

Conduta e Desempenho de Empresas: Uma Aplicação de "Cluster Analysis" à Segmentação da Indústria do Cimento

CLÁUDIO FELISONI DE ANGELO
ANTONIO ZORATTO SANVICENTE

Resumo

O presente trabalho discute a importância de uma adequada classificação de empresas por semelhança para fins comparativos, partindo do paradigma conduta-desempenho da Teoria da Organização Industrial. Para fins ilustrativos, toma como base 36 empresas da indústria do cimento no Brasil, e aplica a metodologia de *cluster analysis* visando identificar, dentro de um subsetor tido como homogêneo, os sub-grupos de empresas mais assemelhadas e que, portanto, permitiriam comparações externas mais apropriadas. Com base em dados relativos a cinco variáveis de desempenho e outras cinco variáveis de conduta, extraídos de demonstrações financeiras publicadas, este estudo inicialmente encontrou, usando a técnica de correlação canônica, resultados significativos e conforme as expectativas para relações entre variáveis de desempenho e conduta. Além disso, foi identificado um grande subgrupo de cerca de 2/3 de empresas homogêneas no sub-setor. Entretanto, isso deixa 14 das 36 empresas como suficientemente distintas para fins de alguma comparação válida com o que seria típico na indústria do cimento.

Palavras-chave: indústria de cimento, organização industrial, classificação setorial, correlação canônica, *cluster analysis*.

Abstract

The present paper discusses the significance of appropriate industrial classification schemes in attempts at comparing different firms, having as departure point the behavior-performance paradigm of Industrial Organization Theory. As an empirical illustration, 36 firms included in the Brazilian cement industry were subjected to cluster analysis, with the goal identifying, in an allegedly homogeneous sector, those sub-groups of more similar firms which could justify external comparisons for analytical purposes. Using data on five performance variables and five behavior indicators taken from published financial statements, we have found, thanks to a canonical correlation study, significant and plausible relationships between groups of performance and behavior variables. In addition, a large homogeneous sub-group of 22 firms was identified, leaving 14 other firms as being too different for justifying valid comparisons with what could be called typical in the cement industry.

Key words: cement industry, industrial organization, industrial classification, canonical correlation, cluster analysis.

Os autores são professores do Departamento de Administração da FEA-USP.

Introdução

A empresa pode ser entendida como uma unidade transformadora, na qual matérias-primas, insumos e fatores são processados para a obtenção de bens e serviços.

A realização destas atividades é sustentada por um sistema gerencial, que analisa os sinais do ambiente e estabelece a orientação estratégica e tática para a organização, com a finalidade de posicioná-la de acordo com o tipo e a intensidade das mensagens percebidas.

Naturalmente, a eficiência e a eficácia da ação empresarial dependem, entre outras coisas, do grau de exatidão na interpretação das evidências ambientais.

Assim, um estudo metuculoso e sistemático do setor em que a empresa atua fornece elementos de fundamental importância para a gestão empresarial.

Este trabalho procura dar uma contribuição nesta direção, focalizando, a título de exemplo, a produção nacional de cal e cimento.

A escolha deste setor foi motivada pelo relativo grau de homogeneidade do produto fabricado nesta atividade industrial. Como argumentamos em trabalho recente (ANGELO & SANVICENTE, 1990), a análise ambiental muitas vezes carece de sentido por tratar um conjunto de empresas com interesses mercadológicos e características operacionais e financeiras diferentes sob a mesma designação setorial.

O artigo analisa, em nível das empresas, suas diferenças em termos econômicos e financeiros, identificando, com base em um quadro teórico de referência, as variáveis responsáveis por estas distinções. Além disto, visamos demonstrar a aplicabilidade da metodologia de *cluster analysis* a problemas desta natureza.

Estruturamos o artigo em seis partes, além desta Introdução. Após a Introdução, é apresentada uma discussão a respeito do universo ("indústria do cimento") do qual foram selecionadas as 36 empresas que fazem parte da amostra. Na seção 2, após um rápido exame da Microeconomia tradicional, apresentamos a base conceitual deste trabalho, que se encontra no campo da Organização Industrial. Dentro desta

abordagem examinamos, na seção 3, a validade do agrupamento de empresas, com a identificação de semelhanças que permitam a classificação das unidades produtivas em subgrupos mais homogêneos. A seção 4 sintetiza a metodologia de *cluster analysis* e relaciona as variáveis utilizadas no modelo. Os resultados são apresentados na seção 5. A última parte é reservada para as conclusões.

1 O Setor Produtor de Cimento no Brasil

A apresentação feita nesta seção baseia-se em dados de Sindicato Nacional da Indústria do Cimento (1990).

A indústria de produtora de cimento desempenha, sem dúvida, uma papel de grande importância na economia brasileira. Esta importância é manifestada pelas freqüentes relações que se estabelecem entre o nível de atividade econômica e a evolução do consumo nacional desse produto.

Existem quatro tipos de cimento: Comum, Alto-Forno, Pozolânico, Branco e Alvenaria. O cimento Comum representa 70,5% da produção nacional, enquanto os tipos Alto-Forno, Pozolânico, Branco e Alvenaria representam 11,7%, 17,5%, 0,2% e 0,1%, respectivamente.

A produção destes produtos está distribuída entre 57 fábricas, das quais 30, representando pouco mais de 50%, localizam-se na Região Sudeste. Das 27 unidades produtivas restantes 13 estão situadas na Região Nordeste, 2 na Região Norte, 7 na Região Sul e 5 na Região Centro-Oeste.

Em termos de volume de produção, a Região Sudeste é responsável por pouco quase 60% da oferta nacional do produto. Em 1989, por exemplo, do total de 25.920 mil toneladas produzidas no país, 15.331 mil toneladas foram fabricadas nos estados do Sudeste.

A exportação e a importação do produto não são expressivas. Tomando como referência novamente o ano de 1989, verifica-se que foram exportadas 64.709 toneladas e importadas 64.332 toneladas, o que representa menos de 3% da produção nacional observada no mesmo ano.

A produção e o consumo do produto são bastante regionalizados, conforme mostra a Tabela 1. Como pode ser verificado, a distribuição regional percentual da produção e do consumo são praticamente idênticas.

**TABELA 1
DISTRIBUIÇÃO PERCENTUAL DA PRODUÇÃO E
DO CONSUMO DE CIMENTO NO BRASIL**

Regiões	Produção	Consumo
Norte	3,0	4,0
Nordeste	13,6	14,6
Sudeste	59,2	56,5
Sul	15,7	17,0
Centro-Oeste	8,5	7,9
Total	100,0	100,0

Fonte: Sindicato Nacional da Indústria do Cimento (1990).

Isto decorre das próprias características do produto, que tornam pouco atrativo o deslocamento da mercadoria para centros consumidores distantes. O peso dos custos de transportes na composição dos custos são importante fator de segmentação de mercado. No Brasil, provavelmente as limitações associadas a sistema de transportes são ainda maiores que em outros países produtores. A razão disso é que a movimentação dessa mercadoria no Brasil é feita, na sua grande parte (mais de 85%), pelo modo rodoviário. Como se sabe, as tarifas cobradas por essa modalidade de transporte são bastante superiores àquelas normalmente observadas no transporte ferroviário e hidroviário.

Quando se divide o mercado em regiões e confrontam-se os dados relativos à distribuição da produção e do consumo com o número de empresas que atuam nas diferentes áreas geográficas, verifica-se que setor apresenta, como característica, uma forte concentração da produção. Esta constatação pode ser reforçada considerando que muitas fábricas pertencem, na verdade, a um mesmo grupo.

Por outro lado, quando se analisa a composição do consumo constata-se que ele ocorre de maneira bastante atomizada. São três as categorias de consumidores: Concreteiras, Consumidores Industriais e Consumidores Finais. Conforme mostra a Tabela 2, a categoria denominada Consumidores Industriais é composta por empresas fabricantes de

artefatos, pré-moldados e chapas de fibro-cimento. Os Consumidores Individuais, por sua vez, são construtoras, pedreiros, empresas privadas, órgãos públicos, prefeituras e indivíduos. Como pode ser observado, esta última categoria, extremamente pulverizada, representa 79 % de todo o consumo.

TABELA 2
DISCRIMINAÇÃO DO CONSUMO DE CIMENTO

Discriminação	%
1. Concretistas	10,0
2. Consumidores Industriais	11,0
Fibrocimento	4,0
Artefatos	4,0
Pré-Moldados	3,0
3. Consumidores Finais	79,0
Consumidores Individuais	31,0
Construtoras/Empreiteiras	27,0
Pedreiros/Pequenos Empreiteiros	12,0
Empresas Privadas	5,0
Órgãos Públicos	2,0
Prefeituras	2,0
Total	100,0

Fonte: Sindicato Nacional da Indústria do Cimento (1990).

Uma vez apontadas algumas das características do mercado de cimento, procurar-se-á identificar subgrupos de empresas que compõem o setor. No tópico seguinte apresentam-se as bases teóricas que orientarão este procedimento.

2. A Análise Microeconômica e o Estudo da Organização Industrial

Os modelos desenvolvidos pela teoria microeconômica procuram analisar o comportamento do consumidor e do produtor partindo de hipóteses bastante simplificadoras. Pretende-se explicar como os recursos escassos são alocados para atender necessidades concorrentes.

A exemplo de outras ciências do comportamento, o campo de observação do economista é toda a realidade econômica, com sua imensa e intrincada complexidade.

CONDUTA E DESEMPENHO DE EMPRESAS

A teoria microeconômica pode ser subdividida em três grandes partes. A primeira e a segunda partes procuram explicar, respectivamente, o comportamento do consumidor e do produtor, e a terceira caracteriza e analisa as estruturas de mercado.

Os três enfoques se assemelham pela generalidade no tratamento dos problemas. A teoria do consumidor estuda como um indivíduo, sujeito a uma restrição orçamentária, distribui seus recursos de forma a maximizar sua satisfação. A teoria da produção parte da hipótese de maximização dos lucros para explicar como a firma aloca os recursos para a obtenção de um determinado volume de produto. Finalmente, a teoria dos mercados estuda o comportamento dos produtores considerando certas condições concorrenciais.

Como afirma Friedman (1953), o valor de um modelo depende não do realismo de suas hipóteses, mas de sua capacidade de explicar ou prever acontecimentos do mundo real. Alguns economistas têm procurado demonstrar que a capacidade preditiva e explicativa da teoria microeconômica é bastante restrita. Para eles, as hipóteses dos modelos tradicionais precisam ser reavaliadas de modo a tornar a abstração uma ferramenta útil de análise econômica.

Esta linha de raciocínio levou Bain (1968) e Scherer (1970) ao desenvolvimento da abordagem conhecida como Organização Industrial, na qual pode-se distinguir três dimensões analíticas: estrutura de mercado, conduta e desempenho.

A estrutura de mercado corresponde às características da organização de mercado que influenciam a estratégia empresarial. A conduta refere-se às políticas que as empresas seguem em suas atitudes de ajuste e adaptação às mudanças e nas condições de mercado. O desempenho diz respeito aos resultados alcançados pelas empresas.

Estas três dimensões estão presentes nos trabalhos de Bain e Scherer. Entretanto, enquanto o primeiro destaca a importância dos aspectos estruturais na determinação do desempenho, o segundo enfatiza a conduta na explicação dos resultados empresariais.

Seguindo esta orientação, diversos trabalhos foram produzidos. Em nível teórico, por exemplo, Porter (1979) formula a hipótese de que

os determinantes da rentabilidade das empresas (desempenho) envolvem a estrutura interna do setor ao qual as empresas pertencem. Em termos empíricos, podemos citar Gronhaug e Fredriksen (1989), que se concentraram no impacto de alterações da estrutura da indústria norueguesa de telecomunicações, entre 1976 e 1981, sobre o desempenho das empresas do setor nesse mesmo período. Segundo esses dois autores, a idéia básica é a de que "*a estrutura da indústria direciona a conduta (estratégia) das empresas do setor, e também pode influenciar seu desempenho*" (p. 257).

A estrutura é definida pelas condições que vigoram na organização do mercado e que influem na natureza da competição. Dentre as variáveis relacionadas a esse aspecto destaca-se o tamanho da unidade produtora.

Esta abordagem nos parece muito apropriada à finalidade deste artigo. Recolocando nosso objetivo de acordo com a terminologia de Organização Industrial, podemos dizer que pretendemos, tomando como um dado a estrutura de mercado de um ramo industrial (produção de cal e cimento), subdividir o setor em subgrupos a partir de algumas dimensões das variáveis associadas à conduta e ao desempenho das empresas.

3. Validade do Agrupamento de Empresas

No contexto dos conceitos de Organização Industrial, discute-se "estrutura" como algo representativo de um setor, isto é, um conjunto de empresas que oferecem produtos afins no sentido de atendimento de necessidades de usuários intermediários ou finais. Em princípio, o mercado sofre influências externas comuns, mas a "conduta" das diversas empresas que compõem o setor não é necessariamente a mesma. Isto é, algumas empresas adaptam-se de um modo, e outras de outro, por exemplo, escolhendo tecnologias diferentes e/ou financiando suas atividades com proporções diferentes entre capital de terceiros e capital próprio.

Essas escolhas, por sua vez, refletem-se em graus de imobilização do ativo e graus de endividamento distintos, para empresas do mesmo setor

Finalmente, em face de opções de conduta diferentes, empresas que atuam no mesmo mercado acabam apresentando desempenho distinto. Como mostra a literatura de Finanças (por exemplo, SANVICENTE, 1987), quanto maior o grau de endividamento, maior tende a ser a variabilidade dos resultados líquidos, no que é conhecido como fenômeno da alavancagem financeira.

Neste trabalho, utilizamos a seguinte lógica: tomamos um conjunto de empresas tradicionalmente classificadas num certo setor, porque atuam no mesmo mercado (produção de cal e cimento), e por isso sujeitas às mesmas influências externas. Em seguida, levantamos indicadores de conduta e desempenho, em termos dos quais as empresas da amostra podem diferir, e tentamos formar, através de *cluster analysis*, agrupamentos de empresas semelhantes. A principal utilidade prática da identificação desses agrupamentos, dentro de um setor, seria a de enumerar aquelas empresas mais diretamente comparáveis para fins de análise, planejamento e tomada de decisões.

O uso de *cluster analysis* para agrupamento de empresas com características semelhantes, que utilizamos em nosso trabalho anterior (ANGELO & SANVICENTE, 1990), já revelou que a classificação industrial convencionalmente usada por anuários de dados sobre empresas (por exemplo, revistas *Visão*, *Exame* e *Balanço Financeiro*) não tem porque levar a conjuntos de empresas comparáveis.

Nos Estados Unidos, a mesma idéia levou Clarke (1989) a examinar a eficácia, para esse fim, da classificação industrial utilizada pelo governo americano (SIC, ou *Standard Industrial Classification*). Nesse caso, a proposta foi a de verificar se empresas classificadas com o mesmo número de código SIC faziam parte de um setor homogêneo. Se isso realmente acontecesse, as empresas teriam o mesmo comportamento quanto a algumas variáveis (por exemplo, crescimento das vendas e rentabilidade).

Os resultados obtidos por Clarke indicaram que a classificação por SICs não oferece uma boa representação por semelhança entre as empresas e, portanto, que o uso de setores convencionais para fins de agregação de dados de empresas em *cross sections* pode não levar a comparações válidas. Ou, nas próprias palavras desse autor (p. 19): "É essencial que as empresas sejam divididas precisamente em setores

econômicos homogêneos para que tais comparações entre indústrias, ou dentro de indústrias no tempo, sejam válidas."

Metodologia

1. Cluster Analysis

Esta técnica de análise multivariada é empregada quando desejamos reduzir o número de objetos (isto é, o número de linhas, numa matriz de observações por variáveis), agrupando-os em *clusters*. Queremos fazer isso de tal modo que os objetos que fiquem reunidos num *cluster* sejam mais parecidos entre si do que com objetos pertencentes a outros *clusters*. Trata-se de uma metodologia de classificação, ou taxonômica, baseada em métodos numéricos.

Segundo Green (1978), as três questões básicas em *cluster analysis* são as seguintes:

- a) como medir a semelhança entre objetos?
- b) supondo que se possa medir a semelhança relativa entre objetos, como então colocar os objetos semelhantes em *clusters*?
- c) após ter efetuado o agrupamento, como descrever os *clusters* e saber se os *clusters* resultantes são reais, e não o produto de um simples artifício estatístico?

No que se refere à primeira questão, a resposta está no conceito de "semelhança": dois objetos são considerados semelhantes se seus perfis, em termos das variáveis utilizadas, são próximos, o que é determinado por medidas de distância, como a distância euclidiana entre pontos, ou através de um processo de *matching*, quando as escalas das variáveis são nominais.

A segunda pergunta é respondida através dos diversos algoritmos disponíveis para *cluster analysis*. Todos os algoritmos, porém, envolvem a determinação de ótimos locais, e não globais. De qualquer maneira, cabe à teoria relevante a definição *a priori* do número correto de *clusters*, e não aos algoritmos.

Por fim, acredita-se que a melhor maneira de descrever os *clusters* resultantes seja através do valor central ou do vetor de médias (centróides) de cada *cluster*. Infelizmente, ainda não foram desenvolvidos procedimentos defensáveis para testar a significância dos *clusters* formados com a utilização de qualquer algoritmo disponível. Na verdade, o principal problema, quando esta metodologia é utilizada, é o de que os *clusters* são formados com base nos próprios dados, o que significa que ainda não há critério estatístico externo de formação de grupos para servir de base de comparação.

Na maioria dos estudos, nos quais deseja-se inferir a partir de um conjunto de dados, supõe-se: a) existir alguma estrutura ou algum agrupamento conhecido nos dados que estão sendo examinados, e b) ser possível estimar ou conhecer certos parâmetros dessa estrutura. Normalmente, a primeira suposição é feita antes do estudo ser realizado, e o problema passa a ser o de estimação dos parâmetros de interesse através de procedimentos estatísticos. Outras vezes, não se conhece a estrutura dos dados antes do estudo ser feito. É o caso dos agrupamentos por maior semelhança entre empresas pertencentes à indústria do cimento, que aqui estamos analisando. Queremos saber, por exemplo, se todas as empresas formam um único grupo homogêneo, ou, no outro extremo, se todas elas são tão distintas que, na verdade, poderiam quase representar setores diferentes de atividade. Portanto, o problema passa a ser o de descobrir relações de semelhança entre as entidades que fazem parte de um conjunto, neste caso empresas aprioristicamente classificadas por alguma publicação em algo denominado "setor de cal e cimento"

Cluster analysis, ou taxonomia numérica, segundo Panton, Lessig e Joy (1976, p. 417), refere-se a "um conjunto de procedimentos cujo objetivo é examinar as relações de semelhança entre entidades dentro de um conjunto de dados"

Partimos de n empresas, cada uma delas descritas por p variáveis. Para se poder especificar a estrutura de semelhança entre as empresas é necessário compará-las em termos de todas as p variáveis ao mesmo tempo. As comparações e a identificação de grupos (*clusters*) podem ser feitas de várias maneiras diferentes. Em *cluster analysis*, usa-se um enfoque denominado *hierarchical clustering*.

Trata-se de um processo de agregação no qual as entidades isoladas (no nosso caso, as 36 empresas do setor) são reunidas em grupos. O processo começa tratando-se cada uma das entidades como um *cluster* separado. Empregando-se uma medida de semelhança, como um coeficiente de correlação, busca-se os dois *clusters* (as duas entidades isoladas, a esta altura) que são mais semelhantes. Ao serem encontradas, são combinadas e tratadas como uma única entidade. Portanto, o número de entidades agora passa a ser igual a $n-1$. Novamente é feita essa busca, e o número total cai para $n-2$, e assim por diante, até que se tenha o número desejado de *clusters*.

Em termos quantitativos, a semelhança é avaliada por "medidas de distância" como se estivéssemos tratando de distâncias entre pontos em algum tipo de espaço métrico. Assim, a distância euclidiana entre dois pontos, num espaço de p dimensões, é freqüentemente utilizada:

$$d_{ij} = \left[\sum_{t=1}^p (X_{it} - X_{jt})^2 \right]^{1/2}$$

onde X_i e X_j são as projeções dos pontos i e j na variável t ($t = 1, 2, \dots, p$). Como as variáveis originais possuem unidades diferentes, sempre se efetua uma conversão das mesmas a seus valores padronizados, isto é, divididos por seus respectivos desvios padrões na amostra.

Em termos práticos, portanto, os métodos hierárquicos se caracterizam pela construção de uma hierarquia, na qual cada entidade começa como um *cluster*, e em seguida outra entidade é a ela reunida, dado que entre elas foi determinada a menor distância euclidiana. Seria o caso, por exemplo, das empresas i e j , que inicialmente eram dois *clusters*, e como d_{ij} é mínimo, consideradas todas as p variáveis e todas as outras empresas da amostra, passam a ser consideradas como uma única "empresa"

4.2. Amostra e Dados

4.2.1. Variáveis Utilizadas

Para 36 empresas pertencentes à Indústria de Cal e Cimento (ver relação no Anexo I), foram calculados os seguintes indicadores de desempenho e conduta:

a) Desempenho:

TURNASS = giro do ativo total, medido pelo quociente entre faturamento e saldo do ativo total. Esta variável procura representar a produtividade média com a qual o capital total é utilizado pela empresa.

LABORPRO = produtividade da mão-de-obra, igual ao quociente entre faturamento e número de empregados.

GROWTH = crescimento, ou seja, média geométrica da taxa anual de crescimento do faturamento da empresa.

PROFIT = rentabilidade, dada pelo quociente entre lucro líquido do exercício e patrimônio líquido.

MKTSHR = participação no mercado, ou seja, a proporção entre o faturamento da empresa e o faturamento total das 36 empresas da amostra.

b) Conduta:

CAPLABOR = relação capital/trabalho, medida pelo quociente entre o valor total do ativo imobilizado e o número de empregados.

LIQUID = índice de liquidez corrente, igual ao quociente entre ativo circulante e passivo circulante, como medida inversamente proporcional ao risco de insolvência, mas diretamente proporcional ao volume relativo de aplicações em ativos de menor rentabilidade.

HOLDING = quociente entre a conta investimentos (participação em coligadas e controladas) e a conta ativo imobilizado, representando a estrutura societária-administrativa escolhida pela empresa.

LEVERAGE = endividamento geral (exigível total/ativo total), uma medida direta do risco de insolvência e indireta da variabilidade potencial do lucro para o acionista.

INVPRO = relação entre investimento e produto, ou seja, quociente entre variação do ativo total e faturamento, representando o nível do fluxo de acréscimo de capital às operações.

As variáveis definidas acima foram escolhidas dentre os indicadores que poderiam ser calculados a partir dos dados publicados e facilmente disponíveis. Inicialmente, portanto, foi considerada a restrição de custo de obtenção de dados.

Sua classificação em grupos de "desempenho" e "conduta" porém, obedece a um critério simples e claro, a nosso ver: "desempenho" inclui tudo aquilo que reflete resultados de decisões e/ou políticas adotadas pelas empresas na utilização de fatores de produção, ou seja, o que mais comumente se chamaria de "rentabilidade" (variável PROFIT), "produtividade" (variáveis LABORPRO, TURNASS), e "crescimento" (variáveis MKTSHR e GROWTH).

Do outro lado ("conduta"), estão indicadores das políticas adotadas quanto à utilização mais ou menos intensiva de certos fatores de produção (variável CAPLABOR), decisões de financiamento (variável LEVERAGE), natureza de decisões de investimento (variáveis LIQUID, HOLDING e INVPRO).

Na literatura, como por exemplo a de Finanças (ver SANVICENTE, 1987), postula-se alguma relação entre algumas dessas variáveis. Assim, diz-se que aplicações concentradas em ativos circulantes oferecem rentabilidade menor do que aplicações em ativo imobilizado, portanto, a correlação entre as variáveis PROFIT e LIQUID deveria ser negativa.

Nessas escolhas, nada é específico, *a priori*, quanto ao setor de cimento. Nossa opção por esse setor, como foi dito na Introdução, deveu-se ao desejo de, ao aplicar *cluster analysis*, mostrar sua utilidade até mesmo num setor considerado homogêneo. Em outras palavras, desejávamos mostrar que até mesmo um setor tido como homogêneo revelaria, após uma análise mais atenta, a existência de subdivisões diferenciadas.

Os dados coletados referem-se aos exercícios de 1983 a 1987, e foram obtidos com o uso de várias edições da publicação "Quem é Quem na Economia Brasileira", da revista *Visão*.

5. Resultados

5.1. Análise de Correlação Canônica: Estudo Exploratório

Com o propósito de identificar, em nível exploratório, possíveis relações de associação entre as variáveis acima, procedemos a um estudo de correlação canônica entre os dois conjuntos de variáveis ("desempenho" e "conduta"). O objetivo específico desta etapa do trabalho foi identificar variáveis para inclusão na etapa seguinte, ou seja, *cluster analysis*.

Descrita em Green (1978), esta técnica nos permite saber se existe correlação entre as variáveis, mas não aos pares de variáveis, e sim entre grupos de variáveis.

O processamento foi feito com o apoio de STATGRAPHICS, e originalmente examinamos a correlação canônica obtida com a inclusão de todas as variáveis.

Os resultados obtidos revelaram a existência de duas correlações significantes (*p-values* iguais a 0.0000 e 0.0201). Considerando-se os coeficientes das variáveis dos dois conjuntos nas variáveis canônicas, percebeu-se que TURNASS (coeficiente igual a -0.95015) e HOLDING (coeficiente igual a -0.63962) eram as variáveis dominantes. Os dois sinais negativos, indicando relação direta entre essas duas variáveis, nos fizeram perceber que era mais recomendável excluir essas duas variáveis da análise por se tratar, na verdade, de uma relação espúria, pois, no caso de uma empresa tipo *holding*, sua base de imobilizado, estando incluída em outras empresas do grupo, acaba forçando um quociente elevado entre faturamento e ativo total (que é a medida usada para TURNASS).

Em vista disso, foi analisada em seguida a correlação canônica entre os dois conjuntos de variáveis, com a exceção de MKTSHR, do primeiro conjunto, e de HOLDING do segundo. Optamos por manter

TURNASS por se tratar de uma medida de desempenho de interesse especial em nosso trabalho ("produtividade do capital").

Neste caso, para as duas primeiras correlações canônicas obteve-se *p-values* iguais a 0.0000 e 0.0455. As variáveis com coeficientes mais elevados, em valor absoluto, na primeira variável canônica foram: TURNASS (coeficiente -0.65644), do primeiro conjunto, e INVPRO (1.16997) do segundo conjunto. Essa relação negativa também pode ser considerada espúria, pois, quanto maior o fluxo de novas aplicações (quanto maior o valor de INVPRO, portanto), maior o nível de ativo total e menor, necessariamente, o valor de TURNASS.

Portanto, numa terceira etapa resolvemos excluir a variável INVPRO do segundo conjunto, e GROWTH, para a qual havia menor interesse e definição teórica menos precisa, do primeiro conjunto.

Os resultados desta etapa são fornecidos na Tabela 3:

TABELA 3
RESULTADOS DE CORRELAÇÃO CANÔNICA EXPLORATÓRIA

Número	Correlação Canônica	Lambda de Wilks	Qui-quadrado	G.L.	Nível de Signif.
1	0.7197	0.4404	25.829	9	0.0022
2	0.2811	0.9136	2.846	4	0.5839
3	0.0893	0.9920	0.252	1	0.6156

Coefficientes para as Variáveis Canônicas (Primeiro Conjunto)

TURNASS	-0.62443	0.83793	0.31323
LABORPRO	0.85933	-0.09923	0.63382
PROFIT	-0.37249	-0.76820	0.56412

Coefficientes para as Variáveis Canônicas (Segundo Conjunto)

CAPLABOR	0.91389	-0.05559	-0.62903
LIQUID	0.13166	1.08259	0.03043
LEVERAGE	0.21634	0.36433	1.10824

Embora naturalmente os níveis de correlação canônica sejam inferiores aos obtidos com a utilização de quatro ou de cinco variáveis em cada conjunto, os resultados desta etapa nos pareceram mais satisfatórios por dois motivos:

- a) instâncias de correlação espúria foram eliminadas;
- b) algumas relações obtidas para pares de variáveis pertencentes a cada conjunto fazem sentido do ponto de vista teórico.

Quanto a este último aspecto, por exemplo, podemos citar:

- a) a relação inversa entre produtividade do capital (TURNASS) e a relação capital/trabalho (CAPLABOR), denotando um efeito de rendimento marginal decrescente (na comparação entre empresas de portes distintos) na utilização do fator capital;
- b) a relação direta entre produtividade da mão-de-obra (LABORPRO) e relação capital/trabalho (CAPLABOR), indicando que quanto mais capital estiver disponível a cada empregado mais produtivo ele tenderá a ser;
- c) a relação inversa entre rentabilidade (PROFIT) e o investimento relativo em ativos circulantes (LIQUID), confirmando o apontado na seção 4.2, com base na teoria de Finanças, e relacionado ao antigo dilema "rentabilidade vs. liquidez";
- d) a relação inversa entre rentabilidade (PROFIT) e grau de endividamento (LEVERAGE), confirmando o efeito desfavorável de alavancagem, previsível num período de redução do nível de atividade econômica.

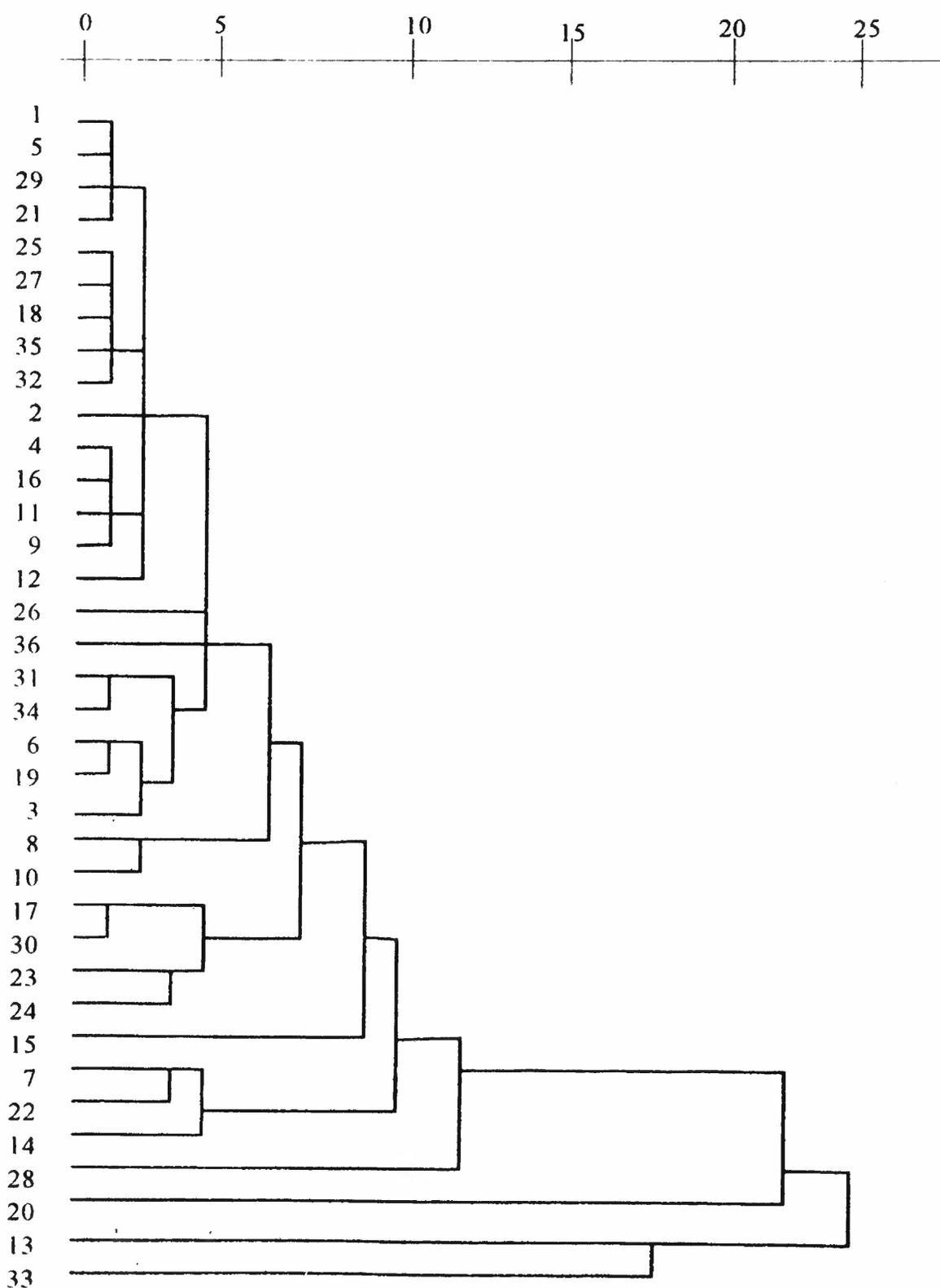
Portanto, concluímos pela utilização, na etapa de *cluster analysis*, das seis variáveis contidas na Tabela 3.

5.2. Resultados de *Cluster Analysis*

Utilizando SPSSPC, comando CLUSTER, procuramos identificar, entre as 36 empresas da amostra, subgrupos do subsetor Cal e Cimento que apresentassem certa homogeneidade.

Os resultados da análise podem ser visualizados por meio de um instrumento denominado "dendrograma" que é apresentado, para esse primeiro estágio, na Figura 1.

FIGURA 1
DISTÂNCIA DE FORMAÇÃO DE CLUSTERS



Foi adotado como critério de identificação de *clusters* a existência de agrupamentos de empresas até uma distância igual a 5. Em função disso, vemos que um grande *cluster* (C1) é formado pelas empresas 1, 5, 25, 27, 4, 16, 29, 18, 11, 35, 31, 34, 6, 19, 32, 21, 17, 30, 12, 2, 3 e 26, ou seja, 22 do total de 36 empresas. Até esse mesmo limite, quatro pares (C2 a C5) bem distintos de empresas são formados, a saber: 17 e 30, 8 e 10, 23 e 24, e 7 e 22 (no total de oito casos, portanto). Por fim, poderíamos dizer que as demais seis empresas são "atípicas", agregando-se ao resto do setor a índices de distância bem mais elevados: essas são as empresas 13, 20, 33, 28, 15 e 14. (ou *clusters* C6 a C11). (Ver Anexo I quanto à relação dos nomes de empresas correspondentes a esta numeração).

É fácil concluir, portanto, que:

- a) as empresas entre as quais uma comparação faz mais sentido, para fins de análise dentro do subsetor, seriam aquelas classificadas em cada *cluster* identificado; por exemplo, a comparação entre as empresas 1 (*cluster* C1) e 20 (*cluster* C7), possuiria fundamentos menos sólidos do que se a comparação fosse feita entre empresas do mesmo *cluster* C1,
- b) o subsetor possui empresas "atípicas", ou seja, aquelas que não são incluídas em *cluster* algum até o nível escolhido do critério de agrupamento. É claro, portanto, que as comparações dessas empresas com o restante do subsetor ficariam bastante prejudicadas. Esse é, particularmente, o caso das empresas "isoladas" nos *clusters* C6 até C11.
- c) tendo em vista que uma das utilidades da aplicação desta metodologia é a classificação, em subgrupos, das empresas com características mais semelhantes, para fins de uma comparação mais válida de desempenho e/ou conduta, é importante que se possa caracterizar o que distingue um *cluster* de outro. Isso pode ser feito, e foi efetuado, neste caso, da seguinte maneira.

Em primeiro lugar, construindo intervalos de confiança (a 5%) para a média de cada variável em C1. como este *cluster* é o maior agrupamento dentro do setor, é de se supor que aí estejam classificadas as

empresas "típicas" Em segundo lugar, comparando os índices obtidos para as empresas agrupadas nos *clusters* C2 até C11 verificando-se quais são as variáveis que se situam fora do intervalo de confiança construído com os parâmetros de C1.

Para dar um exemplo, vamos supor que o intervalo de confiança do índice de liquidez corrente (variável LIQUID), obtido para C1, é dado por $1.86 (2.080 * 5.32) = (0.26, 3.46)$. Ao examinarmos os *clusters* C2 a C11, constatamos que C9 (empresa nº 28, ou seja, Cimento Aratu) possuía índice mais elevado do que o limite superior do intervalo, e o contrário ocorria com C10 (empresa nº 15, ou seja, Cia. Cimento Atol).

Portanto, ao se caracterizar cada *cluster*, pode-se informar o seguinte:

- C2: elevado nível de desempenho nas três medidas, mas sem qualquer causa aparente observável nas variáveis de conduta;
- C3: elevada produtividade do capital;
- C4: rentabilidade e endividamento elevados; portanto, empresas que, tendo conseguido crescimento no faturamento, no período do estudo, alcançaram alavancagem financeira favorável;
- C5: elevado endividamento;
- C6: baixa produtividade do capital, elevada produtividade da mão-de-obra, causada por elevada relação capital/trabalho, mas, rentabilidade reduzida, devido ao alto grau de endividamento;
- C7: altos níveis de produtividade de capital e mão-de-obra, sem causa aparente;
- C8: idem em relação a C7, mas também com rentabilidade reduzida, causada por alto grau de endividamento;
- C9: elevada produtividade da mão-de-obra, mas com lucro reduzido, causado por elevada liquidez;
- C10: liquidez reduzida;
- C11: elevada produtividade da mão-de-obra, causada por elevada relação capital/trabalho, com alto grau de endividamento

Em todos esses casos, sempre que se afirma que há um nível "elevado" ou "reduzido" de alguma variável, está sendo dito que esses níveis estão situados fora do intervalo de confiança construído a partir da média, do desvio padrão e do número de empresas contidas em C1. Uma interpretação útil, portanto, seria a de que esses casos não contidos em C1, e que levam à formação de *clusters*, denotam a existência de empresas que, por um motivo ou outro, alcançaram desempenho e/ou adotaram políticas atípicas em relação ao setor de atividade no qual estão classificadas.

Conclusões

As contribuições deste trabalho podem ser sintetizadas e examinadas considerando dois aspectos. O primeiro refere-se à abordagem econômica e à metodologia estatística adotada. O segundo aspecto diz respeito à colaboração que o artigo traz para a melhor compreensão de um importante setor da economia nacional.

Em termos teóricos, ficou demonstrada a aplicabilidade dos conceitos desenvolvidos no campo da Organização Industrial. Foi possível identificar algumas variáveis representativas da conduta e do desempenho empresarial e estabelecer relações significativas, conforme prevê a mencionada abordagem. Quatro relações foram obtidas: 1) entre produtividade do capital, medida pelo giro do ativo total, e intensidade relativa do capital, em comparação com o volume de mão-de-obra; 2) entre produtividade da mão-de-obra e relação capital/trabalho; 3) entre rentabilidade e endividamento geral; e finalmente 4) entre rentabilidade e liquidez, ou composição dos ativos da empresa.

Com relação à abordagem estatística, o método de análise multivariada conhecido como *cluster analysis* permitiu segmentar a indústria produtora de cal e cimento em agrupamentos relativamente homogêneos, considerando a conduta e o desempenho das empresas pertencentes a esse ramo industrial.

Esta aplicação sugere a utilização desse método estatístico no estudo da concorrência e na formulação de ações estratégicas. A técnica de *cluster analysis* facilita e torna mais preciso o estudo de um setor, à medida que agrupa empresas por semelhança.

Com relação à interpretação dos resultados para a indústria analisada, 22 das 36 empresas consideradas, isto é, cerca de 60% do total, formam um único *cluster*

Os resultados revelam a homogeneidade do setor: no que tange à conduta e ao desempenho, há forte semelhança entre quase dois terços das empresas.

Referências Bibliográficas

- ANGELO, C. F. de & SANVICENTE, A. Z. Aplicação de técnicas de análise multivariada ao agrupamento de empresas por semelhança: uma crítica à análise setorial convencional" *Revista de Administração*, v. 25, n. 2, p. 20-27, abr./jun. 1990.
- BAIN, J.S. *Industrial organization*. 2ª ed. New York: Wiley, 1968.
- CLARKE, R. N. SICs as delineators of economic markets. *Journal of Business*, v. 62, n. 1, p. 17-31, jan. 1989.
- FRIEDMAN, M. The methodology of positive economics. In: *Essays in Positive Economics*. Chicago Ill.: The University of Chicago Press, 1953.
- GREEN, P. E. *Analyzing multivariate data*. Hinsdale Ill.: Dryden Press, 1978.
- GRONHAUG, K. & FREDRIKSEN, T. Concentration ratios, strategy and performance: the case of the Norwegian telecommunications industry. *Managerial and Decision Economics*, v. 9, n. 4, p. 257-262, dez. 1988.
- PANTON, D B., LESSIG, V. P. & JOY, O. M. Comovement of international equity markets: a taxonomic approach. *Journal of Financial and Quantitative Analysis*, p. 415-432, set. 1976.
- PORTER, M. E. The structure within industries and companies' performance. *Review of Economics and Statistics*, p. 214-227, maio 1979.
- SANVICENTE, A. Z. *Administração financeira*. 3ª ed. São Paulo: Atlas, 1987.
- SCHERER, F. M. *Industrial market structure and economic performance*. Chicago, Ill.: Rand McNally, 1970.
- SINDICATO Nacional da Indústria do Cimento. *Anuário Estatístico da Indústria Nacional do Cimento*, São Paulo, 1990

ANEXO 1
EMPRESAS COMPONENTES DA AMOSTRA:
INDÚSTRIA DE CAL E CIMENTO

Número	Denominação Social
1	Cia. Cimento Portland Itaú
2	Cia. Cimento Portland Paraíso
3	Cia. Cimento Portland Rio Branco
4	Itabira Agro Industrial
5	Itapessoca Agro Industrial
6	Camargo Correa Industrial
7	Cimento Cauê
8	Cia. Cimento Portland Poty
9	Cia. Materiais Sulfurosos Matsulfur
10	Cimentos do Brasil - Cibrasa
11	Cia. Cearense Cimento Portland
12	Cimento Santa Rita
13	Ciminas Cimento Nacional de Minas
14	Cimento Tupi
15	Cia. Cimento Atol
16	Cia. Paraíba Cimento Portland Cimepar
17	Cia. Cimento Portland Gaúcho
18	Itapetinga Agro Industrial
19	Cia. Cimento Itambé
20	Cia. Catarinense Cimento Portland
21	Serrana de Mineração
22	Cimento Mauá
23	Cimento Tocantins
24	Cimento Itaú do Paraná Cipsa
25	Cia. Cimento Salvador Cocisa
26	Cimbagé - Cimento Mineração Bagé
27	Ical - Indústria de Calcinação
28	Cimento Aratu
29	Cia. Cimento Portland Maringá
30	Cimento Pirineus
31	Cia. Cimento São Francisco - Cisafra
32	Ibicip Ind. Barbalhense Cim. Portland
33	Cia. Cimento Ipanema
34	Cia. Cimento Portland Ponte Alta
35	Indústria Calcários Sublime
36	Imbralit

(Originais recebidos em junho de 1991. Revistos pelos autores em novembro de 1991).