

UTILIZAÇÃO DA ANÁLISE FATORIAL PARA ORIENTAR A CRIAÇÃO DE POLÍTICAS PÚBLICAS NA ÁREA DO TRABALHO, NAS REGIÕES ADMINISTRATIVAS DE ARAÇATUBA, PRESIDENTE PRUDENTE E MARÍLIA

Luiz Ricardo Nakamura¹, Antônio Assiz de Carvalho Filho², Eduardo Cardoso de Oliveira³, Ricardo Firetti⁴

¹ Programa de Pós-Graduação em Estatística e Experimentação Agronômica – ESALQ/USP

² Departamento de Matemática, Estatística e Computação – FCT/UNESP

³ Programa de Pós-Graduação em Especialização em Estatística com Ênfase em Educação – UEL

⁴ Pesquisador Científico do DDD/APTA, Presidente Prudente, SP

RESUMO

Este trabalho visa, com o auxílio da análise fatorial, resumir e simplificar a interpretação das variáveis relacionadas a número de vínculos empregatícios – na agropecuária, no comércio, na construção civil, na indústria e em serviços – com o intuito de identificar e indicar possíveis propostas de políticas públicas e/ou privadas na área do trabalho. A área temática em questão foi estudada em 3 Regiões Administrativas (R.A.) do Estado de São Paulo – R.A. de Araçatuba, R.A. de Presidente Prudente e R.A. de Marília –, também conhecidas como 9^a, 10^a e 11^a R.A. do Estado de São Paulo, respectivamente. O objetivo principal deste trabalho foi o de formar e caracterizar grupos de municípios homogêneos de acordo com suas similaridades identificadas por variáveis estatísticas, a fim de auxiliar na elaboração de estratégias empresariais e de formulação de políticas públicas, embasadas com informações relevantes para o desenvolvimento sócio-econômico das regiões em estudo.

Palavras-chave: Análise Multivariada; Análise Fatorial; Desenvolvimento Regional; Políticas Públicas.

1 INTRODUÇÃO

Tendo em vista a grande quantidade de informações estatísticas oficiais na área do trabalho em sites oficiais de armazenamento de dados, a aplicação de técnicas estatísticas a esses dados dará um embasamento para análise, interpretação e explicação de vários fenômenos, permitindo a elaboração de metodologias para a obtenção e análise de índices estatísticos para orientar e aproximar os agentes (empresas, fornecedores, instituições públicas ou privadas) oferecendo subsídios para a discussão e análise de diagnósticos e desenhos de diretrizes e ações, tanto no âmbito das políticas públicas, quanto daquelas iniciativas que deveriam ser adotadas pelo setor privado e pelas mais diferentes organizações da sociedade civil, em seu compromisso com o desenvolvimento regional e com a responsabilidade social nas regiões do Estado de São Paulo que serão estudadas.

Sintetizando as idéias de Porter (1999), pode-se afirmar que a localidade já não exerce grande influência na estratégia competitiva. De fato, as principais empresas do mundo se localizam em alguns poucos países, mas devido às facilidades de comunicação e de transporte, acabam por produzirem seus produtos em locais com melhores condições sócio-econômicas.

Como esse tipo de estudo em âmbito mundial, ou até mesmo nacional ou estadual, seria extenso e muito complexo, decidiu-se trabalhar com 3 Regiões Administrativas (R.A.) – R.A. de Araçatuba, R.A. de Presidente Prudente e R.A. de Marília.

Uma vez definida as regiões de estudo, o objetivo foi aplicar as idéias citadas anteriormente nessas regiões; pois, ainda influenciado por Porter (1999), pode-se afirmar que o desenvolvimento dos municípios de Araçatuba, Presidente Prudente e Marília será mais consistente se vier em conjunto com o desenvolvimento dos demais municípios que compõem essa região de influência.

Para realizar este estudo, foram analisadas 5 variáveis que correspondem a vínculos empregatícios nas áreas rural e urbana. Na área rural, foi analisada a variável número de vínculos

empregatícios na agropecuária; e na área urbana, foram analisadas as variáveis: número de vínculos empregatícios no comércio, indústria, construção civil e serviços.

O estudo da análise fatorial para a elaboração e análise de variáveis estatísticas para orientar a elaboração de políticas públicas na área do trabalho, deve servir de base para estratégias que atendam ao compromisso do desenvolvimento regional. Resumindo, o estudo teve por objetivo oferecer um diferencial para a elaboração de políticas públicas e/ou privadas embasadas com informações relevantes para o seu desenvolvimento sócio-econômico.

Os dados trabalhados neste estudo são referentes ao ano de 2008 e foram coletados no armazém de dados oficial da Fundação SEADE.

2 REGIÕES EM ESTUDO

Neste estudo, foram analisadas 5 variáveis em três Regiões Administrativas: R.A. de Araçatuba, R.A. de Presidente Prudente e R.A. de Marília. As três R.A. totalizam 147 municípios, ocupam uma área de 60.856km² e possuem cerca de 2.516.522 habitantes.

As três R.A. estão entre as 5 regiões menos densas demograficamente no Estado e todas possuem sua economia baseada, principalmente, na agropecuária e em algum tipo de indústria.

Nas próximas três seções (2.1, 2.2, 2.3), far-se-á uma breve introdução sobre cada região em estudo. Todas as informações foram retiradas da Fundação SEADE (2009).

2.1 Região Administrativa de Araçatuba

A Região Administrativa de Araçatuba (Figura 1) tem como sede e maior polo regional, o município de Araçatuba, que concentra cerca de 1/4 da população total desta região.



Figura 1 – Região Administrativa de Araçatuba.

A 9ª R.A. do Estado de São Paulo é composta por 43 municípios, abrange 7% do território estadual, possui a 2ª menor taxa de crescimento populacional e a terceira menor densidade demográfica do Estado (37 hab./km², em 2002). A região em questão passou décadas com taxas de crescimento populacional negativo e, por meados da década de 80, começou a crescer de uma forma conservadora (1,4% a.a.), reduzindo essa taxa nos últimos anos. Nos últimos anos ainda, a região vem registrando alterações consideráveis na sua estrutura etária, apresentando menor proporção de crianças, maior população ativa e uma proporção crescente no número de idosos.

A economia da região é baseada principalmente na agropecuária e por atividades industriais, que possuem ampla infra-estrutura em diversos meios de transporte de carga. A atividade que se destaca nessa região é a preparação/confecção de produtos provenientes do couro (18,7% do valor adicionado estadual). Essa atividade é a responsável pelo emprego de 42% da população em idade ativa no setor de indústria regional, sendo assim a maior empregadora nesse setor.

Em relação à dinâmica industrial regional, destaca-se a produção pecuária, que nomeia a capital da R.A. como “capital do boi gordo”.

2.2 Região Administrativa de Presidente Prudente

A Região Administrativa de Presidente Prudente (Figura 2) tem como sede e maior polo regional, o município de Presidente Prudente, que concentra cerca de 24% da população total desta região.



Figura 2 – Região Administrativa de Presidente Prudente.

A 10ª R.A. do Estado de São Paulo é composta por 53 municípios que abrangem cerca de 10% do território estadual, concentrando aproximadamente 2,1% da população do Estado. Essa R.A. possui a 3ª menor taxa de urbanização estadual e apresenta a segunda menor densidade demográfica do Estado (9,6 hab./km², em 2002).

A região manteve um ritmo de crescimento relativamente estável nos últimos anos, apesar de apresentar crescimentos negativos em alguns municípios que a compõem entre 1991 e 2000. Assim como a R.A. de Araçatuba, uma importante mudança na estrutura etária da região ocorreu, onde o número de pessoas menores de 15 anos diminuiu, ao passo que a população em idade ativa aumentou, assim como a população na faixa idosa.

A economia desta região em estudo é baseada, principalmente, na agropecuária e na agroindústria. No ano de 2001, a região era responsável por cerca de 5,3% da produção agropecuária do Estado de São Paulo. Além dessa importante contribuição, a região responde por cerca de 18% da produção total de carne bovina no Estado, garantindo assim, a posição de maior exportadora desse produto em âmbito nacional, além de ser considerada uma das maiores bacias leiteiras do país. Pode-se adicionar ainda, a produção de cana-de-açúcar e de ovos, traçando assim, o perfil agropecuário da região.

Em relação à atividade industrial da região, pode-se citar a fabricação de alimentos e bebidas como maior destaque no setor, tanto no valor adicionado como na geração de empregos.

Considerando-se ainda a geração de empregos, a indústria de confecção de vestuário e acessório desempenha, também, um importante papel.

2.3 Região Administrativa de Marília

A Região Administrativa de Marília (Figura 3) tem como sede e maior polo regional, o município de Marília, que concentra cerca de 23% da população total desta região.



Figura 3 – Região Administrativa de Marília.

A 11ª R.A. do Estado de São Paulo é composta por 51 municípios e possui a 5ª menor densidade populacional do Estado (49 hab./km², em 2002).

A região mantém um nível de crescimento praticamente constante, caracterizando-se há tempos pelo pequeno incremento populacional. Assim como nas demais regiões em estudo, a proporção de crianças vem diminuindo, o número de pessoas com idade ativa está aumentando e uma proporção crescente na faixa etária dos idosos é apresentada.

A economia da região provém basicamente da agropecuária e da indústria de produtos primários. Em 2001, a região era responsável por cerca de 7% da produção agropecuária do Estado. Destas, destacam-se a produção de cana-de-açúcar, carne bovina, soja, milho e ovos.

A atividade industrial que mais se destaca nessa região é a fabricação de alimentos e bebidas, tanto no valor adicionado como na geração de empregos. Graças a esse motivo, Marília, a sede da 11ª R.A. do Estado, foi intitulada como “capital nacional do alimento”.

3 PROCEDIMENTOS METODOLÓGICOS

Para realização deste trabalho, foram utilizadas cinco variáveis relacionadas à área do trabalho: número de vínculos empregatícios na agropecuária (X1), número de vínculos empregatícios no comércio (X2), número de vínculos empregatícios na construção civil (X3), número de vínculos empregatícios na indústria (X4) e número de vínculos empregatícios em serviços (X5). Todas essas variáveis foram retiradas do banco de dados oficial da Fundação SEADE e são referentes ao ano de 2008, onde todos os municípios das 3 R.A. (147 municípios no total), possuíam observações em todas essas variáveis escolhidas. Cada R.A. foi estudada separadamente.

Antes da aplicação de qualquer técnica estatística multivariada, foi realizada uma análise descritiva univariada em cada uma das variáveis, para cada uma das regiões em estudo.

Para facilitar a manipulação e interpretação dos dados coletados, escolheu-se trabalhar com a análise fatorial para resumir a estrutura inicial das variáveis, criando assim um número reduzido de novas variáveis aleatórias (fatores).

3.1 Análise Fatorial

“A análise fatorial é uma técnica de interdependência na qual todas as variáveis são simultaneamente consideradas, cada uma relacionada com todas as outras, empregando ainda o conceito de variável estatística, a composição linear das variáveis” (HAIR et al., 2005).

Sintetizando as idéias de Mardia et al. (1992), pode-se dizer que a técnica tem o poder de resumir a informação contida nas variáveis originais, criando um número reduzido m de novas variáveis aleatórias, que são chamadas de fatores.

Para que a aplicação da análise fatorial seja validada, existem alguns critérios que devem ser seguidos, tais como a Medida de Adequação da Amostra (MSA, do inglês *Measure Sampling Adequacy*), no qual este índice varia entre $0 \leq MSA \leq 1$, sendo 1 o valor ideal. O cálculo dessa medida é dado por:

$$MSA = \frac{\sum_{i \neq j} R_{ij}^2}{\sum_{i \neq j} R_{ij}^2 + \sum_{i \neq j} Q_{ij}^2}$$

onde R_{ij} é a correlação amostral entre as variáveis X_i e X_j , e Q_{ij} é a correlação parcial entre X_i e X_j controlada pelas demais $(p - 2)$ variáveis.

Outro critério também utilizado na validação da aplicação deste tipo de técnica e um pouco menos conceitual do que o MSA, é a análise da matriz de correlação das variáveis. No caso, quando a matriz de correlação possuir um número substancial de correlações maiores do que 0,30, em módulo, a aplicação da análise fatorial pode ser dita aceitável.

Alguns métodos são sugeridos para obtenção dos fatores na literatura. Um dos mais conhecidos segundo Johnson e Wichern (1998), é o Método das Componentes Principais, onde os autovalores da matriz de correlação das variáveis são considerados na criação dos fatores. O interessante deste método, segundo Mingoti (2005), é que a suposição de normalidade dos dados não é necessária para a aplicação da técnica. Neste trabalho, fará-se uso deste método.

Como foi dito anteriormente, um dos objetivos da análise fatorial é o de reduzir o número de variáveis aleatórias em estudo. Então, existem alguns critérios que são utilizados para escolher o número reduzido m de fatores:

- Critério da raiz latente: os fatores que possuem os autovalores maiores do que 1 são considerados significantes, logo todos os fatores que são menores que 1 são descartados. Ou seja, o número m será o número de autovalores maiores do que 1;
- Critério do teste scree: para retermos o número m de fatores, basta observar-se no gráfico o ponto onde os valores obtidos para cada autovalor tende a se estabilizar, então o número de autovalores anteriores a esse ponto, será o número m de fatores;
- Critério de percentagem da variância: este critério baseia-se em manter um número m de fatores que representam uma percentagem $\gamma \times 100\%$ da variância total, onde $0 < \gamma < 1$ é pré-determinado pelo pesquisador.

Na prática, utiliza-se mais de um critério para a escolha do número m de fatores.

Os valores que são observados dentro da matriz fatorial, são chamados de cargas fatoriais, ou *loading*. As cargas fatoriais, nada mais são do que a correlação de cada variável com um determinado fator, ou seja, elas indicam o grau de correspondência entre a variável e o fator, sendo as cargas numericamente maiores responsáveis pela denominação (rótulo) que o fator receberá.

Quando os m fatores são obtidos através de um ou mais critérios citados anteriormente, atinge-se o objetivo de resumir os dados em novas variáveis aleatórias, porém na maioria dos casos,

os fatores obtidos não fornecem informação necessária para uma interpretação adequada das variáveis sob exame. Neste caso, utiliza-se um recurso de transformação destes fatores. Daí vem um conceito muito conhecido dentro da análise fatorial, chamado de rotação de fatores, ou rotação fatorial. Como o próprio nome sugere, a rotação fatorial gira o eixo dos fatores de forma que eles alcancem uma posição onde a interpretabilidade da matriz fatorial é mais simples. Então, a rotação é desejável, uma vez que simplifica a estrutura fatorial, tornando mais fácil a interpretação das novas variáveis aleatórias (fatores); ou seja, com a rotação dos fatores, consegue-se soluções mais simples e teoricamente mais significativas. Utilizou-se neste trabalho a rotação pelo Critério Varimax, que procura minimizar o número de variáveis fortemente relacionadas com cada fator.

Após a interpretação de cada fator, deve-se nomear cada um dos mesmos. As variáveis originais que possuem cargas fatoriais mais significativas em determinado fator, são as que mais contribuem para a nomeação dele. Segundo Hair et al. (2005), uma carga fatorial é dita significativa quando seu valor, em módulo, é superior a 0,60.

Quando os fatores são retidos, interpretados e nomeados, é possível a criação de escores, que serão utilizados para construção de gráficos, auxiliando na distinção de grupos similares e não-similares.

4. ANÁLISE DOS RESULTADOS

4.1 Região Administrativa de Araçatuba

4.1.1 Análise das Variáveis

A tabela a seguir (Tabela 1), nos fornece algumas estatísticas descritivas univariadas acerca das variáveis em estudo na R.A. de Araçatuba.

Tabela 1: Estatísticas descritivas univariadas, na R.A. de Araçatuba.

Variável	Média	Desvio Padrão	Variância	Assimetria	Curtose
X1	365,19	428,77	183841,49	1,63	1,94
X2	698,44	1907,25	3637595,35	4,55	23,05
X3	66,16	203,86	41558,52	4,22	19,59
X4	1225,05	3360,56	11293347,50	5,00	27,67
X5	1224,63	2857,90	8167610,00	4,76	25,50

Através da Tabela 1, observa-se que todas as variáveis possuem uma variância extremamente grande, além de possuírem uma assimetria fortemente positiva, que nos mostra que os dados estão concentrados à esquerda da distribuição. Ainda, excetuando a variável agropecuária, as curtoses são numericamente bem elevadas, o que significa que há um acúmulo grande de observações em um determinado ponto da distribuição das variáveis.

Para uma melhor visualização dessas medidas, a seguir (Figura 4) apresenta-se a matriz de dispersão das variáveis na R.A. de Araçatuba. Valores acima da diagonal principal são as correlações bivariadas; abaixo da diagonal principal observa-se os gráficos de dispersão; e a diagonal principal representa a distribuição de cada variável (histograma).

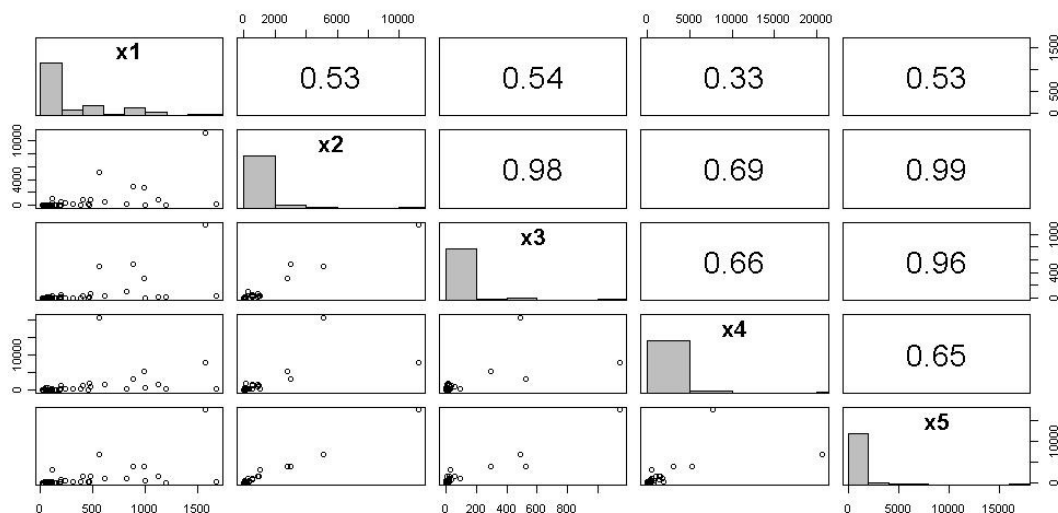


Figura 4 – Matriz de Dispersão das variáveis da R.A. de Araçatuba.

Como observado na Tabela 1, todas as variáveis possuem distribuições com forte assimetria e acúmulo de observações ao lado esquerdo da distribuição; o que já era esperado, pois como visto na Tabela 1, os valores de assimetria e curtose das variáveis eram numericamente grandes. Ainda, pode-se notar que os pares possuem algum tipo de relação linear, excetuando as correlações referentes à variável agropecuária.

Um dos critérios descritos para a correta aplicação da análise fatorial anteriormente era a observação do valor, em módulo, das correlações das variáveis. Observa-se que a matriz de correlação possui todos os seus valores maiores do que 0,30. Portanto, por este critério, pode-se continuar a aplicação da análise fatorial nos dados.

O outro critério citado no texto é a MSA, que no caso dessa matriz de correlação é de 0,67, que segundo Hair et al. (2005) é um valor mediano; sendo assim a aplicação da análise fatorial é aceitável, também, neste critério.

4.1.2 Aplicação da Análise Fatorial

Com a aprovação dos critérios anteriormente, segue-se com a aplicação da análise fatorial.

Para a escolha do número m de fatores a serem retidos no sistema, utilizou-se os critérios da raiz latente (Tabela 2), percentagem de variância (Tabela 2) e teste *scree* (Figura 5).

Tabela 2: Autovalores e variância explicada de cada fator.

Fator	Autovalor	Percentual explicado de variância	Percentual cumulativo de variância
1	3,83	76,64	76,64
2	0,69	13,89	90,54
3	0,43	8,65	99,19
4	0,04	0,77	99,95
5	0,01	0,05	100,00

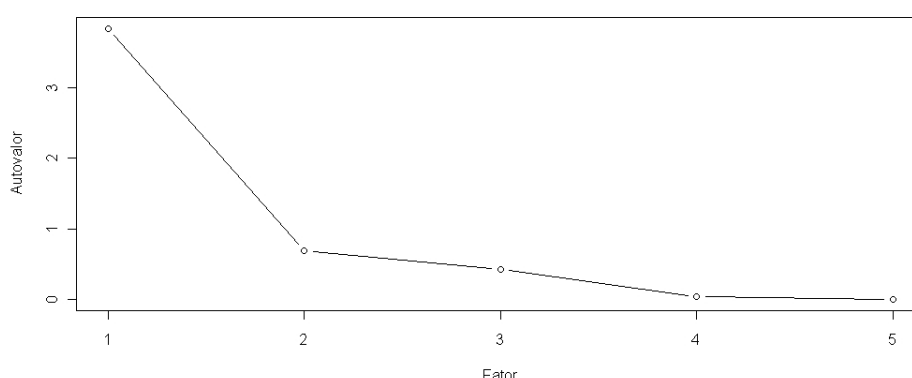


Figura 5 – Scree-plot relacionado a R.A. de Araçatuba.

Observa-se, pelo critério da raiz latente, que apenas o primeiro fator seria mantido no sistema, uma vez que ele o fator 1 é o único que possui autovalor superior a 1. Pela observação do gráfico *scree*, dois fatores seriam mantidos no sistema.

Pela análise da quantidade de variância explicada pelos fatores, observa-se que, quando retêm-se apenas 1 fator no sistema, a percentagem de variância explicada das variáveis originais é de um pouco mais de 76%, o que é considerado um valor relativamente alto. Porém, se o 2º fator é retido também, a percentagem de variância explicada pelos dois fatores é superior a 90%, portanto, por essa variância explicada e os outros dois critérios de retenção de fatores, escolheu-se trabalhar com dois fatores no sistema. Os *loading* dos fatores são apresentados a seguir, na Tabela 3:

Tabela 3: Matriz fatorial sem rotação da R.A. de Araçatuba.

Variável	Fator		Comunalidade
	1	2	
Agropecuária	0,64	0,73	0,95
Comércio	0,98	-0,08	0,97
Construção Civil	0,97	-0,05	0,94
Indústria	0,76	-0,38	0,72
Serviços	0,97	-0,06	0,94

Nota-se que, a matriz fatorial não rotacionada nos retorna cargas fatoriais significativas em ambos os fatores em apenas uma variável (número de vínculos empregatícios na agropecuária). Observa-se também que existe uma certa dificuldade em interpretar e nomear os fatores da forma que estão apresentados. Logo, utiliza-se a rotação fatorial na tentativa de simplificar essa interpretação com a redistribuição das variâncias (Tabela 4).

Tabela 4: Matriz fatorial com rotação pelo critério Varimax da R.A. de Araçatuba.

Variável	Fator		Comunalidade
	1	2	
Agropecuária	0,22	0,95	0,95
Comércio	0,91	0,39	0,97
Construção Civil	0,88	0,41	0,94
Indústria	0,85	0,02	0,72
Serviços	0,88	0,40	0,94

Com a rotação da matriz fatorial, observa-se que agora a interpretação dos fatores está mais simples. Isso vem do fato de que, após a rotação da matriz, a carga fatorial no fator 2 da variável relativa a agropecuária está extremamente elevada (0,95), ao passo que no fator 1, o *loading* desta variável sofreu uma redução brusca (agora 0,22). Este era o grande empecilho na interpretação da matriz e é o motivo pelo qual utilizou-se a rotação fatorial.

Com relação as correlações das variáveis com cada fator, percebe-se que, tanto no fator 1, como no fator 2, quando uma variável aumenta, a outra aumenta também, ou seja, a medida que um setor cresce, os demais crescem conjuntamente.

Seguindo a Tabela 4, pode-se agora denominar os fatores de forma coerente. No caso, ao observar-se a referida tabela, nota-se que o primeiro fator carrega as variáveis: número de vínculos empregatícios no comércio, na construção civil, na indústria e nos serviços. Todas essas variáveis são referentes a vínculos de trabalho urbanos, portanto, pode-se nomear esse fator de “Vínculos Empregatícios Urbanos”. No fator 2, similarmente, pode-se nomear o fator como “Vínculos Empregatícios Rurais”, haja vista que o fator 2 carrega somente a variável número de vínculos empregatícios na agropecuária.

Após esta etapa, calculou-se os escores de cada município nos dois fatores retidos e esboçou-se o gráfico de dispersão desses escores fatoriais (Figura 6).

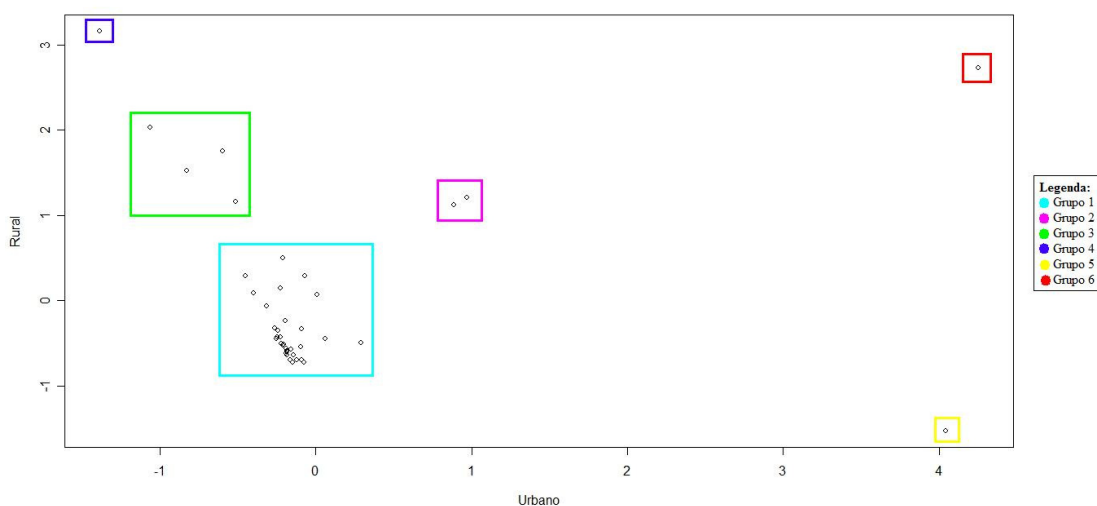


Figura 6 – Gráfico de dispersão dos escores fatoriais na R.A. de Araçatuba.

Como pode-se notar, é possível dividir visualmente os municípios da R.A. de Araçatuba em 6 grupos distintos de uma forma bivariada. Ainda, os municípios da 9ª R.A. do Estado podem ser agrupados em cada fator, formando: 5 grupos de municípios para o fator Vínculos Empregatícios Rurais e 4 grupos para o fator Vínculos Empregatícios Urbanos.

Em relação à divisão bivariada, os grupos ficaram da seguinte forma:

- Grupo 1: Alto Alegre, Auriflama, Avanhandava, Barbosa, Bento de Abreu, Bilac, Braúna, Buritama, Coroados, Gabriel Monteiro, Gastão Vidigal, General Salgado, Glicério, Guaraçaí, Guzolândia, Ilha Solteira, Itapura, Lavínia, Lourdes, Luiziânia, Mirandópolis, Murutinga do Sul, Nova Castilho, Nova Independência, Nova Luzitânia, Pereira Barreto, Piacatu, Rubiácea, Santópolis do Aguapeí, São João de Iracema, Sud Mennucci, Suzanópolis, Turiúba e Valparaíso;
- Grupo 2: Penápolis e Andradina;
- Grupo 3: Santo Antônio do Aracanguá, Guararapes, Castilho e Brejo Alegre;
- Grupo 4: Clementina;
- Grupo 5: Birigui;
- Grupo 6: Araçatuba.

Fazem parte do primeiro grupo, municípios que não possuem muitos vínculos empregatícios, nem na parte urbana, nem na parte rural, quando comparados com o restante dos municípios. Já os municípios que compõem o grupo 2 deste gráfico, possuem um número substancial, tanto nos vínculos de trabalho urbano, como nos vínculos rurais quando comparados com os outros municípios. O grupo 3 agrupa municípios com bom número de contratos de trabalho rurais, porém possui determinada defasagem na parte de vínculos urbanos quando comparados com todos os municípios da região. O grupo 4 é composto por apenas um município, Clementina, que possui alto número de vínculos empregatícios rurais e baixíssimo número de vínculos na parte urbana quando comparado com as demais cidades, percebe-se que este município faz parte dos dois extremos, está em primeiro lugar no que tange aos vínculos rurais e em último nos vínculos urbanos, segundo o escore fatorial obtido. O grupo 5, assim como o anterior, é composto por apenas um município, Birigui, que possui altíssimo número de vínculos empregatícios urbanos e baixo número de vínculos rurais quando comparado com os demais municípios em estudo. O sexto e último grupo, possui como único município, a capital da R.A., Araçatuba, que como esperado no começo deste estudo, está nas primeiras posições (1º lugar no primeiro fator e 2º lugar no segundo fator), no que tange a número de vínculos empregatícios na área urbana e rural.

4.2 Região Administrativa de Presidente Prudente

4.2.1 Análise das Variáveis

A tabela a seguir (Tabela 5) fornece algumas estatísticas descritivas univariadas acerca das variáveis em estudo.

Tabela 5: Estatísticas básicas das variáveis em estudo, na R.A. de Presidente Prudente.

Variável	Média	Desvio Padrão	Variância	Assimetria	Curtose
X1	245,38	284,25	80799,70	2,38	6,14
X2	625,49	1960,02	3841687,83	6,44	44,33
X3	73,55	344,84	118917,64	7,01	50,17
X4	753,46	1894,42	3588849,86	5,79	38,11
X5	1198,87	3443,13	11855143,50	6,75	47,68

Através da Tabela 5, pode-se notar que todas as variáveis possuem uma variância extremamente grande, curtose numericamente elevada, além de possuírem uma assimetria fortemente positiva, o que indica um acúmulo de observações a esquerda da distribuição das variáveis.

Para uma melhor visualização dessas medidas, a seguir apresenta-se a matriz de dispersão das variáveis (Figura 6). A diagonal principal contém os histogramas que representam a distribuição de cada variável; acima dessa diagonal observa-se a matriz de correlação bivariada; e abaixo da diagonal observa-se o gráfico de dispersão.

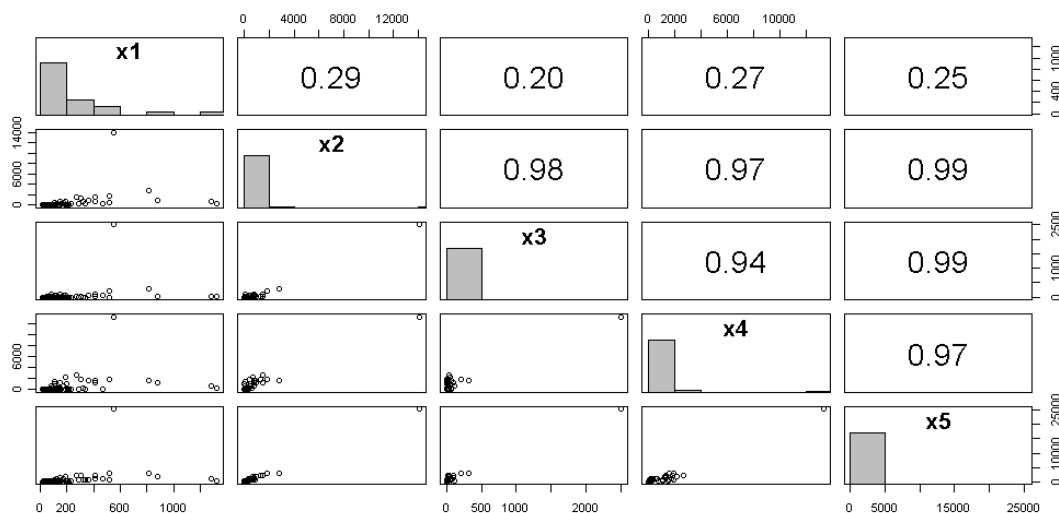


Figura 7 – Matriz de dispersão das variáveis em estudo na R.A. de Presidente Prudente.

Como observado na Tabela 5, todas as variáveis possuem distribuições com forte assimetria positiva e acúmulo de observações ao lado esquerdo da distribuição; o que já era esperado, pois como visto na Tabela 5, os valores de assimetria e curtose das variáveis eram numericamente grandes. Ainda, pode-se notar que a relação que os pares possuem entre si é linear, excetuando correlações não tão significativas ($< 0,30$).

Pelo critério descrito para a correta aplicação da análise fatorial relativo a observação do valor, em módulo, das correlações das variáveis, observa-se que a matriz de correlação possui todas as correlações superiores a 0,30, excetuando a variável X_1 (número de vínculos empregatícios na agropecuária). Portanto, pode-se considerar que há um número substancial de correlações maiores do que 0,30 em módulo, sendo assim, por este critério, pode-se continuar a aplicação da análise fatorial nos dados.

Pelo outro critério, a MSA, que no caso dessa matriz de correlação é de 0,80, tem-se ainda mais indícios de que a aplicação da análise fatorial nos dados é aceitável, uma vez que 0,80 é um valor extremamente alto (segundo as diretrizes de Hair et al. (2005), o nível deste MSA é considerado ótimo).

4.2.2 Aplicação da Análise Fatorial

Com a aprovação dos critérios anteriores, seguiu-se com a aplicação da análise fatorial.

Para a escolha do número m de fatores, utilizou-se os critérios da raiz latente (Tabela 6), percentagem de variância (Tabela 6) e teste *scree* (Figura 8).

Tabela 6: Autovalores e variância explicada de cada fator.

Fator	Autovalor	Percentual Explicado de Variância	Percentual Cumulativo de Variância
1	4,01	80,17	80,17
2	0,92	18,38	98,55
3	0,06	1,17	99,72
4	0,01	0,19	99,92
5	0,01	0,08	100,00

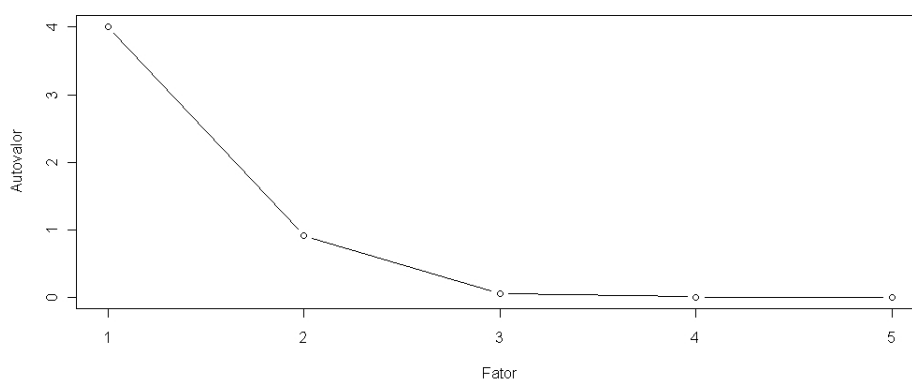


Figura 8 – Scree-plot (Autovalor versus Fator).

Como pode-se observar, pelo critério da raiz latente apenas um fator seria retido no sistema, porém ao observar um pouco mais a fundo, nota-se que o 2º autovalor é muito próximo de 1. Portanto, pelo critério da raiz latente pode-se utilizar 1 ou 2 fatores. Seguindo o critério do teste *scree*, o número de fatores a serem retidos no sistema seria de $m=2$.

Pela análise da quantidade de variância explicada pelos fatores, observa-se que, quando retêm-se apenas 1 fator no sistema, a percentagem de variância explicada das variáveis originais é de um pouco mais de 80%, o que é considerado um valor extremamente grande. Porém, se o 2º fator é retido também, a percentagem de variância explicada pelos dois fatores é superior a 98%, portanto, por essa variância explicada e os outros dois critérios de retenção de fatores, escolheu-se trabalhar com dois fatores no sistema. As cargas fatoriais dos fatores são apresentadas a seguir, na Tabela 7:

Tabela 7: Matriz fatorial sem rotação.

Variável	Fator		Comunalidade
	1	2	
Agropecuária	0,33	0,94	0,99
Comércio	0,99	-0,04	0,99
Construção Civil	0,98	-0,14	0,98
Indústria	0,98	-0,06	0,96
Serviços	0,99	-0,08	0,99

Como pode-se perceber, já somos capazes de nomear os fatores retornados, porém irá-se rotacionar estes fatores (Tabela 8), com o intuito de simplificar ainda mais a interpretação dessa matriz fatorial, otimizando assim a distribuição da variância entre os fatores.

Tabela 8: Matriz fatorial com rotação pelo critério Varimax.

Variável	Fator		Comunalidade
	1	2	
Agropecuária	0,13	0,99	0,99
Comércio	0,98	0,17	0,99
Construção Civil	0,99	0,07	0,98
Indústria	0,97	0,15	0,96
Serviços	0,99	0,13	0,99

Após a rotação da matriz fatorial, observa-se que não houve uma melhoria muito significativa na matriz, porém melhoramos em partes a distribuição da variância nos fatores.

Com relação as correlações das variáveis com cada fator, percebe-se que, tanto no fator 1, como no fator 2, quando uma variável aumenta, a outra aumenta também, ou seja, a medida que um setor cresce, os demais também crescem.

Seguindo a Tabela 8, pode-se agora denominar os fatores de forma coerente. No caso, ao observar a referida tabela, assim como na Região anteriormente estudada, nota-se que o primeiro fator carrega as variáveis número de vínculos empregatícios no comércio, na construção civil, na indústria e nos serviços; no caso, todas as variáveis são referentes a vínculos de trabalho urbanos, portanto, pode-se nomear esse fator como “Vínculos Empregatícios Urbanos”. No fator 2, similarmente, pode-se nomear o fator de “Vínculos Empregatícios Rurais”, haja vista que o fator 2 carrega somente a variável número de vínculos empregatícios na agropecuária.

Após esta etapa, calculou-se os escores de cada município nos dois fatores retidos, e um gráfico de dispersão foi montado com estes escores (Figura 9).

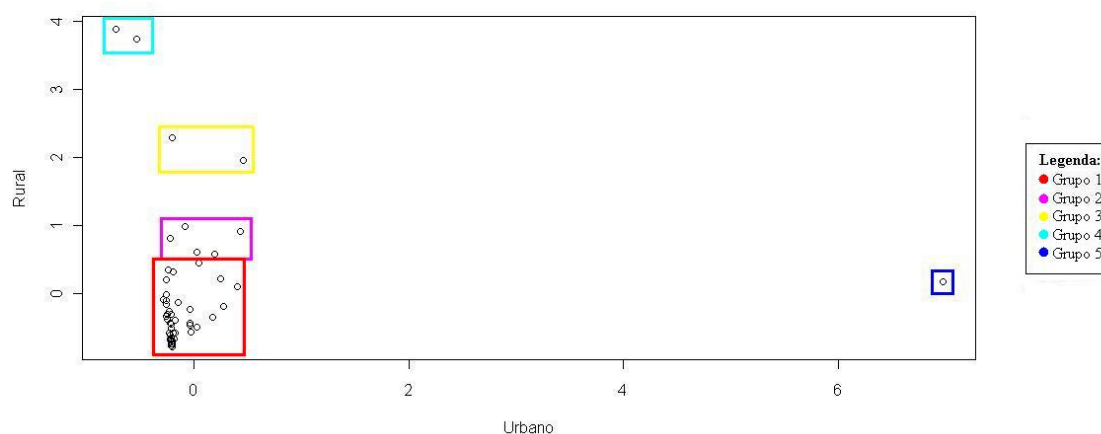


Figura 9 – Gráfico de dispersão dos escores fatoriais obtidos.

Nota-se que é possível fazer a divisão dos municípios em 5 grupos distintos visualmente. Ainda, os municípios da 10ª R.A. do Estado de São Paulo podem ser agrupados em cada fator formando: 4 grupos de municípios para o fator “Vínculos Empregatícios Rurais” e 3 grupos para o fator “Vínculos Empregatícios Urbanos”.

Em relação a divisão bivariada, os grupos ficaram da seguinte forma:

- Grupo 1: Alfredo Marcondes, Álvares Machado, Anhumas, Caiabu, Caiuá, Emilianópolis, Estrela do Norte, Euclides da Cunha Paulista, Flora Rica, Iepê, Indiana, Inúbia Paulista, Irapuru, Junqueirópolis, Lucélia, Marabá Paulista, Mariápolis, Mirante do Paranapanema, Monte Castelo, Nantes, Nandubá, Nova Guataporanga, Osvaldo Cruz, Ouro Verde, Pacaembu, Panorama, Paulicéia, Piquerobi, Pirapozinho, Pracinha, Presidente Bernardes, Presidente Epitácio, Ribeirão dos Índios, Rosana, Sagres, Salmourão, Sandovalina, Santa Mercedes, Santo Expedito, São João do Pau D’alho, Taciba, Tarabai e Tupi Paulista;

- Grupo 2: Adamantina, Martinópolis, Presidente Venceslau, Regente Feijó e Teodoro Sampaio;

- Grupo 3: Dracena e Rancharia;

- Grupo 4: Flórida Paulista e Santo Anastácio;

- Grupo 5: Presidente Prudente.

Fazem parte do primeiro grupo, municípios que não possuem muitos vínculos empregatícios, nem na parte urbana, nem na parte rural, quando comparados com o restante dos municípios. Já os municípios que compõem o grupo 2 deste gráfico, possuem um número substancial no número de

vínculos empregatícios rurais, porém um baixo número de vínculos empregatícios urbanos quando comparados com os demais municípios. O grupo 3 agrupa municípios com um bom número de contratos de trabalho rurais, porém possui um número baixo de vínculos urbanos quando comparados com todos os municípios da região. O grupo 4 é composto pelos dois maiores possuintes de número de vínculos empregatícios na área rural que são, também, os dois piores fornecedores de vínculos empregatícios urbanos, quando comparados com os demais municípios. O 5º e último grupo é composto por apenas um município – Presidente Prudente – que é o maior em número de vínculos empregatícios urbanos, porém possui um número bem baixo de vínculos empregatícios rurais, quando comparado com os demais municípios da 10ª R.A. do Estado de São Paulo.

4.3 Região Administrativa de Marília

4.3.1 Análise das Variáveis

A tabela a seguir (Tabela 9), fornece algumas estatísticas descritivas univariadas acerca das variáveis em estudo na R.A. de Marília.

Tabela 9: Estatísticas básicas das variáveis em estudo, na R.A. de Marília.

Variável	Média	Desvio Padrão	Variância	Assimetria	Curtose
X1	666,24	1031,20	1063370,50	3,22	13,33
X2	805,71	1998,00	3992009,49	3,90	16,68
X3	112,45	431,29	186008,09	6,08	39,76
X4	910,65	1876,50	3521268,63	4,42	23,66
X5	1363,41	3265,48	10663372,80	4,84	26,76

Através da Tabela 9, pode-se observar que todas as variáveis possuem uma variância extremamente alta, além de possuírem uma assimetria fortemente positiva e curtose numericamente elevada, o que indica um acúmulo de observações a esquerda da distribuição das variáveis.

Para uma melhor visualização dessas medidas, a seguir apresenta-se a matriz de dispersão das variáveis (Figura 10). A diagonal principal contém os histogramas que representam a distribuição de cada variável; acima da diagonal principal observa-se a matriz de correlação bivariada; e abaixo da diagonal principal observa-se o gráfico de dispersão das variáveis duas a duas.

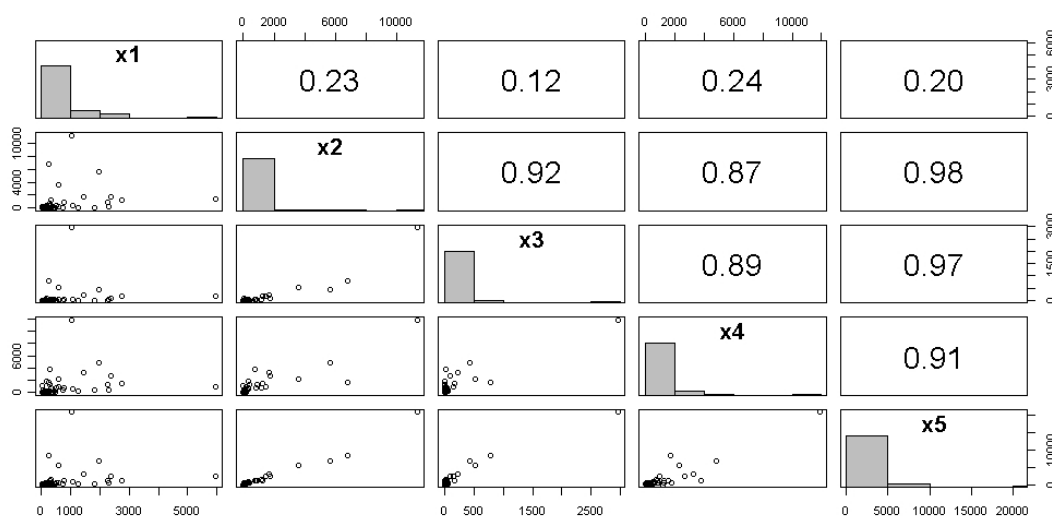


Figura 10 - Matriz de dispersão das variáveis em estudo na R.A. de Marília.

Como observado na Tabela 9, todas as variáveis possuem distribuições com forte assimetria positiva e acúmulo de observações em apenas um ponto da distribuição (curtose numericamente elevada). Nota-se ainda, que os pares possuem algum tipo de relacionamento linear, excetuando correlações não tão significativas ($< 0,30$).

Pelo critério de observação da matriz de correlação das variáveis, pode-se dizer que existe um número substancial de correlações maiores do que 0,30, em módulo, na matriz. Uma vez que apenas as correlações referentes a variável número de vínculos empregatícios na agropecuária são menores do que esse valor. Portanto, por este critério, existem indícios de que a aplicação da análise fatorial é adequada.

Pelo critério da Medida de Adequação de Amostra, considera-se também que a aplicação da análise fatorial nesse conjunto de variáveis é aceitável, uma vez que a MSA dessa matriz é de 0,63, que segundo Hair et al. (2005) é um valor de MSA mediano.

4.3.2 Aplicação da Análise Fatorial

Com a aprovação dos critérios anteriormente, segue-se com a aplicação da análise fatorial.

Para a escolha do número m de fatores a serem retidos no sistema, foram utilizados os critérios da raiz latente (Tabela 10), percentagem de variância (Tabela 10) e teste *scree* (Figura 11).

Tabela 10: Autovalores e variância explicada de cada fator (R.A. de Marília).

Fator	Autovalor	Percentual Explicado de Variância	Percentual Cumulativo de Variância
1	3,83	0,77	0,77
2	0,95	0,19	0,96
3	0,14	0,03	0,99
4	0,07	0,01	1,00
5	0,01	0,00	1,00

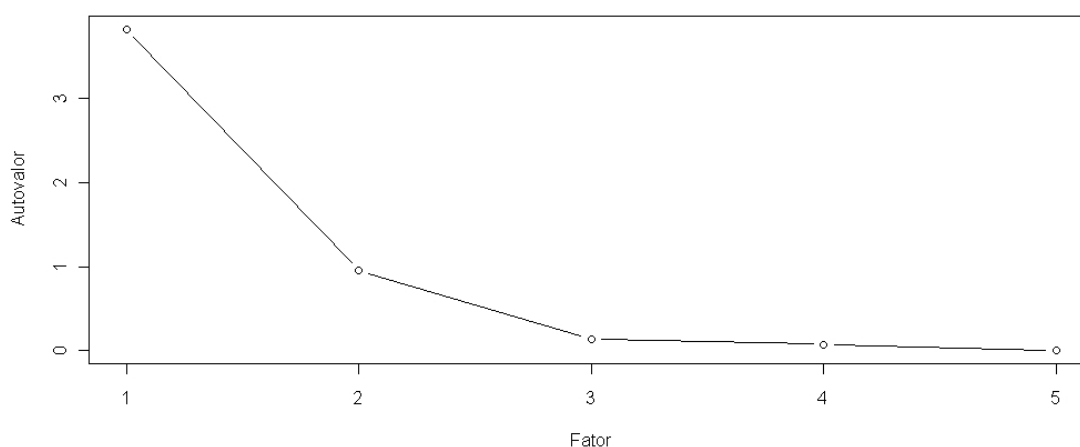


Figura 11 – *Scree-plot* relacionado a R.A. de Marília.

Observa-se que pelo critério da raiz latente, apenas um fator seria mantido no sistema, porém olhando mais a fundo, pode-se notar, que o segundo autovalor está muito próximo de 1. Com a observação do *scree-plot*, reteríamos dois fatores.

Para a escolha do número m de fatores então, observamos a quantidade de variância explicada com 2 fatores retidos. Observa-se na Tabela 10, a quantidade de variância total explicada é superior a 96% com dois fatores mantidos no sistema, o que é um valor extremamente alto, ou seja, a opção

escolhida foi de trabalhar com dois fatores. As cargas fatoriais dos fatores são apresentadas a seguir, na Tabela 11:

Tabela 11: Matriz fatorial sem rotação da R.A. de Marília.

Variável	Fator		Comunalidade
	1	2	
Agropecuária	0,27	0,96	0,99
Comércio	0,97	-0,04	0,94
Construção Civil	0,96	-0,15	0,94
Indústria	0,95	-0,01	0,90
Serviços	0,99	-0,07	0,99

Nota-se que já é possível nomear os fatores retornados, porém por uma questão de procedimento, faremos uma rotação ortogonal pelo critério Varimax, a fim de simplificar ainda mais a interpretação desta matriz fatorial, otimizando a distribuição da variância entre os fatores.

Tabela 12: Matriz fatorial rotacionada pelo critério Varimax da R.A. de Marília.

Variável	Fator		Comunalidade
	1	2	
Agropecuária	0,10	0,99	0,99
Comércio	0,96	0,13	0,94
Construção Civil	0,98	0,02	0,94
Indústria	0,93	0,15	0,90
Serviços	0,99	0,10	0,99

Após a rotação da matriz fatorial, observa-se que não houve uma melhoria muito significativa na interpretabilidade da matriz, e sim, apenas uma redistribuição da variância nos fatores.

Com relação as correlações das variáveis com cada fator, percebemos que, tanto no fator 1, quanto no fator 2, quando uma variável aumenta, a outra aumenta também, ou seja, a medida que um setor cresce, os demais crescem conjuntamente; isso deve-se ao fato das correlações serem todas positivas.

Seguindo a Tabela 12, pode-se agora denominar os fatores de forma coerente. No caso, observando a referida tabela, assim como nas regiões anteriormente estudadas, nota-se que o primeiro fator carrega as variáveis número de vínculos empregatícios no comércio, na construção civil, na indústria e nos serviços; no caso, todas as variáveis são referentes a vínculos de trabalho urbanos, portanto, pode-se nomear esse fator de “Vínculos Empregatícios Urbanos”. No fator 2, similarmente, pode-se nomear o fator de “Vínculos Empregatícios Rurais”, haja vista que o fator 2 carrega somente a variável número de vínculos empregatícios na agropecuária.

Após a rotulação dos fatores, os escores foram calculados para cada município e um gráfico de dispersão foi realizado com estes escores (Figura 12).

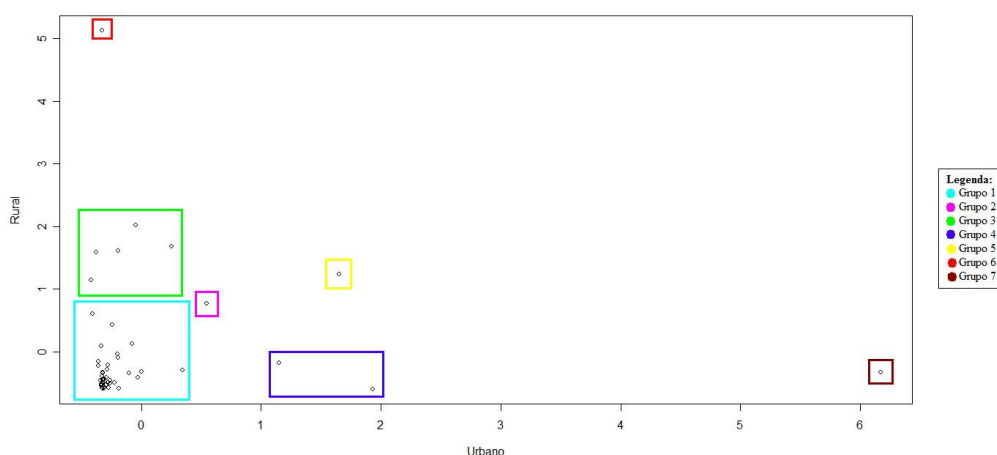


Figura 12 – Gráfico de dispersão dos escores fatoriais na R.A. de Marília.

Ao observar o gráfico, nota-se que é possível dividir visualmente os municípios da R.A. de Marília em 7 grupos distintos de uma forma bivariada. Ainda, os municípios da 11ª R.A. podem ser agrupados em cada fator formando: 3 grupos de municípios para o fator Vínculos empregatícios rurais e 3 grupos para o fator Vínculos empregatícios urbanos.

Em relação à divisão bivariada, os grupos ficaram da seguinte forma:

- Grupo 1: Álvaro de Carvalho, Alvinlândia, Arco Íris, Bernardino de Campos, Borá, Campos Novos Paulista, Cândido Mota, Canitar, Chavantes, Cruzália, Echaporã, Espírito Santo do Turvo, Fernão, Florínia, Gália, Herculândia, Iacri, Ipaussu, João Ramalho, Júlio Mesquita, Lupércio, Lutécia, Maracá, Ocaucu, Óleo, Oriente, Oscar Bressane, Palmital, Parapuã, Pedrinhas Paulista, Platina, Pompéia, Queiroz, Quintana, Ribeirão do Sul, Rinópolis, Salto Grande, São Pedro do Turvo, Timburi e Vera Cruz;

- Grupo 2: Garça;
- Grupo 3: Bastos, Ibirarema, Quatá, Santa Cruz do Rio Pardo e Tarumã;
- Grupo 4: Assis e Tupã;
- Grupo 5: Ourinhos;
- Grupo 6: Paraguaçu Paulista;
- Grupo 7: Marília.

Fazem parte do primeiro grupo, municípios que não possuem muitos vínculos empregatícios, nem na parte urbana, nem na parte rural, quando comparados com o restante dos municípios. Já o município que compõe o grupo 2 deste gráfico – Garça –, possui um número razoável de vínculos empregatícios rurais e urbanos quando comparado com os demais municípios. O grupo 3 agrupa municípios com um bom número de contratos de trabalho rurais, porém possui um número baixo de vínculos urbanos quando comparados com todos os outros municípios da região. O grupo 4, composto por Assis e Tupã, possui um bom número de vínculos empregatícios urbanos, porém um baixo número de vínculos empregatícios rurais quando comparado com os demais municípios que compõem essa região. O 5º grupo, é composto por apenas um município da região – Ourinhos – e tem como característica possuir bons números de vínculos empregatícios, tanto rurais como urbanos, quando comparado com os demais municípios da 11ª R.A. do Estado de São Paulo. O grupo 6, composto somente por Paraguaçu Paulista, é o campeão em número de vínculos empregatícios rurais, porém possui poucos contratos de trabalho relacionados a parte urbana, quando comparado com o restante dos municípios dessa região. O 7º e último grupo é composto pela pelo pólo da região, Marília, que é a campeã em número de vínculos empregatícios urbanos e possui baixo número de vínculos empregatícios rurais, quando comparada com as outras cidades.

5 CONSIDERAÇÕES FINAIS

Nota-se neste estudo que os municípios das 3 Regiões Administrativas estudadas agrupam-se de forma a demandar políticas públicas distintas tendo como referência a utilização da análise fatorial.

Analisando de forma separada cada região, observa-se que um grupo que engloba mais de 3/4 dos municípios foi criado, para cada uma das regiões. No caso, estes grandes agrupamentos correspondem, respectivamente, a aproximadamente 79%, 81% e 78% nas R.A. de Araçatuba, R.A. de Presidente Prudente e R.A. de Marília.

As medidas que podem ser adequadas para desenvolver novos postos de trabalhos nos municípios de Presidente Prudente e Marília não atenderiam com a mesma eficiência o município de Araçatuba, uma vez que, como pode-se observar na Figura 13, a disposição dos dois primeiros municípios citados no gráfico dos escores fatoriais, não são a mesma do município de Araçatuba. Isso decorre do fato que os municípios de Presidente Prudente e Marília possuem um alto número de vínculos empregatícios urbanos, porém poucos contratos de trabalho na parte rural; ao passo que o município de Araçatuba atende bem os dois fatores do estudo.

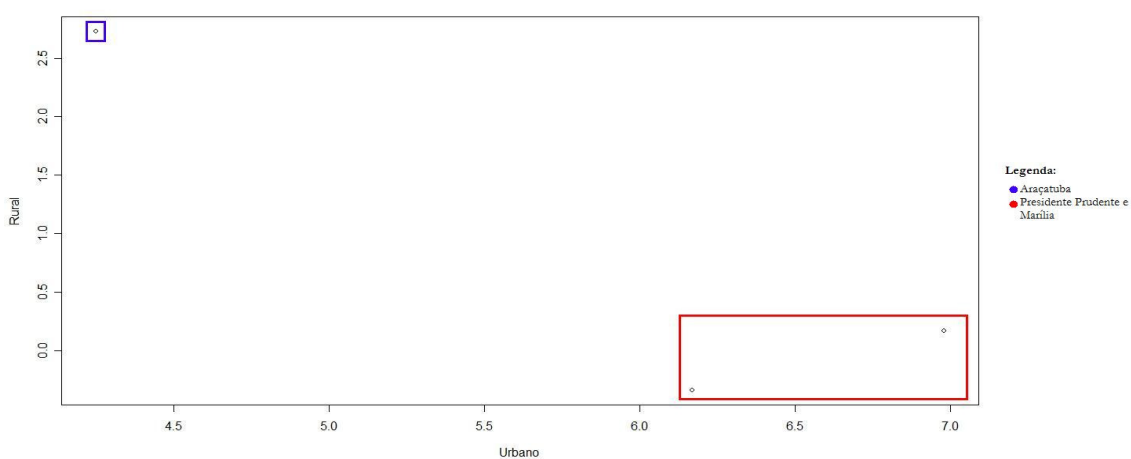


Figura 13 – Escores dos municípios de Araçatuba, Presidente Prudente e Marília

Por outro lado, os grandes agrupamentos, compostos na sua maioria por municípios que têm por característica populações relativamente pequenas, poderiam fazer parte de uma mesma política para desenvolver novos postos de trabalhos para a população economicamente ativa, uma vez que possuem as mesmas características conforme sua região (baixo número de vínculos empregatícios, tanto na parte urbana como na cidade). Podemos comprovar essa afirmação sobrepondo os gráficos dos escores obtidos para cada região, apenas com esses municípios dos grandes grupos (Figura 14).

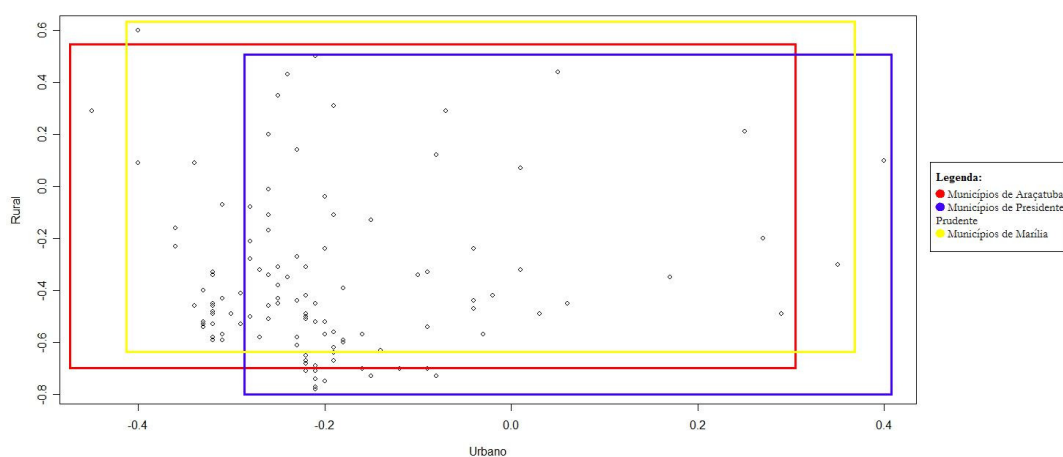


Figura 14 – Escores dos municípios pertencentes ao Grupo 1 de cada região em estudo.

Essa visualização, obtida pelo desenvolvimento teórico aplicado nesse estudo, oferece parâmetros para a elaboração de políticas públicas que atendam de maneira eficiente de acordo com a necessidade de cada agrupamento.

Com este estudo, pode-se notar que possivelmente a elaboração de políticas públicas para determinados grupos de municípios de uma mesma região, seria muito eficaz também se aplicadas em outros municípios das outras regiões; como por exemplo, a criação de políticas públicas para o 1º grupo de cada região, pois como visto na figura anterior (Figura 14), os 3 municípios possuem as mesmas necessidades e similaridades.

Por diversas similaridades como a economia de cada região, a distribuição das populações dos municípios e os gráficos de dispersão dos escores fatoriais durante o estudo dessas três regiões, pode-se afirmar que a aplicação da análise fatorial aplicada diretamente nos 147 municípios em estudo seria, provavelmente, tão eficaz quanto aplicá-la de forma distinta para cada região.

O principal objetivo deste estudo foi alcançado, servindo agora como referencial para a criação de políticas públicas nas R.A. de Araçatuba, Presidente Prudente e Marília.

5 REFERÊNCIAS BIBLIOGRÁFICAS

[1] HAIR, J.F.; TATHAM, R.L.; ANDERSON, R.E.; BLACK, W.C.; trad. Adonai Schlup Sant'Anna e Anselmo Chaves Neto. Análise multivariada de dados. Porto Alegre: Bookman, 5ª ed., 2005.

[2] JOHNSON, R.A.; WICHERN, D.W. Applied multivariate statistical analysis. 4th ed. Upper Saddle River: Ed. Prentice Hall, 1998. 816 p.

[3] MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. Multivariate analysis. London: Academic Press, 1992. 518 p.

[4] MINGOTI, S.A. Análise de dados através de métodos de estatística multivariada: Uma abordagem aplicada. Belo Horizonte: Ed. UFMG, 2005. 297 p.

[5] PORTER, M. E. Competição on competition: Estratégias Competitivas Essenciais. 4ª ed. Rio de Janeiro: Campus, 1999. 515 p.

[6] SÃO PAULO (Estado). Fundação Sistema Estadual de Análise de Dados (SEADE). **Região Administrativa de Araçatuba**. São Paulo: [s.n.], 2002. 7 p. Disponível em: <<http://www.seade.gov.br/produtos/iprs/analises/RAAraçatuba.pdf>>. Acesso em: 18 de março de 2009.

[7] SÃO PAULO (Estado). Fundação Sistema Estadual de Análise de Dados (SEADE). **Região Administrativa de Marília**. São Paulo: [s.n.], 2002. 7 p. Disponível em: <<http://www.seade.gov.br/produtos/iprs/analises/RAMarília.pdf>>. Acesso em: 18 de março de 2009.

[8] SÃO PAULO (Estado). Fundação Sistema Estadual de Análise de Dados (SEADE). **Região Administrativa de Presidente Prudente**. São Paulo: [s.n.], 2002. 7 p. Disponível em: <<http://www.seade.gov.br/produtos/iprs/analises/RAPresprudente.pdf>>. Acesso em: 18 de março de 2009.