

**DANILO ALESSANDRO LÜDKE PEREIRA**

**Aplicação da Análise Multivariada: Índices de Desenvolvimento  
Municipais e Royalties do Petróleo.**

**CURITIBA**

**2014**

**DANILO ALESSANDRO LÜDKE PEREIRA**

**Aplicação da Análise Multivariada: Índices de Desenvolvimento  
Municipais e Royalties do Petróleo.**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciências no Programa de Pós-graduação em Métodos Numéricos em Engenharia, área de concentração em Programação Matemática e Linha de Pesquisa em Métodos Estatísticos Aplicados à Engenharia mantido pelo Setor de Ciências Exatas e pelo Setor de Tecnologia da Universidade Federal do Paraná.

Orientador: Prof. Dr. Eng. Anselmo Chaves Neto

**CURITIBA**

**2014**

Pereira, Danilo Alessandro Lüdke

Aplicação da análise multivariada: índices de desenvolvimento municipais e royalties do petróleo / Danilo Alessandro Lüdke Pereira. – Curitiba, 2014.

122 f. : il.; grafs., tabs.

Dissertação (mestrado) – Universidade Federal do Paraná, Setor de Ciências Exatas e Setor de Tecnologia, Programa de Pós-Graduação em Métodos Numéricos em Engenharia.

Orientador: Anselmo Chaves Neto

1. Petróleo e gás - Royalties. 2. Análise fatorial. 3. Análise multivariada. I. Chaves Neto, Anselmo. II. Título.

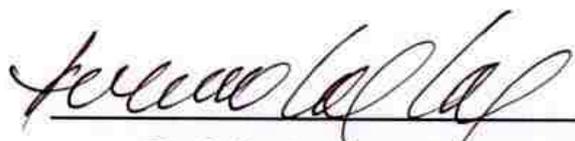
CDD: 519.535

## TERMO DE APROVAÇÃO

DANILO ALESSANDRO LÜDKE PEREIRA

APLICAÇÃO DA ANÁLISE MULTIVARIADA: ÍNDICES DE DESENVOLVIMENTO  
MUNICIPAIS E ROYALTIES DO PETRÓLEO.

Dissertação aprovada como requisito parcial para obtenção do grau de mestre no Programa de Pós-Graduação em Métodos Numéricos em Engenharia, da Universidade Federal do Paraná, pela seguinte banca examinadora:



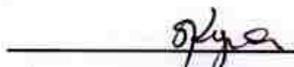
Prof. Dr. Anselmo Chaves Neto.

Orientador – Membro do PPGMNE/UFPR



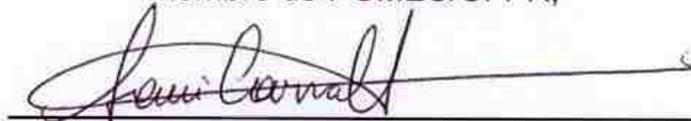
Prof. Dr. José Carlos Pereira Coninck.

Membro da UTFPR – Curitiba/PR



Prof.ª Dr.ª Sachiko Araki Lira.

Membro do PGMEC/UFPR;



Prof.ª Dr.ª Sani de Carvalho Rutz da Silva.

Membro da UTFPR – Ponta Grossa/PR.

Curitiba, 07 de abril de 2014

## **AGRADECIMENTOS**

Sobretudo a Deus por eu chegar até aqui.

À minha esposa Muriel Cristina de Paula e minha filha Maria Luiza de Paula Pereira, que são as pessoas mais importantes na minha vida.

À minha família, que sempre esteve em minha vida, me ajudando e dando força para que eu continuasse.

Ao orientador Prof. Anselmo Chaves Neto, pela amizade, dedicação, compreensão e orientação neste trabalho.

Aos demais Professores do Programa de Pós-Graduação em Métodos Numéricos em Engenharia, que de forma direta ou indireta transmitiram conhecimentos importantes para minha formação profissional e pessoal.

Aos colegas do curso pelo apoio e amizade, especialmente ao Jonilson, Israel, Paulo e Sérgio, companheiros de muitas madrugadas de estudos.

A secretária do CESEC, Jair Anjos, pela enorme ajuda e conselhos durante todo o período do mestrado.

A todas as demais pessoas que, de forma direta ou indireta, contribuíram com a realização do meu trabalho.

## LISTA DE FIGURAS

FIGURA 1 : MUNICÍPIOS QUE RECEBERAM ROYALTIES DO PETRÓLEO EM 2007 .....	15
FIGURA 2: ÁREAS SOB CONCESSÃO .....	24
FIGURA 3: LIMITES INTERESTADUAIS NA PLATAFORMA CONTINENTAL.....	31
FIGURA 4: LINHAS ORTOGONAIS DE CONFRONTAÇÃO DOS MUNICÍPIOS COM A PLATAFORMA CONTINENTAL .....	32
FIGURA 5: EXEMPLO: LINHAS PARALELAS DE CONFRONTAÇÃO DOS MUNICÍPIOS COM A PLATAFORMA CONTINENTAL.....	33
FIGURA 6: LINHAS ORTOGONAIS E PARALELAS .....	33
FIGURA 7: EXEMPLOS DE GRÁFICOS DE PERFIS .....	67
FIGURA 8: INTERFACE DO SOFTWARE MINITAB 16 COM OS DADOS PARA ANÁLISE .....	93
FIGURA 9: MÉTODO DA LIGAÇÃO SIMPLES (SINGLE LINKAGE).....	94
FIGURA 10: MÉTODO DA LIGAÇÃO COMPLETA (COMPLETE LINKAGE) .....	97
FIGURA 11: MÉTODO DO CENTROIDE .....	99
FIGURA 12: MÉTODO DA MEDIANA .....	101

## LISTA DE TABELAS

TABELA 1: COMPARAÇÃO ENTRE OS VALORES DOS ROYALTIES RECEBIDOS COM OS VALORES DOS RECURSOS RECEBIDOS PELO GOVERNO FEDERAL, SEGUNDO MUNICÍPIOS DA AMOSTRA – 2012 .....	19
TABELA 2: DISTRIBUIÇÃO DA PRODUÇÃO DE PETRÓLEO E GÁS NATURAL POR BACIA – 2011 .....	24
TABELA 3: PRODUÇÃO DE PETRÓLEO E GÁS NATURAL POR ESTADO.....	28
TABELA 4: CÁLCULO DE PARTICIPAÇÃO ESPECIAL.....	29

## LISTA DE GRÁFICOS

GRÁFICO 1: PRODUÇÃO NACIONAL DE PETRÓLEO POR ANO (TERRA E MAR) - 2000-2013.....	27
GRÁFICO 2: PERFIL DE MÉDIAS DE DOIS GRUPOS.....	64
GRÁFICO 3: HIPÓTESES $H_{01}$ E $H_{02}$ VERDADEIRAS .....	69
GRÁFICO 4: HIPÓTESES $H_{01}$ VERDADEIRA E HIPÓTESE $H_{02}$ FALSA .....	69
GRÁFICO 5: PERFIS DE MÉDIA DOS GRUPOS 1 E 2 .....	91
GRÁFICO 6: AUTOVALORES DA MATRIZ CORRELAÇÃO .....	104
GRÁFICO 7: DISPERSÃO DOS ESCORES: FATOR1 versus FATOR2: ROTAÇÃO VARIMAX .....	106
GRÁFICO 8: PESOS DOS FATORES: FATOR 1 versus FATOR 2 ...	107

## LISTA DE QUADROS

QUADRO 1: DECISÕES NO TESTE DE HIPÓTESES.....	56
QUADRO 2 : AMOSTRA DE MUNICÍPIOS POR ESTADO .....	83
QUADRO 3: MUNICÍPIOS UTILIZADOS NA COMPARAÇÃO E SUAS RESPECTIVAS POPULAÇÕES EM 2010. ....	84
QUADRO 4: GRUPOS DEFINIDOS PELO MÉTODO DA LIGAÇÃO SIMPLES (SINGLE LINKAGE) .....	95
QUADRO 5: GRUPOS DEFINIDOS PELO MÉTODO DA LIGAÇÃO COMPLETA (COMPLETE LINKAGE) .....	97
QUADRO 6: GRUPOS DEFINIDOS PELO MÉTODO DO CENTROIDE	95
QUADRO 7: GRUPOS DEFINIDOS PELO MÉTODO DO CENTROIDE	100
QUADRO 8: GRUPOS DEFINIDOS PELO MÉTODO DA MEDIANA ....	102
QUADRO 9: AUTO VALORES E VARIÂNCIA EXPLICADA % (63 VARIÁVEIS) .....	105
QUADRO 10: CARREGAMENTOS FATORIAIS .....	105

## LISTA DE SIGLAS

PETROBRAS – Petróleo Brasileiro S.A

IBGE – Fundação Instituto Brasileiro de Geografia e Estatística

ANP - Agência Nacional do Petróleo

CNM - Confederação Nacional de Municípios

FIRJAN - Federação das Indústrias do Estado do Rio de Janeiro

PNUD - Programa das Nações Unidas para o Desenvolvimento

AA - Análise de Agrupamentos

IDHM - Índice de Desenvolvimento Humano Municipal

SIAFI - Sistema Integrado de Administração Financeira

SAEB - Sistema Nacional de Avaliação da Educação Básica

bbl: barris

m<sup>3</sup>: metros cúbicos

boe: barris de óleo equivalente (1 m<sup>3</sup> ≈ 6,29 bbl)

M: milhares

MM: milhões

d: dia

TLD: Teste de Longa Duração

UEP: Unidade Estacionária de Produção

GASA: Gás natural associado

GASN: Gás natural não associado

IDEB: Índice de desenvolvimento da educação básica

## RESUMO

O setor petrolífero brasileiro vem obtendo altos índices de crescimento e em 2006 foi anunciada a autossuficiência em Petróleo. Porém, com o aumento da demanda do recurso nos últimos anos, o país voltou a ser importador de petróleo. Nos últimos 10 anos, mais de 50% das descobertas das jazidas de petróleo do mundo foram em águas profundas, das quais 63% no mar brasileiro. De 1991 a 2011, a Petrobrás teve um crescimento de 164% nas reservas, e possui hoje uma razão reserva/produção de 19,2 anos<sup>1</sup>. Nesse mesmo período, o Brasil aprovou a lei do petróleo (Lei N° 9.478), estabelecendo novos critérios para a transferência de royalties do petróleo para os Municípios brasileiros. Sendo assim, este trabalho tem como objetivo fazer uma avaliação a fim de verificar se os pagamentos de royalties, de alguma forma, têm contribuído para melhorar os indicadores sociais desses Municípios. Então, é feita uma comparação com Municípios de mesmo porte populacional. Para a comparação foram utilizadas as variáveis: IDHM, IFDM, IDEB e incidência de pobreza. Utilizou-se das técnicas de Análise de Perfis, Análise de Agrupamento e Análise Fatorial com objetivo de verificar se existe diferença entre as matrizes de dados referentes aos Municípios que recebem e os que não recebem royalties. No trabalho, concluiu-se que as técnicas aplicadas foram pertinentes aos objetivos propostos e que indicaram que não existe uma diferença significativa entre os indicadores sociais dos Municípios contemplados com as participações governamentais.

Palavras-chave: Análise de Perfil, análise fatorial, análise de cluster, royalties

---

<sup>1</sup> Disponível em: < <http://www.petrobras.com.br/pt/noticias/comemore-conosco-nossos-59-anos/?page=3> > Acesso em: 20 de dezembro de 2013

## **ABSTRACT**

The Brazilian oil industry has achieved high growth rates in 2006 and was announced self-sufficiency in oil. However, with increasing demand in recent years the country has returned to be an oil importer. Over the last 10 years, more than 50% of the discoveries in the world were in deep waters, of which 63% in the Brazilian sea. From 1991 to 2011, Petrobras had a growth of 164% in reserves, and today has a reserve / production ratio of 19.2 years. In the same period, Brazil passed the oil law, establishing new criteria for the transfer of oil royalties to Brazilian municipalities. Thus, this paper aims to make an assessment in order to ascertain whether the royalty payments, somehow, have improved social indicators of these municipalities, making a comparison with the municipalities of the same population size. Variables were used for comparison: IDHM, IFDM, IDEB and incidence of poverty. We used the techniques of Analysis Profiles, Cluster Analysis and Factor Analysis in order to verify the difference between the matrices of data on the municipalities that receive and not receive royalties. At work it was concluded that the techniques applied were relevant to the proposed objectives and indicated that there is no significant difference between the social indicators of the municipalities covered with government participation.

Keywords: Profile Analysis, factor analysis, cluster analysis, royalties

## SUMÁRIO

<b>LISTA DE FIGURAS .....</b>	<b>ii</b>
<b>LISTA DE TABELAS .....</b>	<b>iii</b>
<b>LISTA DE GRÁFICOS .....</b>	<b>iv</b>
<b>LISTA DE QUADROS .....</b>	<b>v</b>
<b>LISTA DE SIGLAS .....</b>	<b>vi</b>
<b>RESUMO .....</b>	<b>vii</b>
<b>ABSTRACT .....</b>	<b>viii</b>
<b>1 INTRODUÇÃO .....</b>	<b>12</b>
1.1 O PROBLEMA .....	12
1.2 OBJETIVOS .....	13
1.3 JUSTIFICATIVA .....	14
1.4 ESTRUTURA DO TRABALHO.....	17
<b>2 REVISÃO BIBLIOGRÁFICA .....</b>	<b>18</b>
2.1 FUNDAMENTAÇÃO TEÓRICA	
2.1.1 ROYALTIES .....	18
2.1.1.1 Royalties de petróleo .....	18
2.1.1.2 A legislação brasileira dos royalties .....	20
2.1.1.3 Campo de petróleo .....	23
2.1.1.4 Metodologia de cálculo dos royalties .....	25
2.1.1.5 O direito dos Municípios aos royalties do petróleo .....	30
2.1.1.6 Áreas geoeconômicas .....	34
2.2 REVISAO DE LITERATURA .....	36

2.3 INFERÊNCIA ESTATÍSTICA .....	38
2.3.1 Variável aleatória .....	39
2.3.2 Vetor aleatório .....	40
2.3.3 Vetor de médias .....	41
2.3.4 Matriz de Covariância .....	41
2.3.5 Matriz de Correlação .....	44
2.3.6 Vetor de médias e a matriz de covariâncias para combinações lineares de variáveis aleatórias .....	45
2.3.7 AMOSTRA ALEATÓRIA E ESTATÍSTICA DESCRITIVA .....	48
2.3.7.1 Amostragem aleatória .....	48
2.3.7.2 Estatística descritiva .....	49
2.3.7.3 Estatísticas descritivas de combinações lineares .....	50
2.3.7.4 Variância amostral generalizada $  S  $ .....	54
2.3.8 TESTES DE HIPÓTESES .....	55
2.3.8.1 Teste de Hipóteses para Médias .....	58
2.3.8.2 Teste de hipóteses para comparar a média de duas populações .....	59
2.3.9 ANÁLISE DE PERFIS .....	61
2.3.9.1 Análise de perfis para dois grupos populacionais .....	64
2.3.9.2 Teste de paralelismo $H_{01}$ .....	65
2.3.9.3 Teste de perfis coincidentes $H_{02}$ .....	67
2.3.9.4 Teste de hipótese de igual efeito de resposta $H_{03}$ .....	69
2.4 ANÁLISE DE AGRUPAMENTO .....	71
2.4.1 Medida de similaridade e dissimilaridade .....	73

2.4.2 Distância Euclidiana .....	74
2.4.3 Distância euclidiana padronizada .....	74
2.4.4 Formação de agrupamentos .....	76
2.4.5 Número de grupos .....	76
2.5 ANALISE FATORIAL .....	77
2.5.1 O modelo matemático da análise fatorial .....	78
<b>3 MATERIAL E MÉTODOS .....</b>	<b>82</b>
3.1 LEVANTAMENTO DOS DADOS .....	82
3.2 TRATAMENTO E ANÁLISE DOS DADOS .....	87
<b>4 RESULTADOS E DISCUSSÃO .....</b>	<b>90</b>
4.1 RESULTADOS .....	90
4.1.1 Análise de Perfis .....	90
4.1.2 Análise de Cluster .....	93
4.1.3 Análise fatorial .....	103
<b>5 CONSIDERAÇÕES FINAIS E CONCLUSÃO .....</b>	<b>108</b>
<b>6 REFERÊNCIAS .....</b>	<b>112</b>
<b>7 ANEXOS .....</b>	<b>117</b>

# 1 INTRODUÇÃO

## 1.1 O PROBLEMA

Dentre os muitos e complexos desafios da administração de um estado ou de um município, está a forma de como é realizado o direcionamento dos recursos econômicos, tendo em vista a melhoria, ampliação e modernização dos padrões sociais, principalmente nas áreas de educação, saúde e segurança pública. Em um cenário nacional, a gestão pública no Brasil é caracterizada atualmente pelo reconhecimento dos limites da sua atuação.

O maior desafio da administração municipal é o de projetar tendências futuras, traçar e operacionalizar estratégias de sobrevivência, alocando da forma mais eficiente e eficaz possível os recursos que estão à disposição. De modo a fornecer à sociedade serviços públicos de qualidade nas áreas sob sua responsabilidade, tais como: educação, saúde e segurança. Por outro lado deve promover a criação de empregos e o desenvolvimento, em geral do município.

O critério de divisão dos royalties provenientes do petróleo propicia que poucos Municípios trabalhem com realidades orçamentárias substancialmente elevadas em relação à realidade estadual e nacional, porém, é uma renda extra que é agregada ao orçamento. De acordo com dados do Inforoyalties<sup>2</sup>, no âmbito nacional, os royalties do petróleo beneficiam 17 estados da federação e 1038 Municípios.

Os royalties são compensações financeiras pagas aos estados e Municípios produtores, derivados de uma atividade finita, uma vez que o petróleo é um recurso natural não renovável. Como os Municípios sofrem forte impacto das atividades de produção de petróleo e gás, e, com vistas a evitar potenciais perdas de bem-estar nestas localidades como, por exemplo, degradação ambiental, o governo federal proporciona a tais Municípios uma

---

<sup>2</sup> Disponível em: < [www. http://inforoyalties.ucam-campos.br/](http://inforoyalties.ucam-campos.br/)> Acesso em: 05 de setembro de 2013

compensação financeira através de um percentual das receitas de royalties do petróleo.

Devido a essa ampla importância da receita dos royalties do petróleo no orçamento desses Municípios, justifica-se a necessidade de desenvolver um estudo sobre o impacto dessa renda extra nos índices de desenvolvimento humano (IDH) desses Municípios, detectando, se necessário, a necessidade de investimentos em atividades econômicas que promovam a melhoria da qualidade de vida e criação de novas alternativas de desenvolvimento sustentável para os Municípios, considerando que o petróleo é uma fonte não renovável.

## 1.2 OBJETIVOS

O crescimento da arrecadação de royalties, aliado ao problema da crescente escassez, característica de diversos recursos não renováveis, tem instigado vários estudos acadêmicos, bem como as discussões na sociedade. O presente trabalho tem como objetivo geral realizar uma investigação sobre as contribuições das receitas extraordinárias, originadas da atividade industrial de petróleo em alguns Municípios brasileiros, verificando se as mesmas contribuem para uma melhoria expressiva da qualidade de vida destes Municípios. Assim, através da comparação de índices como o IDHM pretende-se observar se os royalties têm contribuído para o desenvolvimento humano nos Municípios estudados.

Como objetivos específicos pretendem-se:

- a) Identificar na legislação brasileira quais são os critérios para a distribuição dos royalties;
- b) Utilizar a técnica estatística multivariada conhecida como Análise de Perfis para que, através de uma coleta de dados junto ao Instituto Brasileiro de Geografia e Estatística - IBGE, Agência Nacional do Petróleo - ANP, entre outros, possa ser verificado uma distinção entre

grupos diferentes, com base em índices, como por exemplo o Índice de Desenvolvimento Humano dos Municípios - IDHM.

Para que o referido objetivo possa ser atingido, é necessária a utilização de técnicas de inferência estatística, particularmente, as de testes de hipóteses, tanto as da estatística univariada, como as da estatística multivariada, bem como o estudo sobre as aproximações das distribuições de probabilidades que são utilizadas para testar as hipóteses da análise de perfis.

### 1.3 JUSTIFICATIVA

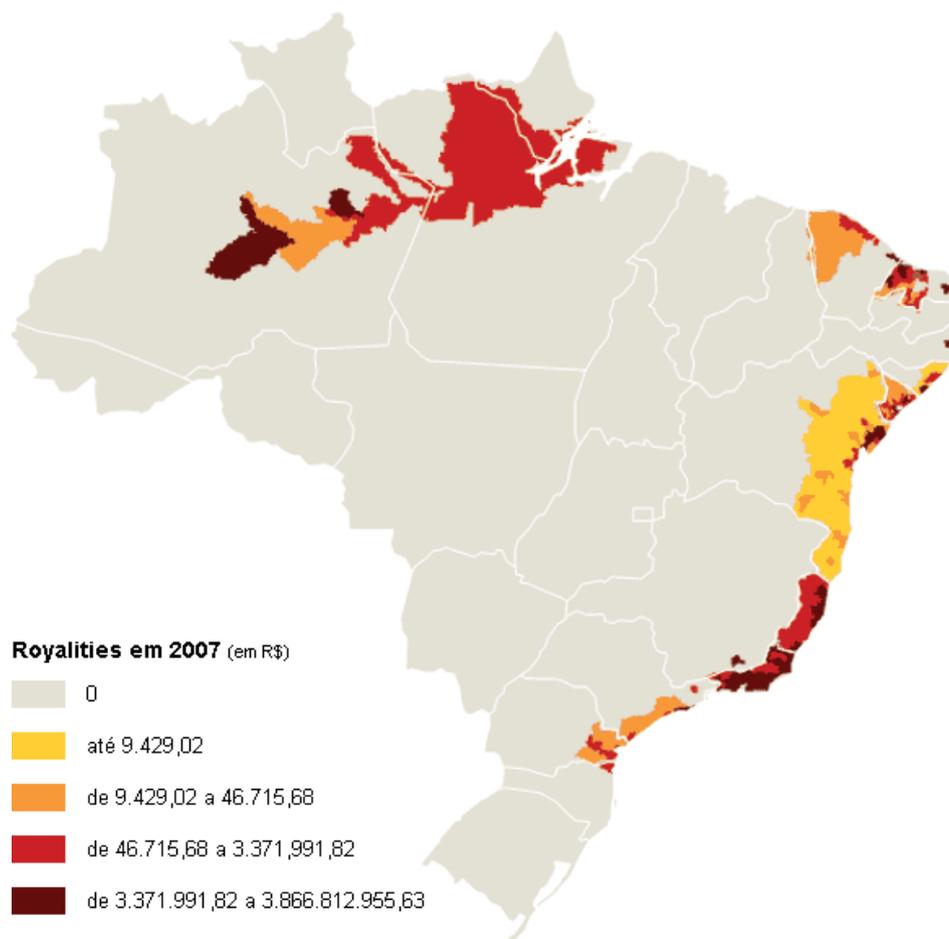
Segundo as informações obtidas da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP, 2013), no que se referem ao âmbito nacional, os royalties do petróleo beneficiam atualmente 17 estados da federação, sendo que o estado do Rio de Janeiro aparece como o maior beneficiado, seguidos dos estados de Rio Grande do Norte e Espírito Santo.

O pagamento dos royalties existe desde a publicação da Lei nº 2.004, de 1953, que criou a Petrobrás. Apesar disto, estes só passaram a ganhar expressividade em termos monetários após a publicação da Lei nº 9.478, de 6/08/97. Esta Lei estabeleceu uma reforma no quadro institucional do setor petrolífero nacional, em que foram estabelecidas as chamadas participações governamentais, sendo elas o bônus de assinatura, pagamento pela ocupação ou retenção da área, royalties e participações especiais.

Destas quatro participações, apenas os recursos originados das duas últimas são repartidos entre União, Estados e Municípios.

Na figura 1 podemos observar a localização dos Municípios que receberam royalties do petróleo no ano de 2007 e os seus respectivos valores em reais.

FIGURA 1: MUNICÍPIOS QUE RECEBERAM ROYALTIES DO PETRÓLEO EM 2007



Fonte: STN/Siafi; ANP 2008

Segundo Barbosa (2001), os royalties de cada município são calculados com base na produção do Estado do qual fazem parte, sendo que o rateio dos royalties devido aos Municípios pertencentes à uma mesma zona de produção será efetuado na razão direta de suas respectivas populações.

A Lei de 1997 regulamentou os royalties para os Municípios afetados pela produção de petróleo e estabeleceu novos critérios de cálculo para a partilha, que resultaram em um aumento significativo de tais recursos para as muitas prefeituras. A incontestável relevância dessas receitas no orçamento desses Municípios justifica a necessidade de desenvolver estudos sobre os impactos dessas em diversos índices dos Municípios em questão, indicando a

necessidade de investimentos em projetos ou atividades econômicas que promovam a melhoria da qualidade de vida e uma diversificação das bases produtivas locais para a criação de novas alternativas de desenvolvimento sustentável para os Municípios, considerando que o petróleo tende a se esgotar.

Utilizando-se dados do Instituto de Pesquisas Econômicas Aplicadas (IPEA), do Instituto Brasileiro de Geografia e Estatística (IBGE) e da Agência Nacional do Petróleo (ANP), foram coletados alguns indicadores sociais de 37 Municípios que mais receberam royalties, entre os anos de 2000 a 2013 tendo como comparativo, os indicadores sociais de 26 Municípios que não têm direito a estas transferências.

Diante de tal contextualização, chega-se a seguinte questão de pesquisa: o pagamento dos royalties aos 37 Municípios produtores, entre 2000 a 2013, tem promovido alguma mudança nos índices destes Municípios? Se existe uma diferença nesses índices, elas realmente são significativas?

Com relação à situação apresentada, deve-se de alguma forma verificar algumas hipóteses com relação a esses índices, verificando se as mesmas são verdadeiras ou falsas. Esse tipo de situação é resolvido por uma técnica estatística específica, chamada Análise de Perfis. Essa técnica é utilizada para testar três hipóteses levantadas a respeito dos dados de ambos os grupos: Municípios que recebem os royalties e Municípios que não recebem esses royalties.

Devido ao fato deste modelo utilizar um número muito grande de operações algébricas matriciais, as hipóteses serão testadas usando, como auxílio, os softwares “STATGRAPHICS”, “MINITAB” e “MATLAB”, onde os modelos serão implementados.

## 1.4 ESTRUTURA DO TRABALHO

A estrutura deste trabalho conta com mais quatro capítulos, além dessa introdução. No capítulo 2, apresenta-se uma revisão de literatura sobre a evolução e as formas de pagamentos dos royalties no Brasil, os tipos de royalties e quais Municípios têm o direito a receber esses valores. Além disso, é feita uma revisão de literatura sobre as técnicas da estatística multivariada utilizadas neste trabalho: análise de perfis, análise de agrupamentos e análise fatorial. No capítulo 3, descreve-se o material utilizado para aplicação das análises, ou seja, as variáveis utilizadas, bem como os métodos aplicados. O capítulo 4 apresenta os resultados das análises feitas com os programas elaborados e, finalmente, o capítulo 5, contém a conclusão do trabalho. A dissertação ainda possui, além das referências, os anexos para melhor compreensão do trabalho.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 FUNDAMENTAÇÃO TEÓRICA

#### 2.1.1 ROYALTIES

##### 2.1.1.1 Royalties de petróleo

Os royalties constituem uma das mais remotas formas de pagamento de direitos e propriedade. Royalty tem origem da palavra inglesa *royal*, que significa “da realeza” ou “relativo ao rei”. Originalmente, designava o direito que o rei tinha de receber pagamentos pelo uso de minerais em suas terras, conceito este que se estendeu no séc. XX a outras atividades extrativas de recursos naturais não renováveis, como o petróleo e o gás natural.

A expressão royalties usualmente designa o fluxo de pagamentos ao proprietário de um ativo não renovável (material ou imaterial) que o cede para ser explorado, usado ou comercializado por outras empresas ou indivíduos.

No caso brasileiro, os royalties do petróleo são uma compensação financeira devida ao Estado pelas empresas que exploram e produzem petróleo e gás natural. É uma remuneração à sociedade pela exploração desses recursos, que são escassos e não renováveis. Este pagamento é feito mensalmente (ANP, 2001).

Com base nesse entendimento, pode-se classificar o fluxo de pagamentos decorrente dos royalties do petróleo como uma receita patrimonial da União, proprietária dos recursos, recebida das empresas que exploram o petróleo. Esta receita é partilhada com estados e Municípios que contam com atividades petrolíferas em seus territórios, ou sofram impactos das mesmas, segundo critérios estabelecidos na legislação.

A Tabela 1 apresenta o total em royalties recebido pelos Municípios da pesquisa em 2012, fazendo uma comparação com os recursos recebidos pelo governo federal.

TABELA 1: COMPARAÇÃO ENTRE OS VALORES DOS ROYALTIES RECEBIDOS COM OS VALORES DOS RECURSOS RECEBIDOS PELO GOVERNO FEDERAL, SEGUNDO MUNICÍPIOS DA AMOSTRA - 2012

Município/ano	Royalties (R\$)	Transferência de Recursos por Estado/Município	Razão entre royalties e transferência (%)
Anchieta (ES)	53.409.715,08	81.331.994,62	65,67
Angra dos Reis (RJ)	96.628.816,91	208.884.630,39	46,26
Aracruz (ES)	34.515.133,81	95.536.148,36	36,13
Armação dos Búzios (RJ)	82.693.743,10	103.515.880,67	79,89
Balneário Barra do Sul (SC)	4.173.000,98	11.971.359,82	34,86
Bertioga (SP)	40.306.364,67	71.882.041,60	56,07
Cabo Frio (RJ)	324.982.556,21	420.215.628,20	77,34
Cachoeiras de Macacu (RJ)	40.184.457,83	82.334.892,77	48,81
Cachoeiro de Itapemirim (ES)	3.292.440,46	103.254.280,37	3,19
Cairu (BA)	7.397.994,99	29.742.606,79	24,87
Campos dos Goytacazes (RJ)	1.354.233.313,47	1.631.119.099,42	83,02
Carapebus (RJ)	38.640.170,21	50.658.310,84	76,28
Casimiro de Abreu (RJ)	112.016.299,75	148.056.500,98	75,66
Guapimirim (RJ)	47.619.544,11	81.308.138,27	58,57
Guararema (SP)	80.865.361,67	100.405.414,34	80,54
Guaratuba (PR) <sup>(1)</sup>	1.670.458,08	16.909.463,06	9,88
Ilhabela (SP)	43.817.856,33	67.658.554,10	64,76
Itapemirim (ES)	171.659.343,84	206.864.896,03	82,98
Linhares (ES)	111.312.698,16	229.078.449,88	48,59
Macaé (RJ)	538.654.219,38	658.266.410,61	81,83
Mangaratiba (RJ)	25.615.850,57	55.678.636,97	46,01
Maricá (RJ)	126.201.392,79	171.913.619,46	73,41
Matinhos (PR) <sup>(1)</sup>	2.125.509,52	14.642.244,78	14,52
Paranaguá (PR) <sup>(1)</sup>	60.774,31	49.453.570,16	0,12
Parati (RJ)	82.067.236,75	105.256.404,59	77,97
Pilar (AL)	3.691.871,75	45.973.922,22	8,03
Presidente Kennedy (ES)	223.513.390,34	281.785.947,54	79,32
Quissamã (RJ)	106.949.297,46	129.353.270,06	82,68
Rio das Ostras (RJ)	348.835.716,27	399.977.779,87	87,21
São Joao da Barra (RJ)	233.262.154,92	256.577.888,39	90,91
São Sebastiao (SP)	93.084.014,67	150.757.533,75	61,74
Silva Jardim (RJ)	36.027.403,59	58.802.377,65	61,27
Itapoá (SC)	4.173.000,98	16.750.612,21	24,91
Araquari (SC)	4.173.000,98	23.366.862,13	17,86
Garuva (SC)	4.173.000,98	16.687.259,36	25,01
Joinville (SC)	4.173.000,98	234.030.620,32	1,78
São Francisco do Sul (SC)	35.731.667,08	64.507.386,84	55,39

Fonte: Infroyalties/Portal da Transparência - 2013

(1) dados referentes ao ano de 2008

A motivação para cobrança dos royalties sobre exploração do petróleo tem como objetivo principal o de fornecer aos governos locais e regionais os recursos necessários para financiar investimentos geradores de riqueza para substituir a riqueza mineral finita.

Assim, o uso das receitas provenientes dos royalties deve estar direcionado para uma diversificação produtiva nas regiões produtoras e em suas áreas de influência, com a finalidade de garantir a capacidade de promover uma dinâmica econômica de desenvolvimento própria.

Compensar os impactos das atividades de exploração e produção, tais como impactos ambientais, custos de manutenção de rodovias entre outras, e aliviar os problemas causados por essas atividades como o aumento demográfico de trabalhadores, a sobrecarga da infraestrutura local e crescimento da demanda de serviços públicos exigem despesas extras que recaem sobre o conjunto da sociedade, sendo justo para esses casos uma compensação financeira adequada à realidade de cada município.

#### 2.1.1.2 A legislação brasileira dos royalties<sup>3</sup>

O que não é possível, nem justificável, é que Estados (petrolíferos) que servem de suporte, de apoio, que têm suas estruturas modificadas pela presença de uma nova tecnologia que lá aporta e, por isso mesmo, modifica os níveis de vida, obrigando que as infraestruturas estaduais sejam reforçadas, com ônus, altos para os seus parques cofres, nada recebam a não ser o orgulho de dizer: temos petróleo. Não é lícito, portanto, que esses Estados, sem nada, ainda arquem com os ônus de suportar essas despesas, sem nenhuma retribuição àquilo que a natureza colocou no confronto de seus territórios. (Discurso do Senador José Sarney, Anais do Congresso Nacional, 1971, vol. 5, pág. 79)

Juntamente com a criação da Petrobras, estabelecida pela Lei n.º 2.004, de 3 de outubro de 1953, foi constituído o direito ao pagamento de royalties

---

<sup>3</sup> Fonte: Guia dos royalties do petróleo e do gás natural. Rio de Janeiro: ANP, 2001

sobre o petróleo. O artigo 27 determinava a realização do pagamento de 4% aos estados e de 1% aos Municípios sobre o valor da produção terrestre de petróleo e gás natural em seus territórios.

Com o início da exploração do petróleo em mar aberto, a Lei n.º 7.453, de 27 de dezembro de 1985, determinou que este tipo de atividade também estivesse sujeita ao mesmo percentual de pagamento de royalties, distribuídos seguindo a seguinte hierarquia: 1,5% aos estados confrontantes com poços produtores; 1,5% aos Municípios confrontantes com poços produtores e àqueles pertencentes às áreas geoeconômicas dos Municípios confrontantes; 1% ao Ministério da Marinha e 1% para constituir o Fundo Especial, a ser distribuído entre todos os estados e Municípios da Federação. (Confederação Nacional de Municípios – 2010)

Em 22 de julho de 1986, a criação da Lei nº 7.525, definiu o que vem a ser uma região geoeconômica e a extensão dos limites territoriais dos estados e Municípios litorâneos na plataforma continental, sendo ambas de competência da Fundação Instituto Brasileiro de Geografia e Estatística (IBGE). A mesma lei atribuiu ao Tribunal de Contas da União a competência para a fiscalização do recurso em questão. Tais conceitos definidos por essa Lei, são os mesmos utilizados hoje na distribuição dos royalties decorrentes da produção marítima de petróleo e gás natural. Em 29 de agosto do mesmo ano, o Decreto 93.189, regulamentou o traçado de linhas de projeção dos limites territoriais dos estados, territórios e Municípios a ser utilizado pelo IBGE para a definição de poços confrontantes.

A distribuição dos royalties foi novamente modificada em 28 de dezembro de 1989 pela Lei nº 7.990, que só foi regulamentada em 11 de janeiro de 1991 pelo Decreto nº 01, distribuindo 0,5% aos Municípios onde se localizassem instalações de embarque e desembarque de petróleo ou de gás natural, 3,5% para os estados quando a lavra ocorresse em terra, e o percentual do Fundo Especial foi reduzido para 0,5%, quando a lavra ocorresse na plataforma continental.

Por fim, em 06 de agosto de 1997, entrou em vigor a Lei nº. 9.478, conhecida como a Lei do Petróleo. A Lei do Petróleo foi fruto da evolução da

indústria do petróleo no país e do crescimento gradativo de sua importância tanto para o mercado interno, quanto para o mercado internacional, em que são determinadas regras especiais para o segmento petrolífero. Essa nova Lei estabelece um papel distinto à Petrobrás, que até então detinha o monopólio estatal. A empresa passa a receber o mesmo tratamento previsto para as demais empresas privadas em processos licitatórios, salvo em situações de empate, apresentando-se como vencedora. A União continua a ter direito de decisão sobre a Petrobrás, pois ainda permanece como acionista majoritária.

A nova Lei não altera o princípio constitucional de que: “a pesquisa e a lavra das jazidas de petróleo e gás natural e outros hidrocarbonetos líquidos, bem como a importação e a exportação de derivados e o seu transporte dutoviário ou marítimo, constituem monopólio da União (Constituição Federal, Art. 177), mas a autoriza a contratar terceiros para a execução de tais atividades” (Postali, 2002).

Para administrar estas diretrizes, o Governo Federal criou a ANP, autarquia federal vinculada ao Ministério das Minas e Energia, tendo como finalidade “promover a regulação, a contratação e a fiscalização das atividades econômicas integrantes da indústria do petróleo” (Postali, 2002).

De maneira geral, a Lei do Petróleo prevê quatro modalidades básicas pelas quais o governo participa dos resultados da produção de petróleo:

- i) Bônus de Assinatura: valor pago para a obtenção da concessão da área;
- ii) Royalties: percentual sobre o valor da produção a ser dividido entre a União, Estados e Municípios;
- iii) Participações Especiais: valor devido em campos de alta rentabilidade, cujo cálculo incide sobre o lucro do petróleo produzido;
- iv) Pagamentos pela ocupação ou retenção de área: valor devido pela utilização das áreas sob concessão.

Sendo que os itens ii e iv são obrigatórios em todos os contratos de concessão.

Analisando os aspectos históricos sobre a Lei do Petróleo, podemos verificar que sua evolução teve como principal finalidade enquadrar o setor nos

princípios e objetivos da política energética nacional, dentre os quais pode-se destacar:

- ✓ preservar o interesse nacional;
- ✓ incrementar, em bases econômicas, a utilização do gás natural;
- ✓ promover a livre concorrência;
- ✓ atrair investimentos na produção de energia; e
- ✓ ampliar a competitividade do país no mercado internacional

### 2.1.1.3 Campo de petróleo

Um campo de petróleo ou de gás natural, definido segundo a Lei do Petróleo, é uma área produtora de petróleo e/ou de gás natural, a partir de um ou mais reservatórios contínuos, a profundidades variáveis, abrangendo instalações e equipamentos destinados à produção. Essa definição considera aspectos geológicos, econômicos e de engenharia.

De maneira geral, um campo de petróleo não produz unicamente o petróleo, mas produz também gás natural, havendo, em alguns, a predominância do petróleo e, noutros, a do gás natural. Na figura 2, estão localizados os campos de petróleos de exploração no mar numerados de 1 a 8 em suas respectivas localizações.

Em seguida, a tabela 2 apresenta a produção de cada campo de petróleo no ano de 2011 segundo o Boletim da Produção de Petróleo e Gás Natural da ANP. Na primeira coluna, temos a produção de petróleo em Barril por dia (bbl/d) para cada campo, na segunda coluna, temos a produção de gás natural em milhares de metros cúbicos por dia ( $Mm^3/d$ ) e na última coluna, temos a produção total do campo em barris de óleo equivalente por dia (boe/d).

FIGURA 2: ÁREAS SOB CONCESSÃO



Fonte: Informações Planimétricas – IBGE

TABELA 2: DISTRIBUIÇÃO DA PRODUÇÃO DE PETRÓLEO E GÁS NATURAL POR BACIA – 2011.

Bacia	Petróleo (bbl/d)	Gás Natural (Mm <sup>3</sup> /d)	Produção Total (boe/d)
Campos ( 7 )	1797729	26725	1965828
Santos ( 8 )	97298	5848	134079
Solimões	34853	11815	109170
Espírito Santo ( 5 e 6 )	49411	9161	107032
Potiguar ( 1 )	57328	1834	68863
Sergipe ( 2 )	45127	3133	64835
Recôncavo	43828	2828	61618
Camamu ( 4 )	318	4173	26567
Alagoas	5760	1557	15556
Ceará	5336	88	5892
Tucano ( 3 )	37	106	703
Total	2137025	67268	2560143

FONTE: ANP/SDP/SIGEP

Para definir se um campo é de petróleo ou de gás natural, a indústria do petróleo utiliza um coeficiente chamado razão gás-óleo, conhecido como RGO,

$$RGO = \frac{\text{Volume de gás produzido}}{\text{Volume de petróleo produzido}} \quad (2.1)$$

ambos medidos nas condições de referência.

Assim, é denominado gás natural associado ao petróleo quando, ao se produzir petróleo, produz-se também, inexoravelmente, o gás. Denomina-se gás natural não associado ao petróleo aquele gás que pode ser produzido de forma independente do petróleo.

Cada campo de petróleo e gás natural é delimitado por uma área geográfica e sua operação é regida por um contrato de concessão celebrado entre a ANP e a concessionária que detém os direitos de produção naquele campo.

Para fins de cálculo dos royalties, cada campo de petróleo ou gás natural é tratado como uma unidade de negócio em separado, ou seja, a cada campo corresponderá uma alíquota de royalties e preços próprios para petróleo e gás natural. Os preços serão utilizados para valorar a produção do campo e a alíquota será aplicada sobre o valor da produção, para calcular os royalties.

#### 2.1.1.4 Metodologia de cálculo dos royalties

Inicialmente, o cálculo para pagamento dos royalties tinha como base de cálculo o preço da refinaria, o qual costumava sofrer intervenções discricionárias do governo.

Devido algumas mudanças nas leis, como salienta Serra (2003), as compensações financeiras para Estados e Municípios são bem anteriores a 1997, mas a lei de 1997 instituiu mudanças importantes: a alíquota de royalties,

que subiu de 5% para 10%<sup>4</sup>, passou a ser aplicada sobre um preço de referência, ou seja, uma cotação média de preços de uma cesta de tipos de petróleo internacionais.

Os valores dos royalties são calculados mensalmente para cada campo produtor, mediante a aplicação da alíquota sobre o valor da produção de petróleo e de gás natural.

O valor de produção é obtido multiplicando-se os volumes produzidos em cada mês pelos preços de referência do respectivo mês, já convertidos para a moeda nacional.

$$\text{Valor da produção/mês} = \text{Volume} \times \text{Preço Referência} \quad (2.2)$$

$$\text{Royalty/mês} = \text{Alíquota} \times \text{Valor da Produção/mês} \quad (2.3)$$

onde:

Royalty: é o valor dos royalties decorrentes da produção do campo no mês, em R\$;

Alíquota: é o percentual que pode variar de um mínimo de 5% a um máximo de 10%;

Volume: é o volume da produção de petróleo do campo no mês, em m<sup>3</sup>;

Preço Referência: é o preço de referência do petróleo produzido no campo no mês, em R\$/m<sup>3</sup>;

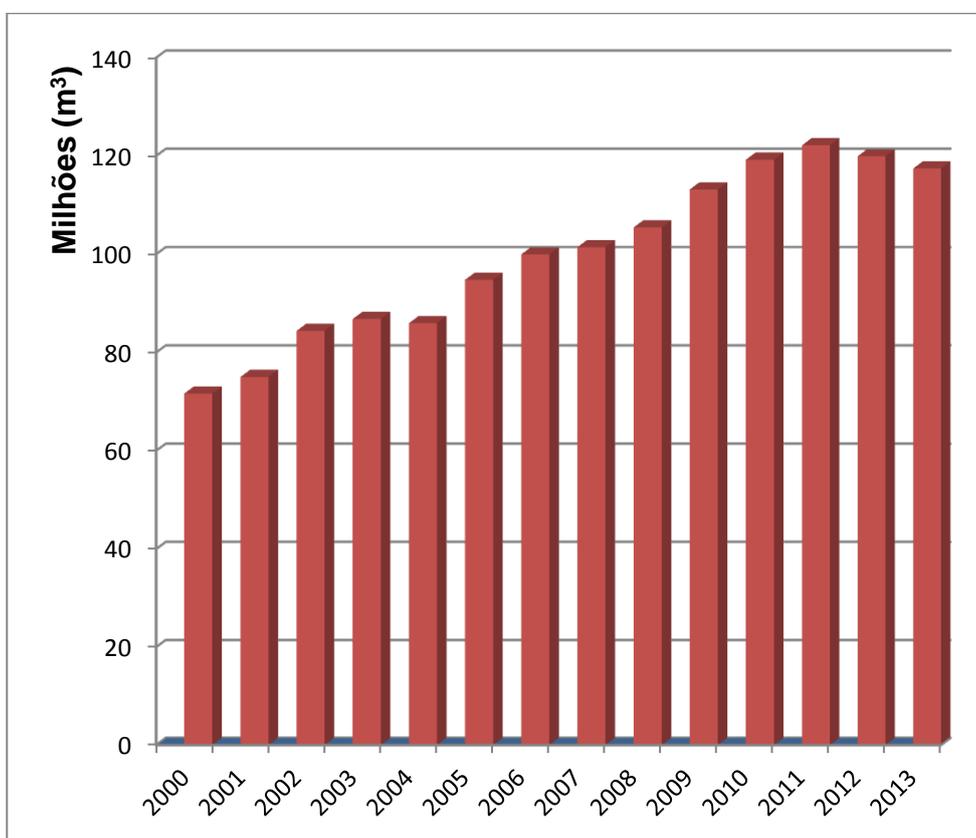
A lei de 1997 possibilitou também uma expansão nas possibilidades para o investimento dos royalties por parte dos Municípios, pois a lei anterior, que só autorizava as prefeituras a aplicarem estas receitas em investimentos em energia, meio ambiente, saneamento e rodovias, foi substituída, não estabelecendo nenhum critério específico, podendo ser destinados a quaisquer investimentos, vedando-se apenas a destinação de tais recursos para o pagamento de dívidas e para folha de pagamento.

---

<sup>4</sup> A lei autoriza a redução da alíquota em até 5 pontos percentuais, em caso de condições difíceis de produção.

Tal conjunto de mudanças derivou em um acréscimo expressivo no valor dos royalties repassados aos Municípios, tornando-o bem mais sensível às oscilações do preço internacional do petróleo e da taxa de câmbio, além de ampliar a margem de manobra para sua aplicação. Verifica-se no gráfico 1 que, com a aprovação das novas regras dos royalties, elevou-se a arrecadação de forma significativa após o ano 2000, tendo sua capacidade de produção quase que dobrada a partir de 2010, bem como a sua distribuição para os Municípios.

GRAFICO 1: PRODUÇÃO NACIONAL DE PETRÓLEO POR ANO (TERRA E MAR) - 2000-2013



Fonte: ANP - Boletim Mensal de Produção, conforme o Decreto n.º 2.705/98. Dados atualizados em 03 de fevereiro de 2014.

Na tabela 3, podemos verificar a produção de petróleo e gás natural por Estado. Nela notamos o domínio da produção pelo estado do Rio de Janeiro responsável por 72% da extração de petróleo. Espírito Santo aparece em segundo com pouco a mais de 16% da extração. Os demais Municípios totalizam os 12% restantes, sendo que Rio Grande no Norte e Bahia são responsáveis por quase metade dessa produção.

TABELA 3: PRODUÇÃO DE PETRÓLEO E GÁS NATURAL POR ESTADO

<b>Estado</b>	<b>Petróleo (bbl/d)</b>	<b>Gás Natural (Mm<sup>3</sup>/d)</b>	<b>Produção Total (boe/d)</b>
Rio de Janeiro	1.524.427	27.265	1.695.926
Espírito Santo	306.544	9.963	369.209
Amazonas	33.357	11.671	106.767
Bahia	43.712	9.122	101.092
Rio Grande do Norte	60.523	1.613	70.668
Sergipe	41.183	2.829	58.978
São Paulo	27.753	4.462	55.817
Alagoas	4.718	1.418	13.636
Ceará	5.546	59	5.915
<b>Total geral</b>	<b>2.047.763</b>	<b>68.402</b>	<b>2.478.009</b>

Fonte: ANP/SDP/SIGEP

Além dos Royalties calculados na produção do petróleo, existe também a chamada participação especial do petróleo, que é uma compensação financeira extraordinária, a qual foi criada pela Lei do Petróleo. Essas cobranças são feitas sobre os lucros realizados nos campos petrolíferos de elevada produção e rentabilidade, de acordo com o estabelecido no decreto presidencial 2.705/1998.

A tabela 4 mostra como essa forma de cobrança pode ser comparada a um imposto de renda, em que a coluna de alíquotas varia de 0% a 40%, incidindo sobre a receita líquida de cada campo de petróleo, dependendo do volume de produção, do tempo de produção e, no caso da plataforma continental, da profundidade em que a extração ocorre.

TABELA 4: CÁLCULO DE PARTICIPAÇÃO ESPECIAL

Milhares de barris/dia	Produção p/ faixa (A)	Alíquota (B)	“Alíquota efetiva (=AxB/Total)”
até 30	30	Isento	0%
30 – 60	30	10%	0,8%
60 – 90	30	20%	1,7%
90 – 120	30	30%	2,5%
120 – 150	30	35%	2,9%
Acima de 150	210	40%	23,3%
<b>TOTAL</b>	<b>360</b>		<b>31,3%</b>

Fonte: CNM

A fim de se obter o cálculo do valor a ser pago das participações especiais, deve-se multiplicar a alíquota efetiva ( $A_e$ ) obtida na última coluna da tabela 3 pela receita líquida ( $R_l$ ), que corresponde à receita bruta ( $R_b$ ) obtida com a venda de petróleo e gás menos os chamados gastos dedutíveis ( $G_d$ ), que incluem os próprios royalties comuns, os investimentos na exploração e os custos operacionais e de depreciação.

$$V = A_e \times R_l \quad (2.4)$$

onde

$$R_l = (R_b - G_d) \quad (2.5)$$

Porém, na prática, cada campo de petróleo está submetido a uma alíquota média ou efetiva diferente. Verifica-se que nos últimos tempos a maior alíquota efetiva é paga pelo campo de Roncador (Bacia de Campos), que produz em média 350 mil barris por dia e paga uma alíquota de 31,3%, sendo que a imensa maioria dos campos de petróleo do país é isenta de pagamento porque sua produção fica abaixo do mínimo estipulado para haver a cobrança. (CNM – 2010)

Quanto à receita da participação especial de petróleo, seja para produção em terra ou mar, a distribuição é feita seguindo uma regra bem

simples. Do total, 50% vai para a União, 40% para os Estados confrontantes e 10% para os Municípios confrontantes.

#### 2.1.1.5 O direito dos Municípios aos royalties do petróleo

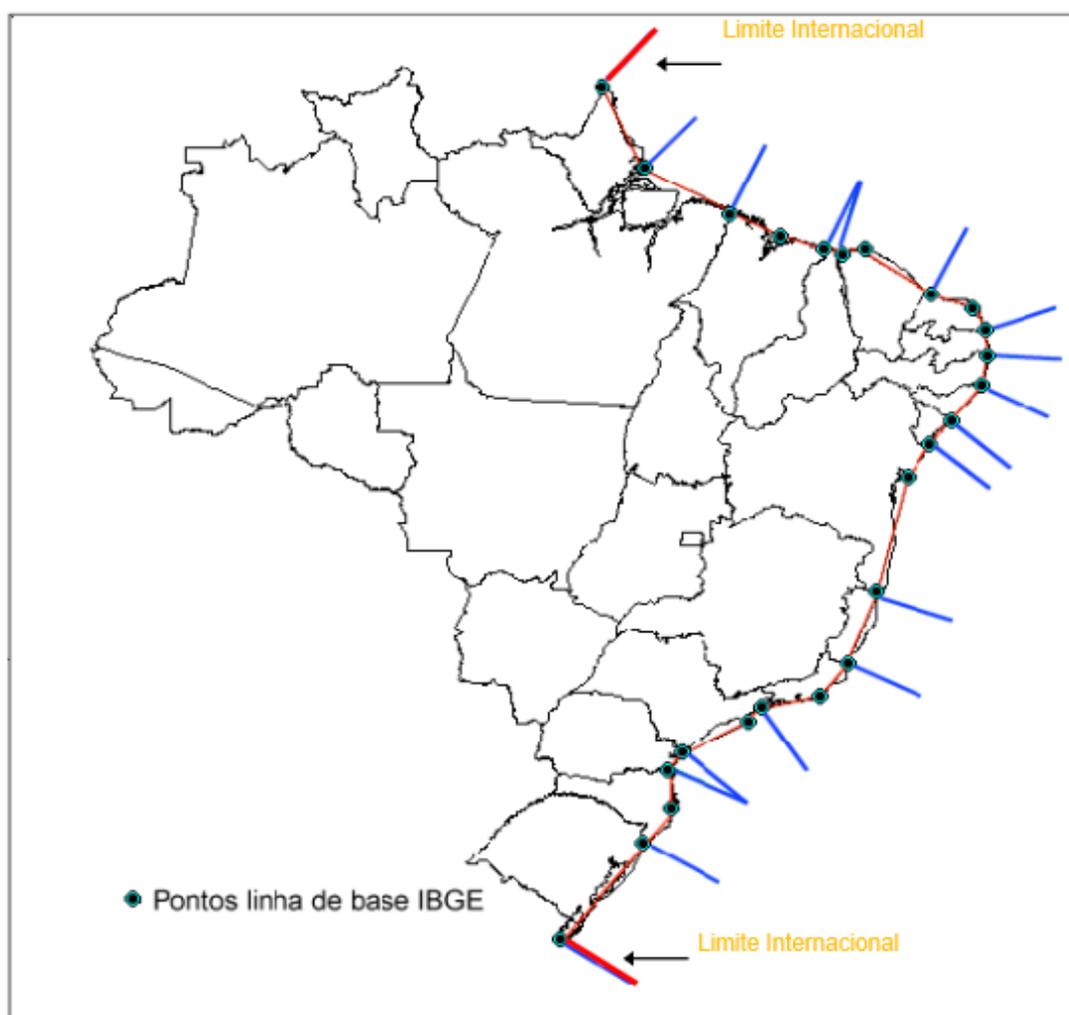
Conforme o artigo 20 da Constituição brasileira, ficou definido que são bens da União, dentre outros, os recursos minerais, inclusive os do subsolo, e é assegurada, aos Estados, ao Distrito Federal e aos Municípios, a participação no resultado da exploração de petróleo ou gás natural, tendo o direito de participar dos lucros da atividade, recebendo royalties e compensações. Na prática, são duas as leis ordinárias que definem qual fatia das rendas do petróleo cabe a cada ente da federação e como devem ser distribuídas.

A Lei 7.990/1989 institui as regras para a distribuição referentes à primeira parcela de 5% dos royalties. A Lei 9.478/1998 define as regras de distribuição da parcela excedente (acima de 5%) e da participação especial. Para ambas as leis existem diferenças nas repartições, que se altera também se a produção é em terra ou mar. Contudo, sabe-se que em média, os Municípios ficam somente com 26,8% da receita gerada pelos royalties, o restante é dividido entre a União com 39,4% dos royalties e participações especiais e os Estados, com 33,8%. Os critérios utilizados para determinar o direito de um Estado ou Município receber royalties são bastante complexos e têm sofrido alterações ao longo do tempo.

Para que um Município ou Estado tenha direito a receber os royalties do Petróleo, o mesmo deve se encaixar no conceito de Estado/Município “produtor”, quando a produção é em terra, sendo definido como o Estado e Município onde está localizado o poço ou campo de petróleo ou Estado/Município “confrontante” que neste caso é para a produção no mar. Neste caso, o Município precisa que, além de estar localizado no litoral, as linhas de projeção sobre a plataforma continental, traçadas a partir de seus limites com a costa, delimitem uma área na qual está inserido algum poço ou campo de petróleo.

Pela Lei nº 7.525/1986, ficou atribuído ao IBGE tratar as linhas de projeção dos limites territoriais dos Estados, Territórios e Municípios confrontantes. Para isso, foi utilizado dois tipos de linhas para demarcação: as linhas ortogonais e as linhas paralelas. No caso dos Estados, a divisão da plataforma continental é feita apenas por linhas ortogonais, como mostra a figura 3. No caso dos Municípios, são consideradas tanto as linhas ortogonais, quanto as linhas paralelas.

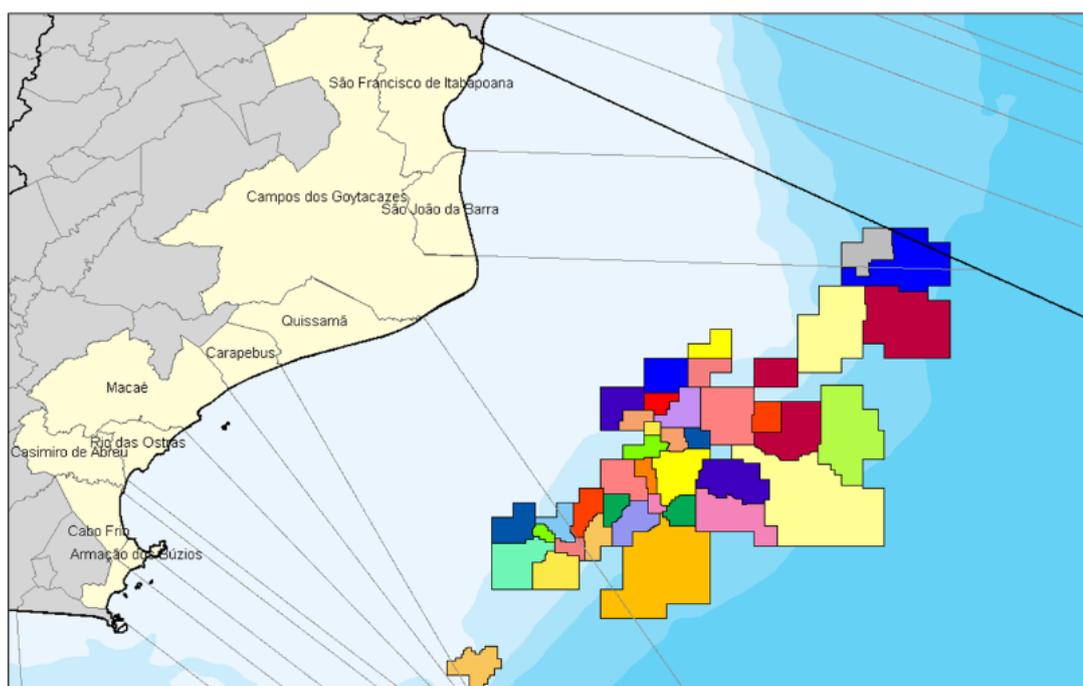
FIGURA 3: LIMITES INTERESTADUAIS NA PLATAFORMA CONTINENTAL



Fonte: ANP

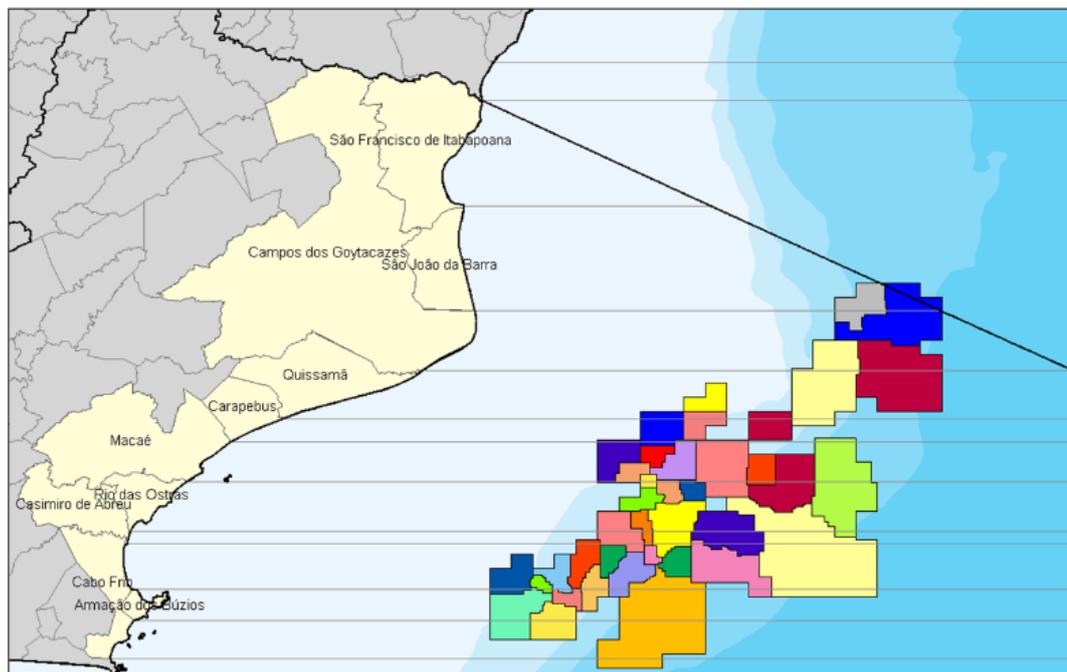
Nas figuras 4 e 5, temos um exemplo de delimitação de Municípios confrontantes com a bacia de Campos pelos critérios das ortogonais e das paralelas respectivamente. Na figura 4, podemos ver que pelo critério das ortogonais, o município de Campos dos Goytacazes possui direito aos royalties de quase todos os poços da bacia, tendo direito a uma grande fatia dessa distribuição. Na figura 5, podemos verificar que o critério das paralelas proporciona uma distribuição para uma maior quantidade de Municípios do estado do Rio de Janeiro.

FIGURA 4: LINHAS ORTOGONAIS DE CONFRONTAÇÃO DOS MUNICÍPIOS COM A PLATAFORMA CONTINENTAL



Fonte: ANP

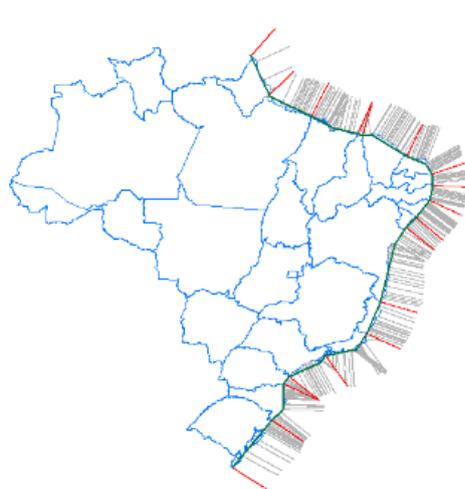
FIGURA 5: EXEMPLO: LINHAS PARALELAS DE CONFRONTAÇÃO DOS MUNICÍPIOS COM A PLATAFORMA CONTINENTAL



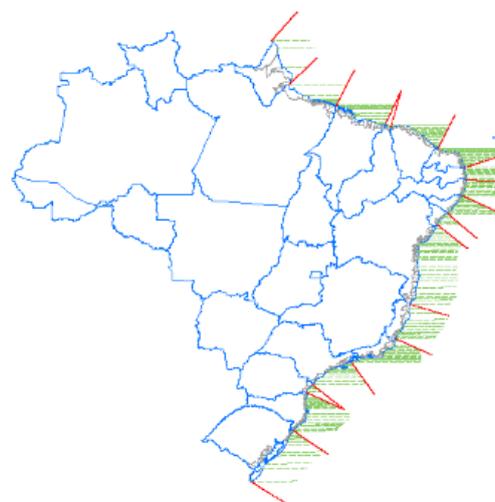
Fonte: ANP

Na figura 6, temos uma visão geral dos dois critérios utilizados para a determinação de Estados e Municípios produtores e confrontantes, pelos critérios das ortogonais e das paralelas, para todo o litoral brasileiro.

FIGURA 6: LINHAS ORTOGONAIS E PARALELAS



**Critério das Ortogonais**



**Critério dos Paralelos**

Fonte: IBGE

### 2.1.1.6 Áreas geoeconômicas

As legislações de 1985 e 1986 instituíram o conceito de área geoeconômica, estabelecendo os critérios para que os royalties de mar fossem distribuídos entre os Municípios. A Lei 7.453/1985 determinou que, não somente os Municípios confrontantes, mas os Municípios integrantes de sua área geoeconômica também recebessem a alíquota de 5% dos royalties. Assim, a Lei 7.525/1986 define as áreas geoeconômicas como aquela formada pelos Municípios pertencentes a três zonas distintas<sup>5</sup>:

- 1. ZONA DE PRODUÇÃO PRINCIPAL:** Municípios confrontantes e Municípios onde estiverem localizadas pelo menos três instalações de processamento, tratamento e armazenamento de petróleo e gás.
- 2. ZONA DE PRODUÇÃO SECUNDÁRIA:** Municípios atravessados por gasodutos ou oleodutos, incluindo as respectivas estações de compressão e bombeio, ligados diretamente ao escoamento da produção.
- 3. ZONA LIMÍTROFE:** Municípios contíguos àqueles integrantes da zona de produção principal ou que sofram as consequências sociais e econômicas da produção e exploração de petróleo e gás.

O Decreto 1/1991 regulamentou a Lei 7.990/1989, classificando como instalação de embarque e desembarque as monoboias, os quadros de boias múltiplas, os píeres de atração, os cais acostáveis e as estações terrestres coletoras de campos produtores e de transferência de óleo bruto e gás natural. Com isso, a partir de 1989, os Municípios onde se localizavam instalações de embarque e desembarque de petróleo, originados da produção terrestre ou marítima, passaram a ter direito a royalties.

A Lei do Petróleo (nº. 9.478/1997) estendeu o direito de receber royalties aos Municípios afetados pelas operações de embarque e desembarque, o que inclui todos os Municípios vizinhos àqueles onde está localizado as instalações. Contudo nesse mesmo ano, o conceito de área geoeconômica foi extinto.

---

<sup>5</sup> Estudos Técnicos CNM / Confederação Nacional de Municípios – Brasília : CNM, 2010

Em resumo, a distribuição de royalties segue a seguinte regra atualmente:

- parcela de 5%, lavra em terra: 70% para os Estados produtores, sendo que um quarto disso (17,5%) é dividido entre seus Municípios, 20% para os Municípios produtores, 10% para os Municípios com instalações de embarque e desembarque de petróleo e gás natural;

- parcela de 5%, lavra em plataforma continental: 30% para os Estados confrontantes com poços, sendo que um quarto (7,5%) é dividido entre seus Municípios, 30% para os Municípios confrontantes com poços, 20% para o Comando da Marinha, 10% para o Fundo Especial (composto por estados e Municípios), 10% para os Municípios com instalações de embarque e desembarque;

- parcela acima de 5%, lavra em terra: 52,5% para os Estados produtores, 25% para o Ministério Ciência e Tecnologia, 15% para os Municípios produtores, 7,5% para os Municípios afetados por operações nas instalações de embarque e desembarque;

- parcela acima de 5%, lavra em plataforma continental: 25% para o Ministério de Ciência e Tecnologia, 22,5% para os Estados confrontantes com campos, 22,5% para os Municípios confrontantes com campos, 15% para o Comando da Marinha, 7,5% para o Fundo Especial e 7,5% Municípios afetados por operações nas instalações de embarque e desembarque.

A participação especial é um tributo adicional em campos com grandes volumes de produção ou elevada rentabilidade sendo que 40% dos recursos seriam destinados ao Ministério de Minas e Energia e 10% para o Ministério do Meio Ambiente, 40% aos estados e 10% aos Municípios.

## 2.2 REVISAO DE LITERATURA

A estatística multivariada, por intermédio de suas técnicas, visa investigar cientificamente fenômenos que englobam diversas variáveis influentes, permitindo dessa forma, estudar as relações entre as variáveis de maneira integrada e evoluída, em geral fornecendo resultados consistentes.

Postali (2007) avaliou se os royalties distribuídos aos Municípios na forma da nova Lei (nº. 9.478/1997) contribuíram para o crescimento do seu Produto Interno Bruto (PIB). Para isso, utilizou-se o estimador de diferenças em diferenças (DD), que consiste em comparar a evolução do PIB per capita nos Municípios afetados e não afetados pela aprovação da nova Lei, assumindo-se que a nova legislação constitui um experimento natural. Os dados referem-se às taxas de crescimento do PIB per capita municipal antes e depois do evento. Os resultados confirmam a chamada “maldição dos recursos naturais” da literatura mundial, mostrando que os Municípios contemplados com royalties cresceram menos que os Municípios que não receberam tais recursos.

Dávila (2007) fez uma comparação dos níveis de satisfação dos pacientes nos Serviços de Cuidados Intensivos e Intermediários e Alojamento Conjunto do Instituto Nacional Materno Perinatal (INMP) de Lima (Peru) utilizando inferência multivariada. Utilizou-se dados gerados no trabalho de Solano, O. et al.(2006), no qual foi analisado os níveis de satisfação nos serviços de Cuidados Intensivos e Intermediários; e outra pesquisa em que se estudaram os níveis de satisfação no Serviço de Alojamento Conjunto no INMP, Solano, O. et al (2006).

Gribskov (1987) utilizou-se da análise de perfil para a detecção de famílias de proteínas através de uma sequência de comparação. A base para a comparação não foi apenas a habitual matriz de mutação “Dayhoff”, mas também os resultados de estudos estruturais e informações implícitas nos alinhamentos das sequências de famílias de proteínas semelhantes. As informações foram expressas numa tabela de posição específica (perfil), criado a partir de um grupo de sequências previamente alinhados por estruturas ou similaridade. A semelhança com outras sequências (alvo) para o grupo de sequências alinhadas (sonda) foi testada comparando-se o alvo para o perfil usando algoritmos de programação dinâmica. Em resumo, os Testes mostraram que a análise de perfil pode distinguir todos os membros dessas famílias e de todas as outras sequências em um banco de dados contendo 3800 sequências de proteínas.

Castro (2011) aplicou o teste  $T^2$  de Hotelling em dados de comunidades de plantas daninhas a fim de investigar a semelhança entre os padrões de infestação contidos nos grupos resultantes da análise de agrupamento por método hierárquico pela qual os dados de infestação de plantas daninhas foram submetidos. O conjunto de dados utilizado foi constituído por 782 registros de plantas daninhas contidos em talhões de cana-de-açúcar. A análise de agrupamento por método hierárquico possibilitou uma divisão das comunidades de plantas daninhas em 12 grupos que diferiram entre si dois a dois pelo teste  $T^2$  de Hotelling ( $p < 0,01$ ) indicando a necessidade de processos de controles individuais para cada grupo de comunidades de plantas daninhas.

Adamczuk (2003) desenvolveu um trabalho revisando as técnicas estatísticas da Análise de Perfis; e, além disso, desenvolveu um software em

linguagem visual aplicando em uma metalúrgica, empregando os métodos a um conjunto de dados obtidos através de ensaios em um fogão a gás, modelo quatro boca. O método foi aplicado em um caso real, em três situações, utilizando para isto o programa computacional desenvolvido, na qual foi possível verificar através dos resultados obtidos o desempenho da metodologia. Para o desenvolvimento do programa utilizou-se a linguagem de programação C/C++.

### 2.3 INFERÊNCIA ESTATÍSTICA

Esta seção baseia-se principalmente na bibliografia de Johnson e Wichern, [1998], Morrison, [1976], Chaves Neto, [2006] e Marques [2013].

A necessidade de se inferir, a partir de dados mensurados de amostra, sobre os fenômenos físicos ou até mesmo sociais é um fato rotineiro para um pesquisador. A pesquisa científica se constitui num processo interativo de aprendizado e segue, muitas vezes, o ciclo PDCA, ou seja, o de planejar, fazer, verificar o resultado, agir e começar novamente o processo. De acordo com Chaves Neto (2006), as inferências dizem respeito a um número finito de parâmetros e dependem da forma especificada para a função densidade de probabilidade da variável aleatória envolvida no estudo. As informações obtidas a partir de uma amostra são úteis para produzir inferências sobre parâmetros da população original.

A fim de esclarecer um fenômeno, um pesquisador em geral coleta e analisa dados com objetivo de validar, ou não, uma hipótese. Ao analisar os dados coletados, sejam eles de uma experimentação ou obtidos através de um questionário, o experimentador deve aplicar a técnica estatística adequada à hipótese em verificação.

A necessidade de compreensão das relações entre diversas variáveis faz com que as análises sejam muitas vezes complexas. Contudo, as técnicas de análise multivariada têm como objetivo:

- ✓ Redução de dados ou simplificação estrutural
- ✓ Ordenação e agrupamento
- ✓ Investigação da dependência entre variáveis
- ✓ Predição
- ✓ Construção e teste de hipóteses.

Modelos multivariados possuem em geral, um propósito através do qual o pesquisador pode testar ou inferir a respeito de uma hipótese sobre um determinado fenômeno, por meio de vetores médios ou matrizes de correlação.

### 2.3.1 Variável aleatória

Sejam  $\mathcal{E}$  um experimento, com  $\mathcal{E} \sim N(0,1)$  e  $S$  um espaço amostral associado ao experimento. Uma função  $X$ , que associe a cada elemento  $s \in S$  um número real,  $X(s)$ , é denominada variável aleatória, ou seja:

$$X(s): S \rightarrow R \quad (2.6)$$

A variável aleatória é chamada de discreta quando o seu contradomínio é um conjunto finito ou infinito enumerável, e é chamada de contínua quando o seu contradomínio é um conjunto infinito. (Meyer, 1978)

### 2.3.2 Vetor aleatório

Um vetor  $\underline{X}$  pode ser representado geometricamente como um segmento de reta, orientado em  $p$  dimensões, com componentes  $X_1$  ao longo do primeiro eixo,  $X_2$  ao longo do segundo eixo, ...,  $X_p$  ao longo do  $p$ -ésimo eixo. O vetor aleatório é aquele cujos elementos são variáveis aleatórias. É representado por:

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \quad \text{ou} \quad \underline{X}' = [X_1 \quad X_2 \quad \dots \quad X_p] \quad (2.7)$$

Da mesma forma que para uma variável aleatória, em que a esperança  $\mu$  e a variância  $\sigma^2$  caracterizam a distribuição, assim também é para o caso multivariado, na qual a distribuição multivariada ficará determinada pelo vetor de médias  $\underline{\mu}$  e pela matriz de covariância  $\Sigma$ .

Uma observação multivariada é uma coleção de medidas em  $p$  variáveis tomadas na mesma unidade amostral sendo que as  $n$  observações amostrais obtidas ficam dispostas em uma Matriz  $X$  de ordem  $n \times p$ . A expressão 2.1 mostra essa matriz  $X$ .

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2.8)$$

Organizada de forma que cada linha de  $X$  representa a  $n$ -ésima observação multivariada. Dado que o conjunto todo de mensurações é muitas vezes uma particular realização de variáveis aleatórias, diz-se que os dados representam uma amostra de tamanho  $n$  de uma população  $p$  variada.

### 2.3.3 Vetor de médias

Seja  $\underline{X}$  um vetor aleatório, de dimensão  $p$ . Então, a esperança ou vetor médio desse vetor aleatório é:

$$E(\underline{X}) = \underline{\mu} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad (2.9)$$

Para a  $h$ -ésima variável aleatória,  $E(X_h) = \mu_h$ , com  $h = 1, 2, \dots, p$ .

### 2.3.4 Matriz de Covariância

Seja  $\underline{X}$  um vetor aleatório de dimensão  $p$ , então cada elemento de  $\underline{X}$  é uma variável aleatória com sua própria distribuição de probabilidade marginal.

As médias marginais  $\mu_i$  e as variâncias  $\sigma_i^2$  ficam definidas como  $\mu_i = E(X_i)$  e  $\sigma_i^2 = E(X_i - \mu_i)^2$ ,  $i = 1, 2, \dots, p$ , respectivamente.

Especificamente, temos que<sup>6</sup>:

$$\mu_i = \begin{cases} \int_{-\infty}^{+\infty} x_i f_i(x_i) dx_i & \text{se } X_i \text{ é uma V.A.C}^* \text{ com fdp } f_i(x_i) \\ \sum x_i p_i(x_i) & \text{se } X_i \text{ é uma V.A.D}^{**} \text{ com função de probabilidade } p_i(x_i) \end{cases} \quad (2.10)$$

$$\sigma_i^2 = \begin{cases} \int_{-\infty}^{+\infty} (x_i - \mu_i)^2 f_i(x_i) dx_i & \text{se } X_i \text{ é uma V.A.C}^* \text{ com fdp } f_i(x_i) \\ \sum (x_i - \mu_i)^2 p_i(x_i) & \text{se } X_i \text{ é uma V.A.D}^{**} \text{ com função de probabilidade } p_i(x_i) \end{cases} \quad (2.11)$$

O comportamento de qualquer par de variáveis aleatórias, tal como  $X_i$  e  $X_k$ , é descrito por a) sua função de probabilidade conjunta  $f(x_i, x_k)$  e b) uma medida da associação linear entre elas é fornecida pela covariância  $\sigma_{ik}$ , em que:

$$\sigma_{ik} = E(X_i - \mu_i)(X_k - \mu_k) \quad (2.12)$$

As médias e covariâncias do vetor aleatório  $\underline{X}$  podem ser expressas como vetor e matriz, respectivamente. O valor esperado de cada elemento está contido no vetor de médias  $\mu = E(\underline{X})$  e as  $p$  variâncias  $\sigma_{ii}$  e as  $\frac{p(p-1)}{2}$  covariâncias distintas  $\sigma_{ik}$  ( $i < k$ ) estão contidas na matriz de variâncias e covariâncias  $\Sigma = E[(\underline{X} - \mu)(\underline{X} - \mu)']$ .

Assim,

---

<sup>6</sup> \* V.A.C. : Variável Aleatória Contínua

\*\* V.A.D. : Variável Aleatória Discreta

$$E[\underline{X} - E(\underline{X})]^2 = E[\underline{X} - \underline{\mu}]^2 = E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'] \quad (2.13)$$

$$= E \left\{ \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 & X_2 - \mu_2 & \cdots & X_p - \mu_p \end{bmatrix} \right\} \quad (2.14)$$

$$= E \left\{ \begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \cdots & (X_p - \mu_p)^2 \end{bmatrix} \right\} \quad (2.15)$$

$$= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) & \cdots & E(X_1 - \mu_1)(X_p - \mu_p) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 & \cdots & E(X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p - \mu_p)(X_1 - \mu_1) & E(X_p - \mu_p)(X_2 - \mu_2) & \cdots & E(X_p - \mu_p)^2 \end{bmatrix} \quad (2.16)$$

$$\sum E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix} \quad (2.17)$$

Em que,  $\underline{\mu}$  e  $\Sigma$  representam o vetor de médias populacionais e a matriz de variâncias e covariâncias populacionais, respectivamente.

A distribuição normal multivariada fica completamente especificada quando o vetor de médias  $\underline{\mu}$  e a matriz de covariâncias  $\Sigma$ , são conhecidos. Por esta razão estas quantidades têm uma importante função em muitos procedimentos multivariados.

### 2.3.5 Matriz de Correlação

Uma melhor interpretação da variabilidade entre as variáveis pode ser obtida usando-se a matriz de correlação ao invés de matriz de covariância, pois a matriz de correlação é obtida de variáveis padronizadas, eliminando assim o efeito de escalas diferentes para as várias variáveis. A medida de associação usada é o coeficiente de correlação populacional,  $\rho_{ik}$ . O coeficiente de correlação  $\rho_{ik}$  é definido por:

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}} \cdot \sqrt{\sigma_{kk}}} \quad (2.18)$$

O coeficiente de correlação mede a quantidade de associação linear entre um par de variáveis aleatórias. Para tanto, consiste em gerar uma matriz de correlação populacional,  $\rho$ , simétrica de ordem ( $p \times p$ ); em que:

$$\rho = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \dots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{1p}}} \\ \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \dots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix} \quad (2.19)$$

ou, de forma simplificada:

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{bmatrix} \quad (2.20)$$

Podemos gerar uma matriz diagonal ( $p \times p$ ) de desvios padrão, designado pela matriz  $V^{\frac{1}{2}}$ , em que:

$$V^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{\sigma_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\sigma_{pp}} \end{bmatrix} \quad (2.21)$$

Então, pode-se verificar que:

$$V^{\frac{1}{2}} \rho V^{\frac{1}{2}} = \Sigma \quad (2.22)$$

e, portanto:

$$\rho = \left(V^{\frac{1}{2}}\right)^{-1} \Sigma \left(V^{\frac{1}{2}}\right)^{-1} \quad (2.23)$$

Assim, tem-se que  $\Sigma$  pode ser obtida de  $V^{\frac{1}{2}}$  e  $\rho$ , enquanto  $\rho$  pode ser obtida de  $\Sigma$ . Estas expressões permitem que os cálculos sejam convenientemente implementados em um computador, pois não admite divisão por zero.

### 2.3.6 Vetor de médias e a matriz de covariâncias para combinações lineares de variáveis aleatórias

Sendo  $c$  uma constante e  $X_1$  uma única variável aleatória quaisquer, tem-se que:

$$E(cX_1) = cE(X_1) = \mu_1 \quad (2.24)$$

e

$$Var(cX_1) = E(cX_1 - c\mu_1)^2 = c^2 Var(X_1) = c^2 \sigma_{11} \quad (2.25)$$

Se  $X_2$  é uma segunda variável aleatória e  $a$  e  $b$  são constantes, então usando propriedades de esperança, temos:

$$\begin{aligned}
 \text{Cov}(a.X_1, b.X_2) &= E[(a.X_1 - a.\mu_1)(b.X_2 - b.\mu_2)] \\
 &= ab E[(X_1 - \mu_1)(X_2 - \mu_2)] \\
 &= ab.\text{Cov}(X_1, X_2) && ( 2.26) \\
 &= ab\sigma_{12}
 \end{aligned}$$

Finalmente para uma combinação linear  $aX_1 + bX_2$ , temos:

$$E(aX_1 + bX_2) = aE(X_1) + bE(X_2) = a\mu_1 + b\mu_2$$

$$\begin{aligned}
 \text{Var}(aX_1 + bX_2) &= E[(aX_1 + bX_2) - (a\mu_1 + b\mu_2)]^2 \\
 &= E[a(X_1 - \mu_1) + b(X_2 - \mu_2)]^2 \\
 &= E[a^2(X_1 - \mu_1)^2 + b^2(X_2 - \mu_2)^2 + 2ab(X_1 - \mu_1)(X_2 - \mu_2)] \\
 &= a^2 E(X_1 - \mu_1)^2 + b^2 E(X_2 - \mu_2)^2 + 2ab \text{cov}(X_1, X_2) \\
 &= a^2 V(X_1) + b^2 V(X_2) + 2ab \text{cov}(X_1, X_2) && ( 2.27 ) \\
 &= a^2 \sigma_1^2 + b^2 \sigma_2^2 + 2ab \sigma_{12}
 \end{aligned}$$

Escrevendo na forma matricial:

Façamos  $\underline{c}' = [a, b]$ , a combinação linear:  $aX_1 + bX_2$  pode ser escrito como:

$$\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \underline{\mathbf{c}}' \underline{\mathbf{X}} \quad (2.28)$$

Similarmente,  $E(aX_1 + bX_2) = a\mu_1 + b\mu_2$  pode ser expresso como:

$$\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \underline{\mathbf{c}}' \underline{\boldsymbol{\mu}} \quad (2.29)$$

e se

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \quad (2.30)$$

Representa a matriz covariância de  $\underline{\mathbf{X}}$ , então:

$$\text{Var}\left(\begin{bmatrix} a, b \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right) =$$

$$\text{Var}(aX_1 + bX_2) = \text{Var}(\underline{\mathbf{c}}' \underline{\mathbf{X}}) = \underline{\mathbf{c}}' \Sigma \underline{\mathbf{c}} \quad (2.31)$$

Assim,

$$\underline{\mathbf{c}}' \Sigma \underline{\mathbf{c}} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \quad (2.32)$$

$$= a^2 \sigma_1^2 + 2ab \sigma_{12} + b^2 \sigma_2^2$$

## 2.3.7 AMOSTRA ALEATÓRIA E ESTATÍSTICA DESCRITIVA

### 2.3.7.1 Amostragem aleatória

Uma única observação multivariada é uma coleção de medidas sobre  $p$  diferentes variáveis cada qual com  $n$  observações, que podem ser arranjadas em uma matriz  $n \times p$  de  $X_{ij}$ :

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \underline{X}'_1 \\ \underline{X}'_2 \\ \vdots \\ \underline{X}'_n \end{bmatrix} \quad (2.33)$$

As linhas de  $X$  representam  $n$  pontos no espaço observável  $p$ -dimensional.

O vetor linha  $X'_j$ , representando a  $j$ -ésima observação, contém as coordenadas de um ponto, sendo que o gráfico de  $n$  pontos no espaço  $p$ -dimensional fornece informação sobre locações e variabilidade.

Se considerarmos os pontos como esferas sólidas, o vetor de média amostral,  $\bar{X}$ , representa o centro de gravidade. A variabilidade ocorre em mais do que uma direção e é quantificada pela matriz de covariância amostral,  $S_n$ . Uma medida numérica simples de variabilidade é fornecida pelo determinante da matriz  $S_n$ , e este determinante significa que quanto maior o seu valor, mais disperso é a distribuição de valores de  $\bar{X}$ .

### 2.3.7.2 Estatística descritiva

Seja  $(X_1, X_2, \dots, X_n)$  uma amostra aleatória do vetor aleatório  $X$  de uma população com esperança  $\underline{\mu} = E(\underline{X})$  e matriz de covariância  $\text{Cov}(X) = \Sigma$ . Então a estimativa não tendenciosa para o vetor de média populacional  $\underline{\mu}$  é o vetor de média amostral  $\bar{X}$ , dado por:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.34)$$

E a estimativa não tendenciosa para a matriz de covariância  $\Sigma$  é a matriz de covariância amostral  $S$ , dada por:

$$S = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \quad (2.35)$$

A variância amostral,  $S_h^2$ , da  $h$ -ésima variável aleatória  $X_h$  é dada por:

$$S_h^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ih} - \bar{X}_h)^2 \quad (2.36)$$

e a covariância amostral entre as variáveis aleatórias  $X_r$  e  $X_v$  é dada por:

$$S_{rv} = \frac{1}{n-1} \sum_{i=1}^n (X_{ir} - \bar{X}_r)(X_{iv} - \bar{X}_v) \quad (2.37)$$

Comumente as amostras são coletadas de  $k$  grupos populacionais independentes. Portanto seja  $(\underline{X}_{1j}, \underline{X}_{2j}, \dots, \underline{X}_{nj})$  uma amostra aleatória do vetor

aleatório  $\underline{X}_j$ , pertencente a  $j$ -ésima população, com esperança  $E(\underline{X}_j) = \underline{\mu}_j$  e matriz de covariância  $\Sigma_j = \Sigma$  comum para todos os grupos.

Porém, como as populações apresentam matriz de covariância comum, tem-se  $\Sigma_1 = \Sigma_2 = \dots = \Sigma$ . Neste caso, a estimativa para a matriz de covariância  $\Sigma$  é a matriz de covariância comum a todos os grupos amostrais, chamada de conjunta, dada por:

$$S_p = \frac{1}{n-k} \sum_{j=1}^k \sum_{i=1}^{n_j} (\underline{X}_{ij} - \bar{\underline{X}}_j)(\underline{X}_{ij} - \bar{\underline{X}}_j)' \quad (2.38)$$

De modo que  $S_p$ , é a estimativa não tendenciosa de  $\Sigma$ ,  $E(S_p) = \Sigma$ . Muitas vezes é conveniente escrever a matriz  $S_p$ , da seguinte forma:

$$S_p = \frac{1}{n-k} [(n_1 - 1) \cdot S_1 + (n_2 - 1) \cdot S_2 + \dots + (n_k - 1) \cdot S_k] \quad (2.39)$$

### 2.3.7.3 Estatísticas descritivas de combinações lineares

Muitos processos de estatística multivariada, tais como na aplicação dos testes de hipóteses da análise de perfis e análise de medidas repetidas, deve-se fazer a transformação dos dados através das duas combinações lineares.

A primeira é do tipo

$$Y = \underline{c}' \cdot \underline{X} = c_1 \cdot X_1 + c_2 \cdot X_2 + \dots + c_p \cdot X_p \quad (2.40)$$

em que:

- ✓  $Y$  é a variável aleatória transformada pela aplicação da combinação linear  $\underline{c}' \cdot \underline{X}$ ;
- ✓  $\underline{c}$  o vetor de coeficientes;
- ✓  $\underline{X}$  o vetor aleatório original.

A segunda é a combinação linear tipo  $\underline{Y} = C' \cdot \underline{X}$ , em que:

- ✓  $\underline{Y}$  é o vetor aleatório transformado obtido da combinação linear  $C' \cdot \underline{X}$ ;
- ✓  $C$  é a matriz formada por  $q$  vetores de coeficientes.

A fim de se efetuar as duas transformações sugeridas acima, deve-se primeiramente obter uma amostra aleatória  $X$  de tamanho  $n$  em  $p$  variáveis, de uma população com vetor de médias  $E(\underline{X}) = \underline{\mu}$  e matriz de covariância  $\text{Cov}(X) = \Sigma$ , estimadas respectivamente através do vetor de médias amostrais  $\bar{\underline{X}}$  e pela matriz de covariância amostral  $S$ .

A amostra aleatória é:

$$X = \begin{bmatrix} \underline{X}'_1 \\ \underline{X}'_2 \\ \vdots \\ \underline{X}'_n \end{bmatrix} \quad (2.41)$$

Seja então a amostra aleatória  $X$  em  $p$  variáveis com vetor de média amostral  $\bar{\underline{X}}$  e matriz de covariância amostral  $S$ , na qual se deseja aplicar a combinação linear do tipo  $Y = \underline{c}' \cdot \underline{X}$  que transforma os dados de  $p$  variáveis em uma única variável, resultando no seguinte conjunto de dados aleatórios,  $\underline{Y}$ :

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \underline{c}' \cdot \underline{X}_1 \\ \underline{c}' \cdot \underline{X}_2 \\ \vdots \\ \underline{c}' \cdot \underline{X}_n \end{bmatrix} \quad (2.42)$$

Então a estimativa não tendenciosa para o vetor de média populacional  $\mu_Y = \underline{c}' \cdot \underline{\mu}$  decorrentes da combinação linear  $\underline{c}' \cdot \underline{X}$  é o vetor de média amostral dado por:

$$\bar{Y} = \underline{c}' \cdot \bar{X} \quad (2.43)$$

cuja estimativa da variância populacional

$$\sigma_Y^2 = \underline{c}' \Sigma \underline{c} \quad (2.44)$$

é dada por:

$$S_Y^2 = \underline{c}' \mathbf{S} \underline{c} \quad (2.45)$$

No qual,  $\bar{X}$  é o vetor de médias amostral e  $\mathbf{S}$  a matriz de covariância amostral, ambos estimados a partir da amostra  $X$ . Assim como o vetor de médias amostral  $\bar{X}$  estima o populacional  $\underline{\mu}$  e a matriz de covariância amostral  $\mathbf{S}$  estima a populacional  $\Sigma$  na matriz de dados originais  $X$ , da mesma forma a média amostral  $\bar{Y} = \underline{c}' \cdot \bar{X}$  estimará a populacional  $\mu_Y = \underline{c}' \cdot \underline{\mu}$  e a variância amostral  $S_Y^2 = \underline{c}' \mathbf{S} \underline{c}$  estimará a populacional  $\sigma_Y^2 = \underline{c}' \Sigma \underline{c}$  no conjunto de dados  $[Y_1 Y_2 \dots Y_n]$  resultantes da transformação  $Y = \underline{c}' \cdot \underline{X}$ .

De forma similar ao exposto acima, seja a mesma amostra aleatória  $X$  originada de uma população  $p$  variada com vetor de média  $\underline{\mu}$  e matriz de

covariância  $\Sigma$ , na qual se deseja aplicar a combinação linear do tipo  $\underline{Y} = C.\underline{X}$ , que transforma os dados de p variáveis em q variáveis, resultando na seguinte matriz de dados aleatórios,

$$Y = (C.X')' = X.C'$$

ou seja:

$$\begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1q} \\ Y_{21} & Y_{22} & \cdots & Y_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nq} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \cdot \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1q} \\ c_{21} & c_{22} & \cdots & c_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pq} \end{bmatrix} \quad (2.46)$$

Neste caso Y é a nova matriz de dados decorrentes da aplicação da combinação linear  $C\underline{X}$ , a qual apresenta parâmetros de média e variância populacional, dados respectivamente por  $\underline{\mu}_Y = C\underline{\mu}$  e  $\Sigma_Y = C\Sigma C'$ . Desse modo a estimativa para a média populacional  $\mu_Y$ , será:

$$\underline{\bar{Y}} = C.\underline{\bar{X}}. \quad (2.47)$$

e a estimativa para a matriz de covariância  $\Sigma_Y = C\Sigma C'$  será:

$$S_Y = CSC' \quad (2.48)$$

Em que,  $\underline{\bar{X}}$  é o vetor de médias e S a matriz de covariância, ambos estimados a partir da amostra original X.

#### 2.3.7.4 Variância amostral generalizada | S |

Com uma única variável, a variância da amostra é usada para descrever a variação nas mensurações dentro desta mesma variável. Quando p-variáveis são observadas em cada unidade da amostra ou do experimento, a variação é descrita pela matriz covariância  $S$ .

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \quad (2.49)$$

A matriz de covariância amostral contém p variâncias e  $\frac{p(p-1)}{2}$  covariâncias, potencialmente diferentes. Algumas vezes, no entanto, deseja-se expressar a variação por um único valor numérico. Uma escolha deste valor é o determinante de  $S$ , o qual reduz à variância amostral usual para o caso de uma única variável ( $p=1$ ). Este determinante é denominado de variância amostral generalizada.

$$\text{Variância amostral Generalizada} = |S|$$

A variância amostral generalizada se constitui numa forma de escrever toda a informação de todas as variâncias e covariâncias como um único valor numérico. Obviamente, quando  $p > 1$  é possível que algumas informações amostrais sejam perdidas no processo.

### 2.3.8 TESTES DE HIPÓTESES

Esta seção baseia-se principalmente na bibliografia de Johnson e Wichern, [1998], Chaves Neto [2006] e Marques [2013].

O teste de hipótese pode ser definido como uma regra de decisão para aceitar ou rejeitar uma hipótese, baseando-se em elementos amostrais podendo se apresentar de duas maneiras: paramétricos, quando formulamos hipóteses com respeito ao valor de um parâmetro populacional ou por aderência, quando formulamos hipóteses sobre a natureza da distribuição da população.

Segundo Chaves Neto (2006), para responder questões acerca de populações, baseadas em informações de amostras, faz-se necessário construir experimentos cujos resultados têm alguma relação com a questão de interesse. O processo para responder a estas questões é chamado teste de hipóteses. Para isso, faz-se sempre necessário, formular duas hipóteses, ou seja:

a) a hipótese nula  $H_0$ : que é a hipótese que se quer testar para verificar se a afirmação levantada sobre a questão é realmente verdadeira;

b) a hipótese alternativa  $H_1$ : caso a afirmação da hipótese nula não se confirme.

Segundo Fonseca e Andrade (1985) os testes de hipóteses também são baseados nas distribuições dos estimadores. Dessa maneira, as distribuições de probabilidade da média amostral, da variância amostral, da frequência relativa serão utilizadas para os respectivos testes sobre a média, a variância e a proporção da população.

Os dois tipos de testes de hipóteses mais difundidos são os utilizados para comparar médias e variâncias. Sendo que no presente trabalho será abordado somente o primeiro deles.

Ao efetuarmos o teste de hipóteses, podemos tomar decisões certas ou erradas ao aceitar ou rejeitar a hipótese testada. Isso pode ser facilmente representado no quadro 1:

QUADRO 1: DECISÕES NO TESTE DE HIPÓTESES

		REALIDADE	
		H <sub>0</sub> Verdadeira	H <sub>0</sub> Falsa
DECISÃO	ACEITAR H <sub>0</sub>	Decisão correta ( 1- $\alpha$ )	Erro do tipo II ( $\beta$ )
	REJEITAR H <sub>0</sub>	Erro do tipo I ( $\alpha$ )	Decisão correta ( 1 – $\beta$ )

Fonte: Spiegel (1993)

A hipótese nula H<sub>0</sub> é a hipótese de igualdade entre o novo e o produto padrão, ou seja, a designação “hipótese nula” decorre da suposição que a diferença entre eles é nula ou zero. A análise de cada situação indicará qual deve ser considerada a hipótese nula e qual a hipótese alternativa.

Geralmente a hipótese alternativa ( $H_1$ ) representa a suposição que o pesquisador quer provar, sendo a hipótese nula ( $H_0$ ) formulada com o exposto propósito a ser rejeitada. Caso rejeita-se  $H_0$ , a hipótese alternativa terá de ser aceita, tendo o pesquisador a decisão correta  $1 - \beta$ . A hipótese nula é sempre a hipótese a ser examinada. Se a aceitarmos, implicitamente estaremos rejeitando  $H_1$  e se rejeitarmos  $H_0$ , então não podemos rejeitar  $H_1$ , devendo esta ser aceita.

Nesse processo estatístico de testar hipóteses, estamos sujeito a dois tipos de erros:

Erro Tipo I ( $\alpha$ ) : A hipótese nula é verdadeira e o pesquisador a rejeita.

Erro Tipo II ( $\beta$ ) : A hipótese nula é falsa e o pesquisador a aceita.

Outro fator importante nos testes de hipóteses é que em razão da variabilidade inerente à amostragem aleatória, as estatísticas amostrais, tendem, ao invés de igualar, a se aproximar dos parâmetros da população. Isto deve ser levado em consideração na análise de uma afirmação de um parâmetro populacional utilizando dados amostrais. Assim, para aceitar ou rejeitar a hipótese nula, necessita-se de três regras básicas:

a) **A estatística do teste:** É o valor calculado a partir da amostra que será usado na tomada de decisão. Consiste em uma forma de verificar o quanto uma estatística amostral observada aproxima-se de uma população com o parâmetro alegado.

b) **O nível de significância do teste:** O tamanho da região crítica é exatamente a probabilidade  $\alpha$  de cometer o erro tipo I. O nível de significância do teste ( $\alpha$ ) é, portanto, a probabilidade com que desejamos correr o risco de cometer o erro tipo I, ou seja, em  $\alpha\%$  dos casos de rejeição de  $H_0$ , estaremos tomando decisão errada.

c) **Comparar a estatística do teste, com o valor associado ao nível de significância:** Se o valor da estatística do teste cair dentro da região crítica, rejeita-se  $H_0$ . Ao rejeitar a hipótese nula ( $H_0$ ) existe uma forte evidência de sua falsidade. Ao contrário, quando aceitamos, dizemos que não houve evidência amostral significativa no sentido de permitir a rejeição de  $H_0$ .

#### 2.3.8.1 Teste de Hipóteses para Médias

Basicamente, com relação à análise de perfis, temos dois testes de hipóteses relativos à média de uma única variável aleatória, sendo eles:

a) teste de hipóteses para comparar se as médias de duas populações são iguais;

b) teste de hipóteses para comparar se as médias de várias populações são iguais, neste caso usamos o teste de ANOVA;

### 2.3.8.2 Teste de hipóteses para comparar a média de duas populações

O  $T^2$  é apropriado para comparar a resposta média de um grupo experimental (população 1) com a resposta média de outro grupo experimental (população 2), distintas entre si. Se possível, as unidades experimentais devem ser sorteadas para cada conjunto de observações de ambas as populações, o que abrandará o efeito da variabilidade entre unidades na comparação entre tratamentos.

Analisando uma amostra aleatória de tamanho  $n_1$  de uma população 1 e uma amostra  $n_2$  de uma população 2, podemos organizar as observações das  $p$  variáveis, por:

Amostra de População  $i$ :  $\underline{X}_{i1}, \underline{X}_{i2}, \dots, \underline{X}_{in_1}$

$$\bar{\underline{X}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \underline{X}_{ij}$$

$$S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\underline{X}_{ij} - \bar{\underline{X}}_i)(\underline{X}_{ij} - \bar{\underline{X}}_i)' \quad (2.50)$$

com  $i = \{ 1, 2 \}$

Realizar a inferência a respeito da diferença de médias populacionais  $(\mu_1 - \mu_2)$  a fim de verificar se esta diferença é não significativa, ou equivalente em assegurar que não existe efeito dos tratamentos. De forma análoga, pode-se fazer tal inferência, testando a hipótese de igualdade dos vetores médias populacionais  $(H_0: \mu_1 = \mu_2)$ . Para a validade dos testes e da inferência, algumas pressuposições devem ser satisfeitas. Entre elas, destaca-se a necessidade de que sejam realizadas amostras aleatórias, de tamanho  $n_1$  e

$n_2$ , de ambas as populações (população 1 com média  $\mu_1$  e covariância  $\Sigma_1$ , e população 2 com média  $\mu_2$  e covariância  $\Sigma_2$ ).

Outra suposição é que as observações da amostra 1 são independentemente obtidas em relação àquelas da amostra 2. Ainda é necessário assumir que ambas as populações sejam normais e que a matriz de covariância amostral seja a mesma,  $\Sigma_1 = \Sigma_2 = \Sigma$ .

As matrizes de covariância  $S_1$  e  $S_2$  são estimadores de  $\Sigma_1$  e de  $\Sigma_2$ , respectivamente. Assim pode-se estimar a variância comum  $\Sigma$  da seguinte ponderação mediante os graus de liberdade:

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 - 1) + (n_2 - 1)}$$

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2} \quad (2.51)$$

Para se testar a hipótese  $H_0: \mu_1 - \mu_2 = \delta_0$ , considera-se os seguintes resultados:

$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$$

$$Cov(\bar{X}_1 - \bar{X}_2) = \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \Sigma \quad (2.52)$$

Como consequência do resultado acima, em que  $S_p$  é um estimador de

$\Sigma$ , então,  $\left( \frac{1}{n_1} + \frac{1}{n_2} \right) S_p$  é um estimador para  $Cov(\bar{X}_1 - \bar{X}_2)$ .

O teste da razão de verossimilhança para a hipótese,  $H_0: \mu_1 - \mu_2 = \delta_0$  fica definido pela distância quadrada  $T^2$  de Hotelling:

$$T^2 = [\bar{X}_1 - \bar{X}_2 - \delta_0]' \left[ \left( \frac{1}{n_1} + \frac{1}{n_2} \right) S_p \right]^{-1} [\bar{X}_1 - \bar{X}_2 - \delta_0] \quad (2.53)$$

sendo que  $H_0$  é rejeitado quando:

$$T^2 > \frac{(n_1+n_2-2)p}{(n_1+n_2-p-1)} F_{p, n_1+n_2-p-1}(\alpha) \quad (2.54)$$

Sendo  $F$  a distribuição F - Snedecor, com  $p$  e  $(n_1 + n_2 - p - 1)$  graus de liberdade ao nível de significância  $\alpha$ .

### 2.3.9 ANÁLISE DE PERFIS

Esta seção baseia-se principalmente na bibliografia de Johnson & Wichern, [1988], Rencher [2007], Chaves Neto [2006] e Tabachnick [2007].

Segundo Tabachnick (2007) em *Using multivariate statistics*, a análise do perfil é uma aplicação especial da análise da variância multivariada (MANOVA) para uma situação em que existem várias variáveis dependentes, todas elas medidas em uma mesma unidade.

Morrison (1976) define a análise de perfis como uma técnica estatística multivariada utilizada em situações em que uma mesma bateria de  $p$  testes é aplicada para  $k$  grupos populacionais com objetivo de responder a três questões principais de interesse:

- I. Os perfis de média dos  $k$  grupos populacionais são paralelos uns em relação aos outros?

Este é comumente conhecido como o teste de paralelismo e é a principal questão abordada pela análise de perfil. Ao usar a abordagem de perfil para medidas repetidas univariadas (ANOVA), o teste paralelismo é o teste de interação.

- II. Se os perfis de médias dos  $k$  grupos são realmente paralelos, eles também são coincidentes? Caso não sejam paralelos, a  $h$ -ésima variável resposta apresenta o mesmo efeito entre os grupos ou tratamentos?

Esta questão se refere a hipótese de mesmo efeito no vetor perfil de médias entre os  $k$  grupos e será resolvida aplicando-se uma técnica desenvolvida particularmente quando se aceita a hipótese de perfis paralelos.

- III. Novamente assumindo paralelismo, então os  $k$  grupos possuem a mesma resposta nas diferentes condições experimentais? Ou, caso a hipótese de perfis paralelos não seja verdadeira, o  $j$ -ésimo grupo apresenta a mesma resposta sob as diferentes condições experimentais?

Esta questão será resolvida testando-se a hipótese de igual efeito de resposta apresentada no teste de medidas repetidas, com uma pequena modificação para  $k$  tratamentos, ou seja, busca responder conjuntamente para os  $k$  grupos através do vetor perfil de media geral.

Os  $p$  testes dizem respeito as diferentes situações experimentais em que uma mesma unidade experimental é testada.

A principal pergunta na análise de perfis é se os grupos realmente possuem perfis diferentes em um conjunto de medidas. A condição essencial para aplicar a análise de perfil é que todas as  $p$  respostas obtidas para os  $p$  testes, devem estar na mesma unidade física, porém não há restrição quanto ao dimensionamento das variáveis.

Geralmente, com o objetivo de melhor visualização do problema, o pesquisador costuma construir um gráfico de perfil de média para cada tratamento. Estes gráficos têm como objetivo indicar qual é o perfil de desempenho sob as diferentes condições experimentais, para o item em estudo. No eixo das abscissas, plotam-se as componentes do vetor de testes, dado por:

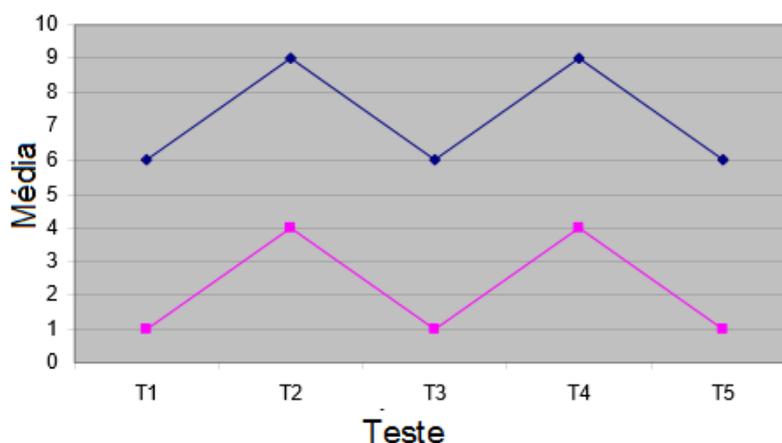
$$\underline{\beta}'_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jp}], \text{ com } j = 1, 2, \dots, p \quad (2.55)$$

e no eixo das ordenadas as componentes do vetor das variáveis respostas médias, ou vetor de perfil de médias dos experimentos, dado por:

$$\underline{\mu}'_j = [\mu_{j1}, \mu_{j2}, \dots, \mu_{jp}], \text{ com } j = 1, 2, \dots, p \quad (2.56)$$

Graficamente, a diferença entre dois pontos é chamada de segmentos na análise de perfil. O gráfico 2 apresenta um exemplo de perfil da média para dois grupos.

GRÁFICO 2: PERFIL DE MÉDIAS DE DOIS GRUPOS



Fonte: San Francisco State University - 2013

### 2.3.9.1 Análise de perfis para dois grupos populacionais

A construção do modelo da análise de perfis é feita de modo que os dados sejam obtidos a partir de uma amostra aleatória de tamanho  $n_1$ , correspondente ao grupo 1, e uma amostra aleatória de tamanho  $n_2$ , correspondente ao grupo 2. As suposições para as observações de perfil as quais irão permitir testar as três hipóteses podem ser resumidos da seguinte forma:

- As respostas devem ser descritas por um vetor aleatório  $\underline{X}'_j = [X_{j1}, X_{j2}, \dots, X_{jp}]$  com distribuição normal p-variada,  $\underline{X}'_j \sim N_p(\underline{\mu}_j, \Sigma)$ ;
- Os dois grupos em análise devem possuir vetores de médias populacionais dado por  $\underline{\mu}_j = [\mu_{j1}, \mu_{j2}, \dots, \mu_{jp}]$
- Ambas as populações devem ter em comum, mas geralmente desconhecida, a matriz de covariância  $\Sigma_1 = \Sigma_2 = \Sigma$ ;

- Também se deve assumir que as amostras para os dois grupos são independentes.

### 2.3.9.2 Teste de paralelismo $H_{01}$

Seja o vetor de contraste  $\underline{\lambda}_j$  com dimensão  $(p-1)$ , formado pelas diferenças sucessivas das  $p$  componentes do vetor de média:

$$\underline{\mu}_j' = [\mu_{j1}, \mu_{j2}, \dots, \mu_{jp}]:$$

$$\underline{\lambda}_j = \begin{bmatrix} \mu_{j1} - \mu_{j2} \\ \mu_{j2} - \mu_{j3} \\ \vdots \\ \mu_{j,p-1} - \mu_{jp} \end{bmatrix} = \begin{bmatrix} \lambda_{j1} \\ \lambda_{j2} \\ \vdots \\ \lambda_{j,p-1} \end{bmatrix} \text{ com } j = 1, 2 \quad (2.57)$$

O mesmo vetor também pode ser escrito através da combinação linear

$$\underline{\lambda}_j = C \cdot \underline{\mu}_j \quad (2.58)$$

em que  $C$  é a matriz de coeficientes ou matriz de transformação com  $q = p - 1$  vetores de coeficientes de ordem  $(p-1) \times p$ , dada por:

$$C = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix} \quad (2.59)$$

Assim, os perfis dos dois grupos serão paralelos se e somente se:

$$H_{01}: \underline{\lambda}_1 = \underline{\lambda}_2$$

$$H_{01}: C \cdot \underline{\mu}_1 = C \cdot \underline{\mu}_2$$

$$H_{01}: \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \vdots \\ \mu_{1,p-1} - \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} - \mu_{22} \\ \mu_{22} - \mu_{23} \\ \vdots \\ \mu_{2,p-1} - \mu_{2p} \end{bmatrix} \quad (2.60)$$

Contudo, para amostras independentes de tamanho  $n_1$  e  $n_2$  determinadas por dois grupos populacionais, a hipótese nula pode ser testada diretamente das observações amostrais decorrentes da combinação linear:

$$\underline{Y} = C \cdot \underline{X} \quad (2.61)$$

Tendo como estimadores da média amostral  $\bar{Y}_j = C \cdot \bar{X}_j$  e da matriz de covariância amostral  $S_{j/Y} = CS_jC'$ . Assim como a matriz amostral  $S_p$  (Spooled) estima a matriz populacional  $\Sigma$ , a matriz de covariância  $\Sigma_Y$  será estimada por:

$$S_{p/Y} = C \cdot S_p \cdot C' \quad (2.62)$$

sendo que o índice Y refere-se a combinação linear  $\underline{Y} = C \cdot \underline{X}$ .

A estatística para testar a hipótese de paralelismo  $C \cdot \underline{\mu}_1 = C \cdot \underline{\mu}_2$  é o teste  $T^2$  de Hotelling utilizada para comparar dois vetores de média. Assim, rejeita-se  $H_{01}: C \cdot \underline{\mu}_1 = C \cdot \underline{\mu}_2$  em um nível  $\alpha$  se:

$$T^2 = (\bar{X}_1 - \bar{X}_2)' C' \left[ \left( \frac{1}{n_1} + \frac{1}{n_2} \right) CS_p C' \right]^{-1} C (\bar{X}_1 - \bar{X}_2) > c^2 \quad (2.63)$$

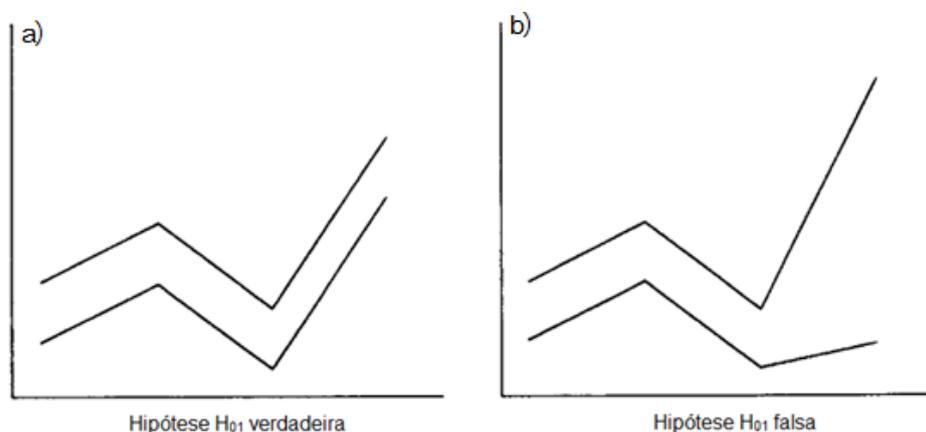
Com

$$c^2 = \frac{(n_1+n_2-2)(p-1)}{n_1+n_2-p} \cdot F_{p-1, n_1+n_2-p}(\alpha) \quad (2.64)$$

A figura 7 apresenta um par de gráficos para auxiliar na visualização de perfis paralelos. O primeiro gráfico apresenta a hipótese  $H_{01}$  aceita, ou seja

$T^2 < c^2$ , logo os perfis são paralelos. No segundo gráfico, a hipótese  $H_{01}$  foi rejeitada, o que implica que os perfis não são paralelos.

FIGURA 7: EXEMPLOS DE GRÁFICOS DE PERFIS



Fonte: RENCHER (2002)

### 2.3.9.3 Teste de perfis coincidentes $H_{02}$

Quando os perfis são paralelos o primeiro está acima do segundo ( $\mu_{1h} > \mu_{2h}$  para todo  $h$ ) ou vice-versa. Sob esta condição, os perfis serão coincidentes se e somente se,  $\mu_{11} + \mu_{12} + \dots + \mu_{1p} = \underline{1}'\underline{\mu}_1$  e  $\mu_{21} + \mu_{22} + \dots + \mu_{2p} = \underline{1}'\underline{\mu}_2$  forem iguais. Assim, a hipótese nula do estágio 2 pode ser escrita de forma equivalente a:

$$H_{02}: \underline{1}'\underline{\mu}_1 = \underline{1}'\underline{\mu}_2 \quad (2.65)$$

Induzido pela combinação linear  $\underline{1}'\underline{\mu}_j$ , deve-se fazer também com os dados originais. Para uma observação em particular  $X_{ij}$ , a transformação é  $R_{ij} = \underline{1}'X_{ij}$  e para as matrizes de dados  $X_1$  e  $X_2$ , a transformação será:

$$\underline{R}_j = \underline{1}' \underline{X}'_j \quad \text{com } j=1,2 \quad (2.66)$$

Pode-se agrupar os dados decorrentes desta transformação em um vetor na seguinte forma  $\underline{R}'_j = [R_{1j}, R_{2j}, \dots, R_{ij}, \dots, R_{nj}]$

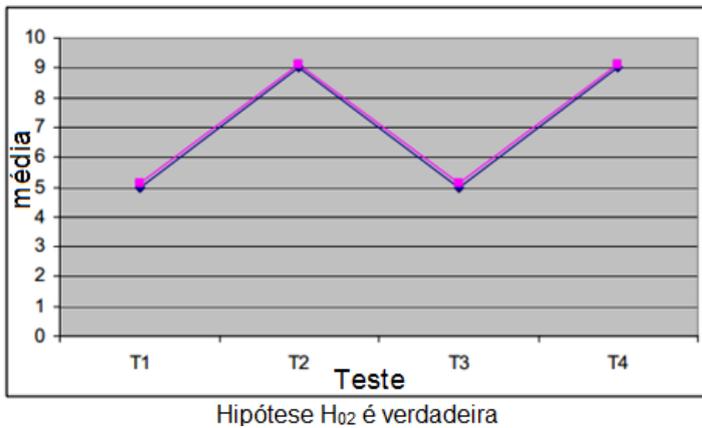
Nesse caso, a média estimada será  $\bar{R}_j = \underline{1}' \bar{X}_j$ , e a variância estimada será  $S_{j/R}^2 = \underline{1}' S_j \underline{1}$ . Quando as matrizes de covariâncias populacionais são iguais, a estimativa de  $\Sigma$  é a matriz de covariância amostral  $S_p$ . Sendo assim, as variâncias das duas populações na transformação  $R_{ij}$  deverão ser também iguais, ou seja,  $\underline{1}' \Sigma_1 \underline{1} = \underline{1}' \Sigma_2 \underline{1} = \underline{1}' \Sigma \underline{1} = \sigma_R^2$ . Com isso, a estimativa para a variância comum  $\sigma_R^2$  será a variância amostral  $S_{p/R}^2 = \underline{1}' S_p \underline{1}$ . Para verificar se os dois grupos têm perfis coincidentes, equivale ao teste t de Student utilizado para testar a igualdade de média de duas populações:

$$t = \frac{\bar{R}_1 - \bar{R}_2}{\sqrt{S_{p/R}^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (2.67)$$

Para fazer a verificação da hipótese de perfis coincidentes, rejeita-se  $H_{02}: \underline{1}' \mu_1 = \underline{1}' \mu_2$  em um nível  $\alpha$  se:

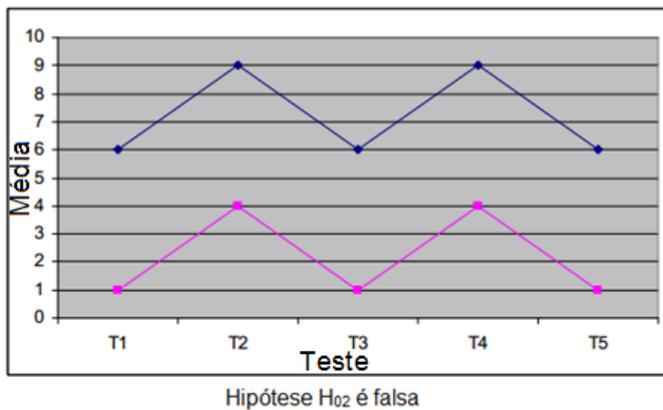
$$\begin{aligned} T^2 &= \underline{1}' (\bar{X}_1 - \bar{X}_2) \left[ \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \underline{1}' S_p \underline{1} \right]^{-1} \underline{1}' (\bar{X}_1 - \bar{X}_2) \\ &= \left( \frac{\underline{1}' (\bar{X}_1 - \bar{X}_2)}{\sqrt{\left( \frac{1}{n_1} + \frac{1}{n_2} \right) \underline{1}' S_p \underline{1}}} \right)^2 > t_{n_1+n_2-2}^2 \left( \frac{\alpha}{2} \right) = F_{1, n_1+n_2-2}(\alpha) \end{aligned} \quad (2.68)$$

Se a hipótese  $H_{01}$  é verdadeiro, a hipótese  $H_{02}$  pode ser representado como no gráfico 3.

GRÁFICO 3: HIPÓTESES  $H_{01}$  E  $H_{02}$  VERDADEIRAS

Fonte: San Francisco State University - 2013

Se a hipótese  $H_{02}$  é falsa, então os dois perfis diferem por uma constante (visto que  $H_{01}$  é verdadeiro), como podemos ver no gráfico 9.

GRÁFICO 4: HIPÓTESES  $H_{01}$  VERDADEIRA E HIPÓTESE  $H_{02}$  FALSA

Fonte: San Francisco State University - 2013

#### 2.3.9.4 Teste de hipótese de igual efeito de resposta $H_{03}$

Por fim, deve-se testar a hipótese  $H_{03}$ , a qual se refere a questão de igual resposta média, sob a condição de que a hipótese de perfis paralelos foi

aceita. Quando os dois perfis são paralelos, pode-se verificar que o perfil do vetor de médias geral, também será paralelo aos demais, então a hipótese de igual resposta média pode ser formulada utilizando o vetor de média geral  $\mu$ . Se a hipótese de paralelismo for rejeitada, o teste deve ser conduzido separadamente para cada um dos grupos.

Nesta situação, o vetor de contraste  $\lambda$ , pode ser obtido através da combinação linear  $\underline{\lambda} = C\underline{\mu}$ . A hipótese nula  $H_{03}$ , de que as  $p$  variáveis respostas são iguais, sob a condição de paralelismo aceita é:

$$\begin{aligned} H_{03}: \underline{\lambda} &= \underline{0} \\ H_{03}: C \cdot \underline{\mu} &= \underline{0} \end{aligned} \quad (2.69)$$

Para testar a hipótese de igual resposta média deve-se utilizar o vetor de média geral  $\mu$ , sendo este estimado por:

$$\bar{X} = \frac{\sum_{j=1}^{n_1} X_{1j} + \sum_{j=1}^{n_2} X_{2j}}{n_1 + n_2} = \frac{n_1}{(n_1 + n_2)} \bar{X}_1 + \frac{n_2}{(n_1 + n_2)} \bar{X}_2 \quad (2.70)$$

Na realização do teste  $H_{03}$ , deve-se também efetuar a transformação nos dados sugeridos pela combinação linear  $\underline{\lambda} = C \cdot \underline{\mu}$ . Aplicando-se a transformação linear  $\underline{Y} = C \cdot \underline{X}$  para cada uma das observações, obtém-se a matriz de dados geral:

$$Y = X \cdot C' \quad (2.71)$$

Para testar a hipótese de igual efeito de resposta, tem-se interesse, principalmente, no vetor de média geral e na matriz de covariância. A média estimada é  $\bar{Y} = C \cdot \bar{X}$  e a matriz de covariância é estimada por  $S_{p/Y} = C \cdot S_P C'$ .

Para fazer a verificação da hipótese de igual resposta média, rejeita-se  $H_{03}: C\mu = 0$  em um nível  $\alpha$  se:

$$(n_1 + n_2)\bar{\underline{X}}' C' [CS_P C']^{-1} C\bar{\underline{X}} > F_{p-1, n_1+n_2-p}(\alpha) \quad (2.72)$$

## 2.4 ANÁLISE DE AGRUPAMENTO

A análise de agrupamento ou “*cluster analysis*” é uma variedade de técnicas e algoritmos, cujo objetivo é agregar observações com base nas características que eles possuem, todavia, não existem impedimentos para realizar o agrupamento de variáveis semelhantes segundo as realizações obtidas pelos objetos amostrados.

Segundo Bussab (1990), o problema da análise de agrupamento pretende, dada uma amostra de  $n$  objetos (ou indivíduos), cada um deles medidos segundos  $p$  variáveis, procurar um esquema de classificação que agrupe os objetos em  $g$  grupos, exigindo-se daí conceitos científicos mais sofisticados de semelhança. Devem ser determinados também o número e as características desses grupos.

A análise de agrupamento adota como princípio que cada observação de uma amostra multivariada pode ser considerada como um ponto em um espaço euclidiano multidimensional.

Os processos de classificação têm por objetivo agrupar esses pontos de modo que evidenciem aspectos marcantes da amostra. Assim, dado um conjunto de observações conhecidas somente por uma listagem de suas

características, objetiva-se encontrar a melhor maneira de descrever seus padrões de similaridade mútuos.

Entretanto, qual o critério a ser utilizado para se decidir até que ponto dois objetos do conjunto de dados podem ser considerados como semelhantes ou não?

Para Mingoti (2005) “responder essa questão é necessário considerar medidas que descrevam a similaridade entre elementos amostrais de acordo com as características que neles foram medidas. (...) Assim, podemos calcular as distâncias entre os vetores de observações dos elementos amostrais e agrupar aqueles de menor distância”.

Segundo Aaker, Kumar e Day (2001), a premissa mais importante da análise de agrupamento é a de que a medida de similaridade ou dissimilaridade na qual o processo de agrupamento se baseia é uma medida válida de similaridade ou dissimilaridade entre os indivíduos. A segunda premissa mais importante é a de que existe uma justificativa teórica para estruturar os indivíduos em grupos. Como em outras técnicas multivariadas, também há teoria e lógica guiando e dando base à análise de agrupamento.

A análise de agrupamento possui um fator interessante, principalmente sob o aspecto descritivo, pois seu resultado final, nos métodos, é um gráfico de esquemas hierárquicos denominado dendograma, quando se utiliza o método hierárquico.

Este representa uma síntese dos resultados, o que ocasiona uma pequena perda de informações. Ainda assim, o resumo da informação torna-se

mais fácil de ser manipulado e armazenado, sendo importante para a comparação, classificação e discussão do material estudado.

Para Aaker, Kumar e Day (2001), a análise de agrupamento pode ser compreendida em cinco etapas principais:

- I. A seleção de indivíduos ou de uma amostra de indivíduos a serem agrupados;
- II. A definição de um conjunto de variáveis a partir das quais serão obtidas informações necessárias ao agrupamento dos indivíduos;
- III. A definição de uma medida de semelhança ou distância entre os indivíduos;
- IV. A escolha de um algoritmo de partição/classificação;
- V. A validação dos resultados encontrados.

Após as etapas I e II citadas acima, uma das questões vitais das técnicas de análise de agrupamento é a definição do coeficiente de similaridade ou dissimilaridade.

#### 2.4.1 Medida de similaridade e dissimilaridade

A escolha de um critério que meça a distância entre dois objetos, ou que quantifique o quanto eles são parecidos é fundamental na utilização das técnicas de análise de agrupamento. Tecnicamente, é possível dividir esta medida em duas categorias: medidas de similaridade e de dissimilaridade.

Na medida de similaridade, quanto maiores os valores observados, mais parecidos serão os objetos. Já para a medida de dissimilaridade, quanto

maiores os valores observados, menos parecidos serão os objetos. O coeficiente de correlação é um exemplo de medida de similaridade, enquanto que a distância euclidiana é um exemplo de dissimilaridade.

Para Bussab et al. (1990), de um modo geral, as medidas de similaridade e de dissimilaridade são interrelacionadas e, facilmente, transformáveis entre si.

#### 2.4.2 Distância Euclidiana

Considere o vetor  $\mathbf{x}$  de coordenadas reais  $(x_1, x_2, \dots, x_p)$  como descritor dos objetos que investigarão os assemelhamentos. A medida mais conhecida para indicar a proximidade entre os objetos A e B é a distância euclidiana  $d(A, B)$ :

$$d(A, B) = \sqrt{\sum_{i=1}^p (x_i(A) - x_i(B))^2} \quad (2.73)$$

ou em linguagem matricial:

$$d(A, B) = [(\mathbf{x}(A) - \mathbf{x}(B))'(\mathbf{x}(A) - \mathbf{x}(B))]^{1/2} \quad (2.74)$$

#### 2.4.3 Distância Euclidiana padronizada

Quando as variáveis são quantitativas, a distância Euclidiana comumente soma distâncias não comparáveis, como cm, kg, anos, milhões,

etc., muito embora, a mudança de uma das unidades possa alterar completamente o significado e o valor do coeficiente.

Essa é uma das razões da padronização das variáveis dos elementos  $x_1, x_2, \dots, x_p$  do vetor  $\mathbf{x}$ . Assim o uso da transformação

$$z_i = \frac{x_i - \bar{x}_i}{s_i} \quad (2.75)$$

em que  $\bar{x}_i$  e  $s_i$ , indicam respectivamente a média e o desvio padrão de  $i$ -ésima coordenada, é um dos modos para evitar essa inconveniência. Feita a transformação, a distância euclidiana passa a ser:

$$d(A,B) = \left[ \sum_{i=1}^p (z_i(A) - z_i(B))^2 \right]^{1/2} \quad (2.76)$$

que é a soma dos desvios padronizados. Com algumas manipulações algébricas podemos verificar que a expressão acima pode ser escrita da seguinte forma, em notação vetorial:

$$d(A, B) = [(\mathbf{x}(A) - \mathbf{x}(B)) \mathbf{D}^{-1}(\mathbf{x}(A) - \mathbf{x}(B))]^{1/2} \quad (2.77)$$

sendo  $\mathbf{D}$  uma matriz diagonal, tendo como  $i$ -ésimo componente a variância  $s_i^2$ , isto é,

$$\mathbf{D} = \text{diag} (s_1^2, s_2^2, \dots, s_p^2) \quad (2.78)$$

Após a escolha das variáveis que serão usadas como critérios de semelhança, uma das questões vitais das técnicas de análise de agrupamento, é a definição do coeficiente de similaridade ou dissimilaridade.

#### 2.4.4 Formação de agrupamentos

A formação de agrupamentos fundamenta-se em duas idéias básicas que são: coesão interna dos objetos e isolamento externo entre os grupos.

As técnicas de agrupar podem ser classificadas em categorias, na qual as técnicas hierárquicas são as mais utilizadas na literatura. Nessas técnicas hierárquicas, os objetos são classificados em grupos, em diferentes etapas, de modo hierárquico, produzindo uma árvore de classificação.

#### 2.4.5 Número de grupos

Determinar o número de grupos para uma base de dados é uma das tarefas mais difíceis no processamento de agrupamento.

O algoritmo euclidiano de agrupamento e o método de formação de grupos produzem conjuntos que constituem uma proposição sobre a organização básica e desconhecida dos dados. Entretanto todo esse procedimento esbarra em uma dificuldade comum que é a determinação do número ideal de grupos.

Para auxiliar na decisão do número de grupos, pode-se adotar, segundo Bussab (1990), a técnica hierárquica que consiste em examinar o *dendograma* em busca de grandes alterações nos níveis de similaridade para as sucessivas fusões.

## 2.5 ANÁLISE FATORIAL

Segundo Hair et al (2005) a Análise Fatorial é um nome genérico dado a uma classe de métodos estatísticos multivariados cujo propósito principal é definir a estrutura subjacente em uma matriz de dados. Assim, a Análise Fatorial é uma técnica estatística que busca, avaliando um conjunto de variáveis, a identificação das interrelações existentes em um conjunto de fenômenos, desvendando estruturas existentes, impossíveis de observá-las por métodos diretos.

Para Chaves Neto (2006), o objetivo da Análise Fatorial é descrever, se possível, a estrutura de covariância dos relacionamentos entre muitas variáveis em termos de poucas variáveis fundamentais, mas não observáveis (latentes), aleatórias chamadas FATORES.

Com o uso da análise multivariada, pode-se identificar as dimensões separadas da estrutura, dimensões estas chamadas de fatores, e assim, determinar o grau em que cada variável é explicada em cada dimensão. Conseguindo realizar esses dois itens, dimensão e grau de explicação, o uso da Análise Fatorial torna-se uma ferramenta muito importante de redução de dados, pois podemos escrever diversas variáveis em um número mínimo de fatores.

Sendo assim, um problema com inúmeras variáveis, na Análise Fatorial, é explicado a partir de dimensões “ocultas” chamadas de fatores.

A ideia principal da Análise Fatorial é que há uma correlação entre as variáveis, porque essas variáveis estão relacionadas pelo mesmo fator.

### 2.5.1 O modelo matemático da Análise Fatorial

Visto que a Análise Fatorial avalia uma correlação que existe entre um grande número de variáveis identificando um possível agrupamento dessas variáveis, reduzindo consideravelmente a quantidade de variáveis e explicando o significado desse agrupamento.

Dizemos que a Análise Fatorial avalia a possibilidade de agrupar  $i$  variáveis ( $X_1, X_2, X_3, \dots, X_i$ ) em um número  $j$  consideravelmente menor de fatores ( $F_1, F_2, F_3, \dots, F_j$ ) com  $j < i$ , se possível.

De acordo com Hair *et al.* (2005), no modelo de análise fatorial, cada uma das variáveis pode ser definida como uma combinação linear dos fatores comuns que irão explicar a parcela da variância de cada variável, mais um desvio que resume a parcela da variância total não explicada por estes fatores.

A parcela explicada pelos fatores comuns recebe o nome de comunalidade, e a parcela não explicada é chamada de especificidade. As comunalidades podem variar de 0 a 1, sendo que valores próximos de 0 indicam que os fatores comuns não explicam a variância e valores próximos de 1 indicam que todas as variâncias são explicadas pelos fatores comuns.

Ainda, Hair *et al.* (2005) estabelecem os seguintes passos para a realização de uma análise fatorial: formulação do problema; construção da matriz de correlação; determinação do método de análise fatorial; determinação

do número de fatores; rotação dos fatores; interpretação dos fatores; cálculo das cargas fatoriais ou escolha de variáveis substitutas e, determinação do ajuste do modelo.

Segundo Corrar (2007), o modelo matemático de análise fatorial pode ser representado por:

$$X_i = \alpha_i F + e_i \quad (2.79)$$

no qual :

$X_i$  = variável “ i “ analisada;

$\alpha_i$  = constante;

F = fator;

$e_i$  = erro.

Pode-se concluir que o fator representa a parcela da variação total dos dados que pode ser explicado de forma conjunta para todas as variáveis que o compõem.

Para Corrar (2007), as variações em uma variável podem ser explicadas a partir de um conjunto de fatores.

$$X_i = \alpha_{i1}F_1 + \alpha_{i2}F_2 + \alpha_{i3}F_3 + \dots + \alpha_{ij}F_j + e_i \quad (2.80)$$

em que:

$X_i$  - são as variáveis padronizadas;

$\alpha_i$  - são as cargas fatoriais;

$F_j$  - São os fatores comuns não relacionados entre si;

$e_i$  - é um erro que representa a parcela de variação não explicada da variável  $i$ .

Assim, as cargas fatoriais, são índices que medem o grau de correlação entre os fatores e seus quadrados. Esses fatores podem ser estimados por uma combinação linear das variáveis originais. Tem-se assim:

$$F_j = \omega_{j1}X_1 + \omega_{j2}X_2 + \omega_{j3}X_3 + \dots + \omega_{ji}X_i$$

$$F_j = \sum_{i=1}^i \omega_{ji}X_i \quad ( 2.81 )$$

Em que:

$F_j$  - são os fatores comuns não relacionados;

$\omega_{ji}$  - são os coeficientes dos escores fatoriais;

$X_i$  - são as variáveis originais envolvidas no estudo.

O escore fatorial é um número resultante do produto dos coeficientes  $\omega_{ji}$  pelas variáveis originais. No caso da existência de mais de um fator, os escores representam as coordenadas da variável em relação aos eixos, que são os fatores.

Com isso, podemos concluir que ao conhecermos o grau de relacionamento entre as variáveis, é possível determinar os dados realmente importantes a partir de apenas algumas variáveis.

Resumindo, a Análise Fatorial possui, basicamente, quatro etapas para a sua elaboração que são:

- i. Cálculo da matriz de correlação das variáveis em estudo para a verificação do grau de associação entre as variáveis, duas a duas. Nesta etapa, é verificada a adequação da aplicação da Análise Fatorial;
- ii. Extração dos fatores mais significativos que representarão os dados, através do método mais adequado. Aqui se sabe o quão bem o modelo representa os dados;
- iii. Aplicação de rotação, nos fatores, para facilitar o entendimento dos mesmos;
- iv. Geração dos escores fatoriais para utilização em outras análises.

### 3 MATERIAL E MÉTODOS

#### 3.1 LEVANTAMENTO DOS DADOS

O estudo das particularidades do processo de produção de recursos não renováveis é antigo, sendo analisado por diversos autores como Postali, Serra, Lima entre outros. A peculiaridade fundamental deste tipo de atividade é a limitação física de seu estoque na natureza, implicando na indisponibilidade do recurso para as gerações futuras caso ele seja extraído no presente.

Com o crescimento da arrecadação de royalties, aliado ao problema da crescente escassez, inerente aos recursos não-renováveis, tem feito surgir uma série de estudos acadêmicos, projetos de alteração da legislação e discussões na sociedade.

O presente trabalho pretende contribuir para esse debate ao propor elementos para a resposta da seguinte questão:

- A aplicação dos recursos recebidos pelos royalties do petróleo tem ocasionado alguma diferença nos índices de desenvolvimento dos Municípios?

Assim, o presente trabalho, tem como principal objetivo apresentar um método de leitura qualitativa dos royalties do petróleo. Esse método poderá servir como ferramenta de avaliação, baseada na verificação da eficiência e da eficácia da aplicação da arrecadação dos royalties de alguns Municípios, por meio do desenvolvimento humano de sua população, quando comparado seu resultado àqueles apurados em outras regiões.

Para o presente estudo foi realizada a coleta dos dados pertinentes a cada indivíduo da amostra a ser analisada. Esta amostra foi composta de 63 Municípios brasileiros, sendo que destes, 37 recebem pagamento das participações governamentais (royalties + participações especiais) e 26 Municípios que não recebem os pagamento das participações governamentais.

O quadro 2 apresenta o número de Municípios amostrados por Estado, bem como o total de Municípios que recebem royalties nesse estado. Como critério de escolha para os Municípios que não receberam royalties, optou-se por escolher Municípios com populações relativamente próximas as dos Municípios que recebem royalties. Podemos observar no quadro 3 os Municípios organizados em ordem alfabética com suas respectivas populações de acordo com o senso de 2010.

#### QUADRO 2 : AMOSTRA DE MUNICÍPIOS POR ESTADO.

ESTADOS	NÚMERO DE MUNICÍPIOS QUE RECEBEM ROYALTIES	AMOSTRA: RECEBEM ROYALTIES	AMOSTRA: NÃO RECEBEM ROYALTIES
ALAGOAS	52	1	1
AMAPÁ	3	-	-
AMAZONAS	20	-	-
BAHIA	269	1	1
CEARÁ	138	-	-
ESPÍRITO SANTO	78	6	6
MINAS GERAIS	5	-	-
PARÁ	17	-	-
PARAÍBA	6	-	-
PARANÁ	38	3	3
PERNAMBUCO	16	-	-
RIO DE JANEIRO	87	16	5
RIO G. DO NORTE	96	-	-
RIO G. DO SUL	24	-	-
SANTA CATARINA	11	6	6
SÃO PAULO	103	4	4
SERGIPE	75	-	-
Total		37	26

Fonte: Organizado pelo autor com base nos dados do Inforoyalties 2013.

Os dados coletados para essa amostra, são referentes as seguintes variáveis:

- Volume da arrecadação dos royalties no período entre 2000 e 2010;
- População demográfica em 2010;
- Incidência de pobreza em 2010;
- Índice de desenvolvimento humano municipal (IDHM) – 2000 e 2010
- Índice Firjan de desenvolvimento municipal (IFDM): Emprego & renda, Educação e Saúde – 2011
- Índice de desenvolvimento da educação básica para as séries iniciais (IDEB) - 2011

QUADRO 3: MUNICÍPIOS UTILIZADOS NA COMPARAÇÃO E SUAS RESPECTIVAS POPULAÇÕES EM 2010.

continua

Município	População	Município	População
Anchieta (ES)	23.902	Silva Jardim (RJ)	21.349
Angra dos Reis (RJ)	169.511	Itapoá (SC)	14.763
Aracruz (ES)	81.832	Araquari (SC)	24.810
Armação dos Búzios (RJ)	27.560	Garuva (SC)	14.761
Balneário Barra do Sul (SC)	8.430	Joinville (SC)	515.288
Bertioga (SP)	47.645	São Francisco do Sul (SC)	42.520
Cabo Frio (RJ)	186.227	Castelo (ES)	34.747
Cachoeiras de Macacu (RJ)	54.273	Iconha (ES)	12.523
Cachoeiro de Itapemirim (ES)	189.889	Colatina (ES)	111.788
Cairu (BA)	15.374	Afonso Cláudio (ES)	31.091
Campos dos Goytacazes (RJ)	463.731	Viana (ES)	65.001
Carapebus (RJ)	13.359	Cariacica (ES)	348.738
Casimiro de Abreu (RJ)	35.347	Sapucaia (RJ)	17.525
Guapimirim (RJ)	51.483	Areal (RJ)	11.423
Guararema (SP)	25.844	Comendador Levy Gasparian (RJ)	8.180
Guaratuba (PR)	32.095	Paraíba do Sul (RJ)	41.084

conclusão

Ilhabela (SP)	28.196	Três Rios (RJ)	77.432
Itapemirim (ES)	30.988	Catanduvas (SC)	9.555
Linhares (ES)	141.306	Lauro Muller (SC)	14.367
Macaé (RJ)	206.728	Joaçaba (SC)	27.020
Mangaratiba (RJ)	36.456	Papanduva (SC)	17.928
Marica (RJ)	127.461	Blumenau (SC)	309.011
Matinhos (PR)	29.428	São Miguel do Oeste (SC)	36.306
Paranaguá (PR)	140.469	Palmeira (PR)	32.123
Parati (RJ)	37.533	Imbituva (PR)	28.455
Pilar (AL)	33.305	Guarapuava (PR)	167.328
Presidente Kennedy (ES)	10.314	Itararé (SP)	47.934
Quissamã (RJ)	20.242	Laranjal Paulista (SP)	25.251
Rio das Ostras (RJ)	105.676	Iperó (SP)	28.300
São João da Barra (RJ)	32.747	Campo Limpo Paulista (SP)	74.074
São Sebastião (SP)	73.942	Ibitiara (BA)	15.508
São Sebastião (AL)	32.010		

Fonte: IBGE - 2013

Para comparação de desempenho de cada município, baseado em seus índices de desenvolvimento, foi criada uma matriz de dados na qual cada coluna refere-se aos seguintes índices:

**INCIDÊNCIA DE POBREZA:** É um índice amplamente empregado para avaliar a dimensão da pobreza nos Municípios brasileiros. É obtido como uma contagem do número de pobres, fazendo um percentual em relação a população do município. Define-se como pobre aquele cidadão cujos rendimentos ficam aquém do estipulado na linha da pobreza fixada. A medida mais comum da pobreza é o índice de incidência da pobreza, o qual constitui a proporção de pobres na população total.

**Índice de desenvolvimento humano municipal:** O IDHM é, assim como o IDH, um índice que mede o desenvolvimento humano de uma unidade geográfica. Como o IDH foi concebido para ser aplicado no nível de países e grandes regiões, sua aplicação no nível municipal passou por algumas

adaptações metodológicas e conceituais<sup>7</sup>. Para o presente trabalho foi utilizado os dados do IDHM de 2000 e de 2010.

Índice FIRJAN de Desenvolvimento Municipal: O IFDM, é um estudo anual do Sistema FIRJAN que acompanha o desenvolvimento de todos os mais de 5 mil Municípios brasileiros em três áreas: Emprego & Renda, Educação e Saúde. Ele é feito, exclusivamente, com base em estatísticas públicas oficiais, disponibilizadas pelos ministérios do Trabalho, Educação e Saúde<sup>8</sup>. Para o referido estudo foram utilizados os dados do IFDM de 2011.

Índice de Desenvolvimento da Educação Básica: O (IDEB) foi criado pelo Inep (Instituto Nacional de Estudos e de Pesquisas Educacionais Anísio Teixeira) em 2007, como parte do Plano de Desenvolvimento da Educação (PDE). Ele é calculado com base na taxa de rendimento escolar (aprovação e evasão) e no desempenho dos alunos no SAEB (Sistema Nacional de Avaliação da Educação Básica) e na Prova Brasil. Ou seja, quanto maior for a nota da instituição no teste e quanto menos repetências e desistências ela registrar, melhor será a sua classificação, numa escala de zero a dez.

Para evidenciar uma aplicação prática das técnicas estatísticas da Análise de Perfis, Análise Fatorial e Análise de Cluster, as mesmas foram empregadas a um conjunto de dados obtidos através de pesquisas junto ao IBGE, ANP e Sistema FIRJAN, a qual tornou possível utilizar estas metodologias. Para o teste de análise de perfis, foi implementado um código na versão do software Matlab, assim, após informada a matriz de dados e um

---

<sup>7</sup> Associação nacional de pós-graduação e pesquisa em administração - ANPAD

<sup>8</sup> <http://www.firjan.org.br>

vetor indicando o tamanho das duas amostras, realiza-se a análise de perfis através das funções programadas.

### 3.2 TRATAMENTO E ANÁLISE DOS DADOS

Para a condução do presente trabalho foram adotados os métodos descritos na revisão de literatura. A análise dos dados foi efetuada utilizando os softwares Excel, Statgraphics Centurion XV versão 15.2.06, Minitab 16.1.1 e Matlab R2012a. O software Excel foi utilizado para digitação e organização dos dados, que em seguida foram exportados para o software Statgraphics e Matlab, através dos quais os dados foram analisados.

O banco de dados foi dividido em dois grupos matriciais, uma matriz denominada M1 com os dados dos Municípios que receberam royalties de petróleo e matriz denominada M2 com os dados dos Municípios que não receberam royalties de petróleo. Ambas matrizes foram utilizadas para a comparação dos dados três testes.

Primeiramente realizou-se o teste de Análise de Perfis, em que foram levantadas as seguintes hipóteses:

I. Hipótese  $H_{01}$ :

Hipótese de que os perfis médios dos Municípios apresentam tendência similar de variação de seus índices, ou seja, de que os perfis de índices são paralelos.

H<sub>01</sub>: “Os perfis de índices dos Municípios são paralelos”

## II. Hipótese H<sub>02</sub>

Hipótese de que os índices dos Municípios são iguais para todos os grupos. Essa hipótese deve ser conduzida de duas formas, em que a forma a ser conduzida depende do resultado da hipótese H<sub>01</sub>. Se a hipótese de perfis paralelos for aceita, o teste é conduzido testando-se a hipótese de perfis coincidentes entre os grupos, ou seja, hipótese de que os perfis coincidentes. Se a hipótese de perfis paralelos for rejeitada, testa-se a igualdade de índices entre os grupos individualmente para cada uma das posições.

H<sub>02</sub>: “Os índices dos Municípios são iguais para todos os grupos”

## III. Hipótese H<sub>03</sub>

Hipótese de que os Municípios apresentam os mesmos índices dentro de seus grupos, ou seja, todos os Municípios que recebem royalties de petróleo possui os mesmos índices? Ou todos os Municípios que não recebem royalties possuem os mesmos índices? Se a hipótese de perfis paralelos for aceita, o teste é executado como o vetor perfil de média geral, em caso contrário, deve-se fazer individualmente.

H<sub>03</sub>: “Os Municípios apresentam os mesmos índices dentro de seus grupos”

Com objetivo de explorar os dados, optou-se por comparar os resultados obtidos pela Análise de Perfis com os resultados obtidos pelos métodos de agrupamento. Foram utilizados os seguintes algoritmos de agrupamento, por serem os mais usados na prática e pela facilidade de serem encontrados nos mais diversos programas computacionais, tais como:

- ✓ Método da Ligação Simples ou do Vizinho mais Próximo (Single Linkage)
- ✓ Método da Ligação Completa ou do Vizinho mais Longe (complete linkage)
- ✓ Método do Centroide
- ✓ Método da Mediana

Por fim, utilizou-se de técnicas de Análise Fatorial que teve como objetivo resumir um grande número de variáveis originais, num pequeno número de variáveis para um melhor entendimento dos dados.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 RESULTADOS

Nesta seção, são apresentados os principais resultados obtidos do problema proposto, utilizando a análise de perfis, resolvidos através do software Matlab, a análise de Cluster, utilizando o software Minitab e a análise fatorial, utilizando o software Statgraphics. Os resultados completos de toda a sequência dos testes, bem como a matriz de dados originais, encontram-se no apêndice.

#### 4.1.1 Análise de Perfis

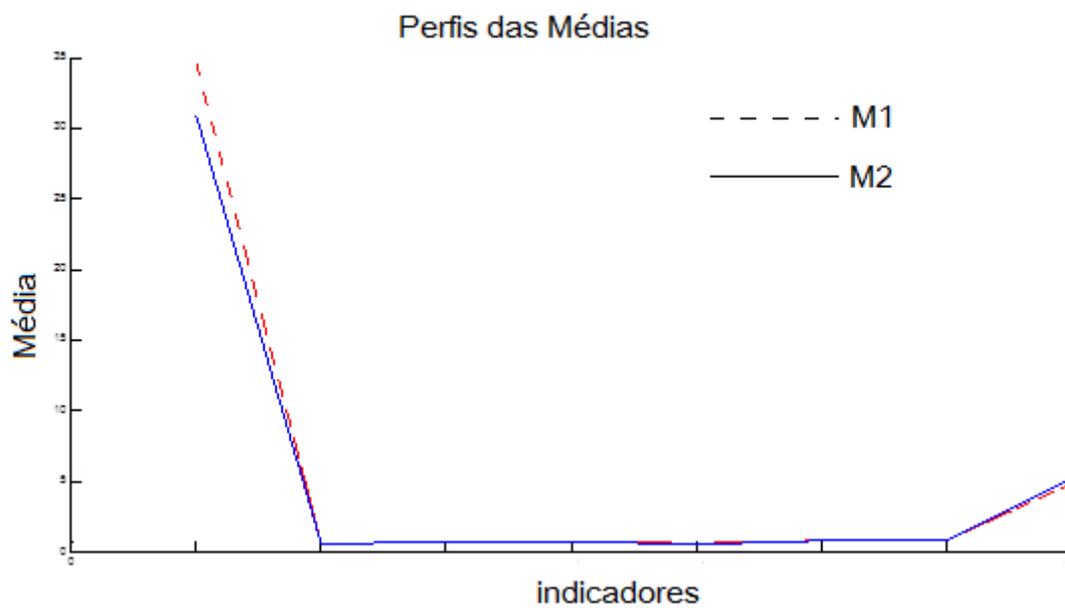
Conforme foi discutido em capítulos anteriores sobre o uso da análise de perfis, vide página 61 – 2.3.9 “ Análise de Perfis”, expomos a necessidade do uso desta ferramenta, bem como a base de sua teoria . Vamos aplicar esta técnica para verificar se o grupo M1 dos Municípios que recebem royalties diferem do grupo M2 composto pelos Municípios que não recebem royalties, quando comparados os seus indicadores sociais.

Para o teste de análise de perfis, inicialmente foram calculadas as médias das duas amostras para então construir o gráfico com os perfis de média de cada um dos modelos para M1 e M2, o qual permite-nos uma análise gráfica preliminar do problema.

Pelo gráfico do perfil das médias, representado no gráfico 5, é possível verificar a existência de paralelismo entre os dois grupos, ou seja, nota-se que os

grupos apresentam desempenhos semelhantes com relação às variáveis observadas.

GRÁFICO 5: PERFIS DE MÉDIA DOS GRUPOS 1 E 2



Fonte: O autor

Na sequência, foram realizados os três testes de hipóteses. O nível de significância adotado para os intervalos de confiança é de  $\alpha=0,05$  para todas as situações propostas.

Como resultado do primeiro teste de hipóteses, hipótese  $H_{01}$  perfis paralelos, tem-se:

$$T^2 = 7.4613$$

$$C^2 = 16.9351$$

Como  $T^2$  (no Matlab representado por T2)  $<$   $c^2$  (no Matlab representado por C2), aceita-se a hipótese, logo:

hipótese H01 aceita: Perfis paralelos

Assim, pelo resultado obtido, aceitamos a hipótese de paralelismo entre os dois grupos. ( $\alpha = 5\%$ )

Pelo fato de ter sido aceita a hipótese de paralelismo ( $H_{01}$ ) entre os dois grupos, testa-se a hipótese  $H_{02}$ , hipótese de que os perfis são coincidentes, ou seja, os índices dos Municípios são iguais para todos os grupos. Como resultado obteve-se:

$$T^2 = 1.3288$$

$$\text{distF} = 3.9985$$

como  $T^2 < \text{distF} (F_{1,n_1+n_2-2}(\alpha))$  aceita-se a hipótese, logo:

hipótese H02 aceita: Perfis coincidentes

Como última etapa da análise de perfis, testa-se a hipótese de que para o j-ésimo município os índices de desenvolvimento são iguais.

$$T03 = 7.6906e+03$$

$$F03 = 2.1813$$

hipótese H03 Rejeitada: Perfis dos grupos não são coincidentes entre si.

Como conclusão das hipóteses testadas acima, podemos constatar que os dois grupos de Municípios, o grupo M1, no qual encontram-se os Municípios que recebem royalties de petróleo e os grupo M2, no qual encontram-se os Municípios que não recebem royalties de petróleo, apresentam perfis paralelos e coincidentes, pois  $H_{01}$  e  $H_{02}$  foram aceitas. Porém dentro dos grupos os perfis não são iguais, visto

que  $H_{03}$  foi rejeitada. Podemos assim afirmar que não há uma diferença evidente quando comparados os perfis dos índices de desenvolvimento dos Municípios que recebem os royalties com os perfis dos índices de desenvolvimento dos Municípios que não recebem royalties.

#### 4.1.2 Análise de Cluster

Na entrada de dados do programa Minitab 16, foram utilizadas todas as variáveis consideradas na aplicação da análise de perfis para fazer os agrupamentos dos Municípios, conforme podemos ver na figura 8.

FIGURA 8: INTERFACE DO SOFTWARE MINITAB 16 COM OS DADOS PARA ANÁLISE

	C1 T	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
	Município/ano	População ( 2010)	Incidência da Pobreza (%)	IDHM - 2000	IDHM - 2010	IFDM	EMPREGO & RENDA	EDUCAÇÃO	SAUDE	IDEB 2011	organização		
1	Anchieta (ES)	23902	21,15	0,627	0,730	0,8088	0,6524	0,8834	0,8905	5,2	1		
2	Angra dos Reis (RJ)	169511	29,08	0,599	0,724	0,8541	0,9354	0,7110	0,9159	5,1	2		
3	Aracruz (ES)	81832	33,72	0,638	0,752	0,8939	0,8600	0,9122	0,9095	5,8	3		
4	Armacao dos Buzios (RJ)	27560	15,32	0,604	0,728	0,6393	0,3776	0,7612	0,7792	4,6	4		
5	Balneio Barra do Sul (SC)	8430	40,76	0,639	0,716	0,7362	0,5051	0,8165	0,8070	5,5	5		
6	Bertioga (SP)	47645	25,03	0,634	0,730	0,7735	0,6526	0,8841	0,7837	5,3	6		
7	Cabo Frio (RJ)	186227	24,01	0,614	0,735	0,7210	0,5661	0,7835	0,8135	4,7	7		
8	Cachoeiras de Macacu (RJ)	54273	41,50	0,580	0,700	0,7438	0,6667	0,7512	0,8135	4,1	8		
9	Cachoeiro de Itapemirim (ES)	189889	30,46	0,631	0,746	0,7874	0,7145	0,8658	0,7818	5,4	9		
10	Cairu (BA)	15374	66,85	0,437	0,627	0,5833	0,4544	0,6185	0,6769	2,9	10		
11	Campos dos Goytacazes (RJ)	463731	33,26	0,618	0,716	0,7556	0,6800	0,7370	0,8498	3,6	11		
12	Carapebus (RJ)	13359	18,80	0,579	0,713	0,7013	0,4741	0,7379	0,8918	4,5	12		
13	Casimiro de Abreu (RJ)	35347	16,92	0,619	0,726	0,6472	0,2453	0,8336	0,8627	5,2	13		
14	Guapimirim (RJ)	51483	46,21	0,572	0,698	0,6120	0,3492	0,6921	0,7947	4,3	14		
15	Guararema (SP)	25844	49,56	0,658	0,731	0,8051	0,6334	0,9297	0,8522	5,1	15		
16	Guaratuba (PR)	32095	45,73	0,613	0,717	0,6544	0,4700	0,7249	0,7684	5,0	16		
17	Ilhabela (SP)	28196	25,36	0,658	0,756	0,7436	0,4438	0,8948	0,8921	5,3	17		
18	Itapemirim (ES)	30988	30,80	0,525	0,654	0,6935	0,5123	0,8189	0,7494	4,7	18		
19	Linhares (ES)	141306	37,12	0,621	0,724	0,8181	0,6645	0,8823	0,9074	5,2	19		
20	Macaé (RJ)	206728	14,65	0,665	0,764	0,8356	0,7966	0,8241	0,8861	5,0	20		
21	Mangaratiba (RJ)	36456	37,20	0,642	0,753	0,7012	0,4302	0,8466	0,8267	5,3	21		
22	Maricá (RJ)	127461	38,72	0,637	0,765	0,7390	0,6062	0,7828	0,8282	4,3	22		
23	Matinhos (PR)	29428	45,80	0,635	0,743	0,6961	0,5222	0,7650	0,8011	4,9	23		
24	Paranaguá (PR)	140469	46,29	0,645	0,750	0,7808	0,7063	0,8220	0,8143	4,3	24		
25	Parati (RJ)	37533	34,45	0,583	0,693	0,7017	0,4807	0,7420	0,8824	4,4	25		
26	Pilar (AL)	33305	66,37	0,418	0,610	0,6281	0,4411	0,6287	0,8144	3,4	26		

Fonte: O autor

Foi realizado, a partir de então, os testes para a verificação de agrupamentos.

## MÉTODO DA LIGAÇÃO SIMPLES OU DO VIZINHO MAIS PRÓXIMO (SINGLE LINKAGE)

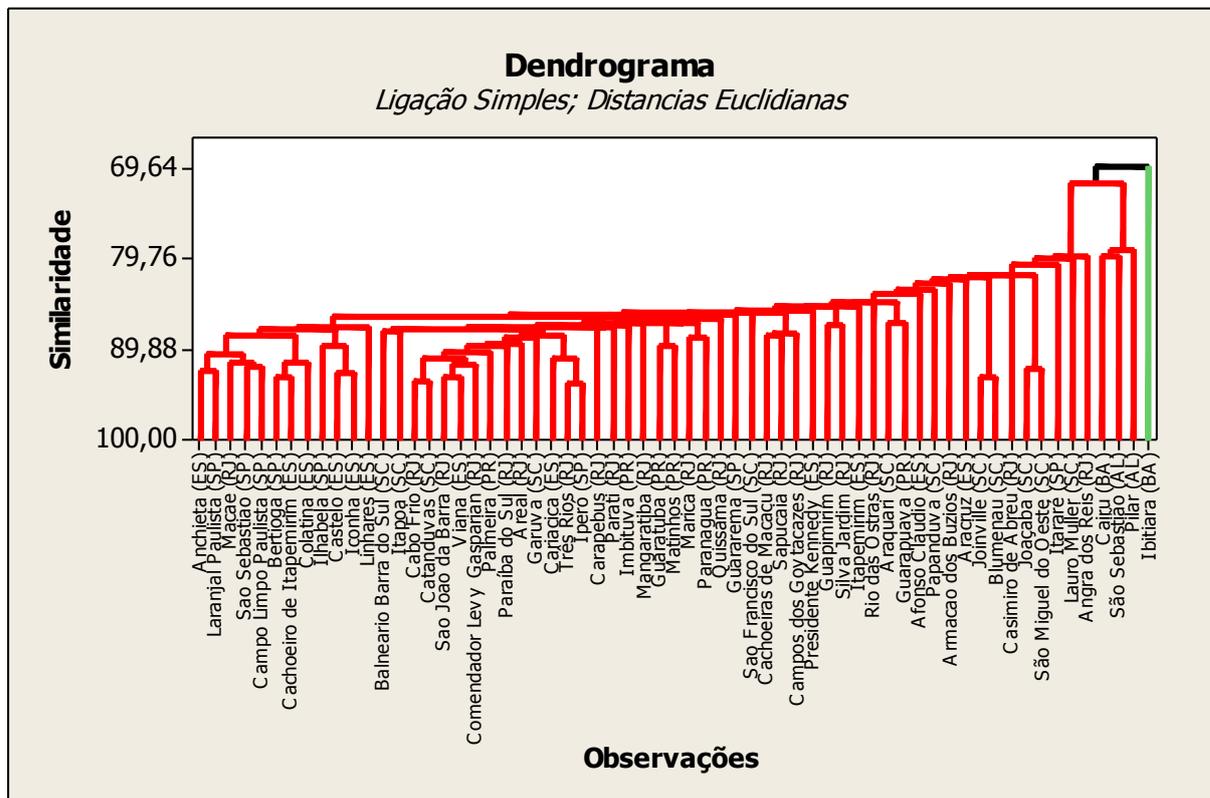
Nas ligações simples, o agrupamento é feito juntando-se **dois grupos** com **menor distância ou maior similaridade**. Uma vez formado o novo grupo, por exemplo, (AB), na ligação simples, a distância entre (AB) e algum outro grupo C é calculado por:

$$d_{(AB)C} = \text{mín}\{d_{AC}, d_{BC}\} \quad (4.1)$$

em que  $d_{AC}$  e  $d_{BC}$  representam as respectivas distâncias.

Os resultados obtidos foram dispostos graficamente em um dendrograma, no qual cada Município foi representado pelo seu respectivo código do IBGE, como mostra a figura 9

FIGURA 9: MÉTODO DA LIGAÇÃO SIMPLES (SINGLE LINKAGE)



Pelo método da ligação simples, os Municípios foram divididos em dois grupos, sendo que um deles contendo apenas o município de Presidente Kennedy (ES) o qual recebe royalties e o segundo grupo contendo os demais Municípios. Para facilitar a visualização, os Municípios foram organizados no quadro 4, no qual os Municípios que recebem os royalties de petróleo foram destacado em caixa alta.

**QUADRO 4: GRUPOS DEFINIDOS PELO MÉTODO DA LIGAÇÃO SIMPLES (SINGLE LINKAGE)**

continua

GRUPO 1	GRUPO 2
<p>ANCHIETA (ES)  ANGRA DOS REIS (RJ)  ARACRUZ (ES)  ARMAÇÃO DOS BÚZIOS (RJ)  BALNEÁRIO BARRA DO SUL (SC)  BERTIOGA (SP)  CABO FRIO (RJ)  CACHOEIRAS DE MACACU (RJ)  CACHOEIRO DE ITAPEMIRIM (ES)  CAMPOS DOS GOYTACAZES (RJ)  CARAPEBUS (RJ)  CASIMIRO DE ABREU (RJ)  GUAPIMIRIM (RJ)  GUARAREMA (SP)  GUARATUBA (PR)  ILHABELA (SP)  ITAPEMIRIM (ES)  LINHARES (ES)  MACAÉ (RJ)  MANGARATIBA (RJ)  MARICA (RJ)  MATINHOS (PR)  PARANAGUÁ (PR)  PARATI (RJ)  QUISSAMA (RJ)  RIO DAS OSTRAS (RJ)  SÃO JOAO DA BARRA (RJ)  SÃO SEBASTIAO (SP)  SILVA JARDIM (RJ)  ITAPOÁ (SC)  ARAQUARI (SC)  GARUVA (SC)  JOINVILLE (SC)  SÃO FRANCISCO DO SUL (SC)  CAIRU (BA)  PILAR (AL)  Castelo (ES)  Iconha (ES)  Colatina (ES)  Afonso Cláudio (ES)  Viana (ES)  Cariacica (ES)  Sapucaia (RJ)  Areal (RJ)  Comendador Levy Gasparian (RJ)  Paraíba do Sul (RJ)</p>	<p>PRESIDENTE KENNEDY (ES)</p>

conclusão

GRUPO 1	GRUPO 2
Três Rios (RJ) Catanduvas (SC) Lauro Muller (SC) Joaçaba (SC) Papanduva (SC) Blumenau (SC) São Miguel do Oeste (SC) Palmeira (PR) Imbituva (PR) Guarapuava (PR) Itararé (SP) Laranjal Paulista (SP) Iperó (SP) Campo Limpo Paulista (SP) Ibitiara (BA) São Sebastião (AL)	

Fonte: O autor

### MÉTODO DA LIGAÇÃO COMPLETA OU DO VIZINHO MAIS LONGE (COMPLETE LINKAGE)

Na ligação completa, o procedimento é muito semelhante ao da ligação simples, com uma única exceção: o algoritmo aglomerativo começa determinando a menor distância  $d_{ik}$ , constrói-se a matriz de distâncias  $D = (d_{ik})$  e os grupos vão se juntando. Se A e B são dois grupos de um único elemento, tem-se (A,B) como novo grupo. A distância entre (A,B) e outro grupo C é dada por:

$$d_{[(A,B),C]} = \max \{d_{(AC)}, d_{(BC)}\} \quad (4.2)$$

Pelo método da ligação completa, foi utilizado as distâncias euclidianas, ficando os Municípios divididos em dois grupos. Observou-se que por esse método os grupos ficaram melhor divididos, sendo esses mostrados no quadro 5. Para facilitar a visualização, os Municípios que recebem os royalties de petróleo foram destacados em caixa alta.

Torna-se visível pelo método “complete linkage” apresentado na figura 9 que os grupos se formam e se aglomeram mais do que no método “Single Linkage”.



conclusão

GRUPO 1	GRUPO 2
<p> <b>MARICA (RJ)</b>  <b>MATINHOS (PR)</b>  <b>PARANAGUÁ (PR)</b>  <b>PARATI (RJ)</b>  <b>PRESIDENTE KENNEDY (ES)</b>  <b>QUISSAMA (RJ)</b>  <b>RIO DAS OSTRAS (RJ)</b>  <b>SÃO JOÃO DA BARRA (RJ)</b>  <b>SÃO SEBASTIÃO (SP)</b>  <b>SILVA JARDIM (RJ)</b>  <b>ITAPOÁ (SC)</b>  <b>ARAQUARI (SC)</b>  <b>GARUVA (SC)</b>  <b>JOINVILLE (SC)</b>  <b>SÃO FRANCISCO DO SUL (SC)</b>            Castelo (ES)            Iconha (ES)            Colatina (ES)            Afonso Cláudio (ES)            Viana (ES)            Cariacica (ES)            Sapucaia (RJ)            Areal (RJ)            Comendador Levy Gasparian (RJ)            Paraíba do Sul (RJ)            Três Rios (RJ)            Catanduvas (SC)            Lauro Muller (SC)            Joaçaba (SC)            Papanduva (SC)            Blumenau (SC)            São Miguel do Oeste (SC)            Palmeira (PR)            Imbituva (PR)            Guarapuava (PR)            Itararé (SP)            Laranjal Paulista (SP)            Iperó (SP)            Campo Limpo Paulista (SP)         </p>	

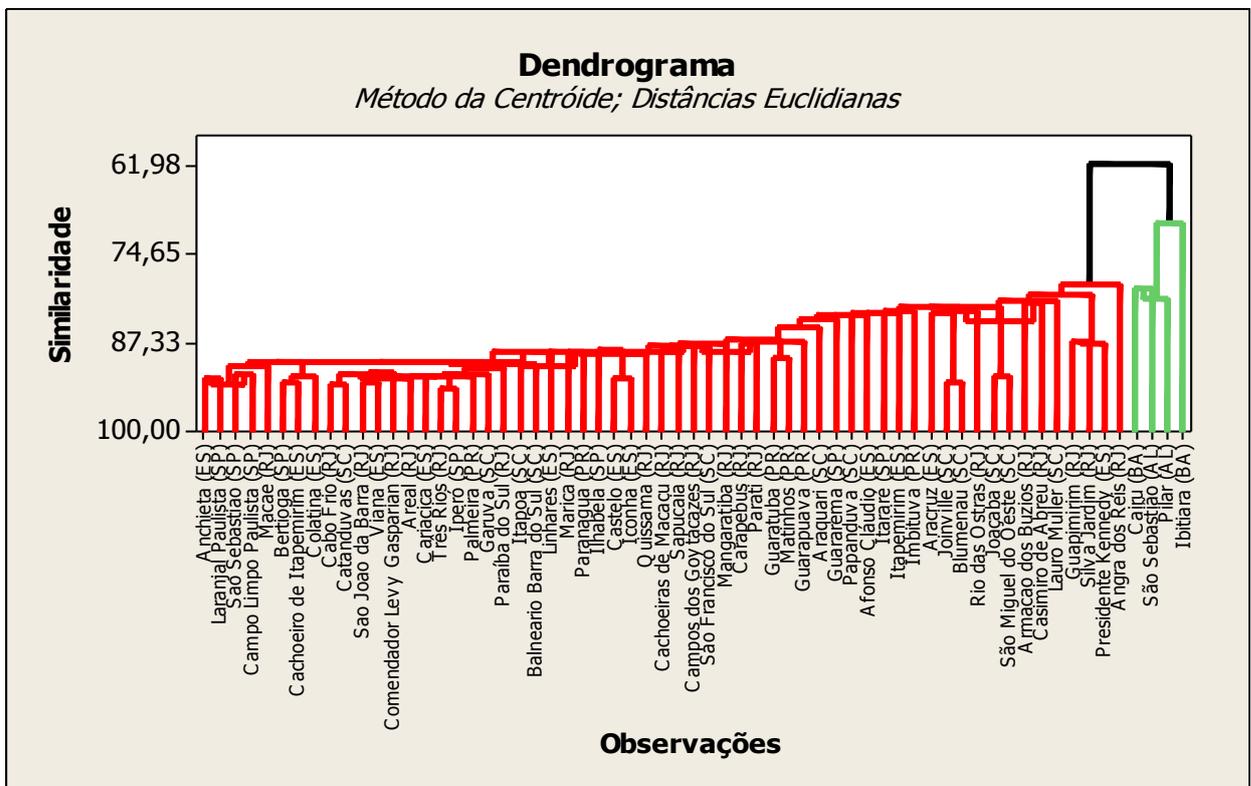
Fonte: O autor

## MÉTODO DO CENTROIDE

O método do centroide foi sugerido por Sokal & Michener (1958) e teve como origem, a caracterização da matriz de dados como pontos do espaço Mahalanobis ( $I_p$ ). Cada agrupamento é considerado um simples ponto, representado pelo seu centro de massa, chamado centroide. O presente método utiliza uma função de agrupamento para medir a distância entre os centros de massa dos dados. Esta técnica é de hierarquização aglomerativa. No método do centroide, a distância entre dois grupos é definida como a distância entre os seus centroides, pontos definidos

pelas médias das variáveis caracterizadoras dos indivíduos de cada grupo, isto é, o método do centroide calcula a distância entre dois grupos como a diferença entre as suas médias, para todas as variáveis. A figura 10 apresenta a divisão dos Municípios após o método da centroide, em que podemos observar uma semelhança com os métodos utilizados anteriormente.

FIGURA 11: MÉTODO DO CENTRÓIDE



Fonte: O autor

Após a aplicação do método da centroide, dividimos os grupos da seguinte forma, como se observa no quadro 6, destacando em caixa alta os Municípios que recebem royalties.

## QUADRO 7: GRUPOS DEFINIDOS PELO MÉTODO DO CENTROIDE

GRUPO 1	GRUPO 2
<p> <b>ANCHIETA (ES)</b>  <b>ANGRA DOS REIS (RJ)</b>  <b>ARACRUZ (ES)</b>  <b>ARMAÇÃO DOS BUZIOS (RJ)</b>  <b>BALNEÁRIO BARRA DO SUL (SC)</b>  <b>BERTIOGA (SP)</b>  <b>CABO FRIO (RJ)</b>  <b>CACHOEIRAS DE MACACU (RJ)</b>  <b>CACHOEIRO DE ITAPEMIRIM (ES)</b>  <b>CAMPOS DOS GOYTACAZES (RJ)</b>  <b>CARAPEBUS (RJ)</b>  <b>CASIMIRO DE ABREU (RJ)</b>  <b>GUAPIMIRIM (RJ)</b>  <b>GUARAREMA (SP)</b>  <b>GUARATUBA (PR)</b>  <b>ILHABELA (SP)</b>  <b>ITAPEMIRIM (ES)</b>  <b>LINHARES (ES)</b>  <b>MACAÉ (RJ)</b>  <b>MANGARATIBA (RJ)</b>  <b>MARICÁ (RJ)</b>  <b>MATINHOS (PR)</b>  <b>PARANAGUA (PR)</b>  <b>PARATI (RJ)</b>  <b>QUISSAMA (RJ)</b>  <b>RIO DAS OSTRAS (RJ)</b>  <b>SÃO JOÃO DA BARRA (RJ)</b>  <b>SÃO SEBASTIÃO (SP)</b>  <b>SILVA JARDIM (RJ)</b>  <b>ITAPOÁ (SC)</b>  <b>ARAQUARI (SC)</b>  <b>GARUVA (SC)</b>  <b>JOINVILLE (SC)</b>  <b>SÃO FRANCISCO DO SUL (SC)</b>            Castelo (ES)            Iconha (ES)            Colatina (ES)            Afonso Cláudio (ES)            Viana (ES)            Cariacica (ES)            Sapucaia (RJ)            Areal (RJ)            Comendador Levy Gasparian (RJ)            Paraíba do Sul (RJ)            Três Rios (RJ)            Catanduvas (SC)            Lauro Muller (SC)            Joaçaba (SC)            Papanduva (SC)            Blumenau (SC)            São Miguel do Oeste (SC)            Palmeira (PR)            Imbituva (PR)            Guarapuava (PR)            Itararé (SP)            Laranjal Paulista (SP)            Iperó (SP)            Campo Limpo Paulista (SP)         </p>	<p> <b>CAIRU (BA)</b>  <b>PILAR (AL)</b>  <b>PRESIDENTE KENNEDY (ES)</b>            Ibitiara (BA)            São Sebastião (AL)         </p>

Fonte: O autor

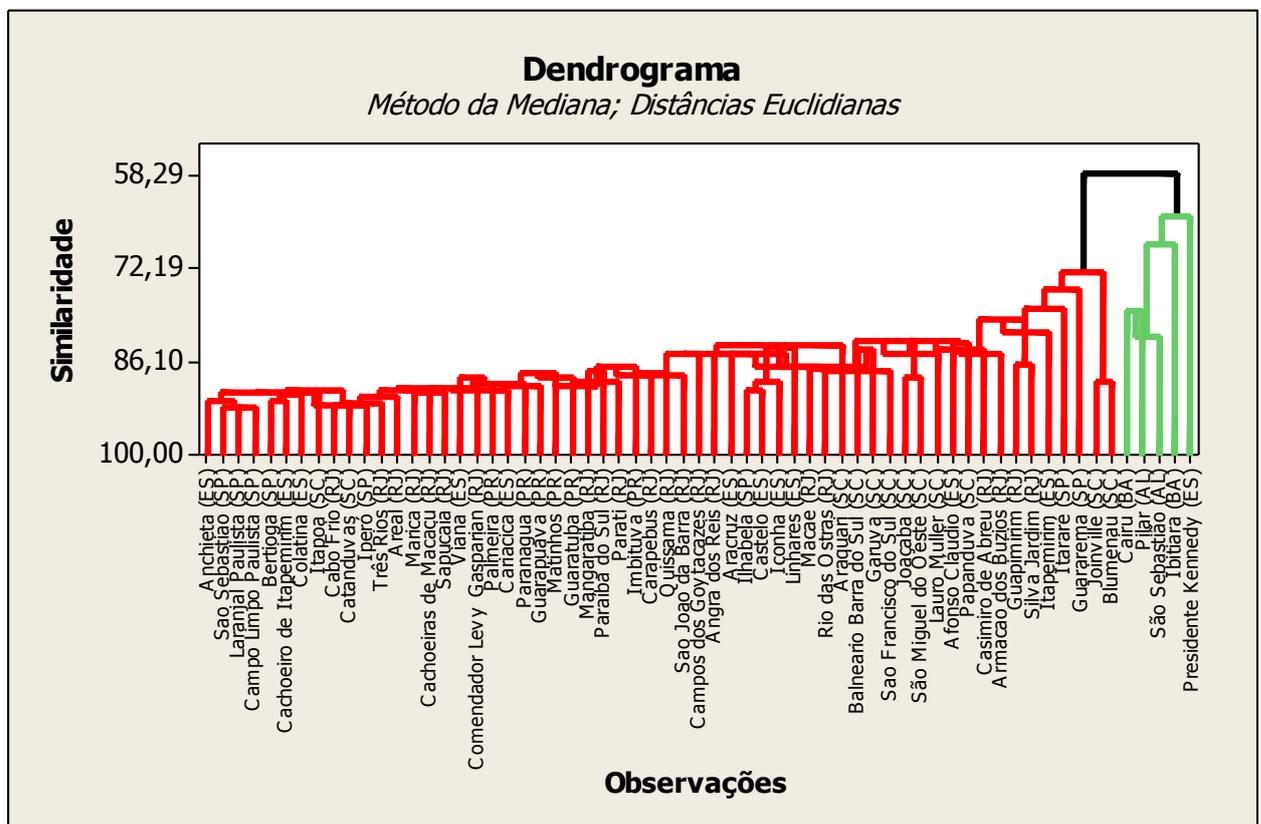
## MÉTODO DA MEDIANA

Este algoritmo é um caso particular do método do centroide. A determinação da distância entre dois agrupamentos por meio do cálculo do centro de massa não considera o número de elementos em cada um dos agrupamentos. Assim, o vetor médio que representa o novo agrupamento, pode eventualmente, ficar situado entre os elementos do agrupamento com maior número de elementos.

Para Barroso & Artes (2003), o método da mediana é uma modificação do método do centroide para a independência da distância do tamanho dos grupos. Se agregarem os grupos, com centroides  $a$  e  $b$ , para formar um novo grupo, a distância desse novo grupo a outro grupo, de centroide  $c$ , é a mediana.

A figura 11 apresenta a divisão dos Municípios após o método da mediana:

FIGURA 12: MÉTODO DA MEDIANA



Fonte: O autor

Observou-se pelo método da mediana que os grupos ficaram divididos igualmente ao método da centroide, como podemos ver no quadro 8, onde os Municípios que recebem royalties foram destacados em caixa alta.

QUADRO 8: GRUPOS DEFINIDOS PELO MÉTODO DA MEDIANA

GRUPO 1	GRUPO 2
<p><b>ANCHIETA (ES)</b>  <b>ANGRA DOS REIS (RJ)</b>  <b>ARACRUZ (ES)</b>  <b>ARMACÃO DOS BÚZIOS (RJ)</b>  <b>BALNEARIO BARRA DO SUL (SC)</b>  <b>BERTIOGA (SP)</b>  <b>CABO FRIO (RJ)</b>  <b>CACHOEIRAS DE MACACU (RJ)</b>  <b>CACHOEIRO DE ITAPEMIRIM (ES)</b>  <b>CAMPOS DOS GOYTACAZES (RJ)</b>  <b>CARAPEBUS (RJ)</b>  <b>CASIMIRO DE ABREU (RJ)</b>  <b>GUAPIMIRIM (RJ)</b>  <b>GUARAREMA (SP)</b>  <b>GUARATUBA (PR)</b>  <b>ILHABELA (SP)</b>  <b>ITAPEMIRIM (ES)</b>  <b>LINHARES (ES)</b>  <b>MACAE (RJ)</b>  <b>MANGARATIBA (RJ)</b>  <b>MARICA (RJ)</b>  <b>MATINHOS (PR)</b>  <b>PARANAGUÁ (PR)</b>  <b>PARATI (RJ)</b>  <b>QUISSAMA (RJ)</b>  <b>RIO DAS OSTRAS (RJ)</b>  <b>SAO JOAO DA BARRA (RJ)</b>  <b>SAO SEBASTIAO (SP)</b>  <b>SILVA JARDIM (RJ)</b>  <b>ITAPOA (SC)</b>  <b>ARAQUARI (SC)</b>  <b>GARUVA (SC)</b>  <b>JOINVILLE (SC)</b>  <b>SAO FRANCISCO DO SUL (SC)</b>  <b>Castelo (ES)</b>  <b>Iconha (ES)</b>  <b>Colatina (ES)</b>  <b>Afonso Cláudio (ES)</b>  <b>Viana (ES)</b>  <b>Cariacica (ES)</b>  <b>Sapucaia (RJ)</b>  <b>Areal (RJ)</b>  <b>Comendador Levy Gasparian (RJ)</b>  <b>Paraíba do Sul (RJ)</b>  <b>Três Rios (RJ)</b>  <b>Catanduvas (SC)</b>  <b>Lauro Muller (SC)</b>  <b>Joaçaba (SC)</b>  <b>Papanduva (SC)</b>  <b>Blumenau (SC)</b>  <b>São Miguel do Oeste (SC)</b>  <b>Palmeira (PR)</b>  <b>Imbituva (PR)</b>  <b>Guarapuava (PR)</b>  <b>Itararé (SP)</b>  <b>Laranjal Paulista (SP)</b>  <b>Iperó (SP)</b>  <b>Campo Limpo Paulista (SP)</b></p>	<p><b>CAIRU (BA)</b>  <b>PILAR (AL)</b>  <b>PRESIDENTE KENNEDY (ES)</b>  <b>Ibitiara (BA)</b>  <b>São Sebastião (AL)</b></p>

Fonte: O autor

Na comparação dos “clusters” obtidos pelos métodos citados, mostramos que as soluções de agrupamento foram muito parecidas, porém em nenhum método houve a separação completa entre Municípios que recebem royalties e os Municípios que não recebem esse incentivo fiscal.

Portanto, considerou-se o quadro 10 como a solução final de agrupamentos mais pertinente, pois dentre todos os métodos utilizados foi o que proporcionou o maior equilíbrio entre as quantidades de Municípios nos grupos, porém, mesmo assim não foi o suficiente para dividir em dois grupos específicos de Municípios contemplados ou não contemplados pelos royalties. Desta forma, conclui-se que os resultados apresentados demonstram que a análise de cluster foi capaz de agrupar os Municípios em dois grupos (*clusters*), porém os dois grupos ficaram com quantidades de Municípios desiguais, não sendo possível a distinção entre eles, quando comparados em relação ao recebimento de royalties do petróleo.

#### 4.1.3 Análise Fatorial

Este tipo de técnica pode ser utilizado de uma forma exploratória, na qual se busca uma estrutura em um conjunto de variáveis ou como um redutor de dados, ou de uma perspectiva confirmatória, quando se tem uma ideia pre-concebida sobre a estrutura real dos dados e deseja-se testar uma hipótese (HAIR et al., 2005).

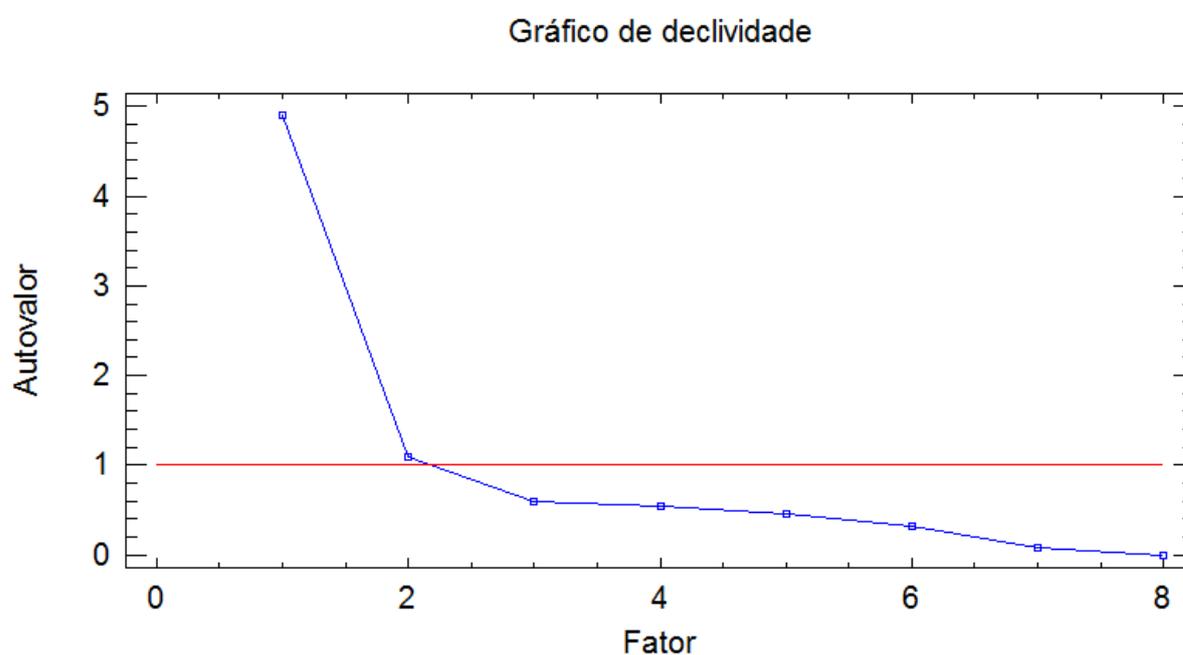
Para essa análise utilizou-se do software estatística Statgraphics Centurion XV. Os dados foram devidamente organizados em colunas, cada uma delas representando uma variável analisada em questão.

Segundo Mingoti (2005) para a seleção do número de fatores ( $m$ ) pode-se utilizar os critérios:

- ❖ **Autovalores:** A quantidade sugerida de fatores ( $m$ ) será igual ao número de autovalores maiores ou igual a 1 ( Critério de Kaiser). Esse critério mantém no sistema dimensões que representam pelo menos a informação de variância de uma variável original.

Para os dados em análise, o valor de  $m$  será igual a 2, pois como podemos verificar no teste, temos dois autovalores que satisfazem essa condição. O gráfico 6 apresenta os valores dos autovalores da matriz de correlação relacionando com a quantidade de fatores sugeridas pelo critério de Kaiser.

GRÁFICO 6: AUTOVALORES DA MATRIZ CORRELAÇÃO



Fonte: O autor

- ❖ **Análise da proporção da variância total:** permanecem aqueles que representam maiores proporções da variância total e, portanto,  $m$  será igual

ao número de autovalores retidos. Malhotra (2001) recomenda que o número de fatores escolhidos corresponda a, no mínimo, 60% da variância.

Considerando-se as 8 variáveis e analisando-se os resultados obtidos verifica-se que o 1º e o 2º autovalor são maiores do que 1 e explicam 74,95% da variabilidade dos dados originais, superando assim a quantidade mínima de 60%. Os demais autovalores são menores do que 1. O quadro 9 apresenta o número de valores com suas respectivas porcentagens de variâncias explicadas.

QUADRO 9: AUTO VALORES E VARIÂNCIA EXPLICADA % (63 VARIÁVEIS)

Número de fatores	$\lambda$ Autovalores	Variância explicada (%)	Variância explicada acumulada (%)
1	4,90441	61,305	61,305
2	1,0918	13,648	74,953
3	0,601358	7,517	82,470
4	0,543134	6,789	89,259
5	0,458348	5,729	94,988
6	0,326146	4,077	99,065
7	0,0748077	0,935	100,000
8	9,52224E-8	0,000	100,000

Fonte: O autor

No quadro 10, verifica-se o resultado dos Carregamentos fatoriais sendo que os valores destacados possuem uma significância maior do que 0,7. Vale ressaltar que para o teste em questão foi utilizado a rotação Varimax.

QUADRO 10: CARREGAMENTOS FATORIAIS

	Fator 1	Fator 2
Incidência da Pobreza _%_	<b>-0,745242</b>	-0,110979
IDHM_ 2000	<b>0,809627</b>	0,413689
IDHM _ 2010	<b>0,717086</b>	0,461713
IFDM	0,521413	<b>0,830699</b>
EMPREGO _ RENDA	0,0782471	<b>0,973728</b>
EDUCAÇÃO	<b>0,831526</b>	0,195904
SAÚDE	<b>0,668185</b>	0,324922
IDEB 2011	<b>0,815202</b>	0,108916

Fonte: O autor

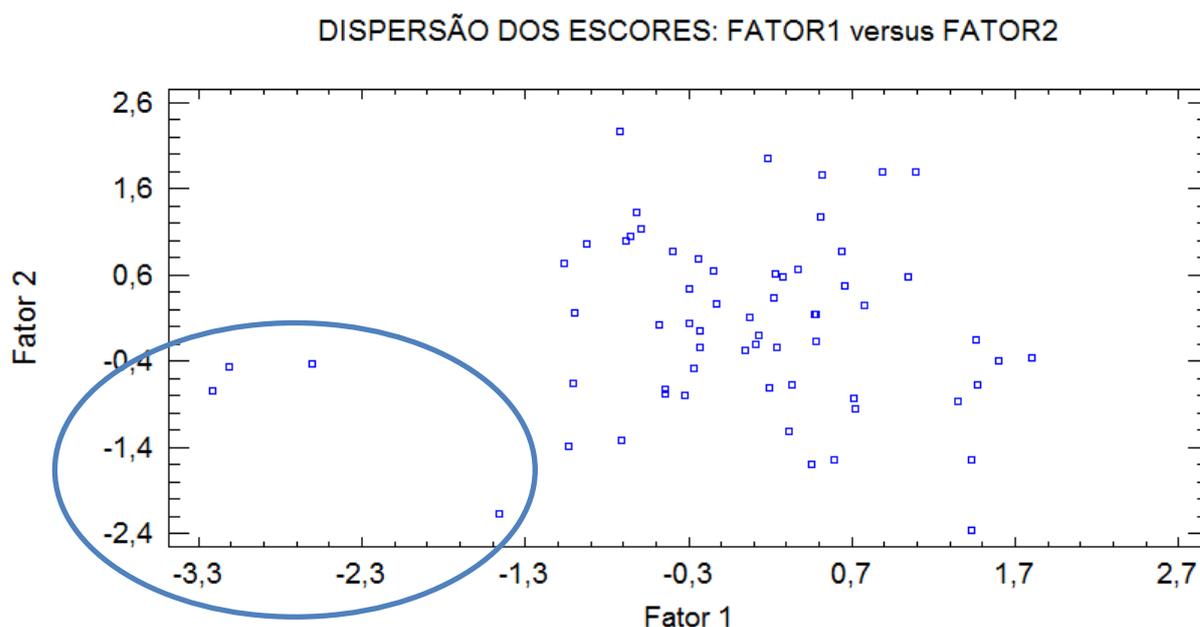
Observamos no quadro 10 que as cargas fatoriais são as correlações entre as variáveis originais e os fatores. Esse é um dos pontos principais da análise fatorial, pois quanto maior a carga fatorial maior será a correlação com um determinado fator. Nota-se que a variável “Incidência da Pobreza” possui uma carga fatorial negativa, o que indica um impacto inverso no fator. Assim, com relação aos resultados mostrados, as dimensões ficaram:

O fator 1 agrega as variáveis Incidência da Pobreza, IDHM-2000, IDHM-2010, Educação, Saúde e IDEB-2011.

O fator 2 agrupou as variáveis IFDM e Emprego & Renda.

Para uma melhor visualização dos resultados apresentados acima, pode ser verificada no gráfico 7 que explicita as duas primeiras componentes principais onde temos fator 1 versus fator 2.

GRÁFICO 7: DISPERSÃO DOS ESCORES: FATOR1 versus FATOR2: ROTAÇÃO VARIMAX

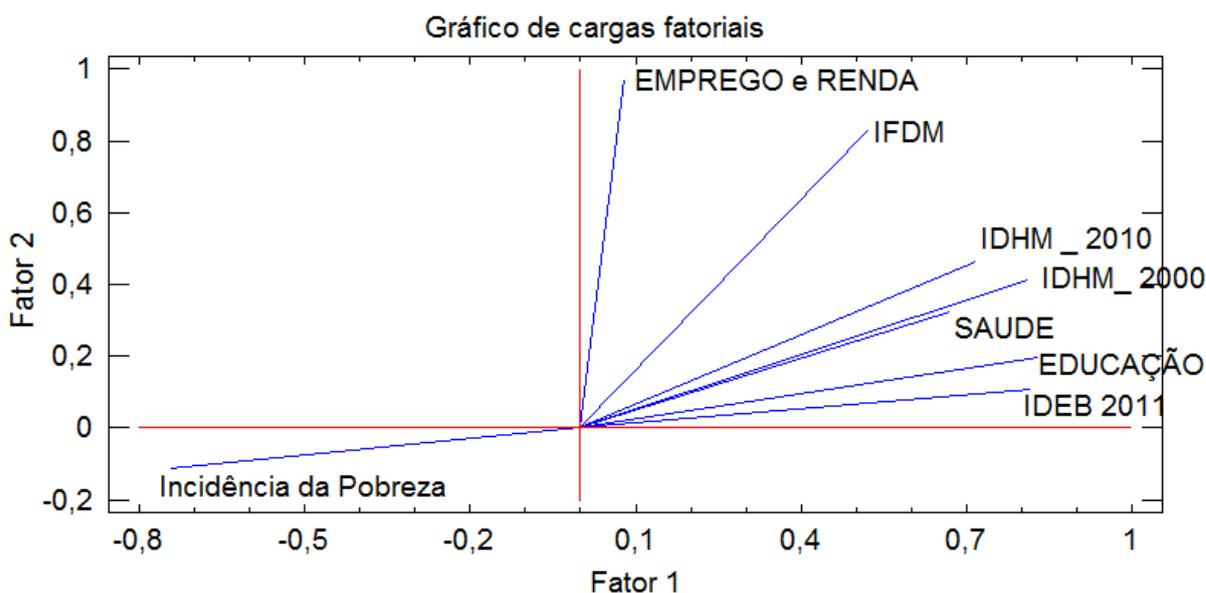


Observando o gráfico, fica visível a distinção entre os dois fatores, pois o mesmo apresenta os grupos distintos. O grupo em destaque no gráfico é composto pelos Municípios de Pilar (AL), Cairu (BA), São Sebastião (AL) e Ibitiara (BA).

É notório a semelhança nos resultados obtidos pela análise fatorial dos resultados obtidos na análise de Cluster, ficando evidente que não existe uma diferença entre os grupos de Municípios que recebem as participações governamentais, como royalties e participações especiais, com os Municípios que não recebem esse benefício financeiro.

O gráfico 8 mostra como ficou a relação entre os pesos dos fatores.

GRÁFICO 8: PESOS DOS FATORES: FATOR 1 versus FATOR 2



## 5 CONSIDERAÇÕES FINAIS E CONCLUSÃO

A distribuição dos royalties petrolíferos entre os Municípios brasileiros é uma questão debatida há alguns anos no Brasil. Uma justificativas para isso é a importância histórica do petróleo no cenário nacional, cabendo também ressaltar o crescimento das rendas petrolíferas nas últimas décadas.

A importância econômica que os royalties representam fez com que o seu repasse assumisse maiores proporções. Aliado a isso, pode-se observar um critério de distribuição dos royalties que gera uma série de divergências entre os beneficiados e os não beneficiados, sendo estes últimos a extrema maioria.

O artigo 20 da Constituição brasileira estabelece que as jazidas de petróleo são bens da União, mas que Estados e Municípios têm o direito de participar dos lucros da atividade, recebendo royalties e compensações. Na prática, são duas as leis ordinárias que definem qual fatia das rendas do petróleo cabe a cada ente da federação e como devem ser distribuídas: a Lei 7.990/1989 e a Lei 9.478/1997.

Pode-se dizer que em média, a União fica com 39,4% das receitas, os Estados, com 33,8%; e os Municípios, com 26,8%. Já os critérios utilizados para determinar o direito de um Estado ou Município receber royalties são bastante complexos e sofreram alterações ao longo do tempo. Podemos resumir em dois conceitos-chave para se entender a distribuição de royalties: o conceito de Estado ou Município “produtor” para produção em terra e Estado ou Município “confrontante” para produção em mar, sendo que neste existem dois tipos de linha que são utilizadas para essa verificação: as linhas ortogonais e as linhas paralelas.

A Estatística Multivariada consiste em uma poderosa ferramenta na análise de dados, pois se trata de métodos estatísticos que permitem analisar simultaneamente um grande número de variáveis. As diversas técnicas de estatística multivariada têm a finalidade de simplificar ou facilitar a interpretação do fenômeno estudado e o seu desenvolvimento tem possibilitado o estudo de diversos fenômenos, sendo cada vez mais complexos. Estas podem ser empregadas com o intuito de se construir índices ou variáveis alternativas e grupos de elementos amostrais, analisar as relações de dependência das variáveis e comparar as populações, facilitado atualmente pelo avanço tecnológico.

Por outro lado, é importante salientar que a esfera pública tem suas responsabilidades e obrigações. Aos que receberam a tarefa de administrar, ainda que temporariamente, precisam ter a noção que a gestão dos recursos públicos não é simplesmente a gestão eficaz de números e a cuidadosa atenção à execução orçamentária como um todo. Uma boa gestão deve focar na população e tentar fornecer a ela um alto nível da qualidade de vida.

Tendo isso em vista, o presente estudo buscou encontrar métodos estatísticos que demonstrassem diferenças entre Municípios contemplados pelos royalties com os Municípios não contemplados, através de um número significativo de variáveis, para isso recorreu-se à aplicação de técnicas exploratórias da análise multivariada, ao conjunto de dados referente, aos índices de desenvolvimento municipais (IDHM).

As três técnicas utilizadas possuem, cada uma, extrema importância para a análise dos dados na situação problema descrita nesse estudo, porém, ao

utilizarmos os métodos como complemento, podemos nos assegurar com relação aos resultados.

Sendo assim, conclui-se que a Análise de Perfil, apesar de ter encontrado grandes dificuldades em se compreender e aplicar a metodologia estatística na forma em que se apresenta nas principais fontes bibliográficas disponíveis, foi possível desenvolver, com o auxílio do software Matlab uma sequência algébrica e matricial de operações em que foi possível testar as três hipóteses pertinentes ao método, sendo que se chegou a conclusão da não existência de diferenças entre os dois grupos de Municípios, logo não podemos assegurar que há uma relação entre o recebimento das participações especiais com altos índices de desenvolvimento municipal.

A técnica de análise de agrupamentos (*cluster*) teve como objetivo a identificação das variáveis pertencentes ao mesmo grupo, ou seja, quais variáveis apresentam maiores similaridades. Foi observado neste estudo, em todos os métodos aplicados, um agrupamento em que não se distinguiu Municípios recebedores de royalties com Municípios que não receberam royalties.

A técnica de análise fatorial foi utilizada na tentativa de reduzir um grande número de variáveis para um número mais significativo, que neste estudo foi reduzido a dois fatores, onde foi possível verificar que os agrupamentos formados pela análise fatorial foram semelhantes aos formados na análise de agrupamentos.

Verifica-se que os agrupamentos e escores fornecidos pela análise de Cluster e Análise Fatorial, respectivamente, podem servir como instrumentos de extrema importância para administradores municipais, podendo servir como forma de direcionar os investimentos, proporcionando maior eficiência nos gastos públicos.

A análise dos orçamentos municipais seria de extrema importância para uma melhor comparação e conclusão da avaliação da aplicação dos royalties, porém pela dificuldade de aquisição desses dados, mesmo entrando em vigor a lei da transparência fiscal, não foi possível a sua realização, ficando como sugestão para futuros trabalhos.

## 6 REFERÊNCIAS

AAKER, D. A.; KUMAR,V.; DAY, G. S. **Pesquisa de marketing**, São Paulo: Atlas, 2001. 745p.

ADAMCZUK, P. C. **Análise de perfis: aplicação a um caso real na indústria metalúrgica**. 218 f. Dissertação (Mestrado em Métodos Numéricos em Engenharia) - Setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2003.

Anais do XXXV Encontro Nacional de Economia ANPEC - Associação de Centros de Pós-Graduação em Economia, 2007.Rev.Bras.Biom., São Paulo, v.26, n.3, p.125-132, jul.-set. 2008SERRA, R.V. 2003. “**Desdobramento Espacial da Exploração e Produção de Petróleo no Brasil**: em busca de um nexos para a distribuição dos royalties entre os Municípios”. Belo Horizonte: X ENANPUR, 2003.

ANP –AGÊNCIA NACIONAL DO PETRÓLEO. **Guia dos Royalties do Petróleo e do gás natural**. Hamilton Barbosa (Coord). Rio de Janeiro: ANP, 2001.

BARBOSA, D. (coord.). **Guia dos Royalties do Petróleo e Gás Natural**. Rio de Janeiro, ANP, 2001.

BARROSO, L. P., ARTES, R. **Análise de Multivariada**. Lavras: UFLA, 2003.

BOCK, R.D. Multivariate statistical methods in behavioral research. McGraw Hill, 1975.CHAVES NETO, A. – “**Confiabilidade e Métodos Estatísticos Aplicados a Sistemas de Engenharia**”, material didático, UFPR, 2006.

BRASIL. Congresso Nacional. **Anais**. Brasília: 1971, v. 5, p. 78-79.

\_\_\_\_\_. **Entenda como as receitas do petróleo são originadas e distribuídas na federação brasileira**. Estudos Técnicos CNM. Brasília : CNM, 2010.

BUSSAB, W. O.; MIAZAKI, E.; ANDRADE, D.F. **Introdução à análise de agrupamentos**. São Paulo: Associação Brasileira de Estatística, 1990.

CASTRO, M. J. F. ; FERRAUDO, A. S. ; FERRAUDO, G. M. ; ALVES, P. L. C. A. ; KUVA, M. A. . **Análise multidimensional como ferramenta de análise em comunidades de plantas daninhas**. In: I Congresso Brasileiro de Fitossanidade - CONBRAAF, 2011, Jaboticabal - SP

CHAVES NETO, A. **Probabilidade e estatística aplicada**: Notas de aula. Departamento de Estatística, Universidade Federal do Paraná, Curitiba, 2006.

Confederação Nacional de Municípios – CNM. **Estudos Técnicos CNM / Confederação Nacional de Municípios – Brasília** : CNM, 2010.

CORRAR, L. J. **Análise Multivariada**. Atlas, 2007. 544p.

DÁVILA, O. S.; TICERAN, D. G.; CRUZ, Y. A.; ROJAS, A. C.; Estudo do nível de satisfação dos pacientes do instituto nacional materno perinatal usando análise multivariada. *Rev. Mat. Estat.*, São Paulo, v.25, n.2, p. 41-57 , abr.-jun. 2007

Estudos Técnicos CNM / Confederação Nacional de Municípios – Brasília : CNM, 2010.

FONSECA, J. S., M.; ANDRADE G., **“Curso de Estatística”**, São Paulo, Ed. Atlas, 1985.

GRIBSKOV, M.; MCLACHLAN, A.D.; EISENBERG, D. Profile analysis: detection of distantly related proteins. *Biochemistry. Proc. Natl. Acad. Sci. USA* Vol. 84, pp. 4355-4358, July 1987. Disponível em:<<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC305087/#reference-sec>>. Acesso em: 13/08/2013.

HAIR JR., J. F. et al. **Análise Multivariada de Dados**. 5a ed. Porto Alegre: Bookman, 2005.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4.ed. New Jersey: Prentice-Hall, inc., 1998.

LIMA, A. S.; Transparência orçamentária: As aplicações dos royalties do Petróleo e a reforma na gestão das finanças públicas. Impactos sociais, ambientais e urbanos das atividades petrolíferas: o caso de Macaé Cap. 1-5: Pag78 Disponível em :<<http://www.royaltiesdopetroleo.ucam-campos.br/index.php/artigos>>. Acesso em: 05/07/2013.

MALHOTRA, N. K. **Pesquisa de marketing**: uma orientação aplicada. Porto Alegre: Bookman, 2001.

MARQUES, J. M. **Análise Multivariada**. Notas de Aula. Curitiba, 2013.

MEYER, P. L. **Probabilidade**: aplicações à estatística. 1. ed. Rio de Janeiro: LTC, 1978.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada**: uma abordagem aplicada. Editora UFMG, 2005. Belo Horizonte, 295 p.

MORRISON, D. F. **Multivariate statistical methods**. New York: McGraw-Hill, 1976.

POSTALI, Fernando Antonio Slaibe. **Renda mineral, divisão de riscos e benefícios governamentais na exploração de petróleo no Brasil**. 2000. Dissertação (Mestrado em Teoria Econômica) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2000. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/12/12138/tde-11102001-143241/>>. Acesso em: 25/10/2013.

POSTALI, F. A. S. **Renda Mineral, divisão dos riscos e benefícios governamentais na exploração de petróleo no Brasil**. Rio de Janeiro: BNDES, 2002.

POSTALI, F. A. S. **Efeitos da distribuição de royalties do petróleo sobre o crescimento dos Municípios no Brasil**. In: XXXV Encontro Nacional de Economia ANPEC - Associação de Centros de Pós-Graduação em Economia, 2007, Recife.

PROFILE ANALYSIS. San Francisco State University: Disponível em: <<http://userwww.sfsu.edu/efc/classes/biol710/manova/Profile-Analysis.pdf>>. Acesso em: 12/05/2013.

PROGRAMA DAS NAÇÕES UNIDAS NO BRASIL (PNUD), **Atlas do Desenvolvimento Humano no Brasil**, disponível em <[www.pnud.org.br](http://www.pnud.org.br)> . Acesso em agosto de 2013.

RENCHER, A. C.; CHRISTENSEN W. F. **Methods of Multivariate Analysis**. 3. ed. New York: John Wiley & Sons, 2007.

SERRA, R.V. “**Desdobramento Espacial da Exploração e Produção de Petróleo no Brasil: em busca de um nexos para a distribuição dos royalties entre os Municípios**”. Belo Horizonte: X ENANPUR, 2003.

SERRA, Rodrigo Valente. **Contribuições para o debate acerca da repartição dos royalties petrolíferos no Brasil**. Tese de Doutorado em Economia, Instituto de Economia, Unicamp, 2005.

SERRA, Rodrigo et al. **Royalties**: ameaças às antigas regras de distribuição. Anais do XI Congresso Brasileiro de Energia. Rio de Janeiro, 2006.

SOLANO, O.; GÓMEZ, D.; CÁRDENAS, A.; ADRIAZOLA, Y. **Diseño muestral para el estudio del nivel de satisfacción de los usuarios del departamento de neonatología del Instituto Especializado Materno Perinatal**. PESQUIMAT. Rev. Invest. Fac. Cienc. Matem. Univ. Nac. Mayor San Marcos, Lima, v.9, n.2, p.63-72, 2006.

SOUZA, A. M.; VICINI, L. **Análise multivariada**: da teoria à prática. Caderno didático. Santa Maria: UFSM, 2005. 215 p.

SOKAL, R. R.; MICHENER, C.D. **A statistical method for evaluating systematic relationships**. Bulletin of the Society University of Kansas, n.38, p.109-1438, 1958.

SPIEGEL, M. R. **Estatística**. 3. ed. São Paulo: Makron Books, 1993.

TABACHNICK, B.; FIDELL, L. **Using multivariate statistics**. 5. ed. Boston: Allyn and Bacon, 2007.

TERRA, D., PONTES, C. **Royalties: ameaças às atuais regras de distribuição**. XI CONGRESSO BRASILEIRO DE ENERGIA. Rio de Janeiro, 20 06. Disponível em: <http://www.royaltiesdopetroleo.ucam-campos.br/>.

\_\_\_\_\_. **Notas sobre os Fundamentos Econômicos da Distribuição Espacial dos Royalties Petrolíferos no Brasil**. In: ANPEC. Anais do XXX Encontro Nacional da ANPEC. Nova Friburgo: ANPEC, 2002.

## **ANEXOS**

QUADRO 14: RECEITAS DE ROYALTIES E PARTICIPAÇÕES ESPECIAIS – 2002 A 2012

Município/Ano	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Anchieta (ES)	0,00	287084,69	571255,49	578753,22	1593659,81	2425514,81	3743449,14	2062649,09	21769020,84	48524821,59	53409715,08
Angra dos Reis (RJ)	11288674,85	13017066,52	13133396,13	20927234,64	27058034,23	51988233,56	87901291,34	66117238,37	82904406,21	83842784,51	96628816,91
Aracruz (ES)	0,00	3858345,96	5474570,21	6335329,83	13542295,58	20996147,95	32884036,11	11466857,18	31986834,64	35955238,51	34515133,81
Armação dos Búzios (RJ)	28069473,82	36247341,45	35680082,57	45892233,09	56413371,92	49524418,30	63761419,24	41897724,65	54433494,59	64834624,60	82693743,10
Balneário Barra do Sul (SC)	0,00	864054,18	1453904,46	4386165,39	1991433,78	1612640,15	2076910,33	1623281,06	1970200,52	3472648,32	4173000,98
Bertioga (SP)	8319367,92	13125666,44	14895737,33	17249115,86	19016650,81	17429784,60	23119834,53	15413841,44	21262377,54	35199936,44	40306364,67
Cabo Frio (RJ)	62859987,71	87556221,75	100454451,79	152964041,87	221054127,55	176280616,46	189691889,99	133778273,24	190041307,95	245856802,43	324982556,21
Cachoeiras de Macacu (RJ)	6747056,13	9129905,06	10496751,27	13308815,59	16070718,14	11196222,10	31193173,81	23645959,81	28002617,64	33694282,76	40184457,83
Cachoeiro de Itapemirim (ES)	1085,40	64769,19	54484,65	69938,94	286128,23	579867,95	2133533,05	1080652,42	1336598,34	2662386,72	3292440,46
Cairu (BA)	0,00	0,00	0,00	0,00	0,00	1926587,64	5159982,72	4783917,41	6751455,82	4239299,75	7397994,99
Campos dos Goytacazes (RJ)	348260944,92	475182177,54	549607310,98	678425571,67	852565850,92	780099183,21	1168642499,45	993167279,11	1016020729,34	1231637184,61	1354233313,47
Carapebus (RJ)	20370812,17	24245167,22	23889909,09	27959213,68	33076809,90	28447909,09	35326724,96	22745005,61	29294670,39	34802768,18	38640170,21
Casimiro de Abreu (RJ)	26387344,44	32072939,22	31873289,34	57424338,18	83354214,53	67110838,43	80218274,15	56490065,54	72130080,65	99712639,25	112016299,75
Guapimirim (RJ)	7346098,42	9964157,88	11369845,15	14362746,28	17570572,99	12491392,98	34381083,68	26032548,23	31293396,73	38944419,97	47619544,11
Guararema (SP)	0,00	10762,35	10620,65	11160,10	12681,52	11752,44	29474830,38	36340120,84	44439108,60	66975516,56	80865361,67
Guaratuba (PR)	194677,90	21889,99	54439,33	63040,72	674795,89	1749885,22	1670458,08	626,95	0,00	0,00	0,00
Ilhabela (SP)	9117905,14	11916059,46	13702058,22	15994806,18	17591344,99	16108845,94	21836503,19	14338744,85	18644987,20	33580143,63	43817856,33
Itapemirim (ES)	759,79	864801,16	4631557,90	3071874,80	2321792,45	10641646,99	17123174,21	11923109,57	23094799,83	118525364,80	171659343,84
Linhares (ES)	11373081,08	17171747,98	16177060,55	19249745,32	20796032,56	24406239,06	56574641,66	28543590,29	49543418,19	95315277,45	111312698,16
Macaé (RJ)	194061125,21	256136554,11	293831875,82	346291967,80	408416988,96	353801517,16	501680924,82	387541067,11	410516330,10	482234671,56	538654219,38
Mangaratiba (RJ)	3634792,07	3584910,86	5134848,47	9922513,43	15433003,52	18989213,11	24900096,62	22722490,42	26603395,84	21060327,94	25615850,57
Maricá (RJ)	0,00	572091,55	3535212,72	5229714,86	5828500,12	5563217,38	7026179,83	22243994,36	36942403,32	68541786,56	126201392,79
Matinhos (PR)	272133,82	1186925,57	5048205,38	5149603,22	3396731,41	1602600,85	2125509,52	57302,41	0,00	0,00	0,00
Paranaguá (PR)	7558,54	30808,14	76618,35	88723,98	67647,18	52542,91	60774,31	970,73	0,00	0,00	0,00
Parati (RJ)	0,00	3143583,09	2407684,38	5888162,02	10936732,00	17081258,81	45723442,87	62914490,61	55686488,82	58808872,37	82067236,75
Pilar (AL)	1658293,75	2031236,29	3212996,24	4296493,36	4722847,01	3726468,49	3747048,48	1820283,36	2348490,81	2562576,18	3691871,75
Presidente Kennedy (ES)	186727,01	4657951,22	12799863,39	11760770,70	12151246,64	19903616,82	79201582,39	79188705,44	98382095,17	208228438,59	223513390,34
Quissama (RJ)	48529278,63	63064726,98	66083508,29	72982583,90	91565639,24	124388669,93	146649730,18	90969685,66	90508817,62	92661381,55	106949297,46
Rio das Ostras (RJ)	158669770,83	201942934,33	221957665,16	265866158,78	311134130,67	240852045,75	331550609,17	284555968,90	249109439,69	330362149,79	348835716,27
São João da Barra (RJ)	5125955,15	30695016,17	38452755,73	44611137,74	54963765,14	45948760,90	161427752,79	168329125,44	203028163,85	251512931,02	233262154,92
São Sebastião (SP)	25324696,84	34223922,26	37749107,00	45874897,47	48513341,63	43024986,89	56062284,96	37004361,09	50827984,09	78887323,39	93084014,67

Município/Ano	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Silva Jardim (RJ)	6265123,55	8477768,99	9746983,32	12358185,91	14922809,72	10396491,96	27966281,66	21199825,71	25105795,13	30208667,29	36027403,59
Itapoa (SC)	0,00	864054,18	1453904,46	4386165,37	1991433,78	1612640,15	2076910,33	1623281,06	1970200,52	3472648,32	4173000,98
Araquari (SC)	0,00	1638321,35	1468991,21	150613,60	1991399,29	1651650,82	2076910,33	1623281,06	1970200,52	3472648,32	4173000,98
Garuva (SC)	0,00	1638321,35	1468991,21	150613,60	1991399,29	1651650,82	2076910,33	1623281,06	1970200,52	3472648,32	4173000,98
Joinville (SC)	0,00	1638321,35	1468991,21	150613,60	1991399,29	1651650,82	2076910,33	1623281,06	1970200,52	3472648,32	4173000,98
Sao Francisco do Sul (SC)	11534206,27	14168866,82	13841909,04	19264370,83	19032877,30	15610523,61	18822015,43	13608558,21	18645585,85	29580724,46	35731667,08

QUADRO 4: PRINCIPAIS INDICADORES SOCIAIS DOS MUNICÍPIOS QUE RECEBEM ROYALTIES DO PETRÓLEO

continua

Município/ano	População 2010	Incidência da Pobreza (%) - 2010	IDHM-2000	IDHM - 2010	IFDM - 2011	EMPREGO & RENDA - 2011	EDUCAÇÃO 2011	SAUDE 2011	IDEB 2011
Anchieta (ES)	23.902	21,15	0,627	0,730	0,8088	0,6524	0,8834	0,8906	5,2
Angra dos Reis (RJ)	169.511	29,08	0,599	0,724	0,8541	0,9354	0,7110	0,9159	5,1
Aracruz (ES)	81.832	33,72	0,638	0,752	0,8939	0,8600	0,9122	0,9095	5,8
Armação dos Búzios (RJ)	27.560	15,32	0,604	0,728	0,6393	0,3776	0,7612	0,7792	4,6
Balneário Barra do Sul (SC)	8.430	40,76	0,639	0,716	0,7362	0,5851	0,8165	0,8070	5,5
Bertioga (SP)	47.645	25,03	0,634	0,730	0,7735	0,6526	0,8841	0,7837	5,3
Cabo Frio (RJ)	186.227	24,01	0,614	0,735	0,7210	0,5661	0,7835	0,8135	4,7
Cachoeiras de Macacu (RJ)	54.273	41,5	0,580	0,700	0,7438	0,6667	0,7512	0,8135	4,1
Cachoeiro de Itapemirim (ES)	189.889	30,46	0,631	0,746	0,7874	0,7145	0,8658	0,7818	5,4
Cairu (BA)	15.374	66,85	0,437	0,627	0,5833	0,4544	0,6185	0,6769	2,9
Campos dos Goytacazes (RJ)	463.731	33,26	0,618	0,716	0,7556	0,6800	0,7370	0,8498	3,6
Carapebus (RJ)	13.359	18,8	0,579	0,713	0,7013	0,4741	0,7379	0,8918	4,5
Casimiro de Abreu (RJ)	35.347	16,92	0,619	0,726	0,6472	0,2453	0,8336	0,8627	5,2
Guapimirim (RJ)	51.483	46,21	0,572	0,698	0,6120	0,3492	0,6921	0,7947	4,3
Guararema (SP)	25.844	49,56	0,658	0,731	0,8051	0,6334	0,9297	0,8522	5,1
Guaratuba (PR)	32.095	45,73	0,613	0,717	0,6544	0,4700	0,7249	0,7684	5
Ilhabela (SP)	28.196	25,36	0,658	0,756	0,7436	0,4438	0,8948	0,8921	5,3
Itapemirim (ES)	30.988	30,8	0,525	0,654	0,6935	0,5123	0,8189	0,7494	4,7
Linhares (ES)	141.306	37,12	0,621	0,724	0,8181	0,6645	0,8823	0,9074	5,2
Macaé (RJ)	206.728	14,65	0,665	0,764	0,8356	0,7966	0,8241	0,8861	5
Mangaratiba (RJ)	36.456	37,2	0,642	0,753	0,7012	0,4302	0,8466	0,8267	5,3
Maricá (RJ)	127.461	38,72	0,637	0,765	0,7390	0,6062	0,7828	0,8282	4,3
Matinhos (PR)	29.428	45,8	0,635	0,743	0,6961	0,5222	0,7650	0,8011	4,9
Paranaguá (PR)	140.469	46,29	0,645	0,750	0,7808	0,7063	0,8220	0,8143	4,3
Parati (RJ)	37.533	34,45	0,583	0,693	0,7017	0,4807	0,7420	0,8824	4,4
Pilar (AL)	33.305	66,37	0,418	0,610	0,6281	0,4411	0,6287	0,8144	3,4
Presidente Kennedy (ES)	10.314	42,24	0,532	0,657	0,6628	0,5020	0,7281	0,7583	4,6
Quissama (RJ)	20.242	19,59	0,561	0,704	0,7707	0,5780	0,8215	0,9127	5,1
Rio das Ostras (RJ)	105.676	14,28	0,620	0,773	0,8624	0,9438	0,7881	0,8554	5,7

Sao Joao da Barra (RJ)	32.747	30,94	0,548	0,671	0,7369	0,6069	0,7604	0,8435	
Sao Sebastiao (SP)	73.942	21,58	0,663	0,772	0,8186	0,7173	0,8582	0,8803	
Silva Jardim (RJ)	21.349	52,09	0,546	0,654	0,6137	0,3445	0,7399	0,7567	4,1
Itapoa (SC)	14.763	41,03	0,634	0,761	0,7684	0,6125	0,8545	0,8382	5,1
Araquari (SC)	24.810	43,67	0,578	0,703	0,7811	0,7806	0,7563	0,8065	5,8
Garuva (SC)	14.761	35,47	0,555	0,725	0,7620	0,6321	0,8072	0,8468	5,2
Joinville (SC)	515.288	27,95	0,711	0,809	0,8727	0,8863	0,8551	0,8766	6,3
Sao Francisco do Sul (SC)	42.520	34,75	0,658	0,762	0,7526	0,7267	0,7714	0,7597	5,2

Fonte: IBGE/PNUD/ FIRJAN

QUADRO 5: PRINCIPAIS INDICADORES SOCIAIS DOS MUNICÍPIOS QUE NÃO RECEBEM ROYALTIES DO PETRÓLEO

Município/ano	População 2010	Incidência da Pobreza (%) - 2010	IDHM-2000	IDHM - 2010	IFDM - 2011	EMPREGO & RENDA - 2011	EDUCAÇÃO 2011	SAUDE 2011	IDEB 2011
Castelo (ES)	34.747	19,1	0,651	0,726	0,7718	0,5083	0,8940	0,9131	5,7
Iconha (ES)	12.523	16,48	0,654	0,729	0,7964	0,5746	0,9074	0,9072	6,1
Colatina (ES)	111.788	32,51	0,657	0,746	0,8017	0,6892	0,9109	0,8052	5,7
Afonso Cláudio (ES)	31.091	33,87	0,553	0,667	0,6943	0,3881	0,8646	0,8303	5,6
Viana (ES)	65.001	32,51	0,592	0,686	0,7332	0,5938	0,7653	0,8404	5
Cariacica (ES)	348.738	35,57	0,613	0,718	0,8010	0,7776	0,7907	0,8348	4,6
Sapucaia (RJ)	17.525	35,26	0,586	0,675	0,7083	0,6155	0,7530	0,7564	4,1
Areal (RJ)	11.423	22,11	0,611	0,684	0,7258	0,6170	0,7349	0,8255	4,6
Comendador Levy Gasparian (RJ)	8.180	27,08	0,597	0,685	0,7464	0,5881	0,8063	0,8449	5,4
Paraíba do Sul (RJ)	41.084	30,83	0,619	0,702	0,7186	0,4837	0,8400	0,8321	4,8
Três Rios (RJ)	77.432	27,01	0,627	0,725	0,7721	0,7248	0,7754	0,8162	4,5
Catanduvas (SC)	9.555	28,63	0,622	0,714	0,7173	0,5681	0,7804	0,8036	5
Lauro Muller (SC)	14.367	27,11	0,640	0,735	0,6895	0,4151	0,7944	0,8588	6,8
Joaçaba (SC)	27.020	27,58	0,741	0,827	0,7601	0,5363	0,8592	0,8848	6,3
Papanduva (SC)	17.928	31,66	0,564	0,704	0,6739	0,4717	0,7819	0,7679	6
Blumenau (SC)	309.011	21,76	0,727	0,806	0,8849	0,8904	0,8832	0,8811	6,1
São Miguel do Oeste (SC)	36.306	29,98	0,705	0,801	0,7700	0,5642	0,8435	0,9023	6,2
Palmeira (PR)	32.123	34,55	0,629	0,718	0,7480	0,5855	0,7759	0,8825	5
Imbituva (PR)	28.455	39,27	0,540	0,660	0,6899	0,4589	0,7209	0,8899	5,1
Guarapuava (PR)	167.328	47,67	0,632	0,731	0,7601	0,7227	0,7325	0,8251	5,2
Itararé (SP)	47.934	38,09	0,588	0,703	0,7464	0,4065	0,8759	0,9569	4,9
Laranjal Paulista (SP)	25.251	21,49	0,661	0,729	0,7912	0,6531	0,8629	0,8575	5,4
Iperó (SP)	28.300	25,94	0,607	0,719	0,7782	0,6958	0,8079	0,8311	4,7
Campo Limpo Paulista (SP)	74.074	19,54	0,678	0,769	0,8185	0,6998	0,8919	0,8638	5,5
Ibitiara (BA)	15.508	38,68	0,424	0,585	0,5589	0,3669	0,7157	0,5942	4,9
São Sebastião (AL)	32.010	59,11	0,351	0,549	0,5937	0,4889	0,6329	0,6594	3,5

Fonte: IBGE/PNUD/ FIRJAN

