira 0903
Figueira da Foz 0605	Figueira de Castelo Rodrigo ... 0904
Góis 0606	Fornos de Algodres 0905
Lousã 0607	Gouveia 0906
Mira 0608	Guarda 0907
Miranda do Corvo 0609	Manteigas 0908
Montemor-o-Velho 0610	Meda 0909
Oliveira do Hospital 0611	Pinhel 0910
Pampilhosa da Serra 0612	Sabugal 0911
Penacova 0613	Seia 0912
Penela 0614	Trancoso 0913
Soure 0615	Vila Nova de Foz Côa 0914
Tábua 0616	<i>Leiria</i> 1000
Vila Nova de Poiares 0617	Alcobaça 1001
<i>Évora</i> 0700	Alvaiázere 1002
Alandroal 0701	Ansião 1003
Arraiolos 0702	Batalha 1004
Borba 0703	Bombarral 1005
Estremoz 0704	Caldas da Rainha 1006
Évora 0705	Castanheira da Pêra 1007
Montemor-o-Novo 0706	Figueiró dos Vinhos 1008
Mora 0707	Leiria 1009
Mourão 0708	Marinha Grande 1010
Portel 0709	Nazaré 1011
Redondo 0710	Óbidos 1012
Reguengos de Monsaraz 0711	Pedrógão Grande 1013
Vendas Novas 0712	Peniche 1014
Viana do Alentejo 0713	Pombal 1015
Vila Viçosa 0714	Porto de Mós 1016
<i>Faro</i> 0800	<i>Lisboa</i> 1100
Albufeira 0801	Alenquer 1101
Alcoutim 0802	Arruda dos Vinhos 1102
Aljezur 0803	Azambuja 1103
Castro Marim 0804	Cadaval 1104
Faro 0805	Cascais 1105
Lagoa 0806	Lisboa 1106
Lagos 0807	Loures 1107
Loulé 0808	Lourinhã 1108
Monchique 0809	Mafra 1109
Olhão 0810	Oeiras 1110

Sintra 1111	Entroncamento 1410
Sobral de Monte Agraço 1112	Ferreira do Zézere 1411
Torres Vedras 1113	Golegã 1412
Vila Franca de Xira 1114	Mação 1413
<i>Portalegre</i> 1200	Rio Maior 1414
Alter do Chão 1201	Salvaterra de Magos 1415
Arronches 1202	Santarém 1416
Aviz 1203	Sardoal 1417
Campo Maior 1204	Tomar 1418
Castelo de Vide 1205	Torres Novas 1419
Crato 1206	Vila Nova da Barquinha 1420
Elvas 1207	Vila Nova de Ourém 1421
Fronteira 1208	<i>Setúbal</i> 1500
Gavião 1209	Alcácer do Sal 1501
Marvão 1210	Alcochete 1502
Monforte 1211	Almada 1503
Nisa 1212	Barreiro 1504
Ponte de Sor 1213	Grândola 1505
Portalegre 1214	Moita 1506
Sousel 1215	Montijo 1507
<i>Porto</i> 1300	Palmela 1508
Amarante 1301	Santiago do Cacém 1509
Baião 1302	Seixal 1510
Felgueiras 1303	Sesimbra 1511
Gondomar 1304	Setúbal 1512
Lousada 1305	Sines 1513
Maia 1306	<i>Viana do Castelo</i> 1600
Marco de Canaveses 1307	Arcoz de Valdevez 1601
Matosinhos 1308	Caminha 1602
Paços de Ferreira 1309	Melgaço 1603
Paredes 1310	Monção 1604
Penafiel 1311	Paredes de Coura 1605
Porto 1312	Ponte da Barca 1606
Póvoa de Varzim 1313	Ponte de Lima 1607
Santo Tirso 1314	Valença 1608
Valongo 1315	Viana do Castelo 1609
Vila do Conde 1316	Vila Nova da Cerveira 1610
Vila Nova de Gaia 1317	<i>Vila Real</i> 1700
<i>Santarém</i> 1400	Alijó 1701
Abrantes 1401	Boticas 1702
Alcanena 1402	Chaves 1703
Almeirim 1403	Mesão Frio 1704
Alpiarça 1404	Mondim de Basto 1705
Benavente 1405	Montalegre 1706
Cartaxo 1406	Murça 1707
Chamusca 1407	Peso da Régua 1708
Constância 1408	Ribeira de Pena 1709
Coruche 1409	Sabrosa 1710

Santa Marta de Penaguião	1711	Oliveira de Frades	1810
Valpaços	1712	Penalva do Castelo	1811
Vila Pouca de Aguiar	1713	Penedono	1812
Vila Real	1714	Resende	1813
<i>Viseu</i>	<i>1800</i>	Santa Comba Dão	1814
Armamar	1801	São João da Pesqueira	1815
Carregal do Sal	1802	São Pedro do Sul	1816
Castro Daire	1803	Sátão	1817
Cinfães	1804	Sernancelhe	1818
Lamego	1805	Tabuaço	1819
Mangualde	1806	Tarouca	1820
Moimenta da Beira	1807	Tondela	1821
Mortágua	1808	Vila Nova de Paiva	1822
Nelas	1809	Viseu	1823
		Vouzela	1824

ANEXO 2

Indicadores socioeconómicos utilizados

- PAI — percentagem de patrões e isolados
- ASS — percentagem de assalariados
- FAM — percentagem de familiares não remunerados
- NWC — percentagem de pessoas sem retrete
- CIC — percentagem de contribuição industrial (CI) paga pela classe C (em relação à CI total)
- CIA — percentagem de contribuição industrial paga pela classe A
- AGR — percentagem de activos na agricultura
- SEC — percentagem de activos nas indústrias transformadoras e de construção, nos transportes e comunicações
- TER — percentagem de activos em serviços e comércio
- EAA — percentagem de empresas agrícolas autónomas (em relação ao total de empresas agrícolas)
- ELE — percentagem de pessoas servidas de electricidade
- PRI — percentagem de pessoas com idades compreendidas entre 0 e 14 anos
- TEI — percentagem de pessoas com idades acima dos 64 anos
- SEI — percentagem de pessoas com idades compreendidas entre 15 e 64 anos
- MIG — atracção-repulsão migratória (1961-70)
- ISA — percentagem de isolados agrícolas (população agrícola)
- OPA — percentagem de assalariados agrícolas pagos sem ser ao mês
- EMT — percentagem de assalariados pagos ao mês
- OPT — percentagem de assalariados pagos sem ser ao mês
- ETE — percentagem de empregados nas indústrias transformadoras e extractivas
- OTE — percentagem de operários nas indústrias transformadoras e extractivas
- IEA — indicador de empresas agrícolas (1968[52-54])
- MIN — mortalidade infantil
- PAS — parto sem assistência
- PPM — pessoas por médico

BIBLIOGRAFIA E FONTES DE DADOS

a) METODOLOGIA

- [1] L. Lebart e Jean-Pierre Fenelon, *Statistique et Informatique Appliquées*, Paris, Dunod, 1971.
- [2] J. P. Benzécri e cols., *L'Analyse des Données*, t. 2, *L'Analyse des correspondances*, Paris, Dunod, 1973.
- [3] F. Costa Nicolau, *Thèse de 3ème Cycle*, Paris-VI, Université de Pierre et Marie Curie, 1977.

b) DADOS

- [4] *Censo Geral da População de 1970*, INE.
- [5] *Estatísticas das Contribuições e Impostos de 1974*, INE.
- [6] *Recenseamentos Agrícolas de 1952, 1953, 1954 e 1968*, INE.
- [7] *Atlas Socioeconómico*, Gabinete de Planeamento do Ministério do Trabalho.

Lisboa, Janeiro de 1977.

Fernando da Costa Nicolau

Centro de Estatística e Aplicação (INIC), Faculdade de Ciências de Lisboa

Jorge de Sá

Fonds National de Recherche Scientifique (Bélgica).