

Níveis de Informalidade na Economia Brasileira

Augusto Carvalho Souza

■ Professor da Universidade Federal de Juiz de Fora e do Instituto Vianna Junior.

Carmem Aparecida Feijó

■ Professora da Universidade Federal Fluminense (UFF).

Denise Britz do Nascimento Silva

■ Pesquisadora do IBGE e professora da Escola Nacional de Ciências Estatísticas (ENCE).

Resumo

Apresenta um retrato do setor informal urbano no Brasil e propõe uma classificação inédita dos níveis de informalidade desse setor baseada em informações estatísticas provenientes dos microdados da pesquisa Economia Informal Urbana de 1997. Os resultados do trabalho constituem subsídios valiosos para a elaboração de políticas públicas para o combate à informalidade no mercado de trabalho e na organização da estrutura produtiva do país. A metodologia utiliza a técnica de Análise de Correspondência, que se mostra bastante adequada para estabelecer os níveis de informalidade na economia brasileira. Nossa contribuição refere-se à descrição da atividade informal considerando o seu elevado grau de heterogeneidade, partindo da concepção de informalidade sugerida pela Organização Internacional do Trabalho, que define a atividade informal com foco na unidade produtiva. Esta diversidade se revela não só pela sua abrangência em termos de número de atividades como também pelas grandes diferenças na geração de receita.

Palavras-chave:

Setor informal; Políticas públicas; Análise de correspondência; Economia-Brasil; Mercado de trabalho-Brasil.

1 – INTRODUÇÃO

A preocupação com o crescimento da informalidade no mercado de trabalho no Brasil intensificou-se nos anos 1990, a partir das transformações na estrutura produtiva decorrentes, em grande medida, dos processos de abertura econômica e das privatizações (REIS; ULYSSEA, 2005; CHAHAD; ZOCKUM, 2003; FEIJÓ, 2003; CACCIA-MALI, 2001; SABÓIA, 1999, dentre outros).

Essas transformações na esfera da produção implicaram mudanças na oferta de empregos, em particular na indústria, e a precarização das condições de trabalho. Entretanto, ainda há na literatura nacional uma lacuna no que se refere à conceituação científica da informalidade – *i.e.*, não há uma definição precisa de informalidade no Brasil nem um perfil detalhado do que seja o setor informal da sua economia. O presente trabalho visa preencher este vazio propondo uma classificação inédita dos níveis de informalidade desse setor baseada em informações estatísticas de microdados provenientes da pesquisa Economia Informal Urbana de 1997 (Ecinf) do Instituto Brasileiro de Geografia e Estatística (IBGE). Neste artigo, informalidade é definida como uma categoria radial, em que o caso prototípico é caracterizado por um conjunto de componentes.

A Organização Internacional do Trabalho (OIT) apresenta uma definição de informalidade¹ com a qual os sistemas de contas nacionais identificam a participação da economia informal no Produto Interno Bruto (PIB). Nesta definição ampla da OIT, há duas abordagens distintas para a informalidade, uma orientada a setor informal (*i.e.*, pela ótica da empresa) e outra orientada à economia informal (*i.e.*, pela ótica da ocupação, do emprego)², que podem ser integradas pelo conceito de trabalho decente³. Tal definição estabelece também que entre a economia formal e a informal não há um limite nítido de separação. O que existe é um ambiente contínuo que abrange desde o trabalho

mais precário até aquele em que todas as características de um trabalho decente são encontradas.

Os autores deste artigo consideram que a forma de identificar o setor informal (*i.e.*, pela ótica da empresa) e de caracterizar a economia informal (*i.e.*, pela ótica da ocupação que inclui a forma de inserção da mão-de-obra em empresas formais e informais) são ambas bastante ricas para descrever estruturas produtivas com elevado grau de heterogeneidade na organização da produção, como ocorre no Brasil. Este duplo enfoque permite que a economia informal e o setor informal, co-existindo e se relacionando, sejam considerados como partes integrantes da estrutura formal, subordinados ao processo de desenvolvimento da economia⁴. Entretanto, no presente trabalho, consideramos apenas a abordagem do setor informal, isto é, analisamos a informalidade sob a ótica da empresa.

Esta estratégia de análise é necessária pela dificuldade de obtenção de estatísticas específicas sobre a economia informal⁵. As formas mais comumente utilizadas no Brasil para identificação das pessoas ocupadas em trabalhos informais são através de pesquisas domiciliares (sobretudo a Pesquisa Nacional por Amostra de Domicílios – PNAD), que identificam as pessoas que possuem ocupação informal por meio de perguntas para caracterizar a informalidade. Isto é, a única opção para uma análise, a ótica da ocupação, é através de *proxies* obtidas com a investigação sobre as características do trabalho dos indivíduos no que se refere à posição na ocupação (trabalhadores por conta própria, trabalhadores não-remunerados, trabalhadores domésticos etc.) e às demais características das relações de trabalho (registro em carteira, por exemplo). Por outro lado, o enfoque do setor informal dispõe dos microdados da Ecinf – uma condição necessária para os objetivos de classificação do presente estudo. A técnica estatística utilizada neste estudo é a Análise de Correspondência (GREENACRE, 1984), uma abordagem pouco comum em economia e ciências sociais que se mostrou bastante adequada para tratar e caracterizar o aspecto multidimensional e heterogêneo do setor informal urbano no Brasil.

¹ Que incorpora o conceito de trabalho decente (OIT, 1991, 1999, 2002). Para uma breve resenha sobre a evolução da conceituação de informalidade no trabalho, ver, por exemplo, Kon (2004) e Bangasser (2000).

² Na *International Classification of Status Employment* (ICSE-93), a OIT propõe que os trabalhadores formais e informais sejam classificados segundo sua posição na ocupação. Para ilustrar como deveria ser a mensuração da informalidade pela ótica da ocupação, desenvolve uma matriz de duas dimensões na qual apresenta como interação as duas principais dimensões (empresa e emprego) da economia informal.

³ Para a OIT (1999), são quatro os componentes principais de um trabalho decente: emprego, seguridade social, direitos trabalhistas e diálogo social.

⁴ Ver a esse respeito, Carneiro (1989).

⁵ Néri (2002) resume as pesquisas existentes no Brasil que podem ser utilizadas como aproximação para análise da informalidade e da qualidade do trabalho.

2 – ANÁLISE DE CORRESPONDÊNCIA

A Análise de Correspondência (BENZÉCRI, 1992 e BEH, 2004) é uma técnica estatística de análise multivariada adequada para exploração de dados categóricos que permite analisar graficamente relações entre variáveis qualitativas, principalmente quando são consideradas simultaneamente muitas variáveis que apresentam um grande número de categorias (ou níveis) de resposta (GREENACRE, 1984). A Análise de Correspondência (AC) é aplicada a tabelas de contingência (i.e., tabulações cruzadas) com o objetivo de determinar o grau de associação global entre as linhas e as colunas de uma tabela, indicando *como* as variáveis estão relacionadas. O método é utilizado para representar graficamente as linhas e as colunas de uma tabela de contingência como pontos em espaços vetoriais de pequena dimensão. Com ele, podemos avaliar se as variáveis de interesse se afastam do pressuposto de independência, a ponto de fornecer evidência suficiente para indicar que são associadas e ainda apontar como se dá esta associação; quais os grupos de e/ou níveis das variáveis são similares. Os níveis das variáveis de linha e de coluna assumem posições nos gráficos de acordo com a associação entre as variáveis e a similaridade entre os níveis⁶ das variáveis. Resumidamente, a AC busca revelar a estrutura de relação entre variáveis categóricas (nominais ou ordinais) através de representação gráfica.

Assim, a referida técnica se aplica ao estudo em questão, no qual dispomos de um grande número de variáveis investigadas na Ecinf e nosso interesse é identificar, dentre o conjunto de variáveis, características comuns que nos permitam classificar as observações de acordo com o propósito de análise da informalidade. Considerando que o uso de Análise de Correspondência não é habitual entre economistas e ci

entistas sociais e que a compreensão do método é fundamental para interpretação dos resultados do trabalho, o Apêndice 1 deste artigo apresenta os conceitos e a descrição formal da referida técnica (para maiores detalhes e exemplos ilustrativos, ver SOUZA, 2004).

⁶ As colunas podem estar definidas de acordo com níveis de uma única variável (e.g.: Ramo de Atividade, onde $Ramo_1$ = “construção”, $Ramo_2$ = “ambulante e feira livre” etc.) ou níveis da combinação de mais de uma variável (e.g.: *Ramo de Atividade x Posição do Proprietário*, sendo $Ramo-Posição_1$ = “conta própria - construção”, $Ramo-Posição_2$ = “empregador - construção”, $Ramo-Posição_3$ = “conta própria - ambulante e feira livre”, $Ramo-Posição_4$ = “empregador - ambulante e feira livre” etc.). A mesma observação é válida para a definição das linhas da tabela.

3 – A PESQUISA ECONOMIA INFORMAL URBANA

A pesquisa Economia Informal Urbana (Ecinf), realizada nas áreas urbanas de todo o Brasil no ano de 1997⁷, buscava “captar o papel e a dimensão do setor informal urbano na economia brasileira” (grifo nosso, ECONOMIA INFORMAL URBANA, 2003). A Ecinf concentrou-se em atividades não-agrícolas das áreas urbanas do país e, dentre outras características, considerou como base para definição do setor informal as unidades econômicas que operavam em pequena escala, com baixo nível de organização e quase inexistência de separação entre capital e trabalho enquanto fatores de produção. Além disto, a investigação da unidade produtiva através de pesquisa domiciliar se deu em função do reconhecimento da existência de empresas individuais e familiares funcionando em ambiente domiciliar. Assim, a Ecinf procurava, a partir de operacionalização da definição estatística de setor informal, identificar “unidades econômicas de propriedade de trabalhadores por conta própria e empregadores com até 5 empregados moradores de área urbana”, fossem elas a atividade principal de seus proprietários ou atividades secundárias. Em relação à metodologia, a Ecinf é uma pesquisa realizada através de uma amostra probabilística cuja definição envolve estratificação e conglomeração das unidades amostrais (para maiores detalhes, consultar ALMEIDA; BIANCHINI, 1998).

4 – APLICAÇÃO DA ANÁLISE DE CORRESPONDÊNCIA AOS DADOS DA ECINF 1997

As tabelas utilizadas nas análises são tabelas de contingência nas quais a interseção da linha i com a coluna j registra o número de empresas do setor informal que apresenta as características i e j , conjuntamente. Duas características de interesse têm destaque ao longo do trabalho, a saber: a área de atuação da empresa do setor informal e seu desempenho em termos da receita. Estas duas características são utilizadas como referência nas análises das demais tabelas de contingência para classificar as atividades por grau de informalidade.

⁷ A pesquisa foi novamente a campo em 2003, com publicação de seus resultados em maio de 2005.

4.1 – Especificação dos Setores, Ramos de Atividade e Níveis de Receita das Empresas do Setor Informal

Para realizar este estudo, as empresas do setor informal foram classificadas em 70 ramos de atividade⁸, que, por sua vez, foram agrupados em 12 setores de atividade. Para definir o ramo de atuação da empresa, utilizamos a variável V4301_1⁹ da Ecinf (ECONOMIA..., 2003). Esta variável informa um “código” que varia de 050 a 903, baseado na Classificação Nacional de Atividades Econômicas (CNAE) em sua versão original de 1994. A descrição da atividade referente a cada “código” é apresentada na metodologia da pesquisa (ECONOMIA..., 1996).

Os ramos de atividade foram construídos com base na comparação desta descrição com a classificação da CNAE, versão 1.0, considerando também a receita média estimada¹⁰ para os grupos definidos por cada “código” (na Ecinf). Neste procedimento, os setores são definidos a partir de agrupamentos¹¹ de ramos que compartilham características comuns (detalhes em SOUZA, 2004). Como resultado final (ver Apêndices A e B de SOUZA, 2004), o número de empresas em cada setor de atividade não é o mesmo que o apresentado nas publicações da Ecinf. Por exemplo, o ramo I3 (Transporte rodoviário de cargas, em geral, e Atividades auxiliares dos transportes terrestres) corresponde ao código 471 da Ecinf e aos códigos 60.26-7 e 63.21-5 da CNAE 1.0.

Cabe destacar que, no setor Q (Serviços diversos, mal-definidos e Atividades em branco ou sem declaração), incluímos as atividades para as quais não foi possível obter, alguma correspondência com a CNAE, principalmente por englobarem atividades muito diversas, como

no caso do código 533 da Ecinf (atribuído às empresas que têm como atividade: “aluguel de fantasia”, “serviço, inclusive municipal, de casa funerária”, “estúdio fotográfico”, entre outras).

A variável de mensuração do desempenho da empresa foi definida, conforme Tabela 1, com base na variável Receita Total (V4903)¹² e na sua categorização de acordo com o salário mínimo vigente em outubro de 1997 (R\$ 120,00). A delimitação dos níveis de receita levou em consideração a distribuição de frequência das empresas, de forma a equilibrar o número de empresas nas diferentes classes. Adicionalmente, as despesas com mão-de-obra (salários etc.), e encargos sociais (Instituto Nacional de Seguridade Social-INSS, Fundo de Garantia por Tempo de Serviço-FGTS etc.), foram descontadas da receita total das empresas de propriedade de empregadores, visando melhorar a comparabilidade entre o desempenho dessas empresas e daquelas cujos proprietários são trabalhadores por conta própria.

4.2 – Considerações Sobre o Efeito do Plano Amostral da Ecinf

A Ecinf é pesquisa cujo plano amostral envolve estratificação, conglomeração e probabilidades distintas de seleção das unidades amostrais, caracterizando-o como um plano amostral complexo (PAC), conforme definido em Pessoa e Silva (1998) e Lehtonen e Pahkinen (1996). Então, faz-se necessária a realização de ajustes na técnica estatística proposta para adequá-la à situação na qual os dados apresentam tal estrutura¹³. Neste caso, o estimador de inércia da AC deve levar em consideração o plano amostral da pesquisa. Por se tratar de uma técnica essencialmente descritiva, pouco conhecemos sobre os efeitos de conglomeração, estratificação e/ou pesos desiguais sobre seus resultados. Nyfjäll (2002a, 2002b, 2002c) apresenta estudos interessantes sobre o assunto, mas não considera planos amostrais com conglomeração e em múltiplos estágios (como é o caso da Ecinf).

Neste trabalho, as tabelas de contingência foram ajustadas para que a estimação dos perfis leve em consi-

⁸ Para uma análise mais detalhada das 70 categorias, ver Souza (2004).

⁹ Variável que investiga qual atividade é realizada na unidade econômica identificando a atividade do empreendimento, “isto é, a finalidade ou ramo do negócio, firma ou empresa do empregador, ou a natureza da atividade exercida para a pessoa que trabalhava por conta própria” (ECONOMIA..., 1996). A ECINF permite ao informante uma segunda identificação da atividade que é registrada na variável V4301_2. Esta segunda variável não foi utilizada por dois motivos principais: (i) a quantidade de informantes que forneceram um segundo ramo de atividade é bastante pequeno em relação ao primeiro e (ii) acreditamos que a primeira lembrança sobre a atividade desempenhada é mais fidedigna do que a segunda.

¹⁰ Procurou-se agrupar atividades com características e receita média semelhantes.

¹¹ Alguns setores não foram subdivididos e contam com apenas um ramo de atividade, como o setor F - Construção.

¹² O lucro não é utilizado como medida de desempenho por causa de comportamentos atípicos. Por exemplo, despesas com “Mercadorias para revenda” (exclusiva do setor de Comércio) apresentam valores expressivos e são sujeitas a efeitos de sazonalidade não observados nas outras categorias, provavelmente devido à formação de estoque para o natal.

¹³ Para uma discussão mais detalhada, ver Souza (2004).

Tabela 1 – Distribuição das empresas do setor informal urbano brasileiro em outubro de 1997 segundo classes de receita total

Códigos	Frequência Relativa (%)	Classes de receita (em salários mínimos)	Classes de receita(em R\$)
1	17,0	0 a 1	0 a 20
2	13,5	1 a 2	121 a 240
3	19,3	2 a 4	241 a 480
4	12,3	4 a 6	481 a 720
5	14,1	6 a 11	721 a 1.320
6	11,7	11 a 21	1.32 a 2.520
7	12,1	21 ou mais	2.521 ou mais

Fonte: Elaboração dos autores a partir de dados da ECINF 1997.

deração o plano amostral da pesquisa. Isto é feito com base nas sugestões de Nyfjäll (2002a, 2002b), no que se refere à utilização da AC em dados obtidos de uma pesquisa amostral com desenho complexo, e Pessoa e Silva (1998), que aborda a questão do cálculo de estatísticas em tabelas de contingência provenientes de planos amostrais complexos. Para estimar as inércias totais incorporando os efeitos do plano amostral, ponderamos cada observação pelo seu respectivo peso¹⁴. Desta maneira, é razoável supor que as tabelas incorporam o plano amostral da pesquisa no cálculo de suas estimativas de totais.

Adicionalmente, esperamos que a inércia calculada da forma tradicional seja adequada, devido aos seguintes fatos: 1) Nyfjäll (2002c) demonstra que, à medida que o tamanho da amostra aumenta, o vício relativo do estimador tradicional da inércia se torna cada vez menor sob qualquer plano amostral; 2) a quantidade de unidades primárias de amostragem (setores urbanos) selecionadas na Ecinf (2.340 setores) é mais que o dobro do tamanho da maior amostra considerada pelo autor (que é de 1.000 unidades) para a qual ele concluiu que o efeito de vício no estimador tradicional para cálculo da inércia já se torna desprezível.

4.3 – Os Componentes Classificatórios do Setor Informal

No enfoque da psicologia e da lingüística, “a categorização é o processo de compreender um objeto social sabendo a que outras coisas equivale, e de que outras coisas difere” (LU ANA, 2005). A teoria clássica da categorização sugere que uma categoria pode ser descrita como “agrupamentos de objetos caracterizados por propriedades bem definidas” (WAZLAWICK, 2005). De for-

ma resumida, a principal característica do enfoque clássico é a pressuposição da existência de atributos ou traços necessariamente comuns a todos os membros de uma categoria. Argumentos contrários a esta teoria têm como base a impossibilidade de especificação de uma lista completa de atributos para definição de uma determinada categoria, pois nem todas as categorias possuem definições tão precisas. Tais questões deram origem a abordagens alternativas, dentre elas a Teoria Prototípica, que define “a organização de categorias em torno de um conjunto de propriedades ou conjuntos de atributos correlacionados que são característicos ou típicos, rejeitando, assim, a noção de atributos definidores” (BATISTA, 2005).

Considerando-se, então, o objetivo deste trabalho, avalia-se que a utilização da abordagem clássica, na qual as categorias são baseadas em propriedades específicas compartilhadas por todos os elementos do grupo, não é adequada para representar a complexidade do sistema em questão – o setor informal –, no qual há um *continuum* de situações. Inspirados na teoria prototípica (LAKOFF, 1990), sugerimos que a informalidade seja reconhecida como uma categoria radial, na qual existe um caso prototípico ou central, mas à qual podem ser associados novos casos, variações não conhecidas, porém motivadas por este caso central.

Por esta abordagem, podemos dizer que determinados setores são fortemente informais, enquanto outros são fracamente informais, sem que haja fronteiras nítidas entre as várias situações. O caso central, prototípico, de informalidade da economia brasileira (que o presente trabalho comprova existir) é definido por um conjunto de 6 componentes/características do setor informal com relação aos ramos de atividade:

¹⁴ O IBGE fornece nos microdados de todas as suas pesquisas, quando necessário, os pesos de cada observação.

1. Nível de Receita;
2. Posição do Proprietário (conta própria ou empregador);
3. Local de Funcionamento;
4. Mercado Consumidor;
5. Controle de Contas;
6. Constituição Jurídica.

A alta informalidade é definida pelas seguintes características: baixo nível de renda; trabalho por conta própria (ao invés de condição de empregador); local precário de atuação da empresa (domicílio, ou fora do domicílio, mas sem local fixo – *e.g.*, em veículos ou em áreas públicas); mercado consumidor formado por pessoas variadas (ao invés de instituições ou clientes fixos); falta de registro contábil; e falta de constituição jurídica (*i.e.*, em condição de ilegalidade). Neste contexto, propomos categorizar os ramos de atividade do setor informal urbano em três níveis: Alta Informalidade, Média Informalidade e Baixa Informalidade (conforme apresentados no Apêndice 3). A partir da teoria prototípica, a classificação aqui apresentada não é elaborada com base em critérios estritamente numéricos (a métrica das nuvens de pontos) – adotamos uma classificação mediante o julgamento qualitativo dos diversos gráficos de correspondência do Apêndice 2. O Quadro 1, a seguir, apresenta os principais eixos de inér-

cia e suas contribuições relativas (t_{α}) para cada um dos 6 componentes analisados.

5 – ANÁLISE DOS COMPONENTES DE INFORMALIDADE

A análise dos Gráficos B.1 a B.10 permite posicionar os ramos de atividade em relação ao caso prototípico de informalidade definido pelas características dos 6 componentes do setor informal. Primeiramente, faz-se necessário ratificar as características do caso prototípico. Depois, utilizando os Gráficos de correspondência (B.1 a B.10), podemos avaliar cada ramo de atividade e construir a tabela classificatória da informalidade (Apêndice 3). O texto a seguir analisa cada um dos 6 componentes. Os códigos de variáveis e dimensões das tabelas estão descritos nas legendas dos Gráficos B.1 a B.10.

O Gráfico Receita X Ramo de Atividade (Gráfico B.1) apresenta o perfil de receita na forma de um arco¹⁵ não-simétrico onde os níveis de receita se distribuem em ordem crescente ao longo do eixo 1. Por sua vez, a projeção dos ramos neste gráfico constitui uma nuvem de pontos que segue o arco de receita. Esta configuração das projeções mostra que o setor informal se distribui conforme os níveis de renda, permitindo caracterizar a alta informalidade pela região de pontos associada à parte inicial do arco de receita. A forma do arco de receita sugere a divisão das regiões como a indicada no Gráfico B.1, em

Componente	Variáveis (linha X coluna) ⁽¹⁾	Gráfico	τ_{α} (%)		
			τ_1	τ_2	τ_3
Nível de Receita	Receita X Ramo de Atividade	B.1	70.96	17.62	<i>nsig.</i>
Posição do Proprietário	Receita X (Ramo de Atividade x Posição)	B.2	68.47	17.39	<i>nsig.</i>
Local de Funcionamento	Local X Ramo de Atividade	B.3, B.4, B.5	46.75	24.05	19.66
	Local X (Ramo de Atividade x Posição) ⁽²⁾	B.6	72.35	15.90	<i>nsig.</i>
Mercado Consumidor	Mercado Consumidor X Ramo de Atividade	B.7, B.8	73.66	21.67	4.67 ⁽³⁾
Controle de Contas	Controle de Contas X Ramo de Atividade	B.9	78.86	20.00	<i>nsig.</i>
Constituição Jurídica	Ramo X (Receita x Constituição Jurídica)	B.10	50.80	21.62	<i>nsig.</i>

Quadro 1 – Relação de gráficos e de eixos principais significativos

Fonte: Elaboração dos autores.

(1) "x" minúsculo indica agregação de códigos.

(2) Também usado na análise da componente *Posição do Proprietário*.

(3) τ_3 não é significativo. Entretanto, a tabela tem 4 linhas, o que determina um espaço de dimensão = 3. Neste caso, optamos por analisar a nuvem completa (*i.e.*, sem perda de representação): eixos (1,2), (1,3) e (2,3); mas o gráfico (2,3) foi descartado por não acrescentar informações relevantes. *nsig.*: não significativo

¹⁵ Estes arcos podem ser indicativos de um tipo de distorção chamado de *arch effect*, o que não é o caso neste trabalho.

que as seguintes classes de receita são adotadas: receita baixa, associada aos níveis 1 e 2, e receita alta, associada aos níveis 6 e 7 (com alguma proximidade do nível 5). Quanto mais próxima estiver a projeção do ramo de atividade de algum nível de receita, maior é a participação relativa deste nível de receita no perfil do ramo. Neste Gráfico B.1, também podemos observar que há desigualdade de receita dentro de um mesmo setor (e.g., pontos D1 a D21 da Indústria de Transformação estão tanto à esquerda como à direita do gráfico).

A afirmação de que o trabalho por conta própria caracteriza a alta informalidade é corroborada pelos Gráficos B.2 e B.6: *Receita X (Ramo de Atividade x Posição)* e *Local X (Ramo de Atividade x Posição)*. O Gráfico B.2 também apresenta a forma de arco (com a distribuição de receita em ordem crescente) e uma divisão clara (indicada pela reta oblíqua) entre pontos Cy e Ey (i.e., entre “conta própria” e “empregador”)¹⁶. Considerando a posição dos trabalhadores por conta própria e dos empregadores de um mesmo ramo para avaliar o desempenho destes, a pior combinação ocorre quando ambos estão do lado esquerdo (quadrantes superior e inferior esquerdo) do Gráfico B.2. Este é o caso de D17 (Fabricação de bebidas e de produtos do fumo) e F (Construção), com uma situação um pouco melhor para o segundo, que tem os dois perfis (EF e CF) no mesmo quadrante. A proximidade de EF e CF indica pouca diferença no desempenho do negócio, se for considerada a posição ocupada pelo proprietário. A melhor combinação (atividade pouco informal) é aquela na qual os trabalhadores por conta própria e empregadores de uma mesma atividade estão na região mais inferior e à direita do Gráfico B.2. Alguns setores têm todas (ou quase todas) as atividades produtivas à direita da reta oblíqua do Gráfico B.2, o que os caracteriza como setores pouco informais e bons geradores de receita, tais como J (Intermediação financeira, seguros, previdência complementar e serviços relacionados) e K (Atividades imobiliárias, aluguéis e serviços prestados às empresas). O Gráfico B.6 mostra uma ordenação ao longo do eixo 1 de local (indo de “fora do domi-

cílio” até “domicílio sem local exclusivo”) e uma concentração de empregadores (Ey) próxima ao vértice “Fora” do triângulo envolvente¹⁷, o que demonstra uma segregação espacial (além da financeira) entre trabalhadores por conta própria e empregadores.

Os Gráficos B.3, B.4 e B.5 *Local X Ramo de Atividade* e *Local X (Ramo de Atividade x Posição)* (Gráfico B.6) ratificam a afirmação de que o caso prototípico de informalidade tem local de funcionamento precário. No primeiro conjunto de gráficos, a nuvem de pontos dos ramos é atraída pela região que concentra os locais com atividades no domicílio (D, DSim e DNão)¹⁸. Estes gráficos também mostram uma proximidade da classe FL com os pontos D, DSim e DNão, o que indica haver informalidade relativamente alta em lojas, oficinas e escritórios. Os Gráficos B.3 e B.4 também revelam que a informalidade divide-se em dois conjuntos: (FL, DSim e DNão) e (FC, FVe, FPub, F). Entretanto, no segundo conjunto, há atividades de alta informalidade em vias públicas ou no domicílio do cliente, destacando-se, aí, os dois ramos mais frequentes: F (Construção) e G5 (Ambulante e Feira livre), que respondem juntos por 28,6%¹⁹ das empresas informais. Quanto ao Gráfico B.6, considerando a observação já feita no parágrafo anterior, temos a alta informalidade deslocada para o vértice DNão, isto é, para a forma mais precária de local (domicílio sem local exclusivo). As projeções de DFora e D apresentam posições que indicam pouca influência sobre os ramos de atividade.

Ainda na questão do componente Local, observa-se que a Indústria de Transformação está entre os ramos com maior nível de informalidade. De fato, de todos os ramos que apresentam uma maior associação com DNão (i.e., funcionamento no domicílio, sem local exclusivo), a maioria pertence a este setor de atividade; apenas G6 (Manutenção e reparação de artigos do vestuário e complementos), N1 (Serviços sociais) e O10 (Lavanderias e tinturarias) são exceções.

Comparando com o Gráfico B.2, percebemos, também, que os setores mais próximos de DNão são quase todos de baixo desempenho, com exceção de ED20 (No

¹⁶ Os pontos Cy estão na região de receita baixa e os pontos Ey, na de receita alta. Isto reflete a diferença na escala de produção, mas, lembrando que das despesas dos empregadores são retiradas aquelas destinadas à mão-de-obra e aos encargos, podemos afirmar, com certa segurança, que os empregadores, de um modo geral, garantem, de forma mais sólida, um nível de receita mais alta.

¹⁷ A existência deste triângulo mostra uma coerência do retrato da informalidade com relação a local, ramo e posição.

¹⁸ Estas regiões são claramente: o quadrante inferior direito do Gráfico B3, a região marcada pela curva pontilhada do Gráfico B4 e a concentração quase unidimensional em torno da seta oblíqua do Gráfico B5.

¹⁹ Situação em 1997, ano da pesquisa ECINF.

lar), e ED17 (Fabricação de bebidas e de produtos do fumo). Por outro lado, os ramos que apresentam um número substancialmente maior de empresas funcionando no domicílio, com local exclusivo, são, em sua maioria, de bom desempenho, já que estão localizados no lado direito do Gráfico B.2 (as únicas exceções são CD10 – Fabricação de acessórios do vestuário e de segurança profissional e preparação de couros e fabricação de artefatos de couro, artigos de viagem e calçados – e CO9 – Atividades de organizações religiosas e outras atividades associativas diversas). Isto indica que a existência de um local exclusivo para desempenho da atividade, quando esta é exercida no domicílio, pode influenciar positivamente o desempenho da empresa.

Notamos também que a separação das empresas de acordo com a combinação ramo de atividade e posição que o proprietário ocupa permite visualizar o efeito que esta segunda característica tem sobre a estrutura de funcionamento da empresa informal. Primeiramente, está claro que os trabalhadores por conta própria que atuam no domicílio tendem a não destinar um local exclusivo para o negócio enquanto os empregadores tendem à outra opção. Como as empresas pertencentes a trabalhadores por conta própria são a maioria, isto faz com que, ao analisar o ramo de atividade como um todo, este seja mais influenciado pelas características das empresas individuais. Para apreender isto basta perceber que os ramos associados ao ponto DNão do Gráfico B5, por exemplo, são basicamente os mesmos ramos dos trabalhadores por conta própria do Gráfico B6, sugerindo que, para estes ramos, enquanto os trabalhadores por conta

própria trabalham no domicílio de forma mais precária, os empregadores tendem a desenvolver o seu negócio fora do domicílio, a partir de uma estrutura de funcionamento aparentemente melhor.

Ainda com relação ao Gráfico B.6, destacamos novamente os ramos D20 (No lar) e D17 (Fabricação de bebidas e de produtos do fumo), por serem os únicos para os quais, tanto os trabalhadores por conta própria como os empregadores estão mais fortemente associados à projeção do ponto de precariedade DNão.

Os Gráficos B7 e B8 (*Mercado Consumidor X Ramo de Atividade*) comprovam a tese de que, no caso prototípico, o mercado consumidor da informalidade é volátil, formado por pessoas diversas (ao invés de instituições ou clientes fixos). Assim, conclui-se que o atendimento a uma clientela de pessoas variadas representa o padrão do consumidor do setor informal. De fato, neste Gráfico, a informalidade se revela por um eixo que vai do mercado de pessoas variadas (PV), até instituições (IñV e IV), com uma grande concentração nos primeiros. O caso de *Mercado X Ramo* (conforme Quadro 1) tem um espaço de dimensão 3, o que permite observar os Gráficos B7 e B8 como planos de uma nuvem tridimensional com a forma ilustrada no Gráfico 1. Por fim, devemos notar que os clientes institucionais (atípicos na informalidade) são fundamentais para alguns ramos específicos, apesar de serem menos importantes para o setor informal como um todo²⁰.

Segundo a Ecinf, uma das características das unidades produtivas do setor informal é “a quase inexistência de

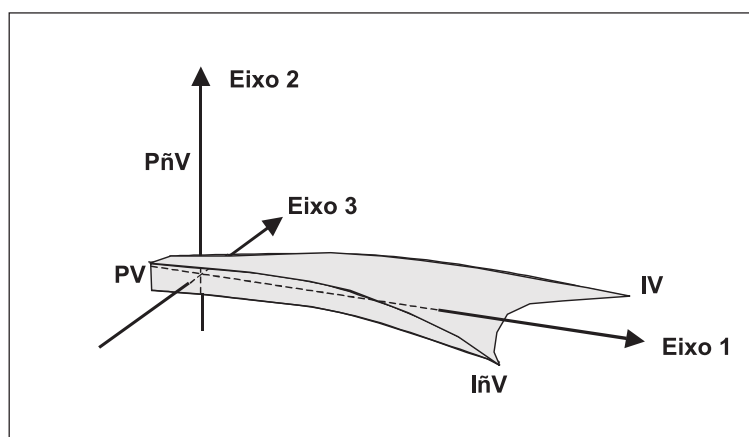


Gráfico 1 – Nuvem tridimensional de Mercado X Ramo

Fonte: Elaboração dos autores.

²⁰ O conjunto de ramos mais fortemente associados a clientes atípicos são os seguintes: PñV = {O10, N1, M, O1}, IñV = {D13, O2}, IV = {O3, J3, D12, O12} e IV ou IñV = {G12, D18, D11, D2, K5, I2}.

separação entre capital e trabalho” (ECONOMIA..., 1996). Assim, o componente “controle de contas do negócio” é um bom indicativo desta separação. As informações sobre se, e como, as contas do negócio são registradas são também úteis para determinar uma escala de informalidade, com a ausência de registro sendo indicativo de alta precariedade no desenvolvimento do empreendimento e o registro formal (por contador), sinal de que há maior qualidade no gerenciamento da atividade econômica.

O Gráfico B.9 *Controle de Contas X Ramo de Atividade* corrobora a afirmação de que o caso prototípico é formado por unidades produtivas sem qualquer tipo de contabilidade. O Gráfico B.9 revela um *continuum* de escala de formalidade, na forma de um arco, que se inicia com a presença de um contador (registro formal), passa pelos perfis associados à condição de registro pelo proprietário e chega ao caso de ausência total de registro. Existe claramente uma tendência à ausência de registro, caracterizando o controle precário de contas pelas atividades de alta informalidade. Devemos também observar que há vários ramos cujas projeções estão próximas à origem do gráfico, revelando que a sua forma de controle de contas não está associada a nenhum nível especificamente (um caso típico é o ramo G11 – Reparação de objetos pessoais e domésticos).

Finalmente, o Gráfico B.10 *Ramo X (Receita x Constituição Jurídica)* atesta a afirmação de que o caso prototípico é de unidades sem constituição jurídica. Este gráfico revela uma linha de base formada pela união dos pontos referentes à não-legalização e um *continuum* de variação de legalidade que se afunila na direção da existência de constituição jurídica (a concentração dos níveis ySim).

Como a Ecinf não considera o *status* legal da empresa como critério de informalidade, não é possível dizer que os ramos com muitas empresas não-legalizadas são “mais informais” do que os que apresentam uma alta participação relativa de empresas legalizadas. Contudo, é possível identificar o grau de legalização de um ramo de atividade através da classificação de informalidade proposta neste artigo (Apêndice 3). Conseqüentemente, podemos identificar os ramos em que há maior chance de as empresas ilegais exercerem concorrência desleal com as possíveis empresas legalizadas de mesma atividade. Os ramos que se encontram nesta última situação são os que têm as seguintes características: (1) têm a projeção do seu perfil localizado, ou associado, próxima aos pontos representantes de altos níveis de receita e

não-legalização – 5Não, 6Não e 7Não – (o caso de baixo nível de receita não é de interesse nesta análise, pois não indica vantagem com a legalização); (2) estão “na direção” de pontos que representam a ocorrência de legalização e baixos níveis de receita – 1Sim, 2Sim e 3Sim. Esta situação é caracterizada pela existência, dentro do ramo de atividade, de uma participação relativamente grande de empresas ilegais com bom desempenho. Assim, quanto mais a projeção do perfil se situa à direita, e entre os pontos 1Sim, 2Sim e 3Sim, maior a possibilidade de haver empresas no ramo de atividade que obtenham vantagens com a ilegalidade. Exemplos deste caso são os ramos de atividade N2 (Atividades de atenção à saúde e Serviços veterinários), D15 (Fabricação de produtos têxteis), G12 (Representantes comerciais e agentes do comércio), G4 (Comércio varejista de artigos do vestuário e complementos), D11 (Fabricação de tintas, vernizes, esmaltes, lacas e produtos afins e fabricação de produtos e preparados químicos diversos), entre outros.

6 – CONCLUSÕES

Este trabalho demonstra que existe uma grande heterogeneidade no setor informal urbano brasileiro em relação às várias características investigadas das empresas. Considerando estas diferenças, mostramos ser possível criar grupos de ramos de atividade no setor informal de modo a expressar uma escala de informalidade, na qual podemos indicar os ramos que apresentam uma proximidade maior ou menor com a formalidade.

Para realizar esta identificação de níveis de informalidade, tomamos como base a situação do ramo segundo o caso prototípico de informalidade, isto é: baixo nível de renda; trabalho por conta própria; local precário de atuação (domicílio ou fora do domicílio sem local fixo – como em veículos ou em áreas públicas); mercado consumidor formado por pessoas variadas (ao invés de instituições ou clientes fixos); falta de registro das contas; e falta de constituição jurídica (*i.e.*, em ilegalidade). O setor informal é, então, dividido em três níveis: alta informalidade, no qual os ramos de atividade se apresentavam em situação bastante precária de funcionamento; média informalidade, nível em que os ramos de atividade apresentavam situação intermediária e baixa informalidade, caso em que se encontram os ramos nos quais as empresas estão, claramente, em melhor situação.

Tanto a tabela classificatória da informalidade (Apêndice 3) como a análise dos 6 componentes de informalidade

(Seção 5) são instrumentos úteis para a elaboração de políticas públicas para o setor informal da economia brasileira.

Ressalta-se que não há trabalhos correlatos na literatura que nos permitam fazer comparações de resultados, no entanto, o tema abordado neste trabalho é de interesse de pesquisadores sobre economia informal e setor informal no Brasil, mormente no que se refere à discussão sobre políticas de combate à informalidade. Referências sobre o assunto estão mencionadas na introdução. Do ponto de vista de metodologia, acreditamos que o desenvolvimento de gráficos tridimensionais de correspondência é um tópico interessante de pesquisa (conforme Gráfico 1). Por fim, estamos convencidos de que devemos investigar melhor uma metodologia para tornar mais automática a classificação dos ramos de atividade com base na métrica do espaço dos perfis das variáveis de informalidade.

Abstract

It presents a picture of the urban informal section in Brazil and it proposes an unpublished classification of the levels of informality of that section based on coming statistical information of the micro data of the research Urban Informal Economy of 1997. The results of the work constitute valuable subsidies for the elaboration of public politics for the combat to the informality in the job market and in the organization of the productive structure of the country. The methodology uses the technique of Analysis of Correspondence, which is shown quite appropriate to establish the informality levels in the Brazilian economy. Our contribution refers to the description of the informal activity considering your high heterogeneity degree, leaving of the informality conception suggested by the International Organization of the Work, which defines the informal activity with focus in the productive unit. This diversity is revealed not only for your inclusion in terms of number of activities as well as for the great differences in the revenue generation.

Key words:

Informal section; Public polices; Correspondence analysis; Economy-Brazil; Market of work-Brazil.

REFERÊNCIAS

BANGASSER, P. E. **The ILO and the informal sector: an institutional history.** Employment paper. Disponível em: <<http://www.ilo.org>>. Acesso em: dez. 2004.

BATISTA, C. G. Formação de conceitos em crianças cegas: questões teóricas e implicações educacionais. **Psicologia: Teoria e Pesquisa**, Brasília, DF, v. 21, n. 1. p. 7-15, jan./abr. 2005

BEH, E. J. Simple correspondence analysis: a bibliographic review. **International Statistical Review**, v. 72, n. 2, p. 257-284, aug. 2004.

BENZÉCRI, J.-P. **Correspondence analysis handbook.** New York: Marcell Dekker, 1992.

BOND, J.; MICHAILIDIS, G. Interactive correspondence analysis in a dynamic object-oriented environment. **Journal of Statistical Software**, v. 2, n. 8, p. 1-30, 1997.

CACCIAMALI, M. C. Padrão de acumulação e processo de informalidade na América Latina: Brasil e México, **Pesquisa e Debate**, v. 12, n. 1 p. 5-42, 2001.

CARNEIRO, F. G. O setor informal urbano no Brasil e o conceito de integração do mercado de trabalho. In: ENCONTRO NACIONAL DOS CENTROS DE PÓS-GRADUAÇÃO EM ECONOMIA, 7. 1989, Brasília, DF. **Anais...** Brasília, DF: ANPEC, 1989. V. 3.

CHAHAD, J.P.Z.; ZOCKUM, M.H. A terceirização do trabalho no Brasil: um estudo de caso. In: CHAHAD, J. P. Z. (Org.); CACCIAMALI, M. C. (Org.). **Mercado de trabalho no Brasil: novas práticas trabalhistas, negociações coletivas e direitos fundamentais no trabalho.** São Paulo: Ltr, 2003. V. 1.

ECONOMIA INFORMAL URBANA 1997 – microdados. Rio de Janeiro: IBGE, 2003. 1 CD-ROM.

ECONOMIA INFORMAL URBANA. Metodologia. Rio de Janeiro: IBGE, 1996.

FEIJÓ, C. Produtividade do trabalho e emprego: o duplo desafio para os próximos anos. In: SICSÚ J. (Org.); OREIRO, J. L. (Org.); DE PAULA, L. F. (Org.). **Agenda Brasil: políticas econômicas para o crescimento com estabilidade de preços.** Barueri: Manole, 2003.

GREENACRE, M. J. **Theory and applications of correspondence analysis**. London: Academic Press, 1984.

HUSSMANN, R. A labour force survey module on informal employment (including employment in the informal sector) as a tool for enhancing the international comparability of data. MEETING OF THE EXPERT GROUP ON INFORMAL STATISTIC (DELHI GROUP), 6. 2002, Rio de Janeiro. **Anais...** Rio de Janeiro: IBGE, 2002.

KON, A. **Diversidade nas condições de informalidade**. ENCONTRO NACIONAL DE ECONOMIA, 32., 2004, João Pessoa. Disponível em: <<http://www.anpec.org.br>>. Acesso em: dez. 2004.

LAKOFF, G. **Women, fire, and dangerous things**: what categories reveal about the mind. Chicago: University of Chicago Press, 1990.

LEHTONEN, R.; PAHKINEN, E. J. **Practical Methods for design and analysis of complex surveys**. Revised edition. London: John Wiley & Sons, 1996.

LU ANA. **O efeito dos acontecimentos**: protótipo na categorização do problema da dependência do álcool e da droga. Disponível em: <<http://www.cultkitsch.org>>. Acesso em: 23 jul. 2005.

NERI, M. C. **Decent work and informal sector in Brazil**. Rio de Janeiro: EPGE/FGV, 2002.

NYFJALL, M. **Aspects on correspondence analysis plots under complex survey sampling designs**. Uppsala: Uppsala University, 2002a.

NYFJALL, M. **On correspondence analysis under complex sampling designs**. 2002. Thesis (Degree of Licentiate of Philosophy (Fil. Lic.) in Statistic) - Uppsala University, Uppsala: 2002b.

NYFJALL, M. **On the estimation of the inertia in correspondence analysis under complex survey sampling designs**. Uppsala: Uppsala University, 2002c.

OIT. **Decent work and the informal economy**. INTERNATIONAL LABOUR CONFERENCE, 90th Session, 2002, Geneva. Disponível em: <<http://www.ilo.org>>. Acesso em: 2 maio 2003

OIT. **Decent work**. INTERNATIONAL LABOUR CONFERENCE, 87th Session, 1999, Geneva. Disponível em: <<http://www.ilo.org>>. Acesso em: 2 maio 2003.

OIT. **Provisional record**. INTERNATIONAL LABOUR CONFERENCE, 78th Session, 1991, Geneva. Disponível em: <<http://www.ilo.org>>. Acesso em: 2 maio 2003.

PESSOA, D. G. C.; SILVA, P. L. N. **Análise de dados amostrais complexos**. SINAPE, 13. 1998, Caxambu. Anias... Caxambu: Associação Brasileira de Estatística, 1998.

REIS, M. C.; ULYSSEA, G. **Cunha fiscal, informalidade e crescimento**: algumas questões e propostas de política. Rio de Janeiro: IPEA, 2005. (Texto para Discussão n. 1068).

SABÓIA, J. Modernização e redução do tamanho dos estabelecimentos da indústria da transformação no passado recente. **Econômica**, Niterói, v. 1, n. 1, p. 53-74, jun. 1999.

SALM, C. **Crescimento sustentado e política de emprego**. FÓRUM NACIONAL: ECONOMIA DO CONHECIMENTO, CRESCIMENTO SUSTENTADO E INCLUSÃO SOCIAL, 16., 2004, Rio de Janeiro. Disponível em: <<http://forumnacional.org.br>>. Acesso em: dez. 2004.

SISTEMA DE CONTAS NACIONAIS: quadros sinóticos 1991/98. Disponível em: <http://www2.ibge.gov.br/pub/Contas_Nacionais/Sistema_de_Contas_Nacionais/Quadros_Sinóticos_1991_98/>. Acesso em: 23 mar. 2005.

SOUZA, A. C. **Análise de correspondência aplicada à Ecinf**: uma visão da diversidade do setor informal urbano no Brasil. Dissertação (Mestrado em Estudos Populacionais e Pesquisas Sociais) – Escola Nacional de Estatística, Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, 2004.

WAZLAWICK, R.S. **Ajustes autônomos e teoria da equilíbrio cognitiva**. Disponível em: <<http://www.labiutil.inf.ufsc.br/wazlawick0.html>>. Acesso em: 24 jul. 2005.

Recebido para publicação em 01.08.2006

APÊNDICE 1 – Conceitos e Equações da Análise de Correspondência

Este Apêndice apresenta, resumidamente, a metodologia estatística denominada Análise de Correspondência (AC), que é uma técnica de análise multivariada desenvolvida para o estudo do relacionamento existente entre variáveis qualitativas. A referida técnica permite representar graficamente as linhas e as colunas de uma tabela de contingência como pontos em espaços vetoriais de pequena dimensão, mediante a representação dos dados originais em um novo sistema de eixos ortogonais (os fatores).

A1.1 – Matriz de correspondência

Uma Tabela de Contingência, também chamada de Matriz de Dados N , contém as ocorrências n_{ij} relativas à i -ésima linha e à j -ésima coluna²¹, tal que $i = 1, \dots, I$ e $j = 1, \dots, J$. A tabela de contingência é normalmente apre-

sentada com as frequências marginais $n_{i\bullet} = \sum_{j=1}^J n_{ij}$ e

$n_{\bullet j} = \sum_{i=1}^I n_{ij}$. e a **Matriz de Correspondência P** é obtida

a partir de N , sendo $P = \frac{1}{n_{\bullet\bullet}} N$ e $p_{ij} = \frac{n_{ij}}{n_{\bullet\bullet}}$.

A1.2 – Perfis

Perfil é um vetor²² de proporções construídas a partir da linha (ou coluna) original de uma tabela (N), ao dividir cada termo n_{ij} pelo total $n_{i\bullet}$ (ou $n_{\bullet j}$). Desta forma, o perfil de uma linha i (\tilde{r}_i – um vetor de dimensão $J \times 1$) e o perfil de uma coluna j (\tilde{c}_j – um vetor $I \times 1$) podem ser definidos, respectivamente, por: $\tilde{r}_i^T = [\tilde{r}_1^i \quad \tilde{r}_2^i \quad \dots \quad \tilde{r}_J^i]$ e

$\tilde{c}_j^T = [\tilde{c}_1^j \quad \tilde{c}_2^j \quad \dots \quad \tilde{c}_I^j]$, sendo $\tilde{r}_j^i = \frac{n_{ij}}{n_{i\bullet}}$, $j=1, \dots, J$ e

²¹ As J colunas podem ser definidas de acordo com os níveis de uma única variável (e.g.: Ramo de Atividade, onde $\text{Ramo}_1 = \text{"construção"}$, $\text{Ramo}_2 = \text{"ambulante e feira livre"}$ etc.) ou níveis da combinação de mais de uma variável (e.g.: *Ramo de Atividade x Posição na Ocupação do Proprietário*, onde $\text{Ramo-Posição}_1 = \text{"conta própria - construção"}$, $\text{Ramo-Posição}_2 = \text{"empregador - construção"}$, $\text{Ramo-Posição}_3 = \text{"conta própria - ambulante e feira livre"}$, $\text{Ramo-Posição}_4 = \text{"empregador - ambulante e feira livre"}$ etc.). A mesma observação é pertinente para a definição das linhas da tabela.

²² \tilde{r}_i é definido como vetor-coluna, com dimensão $(J \times 1)$, tal que \tilde{r}_i^T , de dimensão $(1 \times J)$, é o vetor transposto.

$\tilde{c}_i^j = \frac{n_{ij}}{n_{\bullet j}}$, $i=1, \dots, I$. Demonstra-se, ainda, que o perfil

das marginais (linha e coluna) são os perfis médios das linhas (\mathbf{c}) e das colunas (\mathbf{r}), respectivamente. Assim, os perfis médios são dados por: $\mathbf{c}^T = [\tilde{c}_1 \quad \tilde{c}_2 \quad \dots \quad \tilde{c}_J]$ e $\mathbf{r}^T = [\tilde{r}_1 \quad \tilde{r}_2 \quad \dots \quad \tilde{r}_I]$, sendo $j=1, \dots, J$ e $i=1, \dots, I$.

A técnica de AC emprega os perfis e não as frequências absolutas na análise, uma vez que valores absolutos diferentes podem apresentar perfis iguais. Então, a análise feita a partir de valores absolutos conduzirá a uma diferenciação de elementos que são essencialmente iguais. Generalizando, podemos definir as matrizes dos perfis de linha e coluna, respectivamente por:

$\mathbf{R} = \mathbf{D}_r^{-1} \mathbf{P}$ e $\mathbf{C} = \mathbf{D}_c^{-1} \mathbf{P}^T$, sendo $\mathbf{D}_r = \text{diag}(\mathbf{r})$ a matriz diagonal do perfil da coluna marginal e $\mathbf{D}_c = \text{diag}(\mathbf{c})$, a matriz diagonal do perfil da linha marginal.

A1.3 – Massas

A cada perfil, de linha ou coluna, da Matriz de Correspondência P , pode ser associado um valor chamado de massa, que mede a importância relativa da linha ou coluna. O vetor de massas das linhas é igual ao perfil médio das colunas, \mathbf{r} , e vetor de massas das colunas é igual ao perfil médio das linhas, \mathbf{c} .

A1.4 – Nuvem de pontos e a sua representação gráfica

O perfil de uma linha i pode ser interpretado como sendo o vetor de coordenadas do nível i da variável que compõe as linhas da tabela no espaço vetorial definido pelos J níveis da variável que compõe as colunas (e vice-versa). A dimensão deste espaço pode, entretanto, ser apenas $J-1$, visto que a soma dos valores em um perfil é sempre igual a 1. Desta maneira, uma Tabela de Contingência com 4 colunas terá os perfis de linhas representados em um espaço de 3 dimensões, onde cada linha é representada por um ponto cujas coordenadas são dadas pelos elementos componentes deste perfil (Figura A1.1). Se associarmos a cada ponto (i.e., a cada perfil) o valor da massa, temos que este conjunto de pontos é chamado de nuvem de pontos. Ou seja, a nuvem de pontos é o conjunto de

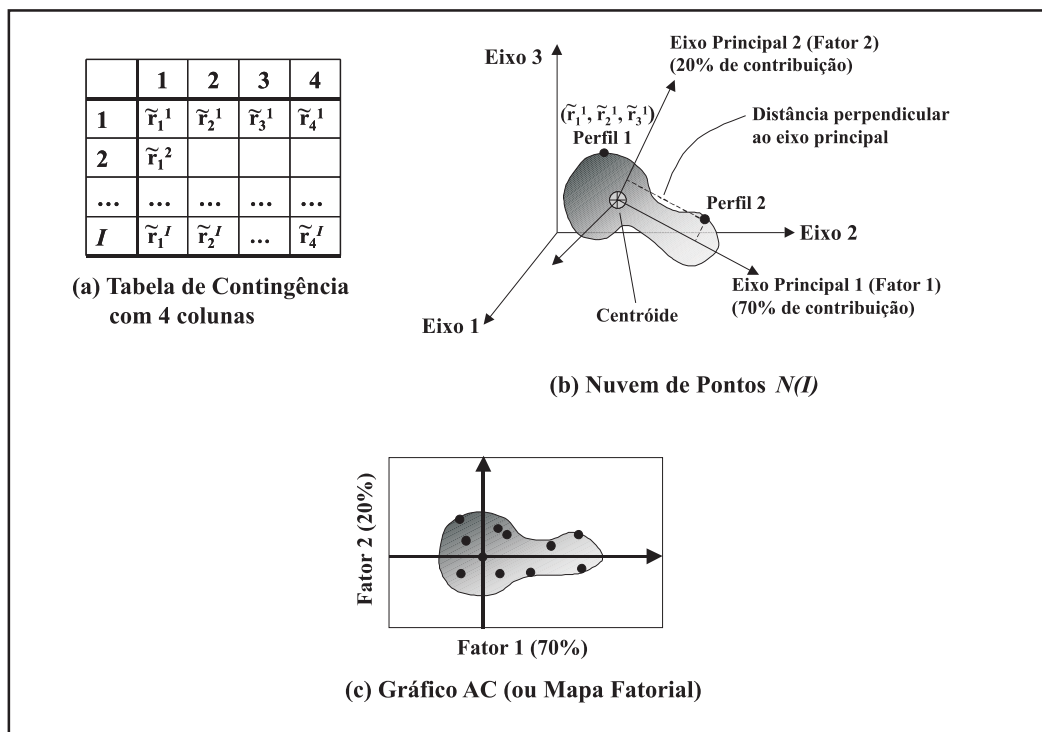


Figura A1.1 – Nuvem de pontos e gráfico AC com eixos principais

Fonte; Elaboração dos autores

todos os pares formados por um perfil, de linha ou coluna, e a massa desta linha ou coluna.

Formalmente, temos:

- A nuvem de pontos $N(I)$, que é o conjunto de pontos $i \in I$, cujas coordenadas são os componentes do perfil \mathbf{r}_i , ponderados pelas respectivas massas;
- A nuvem de pontos $N(J)$, que é o conjunto de pontos $j \in J$, cujas coordenadas são os componentes do perfil \mathbf{c}_j , ponderados pelas respectivas massas.

O objetivo da AC é representar a nuvem de pontos, situada em um espaço n -dimensional, em um espaço de menor dimensão (e.g. o plano), de forma que possa revelar a relação existente entre as variáveis estudadas. Este novo espaço de representação da nuvem de pontos é obtido encontrando-se os eixos fatoriais que melhor se ajustam aos pontos da nuvem. Prova-se que este subespaço ótimo é formado pelos eixos fatoriais que passam pelo perfil médio (ou centro de gravidade) e que minimizam as distâncias dos pontos até o subespaço (BENZÉCRI, 1992). Neste caso, podemos representar graficamente os pontos em relação apenas a dois eixos e, desta maneira, ter uma boa representação (mesmo que aproximada) da nuvem de pontos em um espaço de apenas 2 dimensões.

A1.5 - Distâncias

Uma vez que a AC define como $N(I)$ e $N(J)$ estão dispostos em torno de seus respectivos centros de gravidade, então torna-se necessária a utilização de uma noção de distância. Para medir a distância entre dois pontos dos perfis das linhas i e i' na nuvem $N(I)$, usamos a Distância Euclidiana Ponderada que atribui um peso menor às coordenadas sujeitas a grandes graus de variabilidade em relação àquelas que não apresentam grande variação. A métrica utilizada na AC para medir a distância entre dois perfis de linha é, então, definida por $d^2(\tilde{\mathbf{r}}_i, \tilde{\mathbf{r}}_{i'}) = (\tilde{\mathbf{r}}_i - \tilde{\mathbf{r}}_{i'})^T \mathbf{D}_e^{-1} (\tilde{\mathbf{r}}_i - \tilde{\mathbf{r}}_{i'})$. Uma equação similar pode ser escrita para a distância entre duas colunas j e j' em $N(J)$. Esta métrica denominada como Distância Qui-quadrado por apresentar uma relação entre a distância a ela associada e a estatística de Qui-quadrado de Pearson (GREENACRE, 1984; BENZÉCRI, 1992; BEH, 2004).

A1.6 – Inércia

A inércia é a medida de qualidade do ajuste da nuvem de pontos ao novo espaço. Esta é uma medida de variação total, que objetiva captar a dispersão dos pontos da nuvem em torno do seu centro de gravidade. A Inércia Total da nuvem de pontos formada pelos perfis de linha $N(I)$ e pelos perfis de coluna $N(J)$ são, respectiva-

mente, $in(I) = \text{traço} \left[\mathbf{D}_r (\mathbf{R} - \mathbf{1c}^T) \mathbf{D}_c^{-1} (\mathbf{R} - \mathbf{1c}^T)^T \right]$ e $in(J) = \text{traço} \left[\mathbf{D}_c (\mathbf{C} - \mathbf{1r}^T) \mathbf{D}_r^{-1} (\mathbf{C} - \mathbf{1r}^T)^T \right]$. A inércia total dos perfis de coluna é igual à dos perfis de linha e ambas iguais à conhecida estatística Qui-quadrado para o teste de independência, dividida pelo total de observações n , tal que:

$$in(I)=in(J)=\sum\sum\frac{(p_{ij}-c_ir_j)^2}{c_ir_j}=\frac{\chi^2}{n}=tr\alpha\phi\left[\mathbf{D}_r^{-1}(\mathbf{p}-\mathbf{r}\mathbf{c}^T)\mathbf{D}_c^{-1}(\mathbf{p}-\mathbf{r}\mathbf{c}^T)^T\right]$$

Como o objetivo da AC é reduzir pontos vetoriais localizados em espaços de muitas dimensões a um espaço de dimensão inferior, torna-se importante conhecer a contribuição de cada fator (ou eixo fatorial) para a inércia total. A inércia é decomposta, então, pelos eixos principais de inércia (ou eixos fatoriais, ou simplesmente fatores) que são aqueles que definem o espaço ótimo (ver Figura A1.1-b). Prova-se que a inércia total pode ser obtida através da soma do quadrado dos valores singulares (autovalores) de $\mathbf{A} = \mathbf{D}_r^{-1/2} (\mathbf{P} - \mathbf{r}\mathbf{c}^T) \mathbf{D}_c^{-1/2}$. Sejam \mathbf{U} , \mathbf{D}_μ e \mathbf{V} matrizes²³ obtidas a partir da decomposição do valor singular da matriz \mathbf{A} , então $in(I) = in(J) = \text{traço}[\mathbf{D}_\mu^2] = \lambda_1^2 + \lambda_2^2 + \dots + \lambda_\alpha^2 + \dots + \lambda_r^2$, em que r é o número total de eixos fatoriais da nuvem e λ_α o α -ésimo valor singular de \mathbf{A} . Assim, cada valor singular está associado a um eixo fatorial, de tal forma que o primeiro valor singular (ao quadrado), λ_1^2 , se refere à parcela da inércia total captada pelo primeiro eixo principal de inércia e assim sucessivamente. Com isso, temos que a contribuição relativa do eixo para a inércia total da nuvem é a razão $\tau_\alpha = \lambda_\alpha^2 / \text{traço}[\mathbf{D}_\mu^2]$ e a contribuição relativa do conjunto dos α primeiros eixos para a inércia total é dada por:

$$\text{Porcentagem Acumulada}(\alpha) = \frac{(\lambda_1^2 + \lambda_2^2 + \dots + \lambda_\alpha^2)}{\lambda_1^2 + \lambda_2^2 + \dots + \lambda_\alpha^2 + \dots + \lambda_r^2}.$$

O número de eixos com contribuição claramente significativa determina a dimensão reduzida da análise, ou seja, o subespaço de análise será formado pelos primeiros α eixos. (ver exemplo da Figura A1.1). Esta decomposição é utilizada para descrever a qualidade da representação da nuvem de pontos original no subespaço ótimo.

²³ **U** contém os vetores singulares à esquerda de **A**, **V** os valores singulares à direita de **A** e **D** é uma matriz diagonal de números positivos, valores singulares de μ **A**, em ordem decrescente.

Outras medidas de qualidade são as seguintes: COR – contribuição relativa do eixo α na inércia do ponto; QLT – indica a qualidade da representação de um ponto em um subespaço e é obtido a partir da soma das contribuições relativas dos eixos considerados à inércia do ponto; CTR – a contribuição relativa do elemento i para λ_{α_i} ; INR – a contribuição relativa do elemento i para a inércia total.

A1.7 – Coordenadas dos perfis no novo sistema de eixo fatorial

As coordenadas dos pontos de cada nuvem são definidas separadamente para os perfis de linha e de coluna, com cada conjunto de coordenadas referenciando-se aos seus próprios eixos principais. Assim, as coordenadas principais dos perfis de linha e coluna, respectivamente, são (GREENACRE, 1984):

$\mathbf{F} = \mathbf{D}^{-1/2} \mathbf{U} \mathbf{D} \mathbf{U}^\mu$ e $\mathbf{G} = \mathbf{D}^{-1/2} \mathbf{V} \mathbf{D} \mathbf{V}^\mu$. Como as linhas das matrizes \mathbf{F} e \mathbf{G} definem as projeções das nuvens de pontos N_i e N_j nos eixos formadores dos espaços ótimos e, com base na decomposição do valor singular da matriz formada pela diferença entre as frequências observadas e esperadas ($\mathbf{P}-\mathbf{rc}^T$), podem-se representar os conjuntos de coordenadas \mathbf{F} e \mathbf{G} em um mesmo sistema de eixos coordenados utilizando-se as *fórmulas de transição*: $\mathbf{G} = \mathbf{CFD}^{-1}$ e $\mathbf{F} = \mathbf{RGD}^{-1}$.

Na disposição gráfica conjunta de ambas as nuvens de pontos em um mesmo sistema coordenado utilizando as fórmulas de transição, denominada simétrica, os perfis semelhantes tendem a ter projeções próximas. Entretanto, a projeção conjunta das nuvens apenas indica a tendência das similaridades, a natureza da dispersão e as direções globais da correspondência entre as nuvens. Neste caso, devemos evitar o perigo de interpretar distâncias entre pontos de nuvens diferentes, visto que uma métrica comum não foi explicitamente definida. As distâncias entre pontos dentro de uma mesma nuvem são definidas em termos da distância Qui-quadrado, enquanto a correspondência entre nuvens é governada pela natureza baricêntrica das fórmulas de transição (GREENACRE, 1984). Uma solução para que as distâncias entre as projeções das nuvens tenham significado é normalizar as coordenadas em relação a inércias unitárias, produzindo as chamadas Coordenadas Padronizadas (GREENACRE, 1984). As coordenadas padronizadas dos perfis de linha e de coluna, respectivamente, são dadas por:

$$\Phi = \mathbf{F}\mathbf{D}_r^{-1} = \mathbf{D}_r^{-1/2}\mathbf{U} \quad \text{e} \quad \Gamma = \mathbf{G}\mathbf{D}_c^{-1} = \mathbf{D}_c^{-1/2}\mathbf{V}.$$

Na prática, escolhemos um dos perfis para ser representado por coordenadas padronizadas e o outro por coordenadas principais. Neste caso, temos Projeções Assimétricas (e Gráficos Assimétricos), onde as seguintes interpretações são possíveis:

- Quanto mais próxima estiver uma projeção i de uma projeção j , maior é a participação relativa da categoria j no perfil da linha i .
- Quando $i = j$, a linha i só apresenta observações referentes à coluna j (i.e., na tabela de contingência, a coluna j é nula, exceto na linha i).

APÊNDICE 2 – Gráficos de Correspondência do Setor Informal

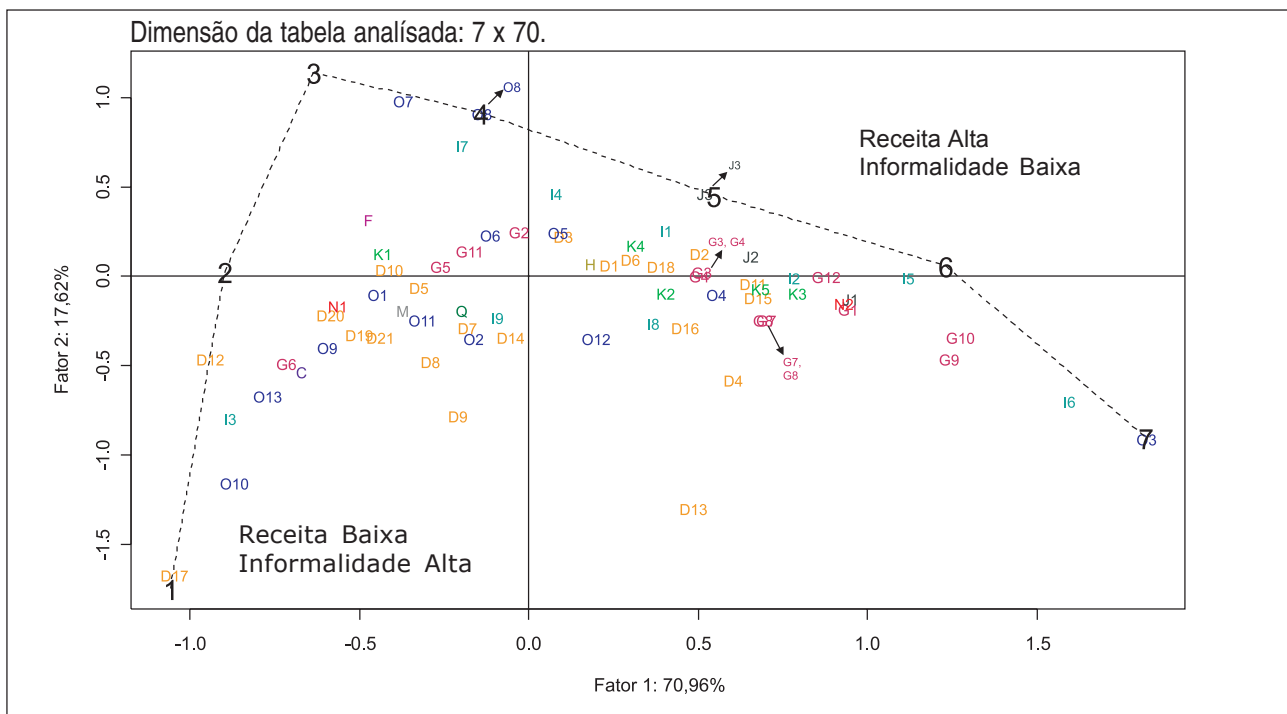


Gráfico B1 – Nível de receita X ramo de atividade

Códigos: Receita (1 a 7) na Tabela 1; Ramos de atividade no Apêndice 3.

Fonte: Elaboração dos autores.

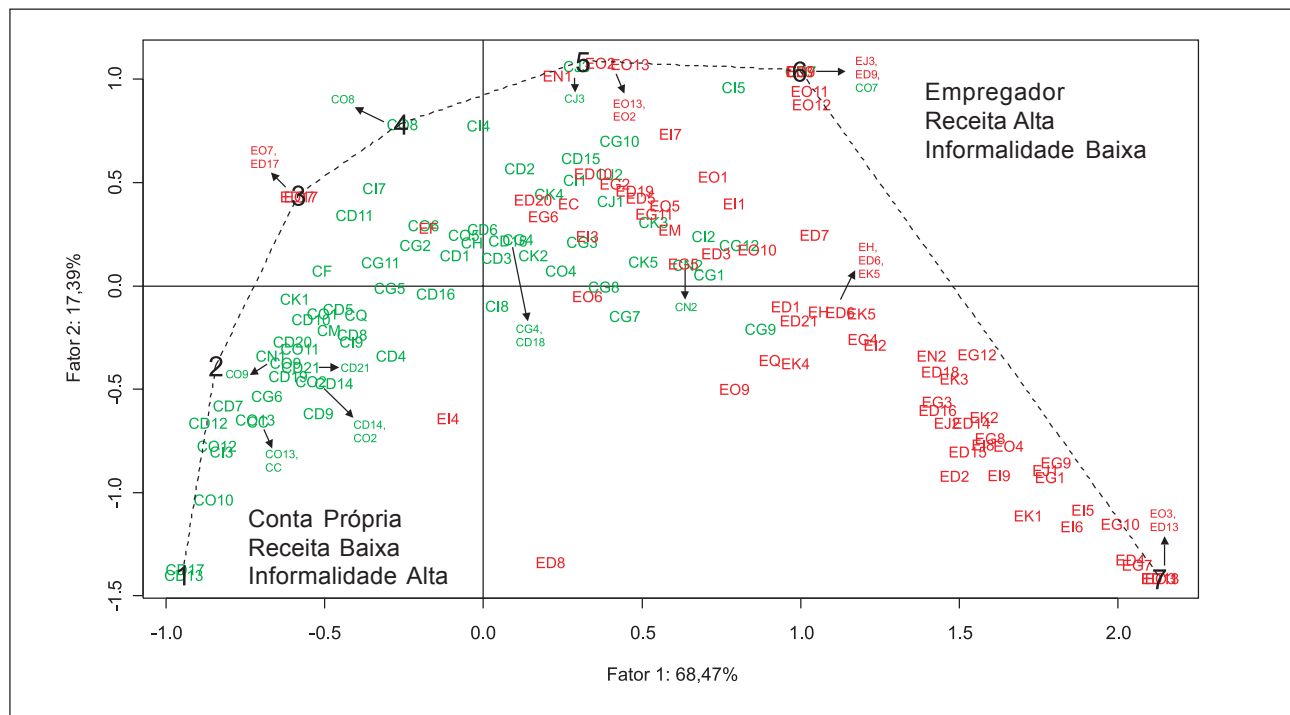


Gráfico B2 – Nível de receita X (ramo de atividade x posição na ocupação).

Códigos: Receita (1 a 7) na Tabela 1; Ramo x Posição: (Cy – conta própria, Ey – empregador, sendo y o código de Ramo de atividade do Apêndice 3).

Dimensão da tabela analisada: 7 x 136.

Fonte: Elaboração dos autores.

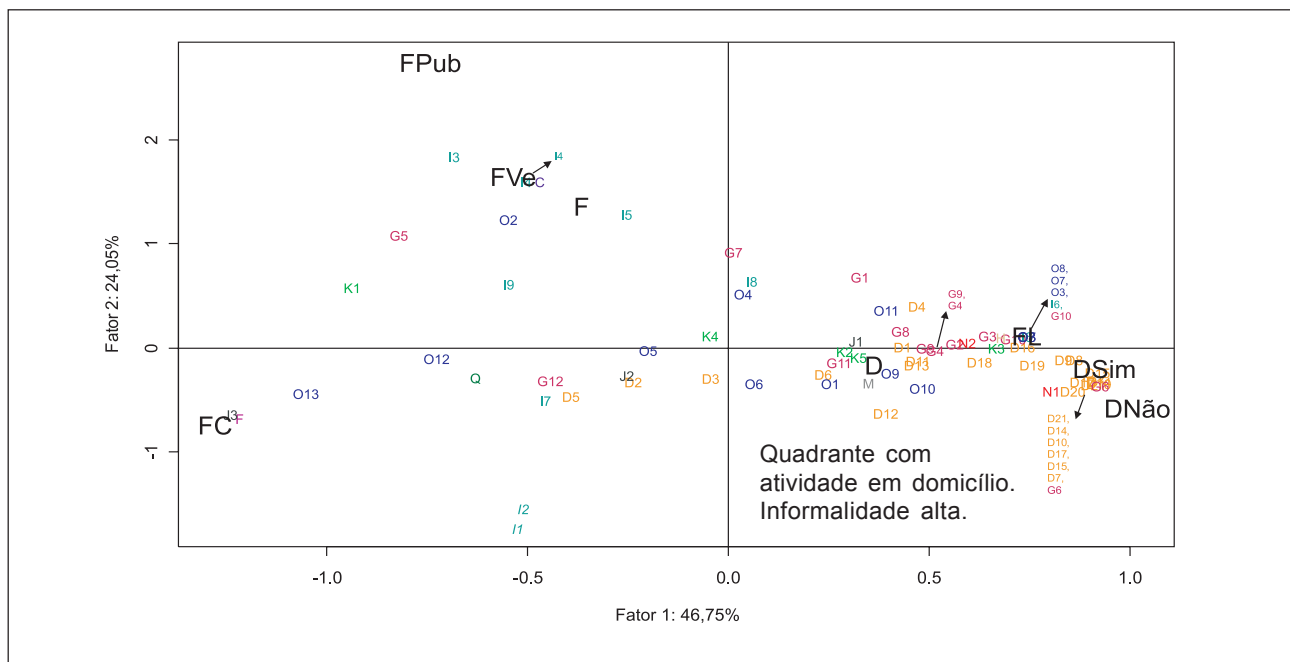


Gráfico B3 – Local X ramo de atividade (eixos 1 e 2).

Códigos: Local (DSim – domicílio com local exclusivo, DNão – domicílio e não tem local exclusivo, D – domicílio ou casa do sócio, FL – loja, oficina ou escritório, FC – domicílio do cliente ou local por ele designado, FVe – veículo automotor, FPub – via ou área pública, F – fora do domicílio); Ramo (os mesmos códigos do Gráfico B1). A variável Local corresponde à agregação das variáveis V4302, V4303 e V4304 do questionário da ECINF.

Dimensão da tabela: 8 x 70.

Fonte: Elaboração dos autores.

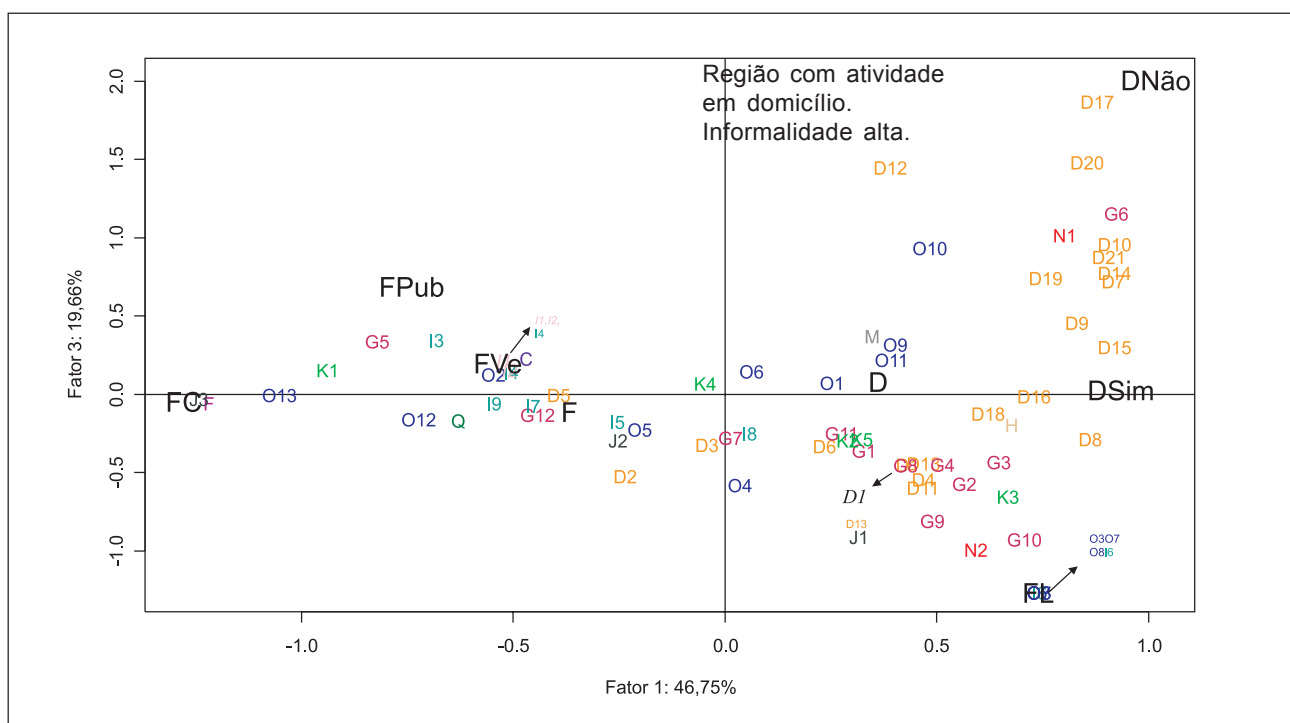


Gráfico B4 – Local X ramo de atividade (eixos 1 e 3).

Códigos e dimensão da tabela: os mesmos do Gráfico B3.

Fonte: Elaboração dos autores.

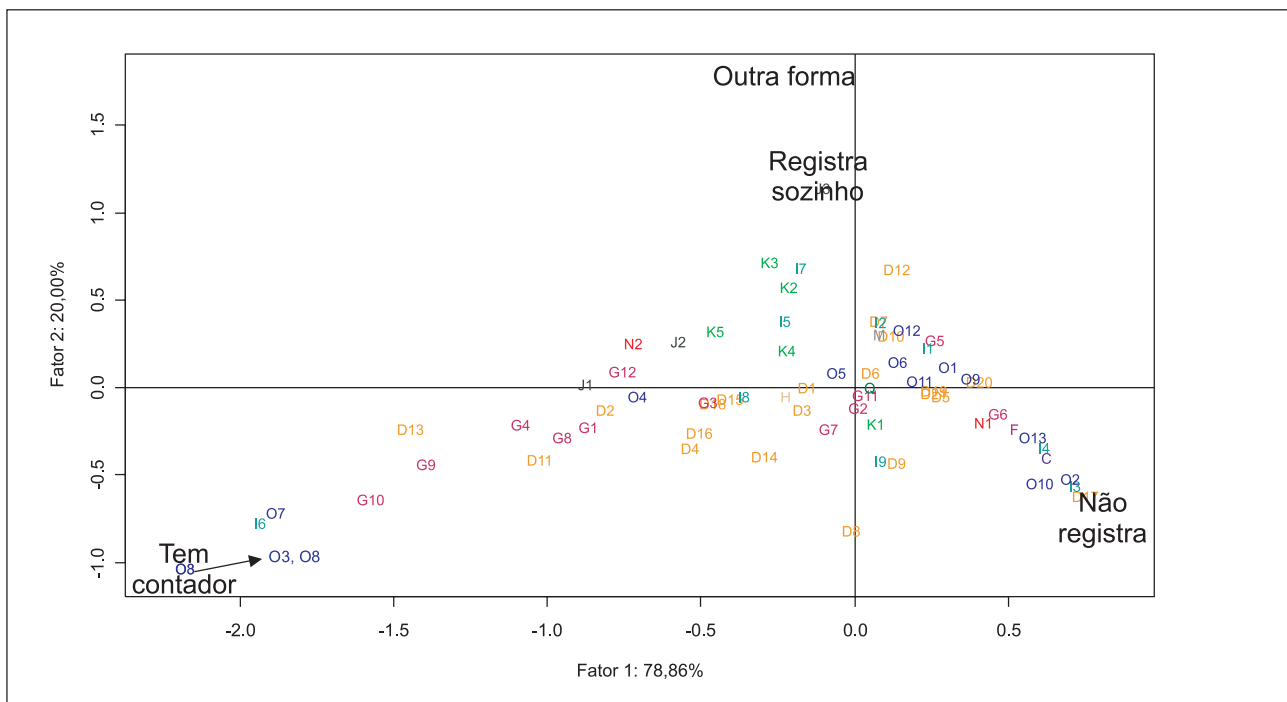


Gráfico B9 – Controle de contas x ramo de atividade.

Códigos: Controle de Contas (Registra sozinho, Tem contador, Outra forma, Não registra); Ramo (os mesmos códigos do Gráfico B1).

Dimensão da tabela 4x70.

Fonte: Elaboração dos autores.

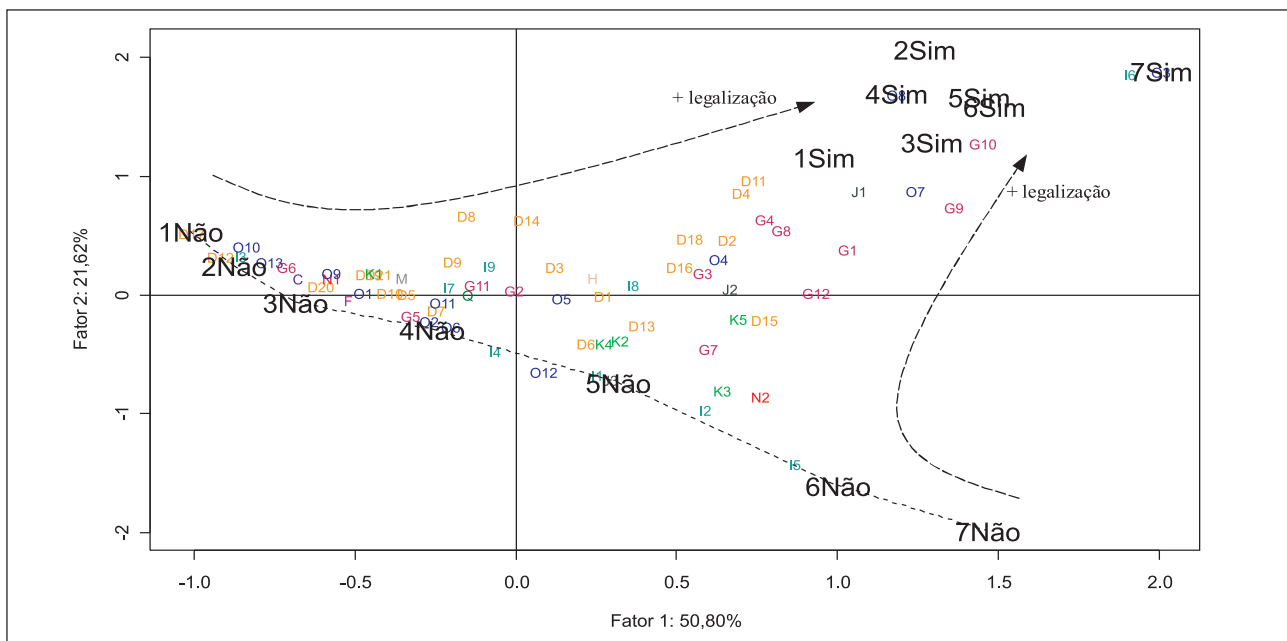


Gráfico B10 – Ramo de atividade x (nível de receita x constituição jurídica).

Nota: Ao contrário dos outros gráficos, os perfis de linha são projetados como Coordenadas Principais e os perfis de coluna como Coordenadas Padronizadas.

Códigos: Ramo (os mesmos códigos da Figura B.1); Receita x Constituição (ySim – nível de receita y tem constituição jurídica, yNão – nível y não tem constituição jurídica; sendo y um dos níveis 1 a 7 da Tabela 1).

Dimensão da tabela 70x14.

Fonte: Elaboração dos autores.

APÊNDICE 3 – Classificação do Setor Informal na Economia Brasileira

Setores e Ramos de Atividade

	Setor C – INDÚSTRIA EXTRATIVA MINERAL
	Setor D – INDÚSTRIA DE TRANSFORMAÇÃO
D5	Fabricação de produtos de madeira
D8	Fabricação de artigos de borracha
D9	Fabricação de produtos de material plástico
D10	Fabricação de acessórios do vestuário e de segurança profissional e preparação de couros e fabricação de artefatos de couro, artigos de viagem e calçados
D12	Fabricação de produtos químicos inorgânicos e inorgânicos
D13	Fabricação de resinas e elastômeros, fabricação de fibras, fios, cabos e filamentos contínuos artificiais e sintéticos e fabricação de produtos farmacêuticos
D14	Fabricação de sabões, detergentes, produtos de limpeza e artigos de perfumaria
D17	Fabricação de bebidas e de produtos do fumo
D19	Fabricação de equipamentos de instrumentação médico-hospitalares, instrumentos de precisão e ópticos, equipamentos para automação industrial, cronômetros e relógios
D20	No lar
D21	Confecção de artigos do vestuário e acessórios
	Setor F – CONSTRUÇÃO
	Setor G – COMÉRCIO
G5	Ambulante e Feira livre
G6	Manutenção e reparação de artigos do vestuário e complementos
	Setor I – TRANSPORTE, ARMAZENAGEM empregador COMUNICAÇÕES
I3	Transporte rodoviário de cargas, em geral, e Atividades auxiliares dos transportes terrestres
	Setor K – Atividades imobiliárias, aluguéis e serviços prestados às empresas
K1	Vigilância e segurança por conta própria
	Setor M – Educação
	Setor N – Saúde e serviços sociais
N1	Serviços sociais
	Setor O – Outros serviços coletivos, sociais e pessoais
O1	Atividades desportivas, Cabeleireiros e outros tratamentos de beleza e Atividades de manutenção do físico corporal
O2	Limpeza urbana e esgoto e atividades relacionadas e Atividades funerárias e serviços relacionados
O9	Atividades de organizações religiosas e Outras atividades associativas diversas
O10	Lavanderias e tinturarias
O11	Exploração de jogos de azar, prostituição e atividades de astrólogos, videntes e similares
O13	Outras atividades de serviços pessoais

Quadro A3.1.1 – Especificação dos ramos e setores de atividade de alta informalidade

Fonte: Elaboração dos autores.

Setores e Ramos de Atividade

Setor D – INDÚSTRIA DE TRANSFORMAÇÃO	
D1	Fabricação de produtos de minerais não-metálicos e de produtos de metal – exceto máquinas e equipamentos
D3	Fabricação de material eletrônico e de aparelhos e equipamentos de comunicações
D4	Fabricação de outros equipamentos de transporte
D6	Fabricação de móveis e indústrias diversas
D7	Fabricação de celulose, papel e produtos de papel e reciclagem
Setor G – COMÉRCIO	
G2	Manutenção e reparação de veículos automotores
G11	Reparação de objetos pessoais e domésticos
G12	Representantes comerciais e agentes do comércio
Setor H – ALOJAMENTO empregador ALIMENTAÇÃO	
Setor I – TRANSPORTE, ARMAZENAGEM empregador COMUNICAÇÕES	
I1	Transporte rodoviário de passageiros, não-regular
I4	Transporte aquaviário
I7	Telecomunicações A - atividades de manutenção operacional das redes de telecomunicações e serviço telefônico público e os postos telefônicos
I9	Atividades auxiliares dos transportes e Manutenção e reparação de veículos automotores
Setor K – Atividades imobiliárias, aluguéis e serviços prestados às empresas	
K2	Atividades imobiliárias
K4	Aluguel de veículos, máquinas e equipamentos sem condutores ou operadores e de objetos pessoais e domésticos
Setor O – Outros serviços coletivos, sociais e pessoais	
O5	Atividades recreativas, culturais e desportivas em geral
O6	Atividades de teatro, música e outras atividades artísticas e literárias
O7	Atividades de organizações empresariais, patronais e profissionais e Atividades de organizações sindicais
O8	Atividades de bibliotecas, arquivos, museus e outras atividades culturais
O12	Atividades de rádio
Setor Q – Serviços Diversos	

Quadro A3.2 – Especificação dos ramos e setores de atividade de média informalidade

Fonte: Elaboração dos autores.

Setores e Ramos de Atividade

	Setor D – INDÚSTRIA DE TRANSFORMAÇÃO
D2	Fabricação de máquinas e equipamentos
D11	Fabricação de tintas, vernizes, esmaltes, lacas e produtos afins e fabricação de produtos e preparados químicos diversos
D15	Fabricação de produtos têxteis
D16	Fabricação de produtos alimentícios
D18	Edição, impressão e reprodução de gravações
	Setor G – COMÉRCIO
G1	Comércio relacionado a veículos
G3	Comércio varejista de produtos alimentícios, bebidas e fumo
	Comércio varejista de artigos do vestuário e complementos
G7	Comércio atacadista de matérias-primas agrícolas, animais vivos; produtos alimentícios para animais
G8	Comércio varejista de outros produtos
G9	Comércio varejista de material de construção, ferragens, ferramentas manuais; vidros, espelhos e vitrais; tintas e madeiras
G10	Comércio varejista não-especializado
	Setor I – TRANSPORTE, ARMAZENAGEM empregador COMUNICAÇÕES.
I2	Transporte rodoviário de cargas, em geral
I5	Transporte aéreo
I6	Correio e outras atividades de entrega
I8	Atividades auxiliares dos transportes
	Setor J – Intermediação financeira, seguros, previdência complementar e serviços relacionados
J1	Intermediação financeira
J2	Seguros de vida e não-vida
J3	Atividades auxiliares da intermediação financeira
	Setor K – Atividades imobiliárias, aluguéis e serviços prestados às empresas
K3	Atividades jurídicas e despachantes
K5	Serviços prestados principalmente às empresas
	Setor N – Saúde e serviços sociais
N2	Atividades de atenção à saúde e Serviços veterinários
	Setor O – Outros serviços coletivos, sociais e pessoais
O3	Limpeza urbana e esgoto e atividades relacionadas
O4	Atividades desportivas e Outras atividades relacionadas ao lazer

Quadro A3.3 – Especificação dos ramos e setores de atividade de baixa informalidade

Fonte: Elaboração dos autores.