

UNIVERSIDADE FEDERAL DE ITAJUBÁ
INSTITUTO DE SISTEMAS ELÉTRICOS E ENERGIA
PROGRAMA DE PÓS-GRADUAÇÃO
EM ENGENHARIA DE ENERGIA

José Almir da Silva

MODELAGEM MULTIVARIADA PARA PREVISÃO DE DEMANDA DE
GASOLINA E ÓLEO DIESEL NO BRASIL

Itajubá

2014

UNIVERSIDADE FEDERAL DE ITAJUBÁ
INSTITUTO DE SISTEMAS ELÉTRICOS E ENERGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ENERGIA

José Almir da Silva

MODELAGEM MULTIVARIADA PARA PREVISÃO DE DEMANDA DE GASOLINA E
ÓLEO DIESEL NO BRASIL

Dissertação submetida ao Programa de Pós-
Graduação em Engenharia de Energia como
parte dos requisitos para obtenção do Título
de Mestre em Ciências em Engenharia de Energia.

Área de Concentração: Planejamento e Gestão de Sistemas Energéticos

Orientador: Prof. Dr. Jamil Haddad

Itajubá
2014

UNIVERSIDADE FEDERAL DE ITAJUBÁ
INSTITUTO DE SISTEMAS ELÉTRICOS E ENERGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ENERGIA

José Almir da Silva

MODELAGEM MULTIVARIADA PARA PREVISÃO DE DEMANDA DE GASOLINA E
ÓLEO DIESEL NO BRASIL

Dissertação aprovada por banca examinadora em 05 de setembro de
2014, conferindo ao autor o título de *Mestre em Ciências em Engenharia da Energia*.

Banca Examinadora:

Prof. Dr. Erik Leandro Bonaldi

Prof. Dr. Luiz Eduardo Borges da Silva

Prof. Dr. José Luiz Gonçalves (Co-Orientador)

Prof. Dr. Jamil Haddad (Orientador) – UNIFE

Itajubá
2014

DEDICATÓRIA

Dedico este trabalho a meus filhos Maria Clara, Camille e Rafael. Dedico em especial a minha avó materna Maria das Dores Ribeiro uma das responsáveis por estar apresentando esta Dissertação.

“A menor distância entre dois pontos é a ciência”

(Autor desconhecido)

AGRADECIMENTOS

A meus filhos Maria Clara, Camille e Rafael, fonte inesgotável de motivação e inspiração.

Aos meus irmãos Sonia (*in-memoriam*), Paulinho, Celeste, Fátima, Adilson, Edson, Djalma e Wilson (Piti) pelo apoio e a atenção nos momentos mais difíceis.

Meus queridos amigos Dra. Helga Gonzaga Martins e Dr. Jose Luiz Gonçalves, pela atenção e valiosas orientações.

Meu querido amigo e mestre Dr. Luiz Eduardo sempre presente nas reflexões e decisões da vida acadêmica.

Ao Professor Dr. Jamil Haddad pela boa vontade, simpatia, orientações, discussões, paciência e muito apoio.

A Agência Nacional do Petróleo (ANP) por meio do Programa de Recursos Humanos para o setor de petróleo e gás PRH16

Ao Centro de Excelência em Eficiência Energética - EXCEN, pela oportunidade de participação em eventos e de desenvolvimento pessoal e profissional.

RESUMO

Vive-se em um mundo de rápidas mudanças e constante renovação de informações e conhecimento, portanto, métodos e técnicas para interpretação de dados são cada vez mais necessários. Quando um evento qualquer depende de muitas variáveis para acontecer, a análise de cada variável de forma isolada não nos proporciona um retrato satisfatório da realidade e, conseqüentemente, não produz conhecimento. Tem que se apropriar das informações fornecidas pelo conjunto das variáveis, suas relações e correlações e o peso que cada variável analisada dispensa ao evento. A análise multivariada proporciona esta condição. A análise Multivariada permite reduzir dados, simplificar estruturas, classificar e agrupar, investigar a dependência entre variáveis, predizer e elaborar hipóteses. Quando as relações e correlações existentes entre todas as variáveis não são consideradas simultaneamente, efeitos desconhecidos entre variáveis irão dificultar a interpretação do fenômeno o qual é objeto estudo.

Nesta dissertação utilizou-se duas técnicas multivariadas complementares, primeiro aplicamos a técnica chamada Análise Fatorial com o objetivo de reduzir um conjunto de variáveis interdependentes em um número reduzido de variáveis compostas. Em seguida aplicamos o produto da Análise Fatorial na técnica multivariada Regressão Múltipla. O objetivo, previamente determinado, é utilizar as referidas técnicas para fazermos inferências da demanda de gasolina e óleo diesel no mercado brasileiro nos próximos anos. A Análise Fatorial é uma técnica de interdependência utilizada para possibilitar o estudo de eventuais inter-relações entre um certo número de variáveis, a fim de condensar a informação contida em um grupo de variáveis originais em um conjunto menor de variáveis, os chamados fatores, sem que haja uma perda considerável de informação. A Análise de Regressão Múltipla é uma técnica multivariada de dependência que utiliza uma variável dependente e duas ou mais variáveis independentes. Cada variável independente é ponderada pelo processo da Análise de Regressão para garantir máxima precisão do resultado.

Palavras-Chave: Multivariada; Análise Fatorial; Regressão Múltipla; Interdependência.

ABSTRACT

We live in a world of rapid change and constant information and knowledge, therefore, methods and techniques for the interpretation of data are increasingly needed. When any event depends on many variables to take place, the analysis of each variable in isolation does not give us a satisfactory picture of reality and, consequently, does not produce knowledge. We must lay hold of the information provided by the set of variables and their relationships, correlations and weight each variable analyzed waiver to the event. Multivariate analysis provides this condition. A Multivariate analysis reduces data, simplify structures, sorting and grouping, to investigate the dependency between variables, predicting and hypothesising. When the relationships and correlations between all variables are not considered simultaneously, unknown effects between variables will make the interpretation of the phenomenon which is the subject study.

In this thesis we use two complementary multivariate techniques first applied the technique called factor analysis with the goal of reducing a set of interdependent variables in a small number of composite variables. Then apply the product of Factor Analysis in Multivariate Multiple Regression technique. The objective, predetermined, is to use these techniques to do inferences demand for gasoline and diesel in the Brazilian market in the coming years. Factor analysis is an interdependence technique used to facilitate the study of possible interrelationships between a number of variables in order to condense the information contained in a set of original variables into a smaller set of variables, called factors, without there is a considerable loss of information. The Multiple Regression Analysis is a multivariate technique that uses a dependency dependent variable and two or more independent variables. Each independent variable is weighted by the process of Regression Analysis to ensure maximum accuracy of the result.

Keywords: Multivariate, Factor Analysis, Multiple Regression, Interdependence

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO	12
1.1 - Introdução	12
1.2 – Objetivos	14
1.2.1 – Objetivo Principal	14
1.2.2 – Objetivos Específicos	14
CAPÍTULO 2 - MODELAGEM MULTIVARIADA	16
2.1 - A análise de dados e modelagem multivariada	16
2.2 - Técnicas Estatísticas de análise multivariada	17
2.3 - Técnica de Interdependência – Análise Fatorial	18
2.3.1 - O que é análise fatorial	18
2.3.2 - Objetivo da análise fatorial	19
2.3.3 - Técnica de Interdependência – Análise Fatorial	19
2.3.4 - Escolha das variáveis	20
2.3.5 - Produto da análise fatorial aplicado em outras técnicas multivariada	20
2.3.6 - Tipos de variáveis usadas em análise fatorial	21
2.3.7 - Suposições na análise fatorial	22
2.3.8 - Adequação da utilização da análise fatorial	22
2.3.9 - Extração dos fatores	24
2.3.9.1 - Divisão da variância de uma variável	24
2.3.9.2 - Métodos de extração de fatores	25
2.3.9.3 - Escolha do número de fatores	26
2.3.10 - Rotação dos fatores	27
2.3.11 - Interpretação de uma matriz fatorial	29
2.3.12 - Uso dos resultados da análise fatorial	30
2.4 - Técnica de dependência – Análise de regressão múltipla	31
2.4.1 - O que é análise de regressão múltipla	31
2.4.2 - Regressão linear simples	31
2.4.2.1 - Equação de regressão simples	32
2.4.3 - Precisão da previsão	33
2.4.4 - Regressão linear múltipla.....	34
2.4.4.1 - Suposições em análise de regressão múltipla	35
CAPÍTULO 3 – SELEÇÃO DAS VARIÁVEIS	37

3.1 – Seleção de variáveis	37
3.2 – Preço da gasolina	38
3.3 – Preço do óleo diesel	39
3.4 – Preço do etanol	40
3.5 – Aumento da frota de carros	40
3.6 – Crescimento da população ocupada	41
3.7 – Crescimento do PIB (Produto Interno Bruto)	42
3.8 – Crescimento de recursos para financiamento de automóveis	43
3.9 – Venda de gasolina pelas distribuidoras	44
3.10 – Produção de gasolina	44
3.11 – Venda de óleo diesel pelas distribuidoras	44
3.12 – Produção de óleo diesel	45
3.13 – Valor de IPI pago para carros populares 1.000 cc	45
3.14 – Valor de IPI pago para carros entre 1.000 e 2.000 cc	45
CAPÍTULO 4 – ANÁLISE DOS RESULTADOS	46
4.1 – Análise fatorial para consumo de gasolina	46
4.2 – Matriz de correlação	47
4.3 – Teste de esfericidade de Bartlett e KMO	47
4.4 – Matriz anti-imagem	49
4.5 – Comunalidades	49
4.6 – Análise fatorial para consumo de óleo diesel	52
4.7 – Análise de regressão	55
4.8 – Previsão do consumo de gasolina	56
4.9 – Previsão do consumo de óleo diesel	59
CAPÍTULO 5 – CONCLUSÃO	62
5.1 – Considerações finais	62
REFERENCIAS BIBLIOGRÁFICAS	64

CAPITULO 1 - INTRODUÇÃO

1.1. INTRODUÇÃO

A análise Multivariada permite reduzir dados, simplificar estruturas, classificar e agrupar, investigar a dependência entre variáveis, prever e elaborar hipóteses. Quando as relações e correlações existentes entre todas as variáveis não são consideradas simultaneamente, efeitos estranhos entre variáveis irão dificultar a interpretação do fenômeno que está sendo analisado.

Em se tratando do mercado de bens e serviços, o cuidado em diminuir custos, elaborar planejamentos estratégicos, a busca constante pela qualidade, decisões com informações e conhecimento mais próximos da realidade se tornou fundamental, uma questão de sobrevivência. Para tamanho desafio, o cuidado no processo de criação e gerenciamento mais eficaz do conhecimento se mostra prioritário. Os avanços da tecnologia permitiram a utilização de técnicas estatísticas, técnicas multivariadas que atuam na análise dos dados, transformando uma massa de dados em conhecimento. No mundo de hoje não há falta de informação, mas escassez do conhecimento “*estamos nos afogando em informações e famintos por conhecimento*” (HAIR, 2009).

Neste trabalho utilizamos duas técnicas multivariadas complementares. Primeiro aplicou-se a técnica chamada Análise Fatorial com o objetivo de reduzir um conjunto de variáveis em um número reduzido de variáveis compostas, variável estatística. Em seguida aplicou-se ao produto da Análise Fatorial a técnica multivariada chamada Regressão Múltipla. O objetivo, previamente determinado, é utilizar as técnicas para fazer inferências da demanda de gasolina e óleo diesel no mercado brasileiro nos próximos anos. A Análise Fatorial é utilizada para possibilitar o estudo de eventuais inter-relações entre certo número de variáveis, a fim de condensar as informações contidas em um conjunto menor de variáveis, os chamados fatores, sem que haja uma perda considerável de informação (LUIZ PAULO, 2009). O principal propósito da Análise Fatorial é descrever as relações de covariância entre as variáveis, identificar fatores não observáveis, também chamados de construtos latentes, que possam explicar a intercorrelação entre as variáveis.

O método de Análise Fatorial tem origem na técnica elaborada por Charles Spearman¹ no início do século XX, na ocasião ele elaborou o que chamou de método tetrádico². Segundo esse método, quando há correlação significativa entre dois ou mais testes escolares aplicados a estudantes, essa correlação é devida a um fator comum denominado por ele de fator “g”. A parte que não se correlaciona corresponderia a fatores específicos a cada teste de inteligência. Portanto, a análise fatorial de Spearman isola dois fatores (fator "g" e fatores específicos). Representando uma inteligência geral, o fator “g”, componente de inteligência, seria uma espécie de aptidão geral ou global que incide no sucesso e no bom rendimento cognitivo e, por ser geral, todos os indivíduos possuem. Além da inteligência geral, haveria inteligências específicas de cada indivíduo. A técnica desenvolvida por Spearman ficou também conhecida como bi fatorial por identificar dois fatores somente para explicar um evento.

Hoje a Análise Fatorial é um nome escolhido para representar um conjunto de métodos estatísticos que tem o propósito primeiro de explicitar uma estrutura inerente entre as variáveis. A técnica analisa as inter-relações entre um grande número de variáveis independentes e explica essas variáveis em termos de suas dimensões inerentes comuns, as quais chamadas de Fatores. As técnicas de análise multivariada proporcionam a expansão da habilidade exploratória da pesquisa e aumenta a eficiência estatística.

A Análise de Regressão Múltipla é uma técnica multivariada de dependência que utiliza uma variável dependente e duas ou mais variáveis independentes. As variáveis independentes são incluídas na análise de regressão com o objetivo de dar maior precisão a variável dependente. Cada variável independente é ponderada pelo processo da Análise de Regressão para garantir máxima precisão do resultado

No Brasil, um país de dimensões continentais e em franco crescimento demográfico e econômico, a previsão de demanda na área energética tem tradição em energia elétrica. O setor elétrico utiliza de métodos econométricos, séries históricas, construção de cenários e outros métodos quantitativos para espiar o futuro da demanda por energia. As possibilidades de fazer previsão são várias, mas, as técnicas de análise multivariadas proporcionam um avanço na qualidade da previsão, diferente das demais técnicas. A multivariada trabalha as variáveis simultaneamente, através de suas correlações.

¹ Charles Edward Spearman (1863 - 1945) foi um psicólogo inglês conhecido pelo seu trabalho na área da estatística, como um pioneiro da análise fatorial

² Que tem os característicos de uma téttrade, O termo téttrade, na harmonia tonal, é referente a qualquer acorde formado por quatro notas. À harmonia musical baseada em téttrade dá-se o nome de tetrafonia, sendo também comum chamá-la harmonia tetrádica.

Procurou-se estruturar algumas condições para que possa vislumbrar cenários possíveis de demanda de gasolina e óleo diesel em médio prazo no Brasil, através da Análise de Fatorial e Regressão Múltipla. Foi utilizada uma série de fontes secundárias para a busca dos dados que têm influência no consumo de gasolina e óleo diesel. Cada variável tem sua participação revelada através da Análise Fatorial mostrando sua influência no evento. Podemos intuir que variáveis como preço do etanol, aumento da frota de carros e caminhões interferem diretamente no consumo de gasolina e óleo diesel. Foram correlacionadas variáveis que interferem, direta ou indiretamente, no consumo de gasolina e óleo diesel de maneira a revelar, através de técnicas multivariadas, uma dimensão inerente às variáveis e, por conseguinte, possibilitar, através da Regressão Múltipla, estabelecer cenários de consumo em médio prazo. Pode-se dizer que a maior correlação entre as variáveis proporciona resultados mais identificados com a realidade, maior qualidade dos resultados.

A opção por utilizar os dois produtos, gasolina e óleo diesel, vem do fato de que são os principais produtos do refino de petróleo em termos de volume produzido em refinarias. Ao considerar os dois produtos para sondarmos a demanda futura, através da modelagem multivariada, assume-se que conseqüentemente estamos avaliando a pertinência e a intensidade dos investimentos no parque do refino de petróleo no Brasil.

Ao falar de investimento em refino, registrou-se uma série de fatores interligados, tais como o tipo de petróleo a ser refinado, o processo de refino, o transporte a ser utilizado para fazer com que o petróleo chegue a refinaria, tecnologias e outros. Um conjunto de ações e decisões que representa custos e visão estratégica afinada com um dos vetores (infraestrutura) de investimentos, e conseqüente desenvolvimento, que faz parte dos esforços do Brasil atual.

1.2. OBJETIVOS

1.2.1. OBJETIVO PRINCIPAL

O presente trabalho tem como objetivo fazer uma análise das correlações entre as principais variáveis que exercem influência no consumo de gasolina e óleo diesel no mercado brasileiro, através de técnicas de análise de interdependência e de dependência de várias variáveis, fazendo projeções para o consumo.

1.2.2. OBJETIVOS ESPECÍFICOS

Dentro do objetivo principal, os objetivos específicos se apresentam da seguinte forma:

- Apresentar um conjunto de variáveis do mercado brasileiro que exercem influência no consumo de gasolina e óleo diesel
- Calcular as correlações entre as variáveis utilizando o coeficiente de correlação de PEARSON³.
- Utilizar uma técnica de análise multivariada chamada Análise Fatorial que tem por objetivo descobrir e analisar a estrutura de um conjunto de variáveis inter-relacionadas. Com a Análise Fatorial, busca-se identificar um número relativamente pequeno de fatores comuns que podem ser utilizados para representar as variáveis inter-relacionadas.

O produto resultante da técnica de Análise Fatorial utilizado em uma técnica de dependência chamada Regressão Múltipla onde estabelecemos a predição do consumo de gasolina e óleo diesel para os próximos anos.

³Karl Pearson Londres, 1857 a 1936 – Professor de matemática e estatística na University College London.

CAPÍTULO 2 – MODELAGEM MULTIVARIADA

2.1. A ANÁLISE DE DADOS E MODELAGEM MULTIVARIADA

O mundo clama por conhecimento. Seja em qualquer área das atividades humanas o conhecimento se torna determinante. Nos mercados de bens e de serviços a produção de conhecimento se torna geralmente dramática, uma questão de sobrevivência num ambiente cada vez mais competitivo. Como se pode perceber não há falta de informação, mas escassez de conhecimento (HAIR, 2009).

Nas últimas décadas as informações disponíveis para tomadas de decisões cresceram de maneira vigorosa, pode-se dizer um crescimento exponencial. A tendência é continuar crescendo com o mesmo vigor no futuro. Tempos atrás muitas dessas informações se perdiam ou eram descartadas. Diferente do passado, hoje as informações são coletadas e armazenadas, ficando disponíveis para serem analisadas e utilizadas como suporte para tomadas de decisão. Em parte, essas informações exigem estatísticas simples para serem compreendidas, mas uma porção cada vez maior exige técnicas estatísticas multivariadas mais complexas para converter tais informações em conhecimentos (COOPER, 2003).

Os avanços tecnológicos ajudam a aplicar as técnicas multivariadas. Com o desenvolvimento de hardware e software, a análise multivariada passou a ser mais utilizada para avaliações de dados, tendências e comportamentos nas mais diferentes áreas do conhecimento (LUIS PAULO, 2009). O desenvolvimento tecnológico apresenta hoje um conjunto de computadores e programas que sem os quais a análise multivariada seria inimaginável (LUIS PAULO, 2009).

As técnicas multivariadas estão cada vez mais sendo utilizadas porque elas permitem que organizações produzam conhecimentos (HAIR, 2009). Quando se fala em Análise Multivariada está se referindo a todas as técnicas estatísticas que analisam simultaneamente múltiplas medidas sobre o objeto de investigação. Portanto, qualquer análise simultânea de mais do que duas variáveis pode ser considerada como multivariada. Importante registrar que para uma análise ser considerada verdadeiramente uma análise multivariada, todas as variáveis devem ser aleatórias e inter-relacionadas de tal maneira que seus efeitos não podem ser significativamente interpretados separadamente (HAIR, 2009).

Com origens nas estatísticas univariadas e bivariadas a análise multivariada introduz conceitos adicionais. Um conceito importante é a chamada variável estatística. Trata-se de um constructo da análise multivariada, uma combinação linear de variáveis originais com seus

devidos pesos. As variáveis originais são escolhidas pelo pesquisador e os pesos determinados pela técnica multivariada para atingir um objetivo específico (ANDERSON, 2007).

Uma variável estatística de n variáveis ponderadas pode ser enunciada matematicamente de acordo com a Equação 2.1:

$$Y = p_1X_1 + p_2X_2 + \dots + p_nX_n \quad (2.1)$$

Sendo, Y o valor da variável estatística, X as variáveis observadas e p é o peso determinado pela técnica multivariada.

Tabela 2.1. Categorias dos objetivos das técnicas estatísticas de análise multivariada

CATEGORIA	CLASSIFICAÇÃO	TÉCNICAS RELACIONADAS
1	Investigação da dependência entre as variáveis; Todas as variáveis são mutuamente independentes, ou uma ou mais variáveis são dependentes de outras	Análise Discriminante/Análise de Regressão/Correlação Canônica/Regressão Logística/Análise Conjunta/Manova
2	Predição; As relações entre as variáveis devem ser determinadas com o objetivo de prever o valor de uma ou mais variáveis com base nas observações de outras variáveis.	Análise Discriminante/Análise de Regressão/Análise de Regressão/Logística
3	Construção de hipóteses e testes; Hipóteses estatísticas específicas, formuladas em termos de parâmetros da população multivariada, são testadas. Isso pode ser feito para validar premissas ou para reforçar convicções prévias.	Modelagem de Equações Estruturais/Análise Fatorial Confirmatória
4	Redução dos dados ou simplificação estrutural; O fenômeno em estudo é representado de um modo tão simples quanto possível, sem sacrificar informações importantes.	Análise Fatorial Exploratória
5	Agrupamento de objetos ou variáveis; Grupos de objetivos ou variáveis “similares” são criados com base nas medidas características	Análise de Conglomerados/Análise de Correspondência/ Escalonamento Multidimensional

Como exemplo de objetivo da técnica multivariada aplicada em relação a variável estatística, em regressão múltipla, a variável estatística é determinada de modo a maximizar a correlação entre as variáveis independentes múltiplas e a única variável dependente. Em análise fatorial, variáveis estatísticas são formadas para melhor representar a estrutura subjacente a um conjunto de variáveis originais inter-relacionadas.

A escolha das técnicas e tipos de análises empregados nos trabalhos científicos deve ser determinada pelo problema de pesquisa. Nesse sentido, (JOHNSON, 1999) propõe uma classificação dos objetivos para atendimento do problema em cinco categorias, de acordo com a Tabela 2.1.

2.2. TÉCNICA DE INTERDEPENDÊNCIA – ANÁLISE FATORIAL

2.2.1. O QUE É ANÁLISE FATORIAL

O propósito maior da técnica de Análise Fatorial é definir a estrutura inerente entre as variáveis originais em análise. Trata-se de uma abordagem estatística que pode ser usada para analisar inter-relações entre um grande número de variáveis e explicar essas variáveis em termos de suas dimensões inerentes comuns, os chamados fatores (HAIR, 2009).

Quando se tem um conjunto de variáveis correlacionadas, a análise fatorial é um caminho para gerenciar essas variáveis. Quando se tem a percepção que permite compreender as relações entre as variáveis, então as dimensões inerentes, o fator extraído, tem significado para aquilo que elas coletivamente representam. Essas dimensões podem corresponder a conceitos que não podem ser descrito adequadamente por uma única variável. A técnica Análise Fatorial fornece um conjunto de variáveis que representam as variáveis originais, o que permite que elas sejam utilizadas em outras técnicas multivariadas (ANDERSON, 2003).

2.2.2. OBJETIVOS DA ANÁLISE FATORIAL

Como qualquer outro processo de pesquisa, a técnica de Análise Fatorial tem como ponto de partida determinar o problema da pesquisa. A técnica análise fatorial tem como propósito geral condensar as informações contidas em um conjunto de variáveis originais em um número menor de variáveis, as chamadas variáveis estatísticas (fatores) com perda mínima de informações. Ao fazer esta condensação, define os constructos fundamentais ou dimensões assumidas como inerentes às variáveis originais (HAIR, 2009).

2.2.3. ANÁLISE FATORIAL R OU Q

A chamada Análise Fatorial R tem como objetivo resumir as características de um conjunto de variáveis. De tal forma que analisa relações entre variáveis para identificar grupos

de variáveis que formam dimensões latentes, fatores, a partir de uma matriz de correlação das variáveis. Os fatores são formados para explicar um conjunto de variáveis, agrupando as variáveis originais pelo que representam coletivamente na expressão de um conceito (LUIZ PAULO, 2009).

A técnica de Análise Fatorial Q é especialmente aplicada a uma matriz de correlação que tem origem em questionários individuais e é baseada nas características dos mesmos. Esta técnica condensa grandes números de pessoas que respondera aos questionários em diferentes grupos. Tem como objetivo reduzir as informações, identificação de variáveis representativas a partir de um número maior de variáveis, criando um novo conjunto de variáveis para substituir o original (HAIR, 2009).

As duas saídas distintas, mas que guardam relação, da análise fatorial fornecem um resumo dos dados ou redução dos dados. No caso do resumo dos dados, a análise fatorial obtém resultados que refletem as dimensões inerentes ao conjunto original das variáveis originais. Estes fatores extraídos, quando interpretados e compreendidos descrevem os dados em número muito menor de variáveis, de conceitos, do que as variáveis originais. Em se tratando de redução, os fatores têm a função de simplesmente substituir as variáveis originais (ANDERSON, 2003).

2.2.4. ESCOLHA DAS VARIÁVEIS

A Análise Fatorial permite dois caminhos na análise dos dados, são duas saídas distintas, mas relacionadas, o resumo dos dados e a redução dos dados. No resumo dos dados o conceito fundamental é a definição da estrutura inerente ao conjunto de variáveis originais, onde variáveis individuais são agrupadas e a partir daí vistas não por aquilo que elas representam individualmente, mas por aquilo que elas representam coletivamente na expressão de um conceito. No resumo de dados a Análise Fatorial obtém dimensões inerentes que, quando interpretadas e compreendidas, descrevem os dados em um número muito menor de conceitos (HAIR, 2009).

Pode-se aplicar a Análise Fatorial simplesmente para a redução de dados que podem ser usadas para identificar variáveis representativas a partir de um conjunto maior de variáveis. Pode-se também aplicar a técnica promovendo a criação de um conjunto inteiramente novo de variáveis, muito menor, para substituir parcialmente ou completamente o conjunto original de variáveis. Em ambos os casos citados é importante registrar que o propósito, no caso de redução de dados, é manter a natureza e o caráter das variáveis originais.

Na técnica de Análise Fatorial sempre se deve considerar as bases conceituais das variáveis e sempre refletir sobre a pertinência ou não de uma determinada variável para os objetivos do evento em que se está trabalhando. Deve-se implicitamente especificar as dimensões em potencial que podem ser identificadas por meio do caráter e da natureza das variáveis originais. A análise fatorial sempre irá produzir fatores, o que pode levar a produção de fatores sem nenhum sentido prático. A inclusão indiscriminada de variáveis originais sem um “nexo causal” faz com que a aplicação da técnica produza resultados pobres. A qualidade e o significado dos fatores obtidos refletem as bases conceituais das variáveis incluídas na análise (ANDERSON, 2003)

2.2.5. O PRODUTO DA ANÁLISE FATORIAL APLICADO EM OUTRAS TÉCNICAS MULTIVARIADAS

Uma reflexão sob a perspectiva de resumo de dados mostra que a aplicação da técnica de análise fatorial proporciona uma boa compreensão sobre quais variáveis podem atuar juntas e quais variáveis terão de fato impactos na análise. A técnica fornece uma visão ampla das inter-relações entre as variáveis que permite obter uma estrutura adjacente de dados, tornando o produto da técnica um excelente ponto de partida para outras técnicas multivariadas. O produto, resultado da aplicação da técnica análise multivariada, mais particularmente o resumo de dados, podem ser diretamente incorporado em outras técnicas multivariadas (HAIR, 2009).

A análise fatorial, em resumo, fornece a base para a criação de um novo conjunto de variáveis que incorpora a natureza das variáveis originais em um número bem menor de novas variáveis representativas, escores fatoriais ou escalas múltiplas. Dessa maneira, problemas associados com grande número de variáveis ou altas correlações entre variáveis podem ser substancialmente reduzidos pela substituição de novas variáveis (HAIR, 2009). Estas novas variáveis, em número bem menor, realmente se tornam o excelente ponto de partida para outras técnicas multivariadas.

2.2.6. TIPOS DE VARIÁVEIS USADAS EM ANÁLISE FATORIAL

Uma importante restrição à técnica é que as variáveis possam permitir que sejam calculadas as correlações entre todas as variáveis. Neste trabalho são utilizadas as variáveis métricas, as quais são facilmente medidas suas correlações. As variáveis não métricas por outro lado apresentam mais dificuldades de definição de suas correlações. Se a pesquisa exigir a inclusão de uma variável não métrica, pode-se usar como saída a definição de variáveis

dicotômicas, codificada em 0 ou 1 para representarem as referidas variáveis, um método que pode auxiliar nas análises dos dados (ANDERSON, 2003).

Deve-se tentar minimizar o número de variáveis a ser incluído na aplicação da Análise Fatorial, mas deve-se manter um número razoável por fator. No caso de se estar trabalhando para promover o resumo das variáveis originais, deve-se certificar se as variáveis podem ser representadas pelos fatores proposto. A força da técnica reside em encontrar padrões entre grupos de variáveis (FERRANDO, 1993).

Outra questão importante na definição das variáveis é o tamanho da amostra. Pela prática de aplicação da técnica pode-se afirmar que, dificilmente, em uma análise fatorial com consistência deve apresentar menos de cinquenta observações. Como regra geral, o recomendado é ter pelo menos dez vezes mais observações do que o número de variáveis originais (HAIR, 2009).

Espera-se da técnica de Análise Fatorial uma visão sobre a estrutura dos dados, mas a estrutura revelada depende da decisão que se toma em respeito às variáveis incluídas na análise e o tamanho da amostra (HAIR, 2009).

2.2.7. SUPOSIÇÕES NA ANÁLISE FATORIAL

Na aplicação da análise fatorial fazem-se mais suposições conceituais do que estatísticas. Normalmente, deve-se estar sempre preocupado em atender as exigências estatísticas para qualquer técnica multivariada. Quando se aplica a técnica Análise Fatorial as preocupações se concentram muito mais no caráter e na composição das variáveis incluídas na análise do que em suas qualidades estatísticas. A correlação entre as variáveis e o fato de extrair fatores não garante relevância no resultado, mesmo que satisfaça as exigências estatísticas. Portanto, cabe ao pesquisador garantir que os padrões observados sejam conceitualmente válidos para a aplicação da técnica. Não é possível, por exemplo, misturar variáveis dependentes com variáveis independentes e extrair fatores, estes fatores extraídos não terão consistência nenhuma (FERRANDO, 1993).

Entre as questões estatísticas podem-se relacionar os desvios de normalidade, da homocedasticidade e da linearidade que se aplicam, porque elas diminuem as correlações observadas (ALBUQUERQUE, 2008).

2.2.8. ADEQUAÇÃO DA UTILIZAÇÃO DA ANÁLISE FATORIAL

Definidas as variáveis originais que atendam às exigências conceituais, o próximo passo é verificar se as variáveis são suficientemente correlacionadas uma com as outras para

garantir fatores representativos. Existem várias medidas empíricas para ajudar no diagnóstico da fatorabilidade da matriz de correlação (BARRETO, 2005).

A matriz de dados deve apresentar correlações suficientes para poder ser utilizada na técnica de Análise Fatorial. Duas situações podem ser verificadas na matriz de dados, uma em que as correlações entre as variáveis possam ser muito pequenas ou então que as correlações entre as variáveis são iguais. Nesses casos não é aconselhado aplicar a técnica. No caso de se ter correlações iguais, significa que não existe qualquer estrutura para agrupar as variáveis. A primeira forma de verificar as correlações é fazer uma inspeção visual na matriz de correlação. Se a inspeção visual da matriz de correlações não revelar um número substancial de valores superiores a 0,30 há fortes indícios de que a utilização da técnica não é adequada (LUIZ PAULO, 2009).

As correlações entre as variáveis podem ser também observadas analisando a chamada correlação parcial. A correlação parcial é aquela que não é explicada quando os efeitos de outras variáveis são levados em consideração. A correlação parcial deverá ser pequena quando existirem fatores verdadeiros na matriz de dados, isso mostra que a variável pode ser explicada pelas variáveis que compõem os fatores. Se as correlações parciais são altas, isto indica a ausência de fatores inerentes ao conjunto de variáveis (HAIR, 2009). Alguns softwares estatísticos, como o SPSS (*Statistical Package for the Social Sciences*) que é utilizado neste trabalho, fornece a matriz de correlação anti-imagem que apresenta os valores negativos das correlações parciais.

O teste Kaiser-Meyer-Olkin (KMO), de acordo com a Equação 2.2, é uma estatística que indica a proporção da variância dos dados e que pode ser considerada comum a todas as variáveis, ou seja, que pode ser atribuída a um fator comum, então: quanto mais próximo de 1 (unidade) melhor o resultado, ou seja, mais adequada é a amostra à aplicação da análise fatorial (LUIZ PAULO, 2009)

$$KMO = \frac{\sum \cdot \sum r^2}{\sum \cdot \sum r^2 + \sum \cdot \sum a^2} \quad (2.2)$$

r = coeficiente de correlação entre variáveis;

a = coeficiente de correlação parcial.

Outra maneira de determinar a pertinência de utilizar a análise fatorial é examinar a matriz de correlação inteira através do teste estatístico chamado Teste de Esfericidade de

Bartlett. Este teste indica ou não, a presença de correlações entre as variáveis. Possibilita avaliar a hipótese de que a matriz das correlações pode ser a matriz identidade. Se a matriz de correlação for igual à matriz identidade, significa que as intercorrelações entre as variáveis são iguais a 0, neste caso, deve-se reconsiderar a utilização da análise fatorial. Considera-se a hipótese nula como sendo a matriz de correlação uma matriz identidade. Se a hipótese nula não for rejeitada, significa que as variáveis não estão correlacionadas. Por outro lado, se a hipótese nula for rejeitada, haverá indícios de que existem correlações significativas entre as variáveis originais (ANDERSON, 2007).

A Medida de Adequação da Amostra (Measurement of Sampling Adequacy - MSA) para cada variável é outra medida para quantificar o grau de intercorrelações entre as variáveis, mostrando a pertinência da utilização da análise fatorial. A medida varia de 0 a 1, sendo 1 quando cada variável é perfeitamente prevista sem erros pelas outras variáveis. Esta medida é similar ao cálculo KMO, inclui apenas os coeficientes que se deseja analisar. A diagonal principal da matriz anti-imagem gerada pelo software SPSS fornece os valores de MSA (HAIR, 2009).

Importante observar que, em alguns casos a baixa correlação de determinada variável com as demais não necessariamente implica sua eliminação, uma vez que esta pode representar um fator isolado (LUIZ PAULO, 2009).

2.2.9. EXTRAÇÃO DOS FATORES

2.2.9.1. Divisão da Variância de uma Variável

Pode-se escolher entre dois métodos semelhantes para a extração dos fatores que representam as estruturas das variáveis originais (FERRANDO, 1993). O que determina a escolha do método é o objetivo para que se está aplicando a Análise Fatorial. Deve-se ter certa compreensão da variância para uma variável e como ela é dividida ou particionada.

A variância de uma variável é um valor - o quadrado do desvio padrão, que representa a quantia total de dispersão de valores para uma única variável em torno de sua média. Quando uma variável é correlacionada com outra, significa que ela compartilha variância com a outra variável, e essa quantia de partilha entre apenas duas variáveis é simplesmente a correlação ao quadrado (HAIR, 2009).

Na técnica de Análise Fatorial agrupam-se as variáveis por suas correlações. O fator é formado por um grupo de variáveis altamente correlacionadas entre si. Portanto, é importante entender o quanto da variância de uma variável é compartilhado com outras variáveis. A variância de uma variável pode ser particionada em três partes (HAIR, 2009).

Variância Comum é aquela variância em que uma variável que é compartilhada com todas as outras variáveis. A estimativa de uma variância de uma variável compartilhada entre as variáveis representadas por um fator é chamado de Cumunalidade;

Variância Específica também conhecida como variância única, é associada apenas a uma avaliável específica. Não pode ser explicada pelas correlações com outras variáveis;

Variância de Erro também não pode ser explicada por correlações com outras variáveis. É resultado da falta de confiabilidade no processo de coleta de dados, de erro de medidas ou de qualquer outro componente aleatório no fenômeno medido.

A Variância Total de uma variável é composta de Variância Comum, Única e de Erro, como segue na Figura 2.1.

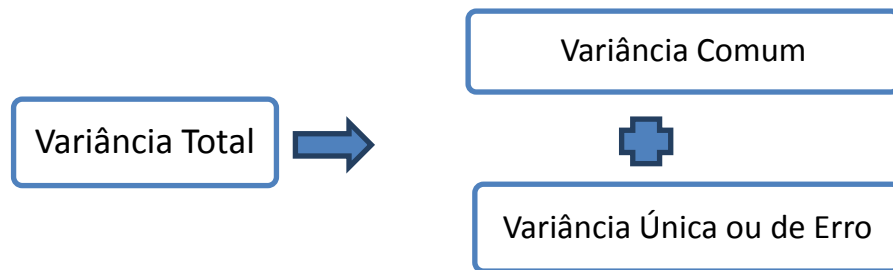


Figura 2.1. Relação das Variâncias.

2.2.9.2. Métodos de Extração de Fatores

Existem dois métodos principais que podem ser utilizados para a obtenção de fatores, Análise de Componentes Principais e Análise dos Fatores Comuns. A diferença entre os dois métodos é que o primeiro considera a variância total dos dados e o segundo os fatores são estimados com base na variância comum (HAIR, 2009).

A Análise de Componente Principal é utilizada quando o objetivo da aplicação da Análise Fatorial é resumir a maior parte das informações originais a um número mínimo de fatores. Este método considera a variância total e deriva fatores que contêm pequenas proporções de variância única ou de erros. Os primeiros fatores não contem variância única ou de erros o suficiente para distorcer a estrutura fatorial geral. A Análise dos Fatores Comuns é usada para identificar fatores ou dimensões latentes que refletem o que as variáveis originais têm em comum.

A Análise de Componentes Principais procura uma combinação linear das variáveis originais, de maneira a maximizar a variância explicada. Um conjunto de variáveis originais

altamente correlacionadas será combinado de modo a formar um fator que explique a maior quantidade de variância na amostra. O segundo fator terá a segunda maior quantidade de variância e não será correlacionado com o primeiro e, assim sucessivamente, os fatores vão sendo extraídos de forma que expliquem quantias cada vez menores de variância até que toda ela seja explicada. Portanto, o primeiro fator pode ser visto como o melhor resumo de relações lineares exibidas nas variáveis originais. O segundo fator é a segunda melhor combinação linear das variáveis originais, sujeita a restrição de que é ortogonal ao primeiro fator. Para ser ortogonal, o segundo fator deve ser extraído da variância remanescente depois que o primeiro fator foi obtido. Um conjunto de variáveis correlacionadas é transformado em um conjunto de variáveis não correlacionadas. Este produto da aplicação da análise fatorial pode ser utilizado como entradas em outras técnicas multivariadas (FERRANDO, 1993).

2.2.9.3. Escolha do Número de Fatores

Os primeiros fatores extraídos na análise fatorial explicam uma parcela substancial de da variância ao longo de todo o conjunto de variáveis originais. Espera-se, com a aplicação da análise fatorial, utilizar apenas um pequeno número de fatores que possa representar adequadamente o conjunto inteiro das variáveis originais. A questão que se coloca é quantos fatores devem ser extraídos? Ainda não foi desenvolvido um método que tenha uma base quantitativa exata para decidir o número de fatores que deva ser extraído (HAIR, 2009).

Alguns critérios de determinação de parada de excluir fatores têm sido utilizados, conforme a seguir.

A técnica chamada de “Critério da Raiz Latente” - é simples de aplicar na análise de componente principal. Escolhe-se o número de fatores a reter em função do número de valores próprios acima de um. Os valores próprios, também chamados de autovalores ou *eigenvalues*, é a soma em coluna das cargas fatoriais ao quadrado para um fator, também conhecida como raiz latente. Representa a quantia de variância explicada por um fator. A escolha de componentes que apresentam o autovalor maior que um vem do fato de que, no mínimo, o componente deve explicar a variância de uma variável utilizada no modelo, uma vez que se trabalha com variáveis padronizadas de média zero e variância igual a um (MARTINS, 2005).

Outro critério utilizado é o de “Porcentagem de Variância”. Trata-se de um critério que leva em consideração o acúmulo de variância total extraída por fatores sucessivos. O objetivo

é garantir significância para os fatores extraídos, garantindo que expliquem um montante de variância previamente especificado (LUIZ PAULO, 2009).

Critério Teste Scree - Na aplicação da análise fatorial utilizando a análise de componente principal, podemos observar no gráfico do teste Scree que os últimos fatores extraídos contêm tanta variância comum como a variância única. Mas, a variância única nos últimos fatores é substancialmente maior que a variância comum. O critério do Teste Scree é usado para identificar o número ótimo de fatores que podem ser extraídos antes que a quantidade da variância única comece a dominar a estrutura de variância comum (LUIZ PAULO, 2009).

O gráfico Scree é feito por meio da plotagem dos valores da raiz latente no eixo Y e o número de fatores no eixo X, de acordo com a ordem de extração e a forma da curva resultante é usada para avaliar o ponto de corte. Observa-se na Figura 2.2 que ao utilizar o teste scree podem-se considerar quatro fatores.

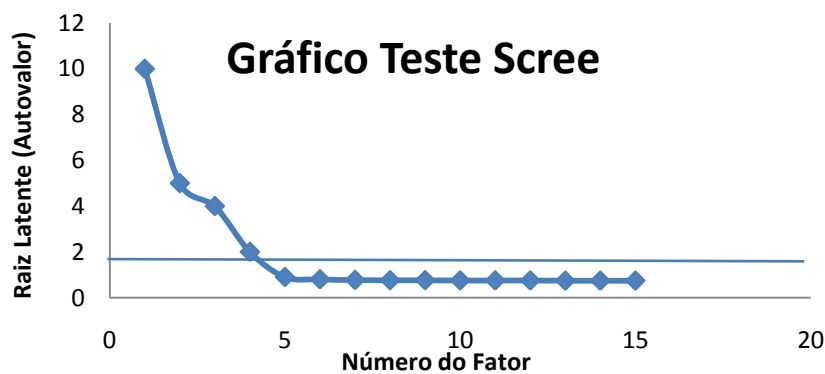


Figura 2.2. Exemplo de um Gráfico de Teste de Scree.

2.2.10. ROTAÇÃO DOS FATORES

Alguns autores afirmam que talvez a ferramenta mais importante na interpretação de fatores seja a Rotação Fatorial (LUIZ PAULO, 2009). O método consiste em promover uma rotação dos eixos de referência dos fatores em torno da origem. Em geral os fatores produzidos na fase da extração nem sempre são facilmente interpretados. O efeito final de rotacionar a matriz fatorial é distribuir a variância dos primeiros fatores para os últimos, com o objetivo de atingir um padrão fatorial mais simples e teoricamente mais significativo (HAIR, 2009). Segundo os autores (CORRAR, 2007), a rotação dos fatores é possível, pois as cargas fatoriais podem ser representadas como pontos entre eixos, sendo estes eixos o próprio fator, de acordo com a Figura 2.3.

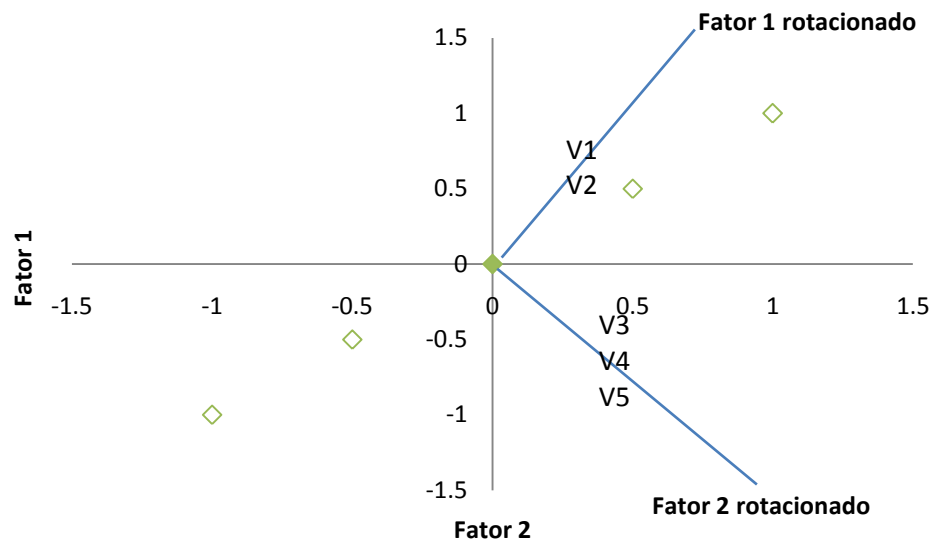


Figura 2.3. Exemplo de rotação de eixo.

Esses eixos podem ser girados sem alterar as distâncias entre os pontos. No entanto, as coordenadas do ponto em relação aos eixos são alteradas, ou seja, a carga fatorial que representa a relação entre o fator e a variável é alterada na rotação. A carga fatorial para a variável V2 no Fator 2 não rotacionado é determinado desenhando-se uma linha horizontalmente a partir da variável V2 até o eixo vertical do Fator 2. De modo similar, uma linha vertical a partir da variável 2 até o eixo horizontal do Fator 1 não rotacionado para determinar a carga da variável V2.

De acordo com a Figura 2.3, podem-se perceber dois agrupamentos de variáveis, variáveis V1 e V2 e o outro agrupamento são as variáveis V3, V4 e V5. No entanto, tal agrupamento de variáveis pode não ser tão obvio a partir das cargas não rotacionadas. A rotação acontece com os fatores mantidos em 90 graus, isto significa que os fatores são matematicamente independentes e a rotação é ortogonal. Pode-se observar que após rotacionar os eixos fatoriais, as variáveis V3, V4 e V5 tem cargas altas no Fator 1, e as variáveis V1 e V2 tem cargas elevadas no Fator 2.

Existem os métodos ortogonais e os oblíquos. Quando se utilizam os métodos ortogonais, preserva-se a condição com que os fatores não possuem correlação entre si. Os fatores ortogonais, resultado da rotação ortogonal, são interpretados a partir de suas cargas. A carga, também chamadas de carga fatorial, é a correlação entre as variáveis originais e os fatores, é a chave para o entendimento da natureza de um fator em particular. As cargas fatoriais ao quadrado indicam qual percentual de variância em uma variável original é explicada por um fator. Por outro lado, na rotação oblíqua, os fatores estão correlacionados,

para interpretar os fatores rotacionados deve ser considerada simultaneamente a correlação e a carga (HAIR, 2009).

Dos métodos rotacionais destacam-se o Varimax, Quartmax e o Equamax. O mais utilizado é o Varimax, que busca minimizar o número de variáveis que tem altas cargas em um fator, simplificando a interpretação dos fatores. Sobre o método Varimax, o autor afirma que é um método ortogonal onde, para cada componente principal existam apenas alguns pesos significativos e todos os outros sejam próximo de zero, isto é, o objetivo é maximizar a variação entre os pesos de cada componente principal, daí o nome Varimax (REIS, 2001).

A rotação ortogonal tem como objetivo extremar os valores das cargas, de modo que cada variável associe a apenas um fator. A rotação transforma a matriz inicial das cargas fatoriais em outra mais fácil de ser interpretada.

2.2.11. INTERPRETAÇÃO DE UMA MATRIZ FATORIAL

Nessa fase, a matriz de cargas fatoriais será interpretada com o objetivo de identificar a estrutura inerente às variáveis originais. O objetivo é interpretar e nomear os fatores por meio das cargas, para tanto, deve-se decidir quais cargas fatoriais valem a pena ser consideradas.

Quando se move do primeiro fator para os posteriores, a Variância Única e a Variância de Erro começam a surgir cada vez com mais força. O número de variáveis em análise é importante na decisão de quais cargas significantes no processo. À medida que o número de variáveis originais aumenta o nível aceitável para considerar uma carga significante diminui (HAIR, 2009).

Importante ressaltar que o valor da carga fatorial ao quadrado representa a quantia de Variância Total da variável explicada pelo fator. A comunalidade é a proporção da Variância de cada variável explicada pelo conjunto de fatores selecionados, sendo calculada pela somatória dos quadrados das cargas de cada fator na variável (HAIR, 2009). A Tabela 2.2 apresenta orientações para identificar as cargas fatoriais significantes, ao nível de 5% de significância, com base no tamanho da amostra (HAIR, 2009). Utilizando esta condição descrita na Tabela 2.2 têm-se as variáveis distintas para cada fator, com isso, procura-se uma correspondência com fundamentação conceitual ou as expectativas administrativas depositadas na análise para avaliar a significância prática.

Ao examinar a matriz fatorial de cargas, deve-se identificar as cargas significantes para cada variável. Começando com a primeira variável na matriz no primeiro fator e se mover horizontalmente da esquerda para a direita, procurando a carga mais alta daquela

variável em qualquer fator. Repete-se o procedimento para a segunda variável e assim sucessivamente até que todas as variáveis tenham sido revistas.

Quando todas as variáveis têm uma carga significativa em um determinado fator, aquele fator representa o total das variáveis originais. As variáveis com maior carga terão papel importante na escolha do nome que será dado para representar o fator em questão. Depois de examinar todas as variáveis significantes para um fator em particular, o desafio é definir um rótulo ou nome para o fator que expresse com maior precisão possível as variáveis com cargas no fator em questão. Este nome ou rótulo deve ser desenvolvido intuitivamente expressando as dimensões latentes de um fator em particular. Portanto, o processo de interpretação de fatores envolve julgamentos tanto objetivos quanto subjetivos.

Tabela 2.2. Identificação das cargas fatoriais significantes.

Carga Fatorial	Tamanho da Amostra
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

Fonte: Hair (2009)

2.2.12. USO DOS RESULTADOS DA ANÁLISE FATORIAL

A utilização do produto da análise fatorial depende do objetivo da aplicação da técnica. Dependendo deste objetivo podem-se tomar dois caminhos: parar com a interpretação ou partir para um dos métodos de redução de dados.

Quando o objetivo é somente identificar combinações lógicas entendendo melhor as inter-relações entre as variáveis, a interpretação fatorial é o suficiente.

Quando o objetivo está em identificar variáveis adequadas para aplicação em outras técnicas estatísticas, então alguma forma de redução de dados será necessariamente empregada. Uma opção é selecionar a variável com maior carga fatorial em um determinado fator e, a variável, passa a ser representante da dimensão fatorial. A outra opção é substituir o

conjunto de variáveis originais por um conjunto menor criado a partir de escalas múltiplas ou escores fatoriais (HAIR, 2009).

As escalas múltiplas são formadas pela combinação de diversas variáveis formando uma única medida composta. Todas as variáveis com cargas elevadas em um determinado fator são combinadas, normalmente o valor médio das variáveis é usado como uma variável substituta. Um dos benefícios da escala múltipla é representar múltiplos aspectos de um conceito com uma medida única (HAIR, 2009).

O chamado escore fatorial é mais uma opção para criar um conjunto menor de variáveis para substituir o conjunto original. Os escores fatoriais também são medidas compostas de cada fator. O que diferencia o escore fatorial de uma escala múltipla é que o escore fatorial é calculado computando todas as cargas fatoriais, enquanto que a escala múltipla é calculada considerando somente variáveis selecionadas (HAIR, 2009).

2.3. TÉCNICA DE DEPENDÊNCIA – ANÁLISE DE REGRESSÃO MÚLTIPLA

2.3.1. O QUE É ANÁLISE DE REGRESSÃO MÚLTIPLA

A Análise de Regressão Múltipla é a técnica de dependência mais utilizada e versátil em relação às demais técnicas de dependência (MC CLARE, 2009). Trata-se de uma técnica estatística utilizada para analisar a relação entre uma única variável dependente e outras variáveis independentes. Utilizam-se as variáveis independentes, as quais os valores são conhecidos, para prever o valor da variável dependente selecionada (MC CLARE, 2009).

A regressão linear múltipla apresenta a mesma lógica apresentada pela regressão simples, porém com a inclusão de mais de uma variável explicativa no modelo.

2.3.2. REGRESSÃO LINEAR SIMPLES

O objetivo da análise de regressão é partindo de uma ou mais variáveis independentes prever uma única variável dependente. Para apresentarmos a técnica de Análise de Regressão múltipla veremos primeiro a regressão simples, com uma variável independente, para facilitar a apresentação da técnica multivariada.

A primeira ação é o estabelecimento ou cálculo de um ponto de referência da variável dependente (LUIZ PAULO, 2009). O cálculo do ponto de referência serve para compararmos a capacidade preditiva do modelo de regressão que está sendo aplicado. O ponto de referência deve representar a melhor previsão da variável dependente sem a participação das variáveis independentes (LUIZ PAULO, 2009). Na técnica de regressão, o ponto preditor de referência comumente utilizado é a média simples da variável dependente. Como este ponto, a média,

não irá prever perfeitamente cada valor da variável dependente, examina-se os erros da variável dependente a partir da média e com isso está avaliando também a precisão de qualquer previsão (LUIZ PAULO, 2009).

Uma forma útil de analisar os erros é calculando o quadrado de cada erro e somando o resultado, este total resultante, chamado de Soma de Quadrados dos Erros (SSE), fornece uma medida de precisão de previsão. O objetivo é obter a menor soma possível de “erros quadrados” como medida de precisão da previsão (HAIR, 2009).

Quando não se utiliza as variáveis independentes o melhor preditor é a média da variável dependente (LUIZ PAULO, 2009). Avançando na análise, o objetivo é incorporar uma ou mais variáveis independentes que possam melhorar este valor de referência. Quando inclui somente uma variável na análise chamamos de regressão simples. O objetivo na regressão simples é encontrar uma variável independente que melhore a precisão de referência (MC CLARE, 2009).

Podemos ter inúmeras variáveis para ser escolhida como variáveis independentes no modelo, mas, qual variável devemos escolher para obter a melhor previsão. Para escolha da variável contamos com o chamado coeficiente de correlação. Este coeficiente representa o conceito de associação e é fundamental na análise de regressão, representando a relação entre duas variáveis. Duas variáveis são correlacionadas quando a mudança de uma é associada a mudança na outra. Se acharmos uma variável que tenha comportamento similar, possuem correlação com a variável dependente, podemos melhorar nossa previsão feita a princípio com o uso somente da média. Quanto maior a correlação entre a variável dependente com a variável independente, melhor será a previsão (MC CLARE, 2009).

Além da escolha da variável preditora mais adequada é possível melhorar a previsão usando uma constante conhecida como intercepto. Esta constante representa o valor na variável dependente quando todas as variáveis independentes têm valor nulo. O nome intercepto decorre que na representação gráfica ele representa o ponto no qual a reta que descreve o modelo de regressão cruza o eixo do Y (MC CLARE, 2009).

2.3.3. EQUAÇÃO DE REGRESSÃO SIMPLES

Como já visto, selecionamos a melhor variável independente com base no coeficiente de correlação, quanto maior o coeficiente de correlação, maior a precisão preditiva. Na equação representamos o intercepto como b_0 . A quantia de mudança que a variável independente imprime na variável dependente é representada pelo termo b_1 , conhecido como coeficiente de regressão. Utilizando um procedimento matemático chamado “mínimos

quadrados” estimamos os valores de b_0 e de b_1 , de tal maneira que a soma dos erros quadrados de previsão seja minimizada. Chamamos de erros de previsão a diferença entre os valores reais e o previsto da variável dependente, também chamado de resíduo e (MC CLARE, 2009).

$$Y = b_0 + b_1X_1 + e \quad (2.2)$$

b_0 - o intercepto;

b_1 - o coeficiente de regressão;

e - o resíduo.

O coeficiente de regressão indica a extensão na qual a variável independente se associa com a variável dependente (MC CLARE, 2009).

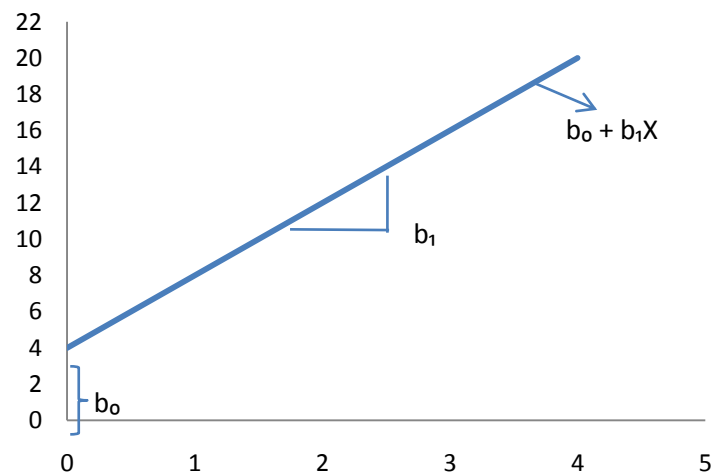


Figura 2.4. Reta e coeficientes de uma regressão linear.

Os cálculos dos parâmetros b_0 e b_1 são feitos por meio das seguintes expressões:

$$b_0 = \bar{Y} - b_1\bar{X} \quad (2.3)$$

$$b_1 = \frac{cov(X,Y)}{var(X)} \quad (2.4)$$

Sendo \bar{Y} e \bar{X} as médias aritméticas das variáveis Y dependente e X independente (LUIZ PAULO, 2009).

2.3.4. PRECISÃO DA PREVISÃO

Da mesma maneira que temos representado uma medida de erros de previsão na soma de quadrados de erros (SSe), podemos expressar uma medida de sucesso de previsão através da chamada soma de quadrados da regressão (SSr). Somando as duas teremos a soma total de quadrados (SSt), da nossa previsão de referencia (HAIR, 2009).

$$\Sigma(Y - \bar{Y})^2 = \Sigma(Y - \hat{Y})^2 + \Sigma(\hat{Y} - \bar{Y})^2 \quad (2.5)$$

$$SSt = SSe + SSr \quad (2.6)$$

A soma total de quadrados é igual à soma de erros quadrados mais a soma de quadrados de regressão.

\bar{Y} = média de todas as observações;

Y = valor da observação individual;

\hat{Y} = valor previsto da regressão.

Dividindo a soma dos quadrados de regressão pela soma total dos quadrados nos permite estimar o quão bem a variável estatística de regressão descreve a variável dependente (HAIR, 2009). Como vimos nossa previsão de referência é a média da variável dependente, e que, é a melhor estimativa disponível sem a presença de nenhuma variável independente. Podemos quantificar a melhoria da precisão preditiva quando a equação de regressão é usada no lugar da previsão de referencia. Para que possamos melhorar a previsão além da referência utilizamos o coeficiente de determinação R^2 para expressar o nível de precisão de previsão. Este coeficiente é calculado pela razão entre a soma de quadrados da regressão pela soma total de quadrados (HAIR, 2009).

$$R^2 = \frac{SSr}{SSt} \quad (2.7)$$

Caso o modelo de regressão preveja perfeitamente a variável dependente, $R^2 = 1$. Se o modelo não forneceu valores melhores do que a previsão de referência, $R^2 = 0$. Portanto, o R^2 é a proporção da variação amostral da variável dependente explicada pelas variáveis independentes, e é utilizado como uma medida do grau de ajuste (WOOLDRIGE, 2003).

2.3.5. REGRESSÃO LINEAR MULTIPLA

A regressão linear múltipla apresenta a mesma lógica da apresentada para a regressão linear simples, agora com a inclusão de mais de uma variável independente no modelo (MC CLARE, 2009). Importante registrar que a utilização de muitas variáveis explicativas depende de experiência e do bom senso de quem está aplicando o método (LUIZ PAULO, 2009). O objetivo é expandir o modelo de regressão simples acrescentando variáveis independentes que tenham o maior poder preditivo adicionais. O potencial de uma variável independente adicional de melhorar a previsão da variável dependente está relacionado não apenas com a sua correlação com a variável dependente, mas também com a correlação da variável independente adicional com a(s) variável(eis) independente(s) já incluídas na equação de regressão (HAIR, 2009). A correlação entre duas variáveis independentes é a chamada colinearidade e a multicolinearidade é a correlação entre três ou mais variáveis independentes (HAIR, 2009).

A correlação entre as variáveis independentes pode ter um forte impacto sobre o modelo de regressão (MC CLARE, 2009). A multicolinearidade reduz o poder preditivo de qualquer variável independente quando esta variável está sendo incluída no modelo e ela tem forte correlação com outras variáveis do modelo (HAIR, 2009). Aumentando a colinearidade, o percentual de previsão compartilhada entre as variáveis independentes aumenta. Como esta previsão compartilhada só pode ser considerada apenas uma vez, a previsão aumenta mais vagarosamente. Portanto, devemos incluir no modelo variáveis independentes com baixa multicolinearidade com as outras variáveis independentes, mas que tenha alta correlação com a variável dependente (HAIR, 2009).

Utilizamos o conceito *ceterisparibus* (mantidas as demais condições constantes) na análise da regressão múltipla. A interpretação de cada variável independente a ser incluída no modelo deverá ser feita de forma isolada. Um modelo, por exemplo, que apresente duas variáveis independentes X_1 e X_2 , os respectivos coeficientes serão analisados de forma a considerar todos os outros fatores constantes (LUIZ PAULO, 2009).

De forma análoga, a regressão simples o b_0 é calculado da seguinte forma:

$$b_0 = \bar{Y}_i - b_1\bar{X}_1 \dots - b_n\bar{X}_n \quad (2.8)$$

$$b_1 = \frac{cov(X_1, Y)Var(X_2) - cov(X_2, Y)cov(X_1, X_2)}{var(X_1)Var(X_2) - (cov(X_1, X_2))^2} \quad (2.9)$$

Para mais de duas variáveis, basta repetir o procedimento (LUIZ PAULO, 2009).

2.3.6. SUPOSIÇÕES EM ANÁLISE DE REGRESSÃO MÚLTIPLA

As técnicas estatísticas multivariadas são baseadas em um conjunto de suposições respondendo as exigências teóricas da estatística inerente (MC CLARE, 2009). Elencamos algumas suposições com alto potencial de afetar todas as técnicas estatísticas univariadas e multivariadas.

A primeira suposição é a normalidade. Trata-se da suposição mais fundamental da análise multivariada, se refere à forma de distribuição de dados para a variável métrica individual e sua correspondência com a chamada distribuição normal (MAC CLARE, 2009). A normalidade na distribuição de uma variável pode ser verificada através de gráfico, mas também pode ser verificada através de testes estatísticos. Os dois testes mais comuns são o teste Shapiro-Wilks e o teste Kolmogorov-Sminov (HAIR, 2009). Os referidos testes calculam o nível de significância para as diferenças em relação a uma distribuição normal.

A segunda suposição é a homocedasticidade. Na regressão linear múltipla, a suposição de homocedasticidade remete a suposição de que as variáveis dependentes exibem níveis iguais de variância ao longo do domínio das variáveis preditoras (MC CLARE, 2009). Como na regressão estabelecemos uma relação de dependência entre uma variável dependente e variáveis independentes, esta explicação de dependência não deve se limitar a um determinado domínio das variáveis independentes (ANDERSON, 2007). Para que a explicação da relação de dependência seja devidamente capturada, a dispersão dos valores da variável dependente deve ser relativamente semelhante em cada valor da variável preditora (HAIR, 2009).

Por último registramos a suposição de que as correlações apresentam associações lineares entre variáveis, esta, portanto é uma suposição de fundamental importância para aplicação da análise de regressão e de todas as técnicas baseadas na correlação entre variáveis (HAIR, 2009).

CAPÍTULO 3 – SELEÇÃO DAS VARIÁVEIS

3.1. SELEÇÃO DAS VARIÁVEIS

Na construção do banco de dados e a definição das variáveis, levou-se em consideração a pertinência e a adequação, de cada variável, para a aplicação da Análise Fatorial. Ao escolher o conjunto de variáveis que serão submetidas à Análise Fatorial, implicitamente o pesquisador especifica as dimensões potenciais que podem ser identificadas por meio do caráter e da natureza das variáveis (HAIR, 2009).

As variáveis definidas para aplicação da Análise Fatorial neste estudo foram fruto de uma avaliação quanto ao caráter e natureza das mesmas, além disso, observou-se a facilidade de acesso aos dados bem como sua confiabilidade. Utilizaram-se fontes secundárias e de reconhecida legitimidade por parte da sociedade brasileira. São valores registrados no período de dez anos que vai do ano de 2002 ao ano de 2011 considerando registros mês a mês, totalizando cento e vinte registros para cada variável.

A seguir é apresentado o conjunto de variáveis com seus valores conforme definido (Tabela 3.1).

Tabela 3.1. Resumo das variáveis a serem usadas.

01	Preço da Gasolina
02	Preço do Óleo Diesel
03	Preço do Etanol
04	Aumento da Frota de Carros
05	Aumento da Frota de Caminhões e Ônibus
06	Crescimento do Produto Interno Brasileiro PIB
07	Crescimento de Recursos Para Financiamento de Automóveis
08	Crescimento da População Ocupada
09	Produção de Gasolina
10	Produção de Diesel
11	Consumo de Gasolina
12	Consumo de Diesel
13	IPI Para Carros Populares até 1000 cc
14	IPI Para Carros de 1000 até 2000 cc

3.2. PREÇO DA GASOLINA

O preço da gasolina no Brasil é livre nas bombas dos postos de gasolina e as distribuidoras de combustível são legalmente impedidas de exercer qualquer influência sobre

eles. O mercado da gasolina no Brasil hoje é regulamentado pela Agência Nacional do Petróleo (ANP) e pela Lei Federal 9.478/97 (Lei do Petróleo). Esta lei flexibilizou o monopólio do setor de petróleo e gás natural, até então exercido pela Petrobras, tornando aberto o mercado de combustíveis no País. Dessa forma, desde janeiro de 2002 as importações de combustíveis foram liberadas e o preço passou a ser definido pelo próprio mercado. O preço final ao consumidor varia em função de múltiplos fatores como: carga tributária (municipal, estadual, federal), concorrência com outros postos na mesma região e a estrutura de custos de cada posto (encargos trabalhistas, frete, volume movimentado, margem de lucro, por exemplo). O preço do litro da gasolina, sem adição de etanol, vendida pela Petrobras é de R\$ 1,054 vendida às distribuidoras na porta das refinarias. Esse valor remunera em seus custos de produção, refino, logística e margem de lucro. Os preços de gasolina utilizados neste estudo tem sua origem em séries históricas produzidas pela Agência Nacional do Petróleo (ANP) conforme a Tabela 3.2. São preços registrados todos os meses, uma média dos preços praticados em todo Brasil direto das bombas dos postos de gasolina.

Tabela 3.2. Preço da gasolina médio entre 2002 e 2011⁵.

Ano	jan	fev	março	abril	maio	jun	jul	agos	set	out	nov	dez
2002	1,588	1,509	1,574	1,713	1,722	1,711	1,767	1,751	1,742	1,761	1,978	2,001
2003	2,160	2,223	2,215	2,195	2,113	2,026	1,971	1,975	2,003	1,997	1,993	1,998
2004	2,007	2,003	1,981	1,972	2,086	2,157	2,203	2,226	2,232	2,262	2,214	2,303
2005	2,296	2,285	2,291	2,302	2,283	2,256	2,267	2,273	2,401	2,470	2,478	2,483
2006	2,511	2,523	2,587	2,595	2,582	2,545	2,557	2,559	2,554	2,547	2,538	2,531
2007	2,523	2,516	2,518	2,532	2,540	2,527	2,507	2,487	2,471	2,475	2,487	2,509
2008	2,505	2,516	2,518	2,532	2,540	2,527	2,507	2,487	2,471	2,475	2,487	2,509
2009	2,516	2,518	2,514	2,500	2,487	2,488	2,492	2,498	2,488	2,528	2,551	2,556
2010	2,586	2,611	2,578	2,555	2,550	2,534	2,534	2,542	2,544	2,572	2,589	2,602
2011	2,612	2,622	2,670	2,824	2,842	2,738	2,735	2,736	2,742	2,750	2,746	2,750

A formação do preço da gasolina, considerando a composição de 80% (oitenta por cento) de gasolina A e 20% (vinte por cento) de etanol anidro: realização da Petrobras 34% (trinta e quatro por cento), CIDE, PIS/PASEP e COFINS 11% (onze por cento), ICMS 28% (vinte e oito por cento), Custo do Etanol Anidro 10% (dez por cento) e distribuição e revenda 17% (dezessete por cento)⁶.

⁴ Dado fornecido por - <http://fatosedados.blogspotpetrobras.com.br>, consultado em agosto 2012

⁵ Fonte - http://www.anp.gov.br/preco/prc/historico/Resumo_Mensal_Combustiveis.asp

⁶ Fonte - http://www.anp.gov.br/preco/prc/historico/Resumo_Mensal_Combustiveis.asp

3.3. PREÇO DO DIESEL

Desde janeiro de 2002, vigora no Brasil o regime de liberdade de preços. Não há qualquer tipo de tabelamento, valores máximos e mínimos, nem necessidade de autorização prévia da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) para reajustes dos preços dos combustíveis. Apesar da liberalização do mercado, da eliminação dos controles formais de preços e da completa abertura às importações, ocorrida em 2002, a Petrobrás ainda exerce, na prática, o monopólio do transporte e do refino do óleo diesel no mercado brasileiro. O governo é detentor da maioria das ações ordinárias da Petrobrás e, por isso, exerce grande influência na política de preços da empresa. Essa política tem grande impacto social e econômico, especialmente no caso do óleo diesel. O preço do óleo diesel, composição de 95% (noventa e cinco por cento) de diesel e 5% (cinco por cento de biodiesel), é formado da seguinte maneira: 56% (cinquenta e seis por cento) realização da Petrobras, CIDE, PIS/PASEP e COFINS 9% (nove por cento), ICMS 14% (quatorze por cento), custo do biodiesel 6% (seis por cento), distribuição e revenda 15% (quinze por cento)⁷.

Tabela 3.3. Preço médio do diesel entre 2002 e 2011⁸.

Ano	jan	fev	març	abril	maio	jun	jul	agos	set	out	nov	dez
2002	0,884	0,878	0,889	0,958	0,976	0,988	1,064	1,068	1,066	1,068	1,264	1,392
2003	1,527	1,539	1,537	1,529	1,456	1,429	1,416	1,408	1,401	1,395	1,391	1,390
2004	1,393	1,392	1,390	1,387	1,440	1,494	1,547	1,548	1,548	1,580	1,591	1,692
2005	1,690	1,687	1,684	1,683	1,697	1,695	1,694	1,694	1,830	1,885	1,888	1,888
2006	1,889	1,889	1,889	1,889	1,886	1,883	1,883	1,882	1,880	1,879	1,879	1,878
2007	1,878	1,877	1,878	1,879	1,877	1,875	1,873	1,875	1,873	1,872	1,876	1,880
2008	1,891	1,877	1,878	1,879	1,877	1,875	1,873	1,875	1,873	1,872	1,876	1,880
2009	2,125	2,126	2,125	2,123	2,122	2,078	2,016	2,009	1,998	2,000	2,002	1,999
2010	2,005	2,008	2,007	2,005	2,003	2,002	2,002	1,999	1,998	2,000	2,000	2,000
2011	2,009	2,016	2,025	2,032	2,030	2,026	2,028	2,027	2,027	2,029	2,030	2,032

3.4. PREÇO DO ETANOL

A introdução da tecnologia flex-fuel⁹ teve como consequência imediata um expressivo aumento do consumo de etanol no Brasil, tornando-o principal concorrente da gasolina nos postos de combustíveis. Apesar disso, diferentemente da gasolina, a oferta de etanol sofre

⁷ Fonte - <http://www.petrobras.com.br/pt/produtos/composicao-de-precos>

⁸ Fonte - http://www.anp.gov.br/preco/prc/historico/Resumo_Mensal_Combustiveis.asp

⁹ motor de combustão interna a quatro tempos (Ciclo Otto) que tem a capacidade de ser reabastecido e funcionar com mais de um tipo de combustível, misturados no mesmo tanque e queimados na câmara de combustão simultaneamente.

com a sazonalidade da safra de cana de açúcar, mostrando-se insuficiente para atender a demanda, apresentando altos preços nos períodos de entressafra. Os preços de etanol combustível ao consumidor são menores quanto mais próximos das regiões produtoras e maiores conforme se distanciam, até se tornarem desvantajosos para o consumidor de alguns estados na medida em que a paridade de preços frente à gasolina supera o limite técnico de 70%. O álcool que é misturado diretamente na gasolina a uma proporção de 20% (vinte por cento) é o chamado álcool anidro que é composto por 99,5% de álcool puro e 0,5% de água. Este tipo de álcool é menos poluente e, se for adicionado na proporção correta, não afeta o desempenho dos motores. O etanol, também chamado álcool etílico, é hidratado com 94% de etanol e 6% de água.

Tabela 3.4. Preço ao consumidor do etanol preço praticado no Brasil- unidade R\$/L¹⁰.

Ano	jan	fev	março	abril	maio	jun	jul	agos	set	out	nov	dez
2002	1,019	1,010	1,010	1,010	1,002	0,991	0,940	0,912	0,935	1,014	1,298	1,313
2003	1,386	1,569	1,586	1,562	1,506	1,348	1,203	1,208	1,233	1,201	1,176	1,191
2004	1,205	1,170	1,055	1,032	1,289	1,359	1,381	1,462	1,487	1,534	1,445	1,478
2005	1,460	1,448	1,444	1,451	1,354	1,210	1,256	1,271	1,308	1,434	1,450	1,536
2006	1,693	1,732	1,950	1,943	1,710	1,555	1,571	1,576	1,515	1,466	1,448	1,448
2007	1,544	1,545	1,540	1,630	1,626	1,463	1,351	1,305	1,279	1,268	1,360	1,467
2008	1,464	1,545	1,540	1,630	1,626	1,463	1,351	1,305	1,279	1,268	1,360	1,467
2009	1,477	1,499	1,478	1,425	1,388	1,340	1,374	1,400	1,447	1,616	1,675	1,703
2010	1,890	1,943	1,763	1,629	1,545	1,475	1,501	1,549	1,566	1,673	1,721	1,778
2011	1,835	1,872	2,083	2,301	2,032	1,859	1,932	1,946	2,006	2,005	2,030	2,055

3.5. AUMENTO DA FROTA DE CARROS

O levantamento dos números referentes ao aumento da frota brasileira de carros foi dividido em: automóveis (compreende automóveis e comerciais leves) e pesados (caminhões e ônibus). O crescimento da frota de veículos novos vem sendo puxada, nos últimos anos, pela chamada nova classe média (A Secretaria de Assuntos Estratégicos, SAE, do Governo Federal define nova classe média, ou classe C, como indivíduos que vivem em famílias com renda per capita entre R\$ 291 e R\$ 1.019). A nova classe média brasileira, conseqüentemente, está sendo responsável pela expansão no consumo de combustíveis, segundo o relatório anual divulgado pela ANFAVEA. Tivemos no Brasil um aumento de consumo de combustível devido ao crescimento da frota de veículos no Brasil. Nossa principal demanda é a do

¹⁰ Fonte - http://www.anp.gov.br/preco/prc/historico/Resumo_Mensal_Combustiveis.asp

consumo dos automóveis, o diesel está mais vinculado ao Produto Interno Bruto [PIB], porque 86% do transporte no Brasil são feitos por caminhões¹¹.

Tabela 3.5. Número de carros e caminhões licenciados mês a mês entre 2002 e 2011¹².

2011	janeiro	fevere	março	Abril	maio	junho	julho	agosto	setem	outu	novem	dezem
Automóveis	202.022	205.056	262.523	250.323	226.035	217.896	220.900	234.591	214.891	193.102	222.972	235.423
Caminhão	12.325	10.825	12.562	16.252	17.753	17.130	17.946	19.153	17.719	16.436	15.921	18.758
2010												
Automóveis	201.701	211.348	337.346	261.878	235.753	247.480	285.210	296.594	291.434	287.575	311.397	361.230
Caminhão	11.611	9.609	16.392	15.965	15.334	15.278	17.122	16.189	15.623	15.597	17.076	20.322
2009												
Automóveis	189.712	191.335	260.924	224.379	237.378	289.780	273.581	247.503	296.652	281.270	238.413	277.815
Caminhão	7.742	8.031	10.518	10.011	9.600	10.377	11.835	10.626	12.066	13.196	13.285	15.211
2008												
Automóveis	205.366	191.132	220.873	247.944	229.945	242.836	272.924	231.030	254.125	224.712	166.250	183.854
Caminhão	9.593	9.712	11.274	13.302	12.065	13.169	15.213	13.741	14.561	14.524	11.573	10.632
2007												
Automóveis	145.398	139.551	183.692	170.005	200.197	188.694	206.264	223.828	194.064	232.503	225.774	231.161
Caminhão	7.533	7.214	9.758	9.315	10.948	10.070	11.101	11.428	9.962	11.948	11.241	11.079
2006												
Automóveis	125.747	121.429	148.159	124.421	155.527	140.857	157.962	169.923	151.179	166.776	173.996	195.927
Caminhão	7.143	6.452	8.629	6.729	8.605	7.527	7.845	8.585	8.189	8.479	8.736	8.916
2005												
Automóveis	98.001	112.166	146.016	136.414	136.545	137.377	132.880	146.287	138.689	131.983	147.548	151.679
Caminhão	7.297	8.405	9.037	8.707	8.895	8.764	7.339	8.546	7.923	6.947	7.043	7.484
2004												
Automóveis	111.050	120.343	131.570	116.206	115.036	120.953	123.153	128.563	135.045	124.393	134.863	152.444
Caminhão	6384	7722	8424	8522	8329	8396	8701	9061	9178	8305	8578	8135
2003												
Automóveis	97.602	116.095	98.266	211.756	99.447	90.497	97.110	97.356	112.495	127.662	129.929	125.388
Caminhão	6.276	7.378	6.842	6.175	7.276	6.416	6.947	6.702	7.783	7.726	6.748	7.464
2002												
Automóveis	104.282	104.332	124.899	127.137	110.824	104.592	106.108	122.726	120.330	140.954	117.369	121399
Caminhão	5774	6674	8039	7967	7371	6182	5739	6527	7436	7767	6244	5108

3.6. CRESCIMENTO DA POPULAÇÃO OCUPADA

Pessoas ocupadas são aquelas que, num determinado período de referência, trabalhou ou tinham trabalho, mas não trabalharam (por exemplo, pessoas em férias). No nosso caso o período de referência será mês a mês no intervalo de 2002 até 2011 As pessoas ocupadas são classificadas em: EMPREGADOS - aquelas pessoas que trabalham para um empregador ou mais, cumprindo uma jornada de trabalho, recebendo em contrapartida uma remuneração em dinheiro ou outra forma de pagamento (moradia, alimentação, vestuário, etc.). Incluem-se, entre as pessoas empregadas, aquelas que prestam serviço militar obrigatório e os clérigos. Os empregados são classificados segundo a existência ou não de carteira de trabalho assinada. CONTA PRÓPRIA - aquelas pessoas que exploram uma atividade econômica ou exercem uma profissão ou ofício, sem empregados. EMPREGADORES - aquelas pessoas que

¹¹ <http://www.anfavea.com.br/tabelas.html>

¹² <http://www.anfavea.com.br/tabelas/autoveiculos/licenciamento.xls>

exploram uma atividade econômica ou exercem uma profissão ou ofício, com auxílio de um ou mais empregados. No levantamento utilizou-se as pesquisas feitas pelo Instituto Brasileiro de Pesquisas (IBGE), em seis regiões metropolitanas: Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre.

Tabela 3.6. Número da população ocupada (PO) nas regiões metropolitanas de Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre em (1.000)

Ano	jan	fev	mar	abr	mai	jun	jul	agos	set	out	nov	dez
2002	17.358	17.379	17.598	17.564	17.654	17.621	17.750	17.828	17.802	17.802	17.686	18.080
2003	18.242	18.209	18.227	18.337	18.324	18.402	18.420	18.480	18.704	18.633	18.845	18.902
2004	18.500	18.500	18.630	18.740	18.902	18.977	19.148	19.220	19.415	19.422	19.523	19.501
2005	19.490	19.455	19.580	19.590	19.825	19.750	19.800	19.890	20.070	20.080	20.030	20.150
2006	20.040	19.920	19.950	19.940	20.030	20.120	20.240	20.470	20.708	20.710	20.705	20.790
2007	20.495	20.450	20.610	20.550	20.410	20.780	20.809	21.010	21.280	21.290	21.370	21.380
2008	21.340	21.310	21.370	21.390	21.490	21.720	21.730	21.880	22.030	22.210	22.190	22.180
2009	21.210	20.910	21.020	20.930	21.010	21.170	21.360	21.420	21.510	21.520	21.603	21.815
2010	21.603	21.660	21.740	21.820	21.870	21.870	22.020	22.130	22.280	22.340	22.390	22.450
2011	22.080	22.180	22.280	22.310	22.430	22.390	22.470	22.620	22.650	22.680	22.830	22.730

3.7. CRESCIMENTO DO PIB (PRODUTO INTERNO BRUTO)

O produto interno bruto é equivalente à soma dos consumos anuais de bens e serviços produzidos pelas unidades produtoras residentes no país valorados a preço de mercado.

Tabela 3.7. Produto Interno Bruto (PIB), valores atualizado a dezembro de 2011¹³.

Ano	jan	fev	mar	abr	mai	jun	jul	agos	set	out	nov	dez
2002	226,243	222,037	219,536	225,231	230,694	237,493	236,346	233,871	232,095	235,196	236,367	232,883
2003	225,015	218,588	214,352	216,504	218,497	221,239	223,058	224,666	225,899	228,864	232,790	236,749
2004	232,169	227,091	223,517	229,673	236,133	242,548	243,252	243,346	244,006	248,026	252,046	255,319
2005	248,438	241,841	236,358	241,395	246,757	252,101	252,111	251,496	251,458	256,105	260,734	264,596
2006	257,673	252,166	245,475	249,551	253,979	258,358	262,482	266,095	270,062	275,434	280,012	285,782
2007	280,347	275,333	270,561	275,764	281,449	287,123	287,836	288,517	288,424	292,665	296,672	300,269
2008	295,339	291,403	288,096	295,415	302,298	308,031	309,036	311,081	314,288	314,312	313,280	312,480
2009	303,342	293,305	284,320	291,355	297,284	302,971	306,593	310,744	315,393	323,655	331,627	339,107
2010	333,230	325,322	318,072	324,590	330,952	338,174	342,999	347,645	352,329	357,742	361,602	364,946
2011	355,351	344,799	335,731	342,931	349,735	356,990	356,552	356,914	355,705	359,027	362,733	365,532

¹³ Tabela do Índice Nacional de Preços ao Consumidor (INPC) em R\$1.000.000,00. Fonte: <http://seriesestatisticas.ibge.gov.br/series.aspx?vcodigo=ST17&sv=78&t=produto-interno-bruto-brvalores-correntes>

Utilizando a tabela de Índice de Preços ao Consumidor atualizou-se os valores da tabela do PIB a valores correntes. Os valores do PIB apresentados na Tabela 3.7 são valores atualizados para dezembro de 2011.

Existe hoje no mercado de financiamento de veículos no Brasil, cerca de cinquenta instituições que atuam na área (sítio da ANFAVEA). É um mercado muito competitivo e o financiamento de automóveis é o tipo de crédito (dinheiro) com garantia, mais barato do mercado.

Ano	jan	fev	mar	abr	mai	jun	jul	agos	set	out	nov	dez
2002	45.510	46.020	46.430	46.920	47.810	48.020	48.723	49.009	49.210	49.600	50.010	50870
2003	51.120	51.560	51.923	52.045	52.325	52.835	53.080	53.420	54.325	54.834	55.120	55.902
2004	56.195	57.368	58.280	59.774	61.353	61.628	63.480	64.922	66.800	68.167	70.701	72.955
2005	75.332	76.597	78.852	80.946	82.751	85.388	86.802	88.438	90.304	92.894	94.862	96.989
2006	98.308	100.097	101.557	103.487	106.599	108.045	110.207	112.203	114.791	114.443	115.836	117.850
2007	119.001	113.871	123.373	126.205	130.068	133.264	136.466	140.657	144.025	149.767	155.317	159.849
2008	163.025	167.274	172.996	176.401	179.812	183.827	187.982	191.173	195.629	200.451	199.886	200.345
2009	196.659	203.198	203.939	204.159	205.091	206.722	208.057	210.433	215.345	219.384	222.894	227.213
2010	227.338	229.444	233.387	237.319	242.579	248.244	254.351	258.558	264.651	268.487	273.264	276.311
2011	276.079	281.344	285.076	287.317	291.287	295.205	298.739	301.650	305.439	307.948	310.821	319.225

Tabela 3.8. Montante de operações de crédito setor automotivo em milhão R\$¹⁴.

3.8. DESCRIÇÃO DE OUTRAS VARIÁVEIS

A seguir são apresentadas outras variáveis usadas para o modelo de previsão que podem estar relacionadas ao consumo de combustíveis.

Tabela 3.9. Vendas de gasolina pelas distribuidoras – em 1.000m³¹⁵.

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Jan	1.921	1.827	1.883	1.843	1.926	1.976	2.034	2.039	2.448	2.521
Fev	1.805	1.590	1.742	1.805	1.874	1.844	1.926	1.914	2.385	2.534
Mar	1.876	1.578	1.933	2.062	2.040	2.083	2.031	2.071	2.663	3.097
Abril	1.893	1.705	1.976	1.917	1.947	1.973	2.062	2.084	2.412	3.155
Mai	1.885	1.839	1.827	1.963	2.007	2.042	2.044	1.970	2.332	2.842
Jun	1.757	1.739	1.835	1.956	1.934	2.007	1.993	2.023	2.323	2.775
Jul	1.848	1.879	1.980	1.940	1.928	1.992	2.146	2.127	2.424	2.845
Agos	1.857	1.800	1.941	2.053	2.062	2.069	2.096	2.020	2.415	3.022
Set	1.839	1.836	2.012	1.984	2.052	1.923	2.187	2.098	2.465	3.063
Out	2.104	1.986	1.942	1.873	2.022	2.121	2.239	2.296	2.487	3.009
Nov	1.736	1.819	1.876	1.920	1.961.	2.016	2.011	2.138	2.525	3.066
Dez	2.083	2.187	2.221	2.229	2.246	2.273	2.398	2.623	2.959	3.556

¹⁴ <https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries>

¹⁵ Fonte - http://www.anp.gov.br/preco/prc/historico/Resumo_Mensal_Combustiveis.asp

Tabela 3.9. Produção de gasolina – em 1.000 m³¹⁶.

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Jan	1.680	1.555	1.534	1.699	1.765	1.657	1.811	1.624	1.840	2.006
Fev	1.534	1.312	1.328	1.376	1.603	1.619	1.650	1.474	1.709	1.832
Mar	1.574	1.477	1.568	1.534	1.739	1.856	1.697	1.722	1.730	2.049
Abril	1.595	1.413	1.469	1.565	1.705	1.711	1.643	1.609	1.738	1.930
Mai	1.617	1.545	1.349	1.669	1.695	1.836	1.740	1.687	1.654	2.201
Jun	1.461	1.568	1.366	1.436	1.720	1.846	1.749	1.677	1.699	1.833
Jul	1.610	1.493	1.494	1.658	1.751	1.750	1.824	1.747	1.823	1.788
Agos	1.573	1.459	1.351	1.711	1.715	1.689	1.609	1.774	1.768	1.866
Set	1.428	1.543	1.491	1.565	1.647	1.646	1.647	1.593	1.712	2.061
Out	1.558	1.497	1.577	1.743	1.655	1.757	1.601	1.555	1.865	1.973
Nov	1.552	1.397	1.469	1.538	1.558	1.656	1.564	1.641	1.877	1.918
Dez	1.581	1.401	1.601	1.634	1.835	1.681	1.676	1.667	2.084	2.046

Tabela 3.10. Vendas de óleo diesel pelas distribuidoras - em 1.000 m³¹⁷

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Jan	2.899	2.724	2.823	2.820	2.935	3.047	3.366	3.158	3.349	3.556
Fev	2.733	2.749	2.699	2.869	2.824	2.970	3.403	3.102	3.544	3.840
Mar	3.252	2.794	3.402	3.398	3.421	3.640	3.697	3.638	4.275	4.297
Abril	3.076	2.901	3.253	3.269	3.031	3.235	3.712	3.568	3.966	4.091
Mai	3.121	3.154	3.090	3.204	3.229	3.424	3.738	3.531	4.087	4.433
Jun	3.029	2.930	3.215	3.293	3.205	3.447	3.836	3.700	4.128	4.385
Jul	3.231	3.254	3.426	3.319	3.262	3.496	3.873	3.898	4.329	4.519
Agos	3.302	3.234	3.600	3.641	3.547	3.834	3.885	3.833	4.427	4.863
Set	3.265	3.350	3.556	3.453	3.460	3.523	4.052	3.932	4.328	4.794
Out	3.683	3.563	3.488	3.357	3.546	3.907	4.134	4.265	4.413	4.652
Nov	3.124	3.105	3.391	3.360	3.400	3.663	3.603	3.861	4.297	4.542
Dez	2.947	3.090	3.276	3.177	3.143	3.365	3.458	3.806	4.091	4.287

Tabela 3.11. Produção de Óleo Diesel – em 1.000 m³¹⁸.

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Jan	2.644	2.640	3.128	3.175	3.252	3.257	3.244	3.293	3.405	3.423
Fev	2.545	2.358	3.086	2.914	2.934	2.943	3.551	3.152	3.114	3.272
Mar	2.952	3.077	3.424	3.521	3.474	3.563	3.086	3.634	3.220	3.820
Abril	2.818	2.876	3.464	2.976	3.262	3.333	3.320	3.382	3.305	3.487
Mai	2.780	2.888	3.308	3.171	3.488	3.149	3.529	3.469	3.358	3.680
Jun	2.487	2.761	3.013	3.072	3.284	3.237	3.394	3.667	3.499	3.601
Jul	2.804	2.956	2.932	3.363	3.283	3.376	3.549	3.677	3.853	3.280
Ago	2.910	2.902	2.947	3.427	3.214	3.430	3.617	3.733	3.449	3.271
Set	2.792	3.029	3.270	3.328	3.182	3.342	3.619	3.649	3.453	3.545
Out	3.022	3.171	3.548	3.392	3.324	3.297	3.387	3.872	3.413	3.951
Nov	2.921	2.886	3.005	3.173	3.177	3.116	3.401	3.603	3.659	3.478
Dez	2.529	2.837	3.380	3.230	3.237	3.532	3.432	3.762	3.694	3.718

¹⁶ Fonte - http://www.anp.gov.br/preco/prc/historico/Resumo_Mensal_Combustiveis.asp¹⁷ Fonte - http://www.anp.gov.br/preco/prc/historico/Resumo_Mensal_Combustiveis.asp¹⁸ Fonte - http://www.anp.gov.br/preco/prc/historico/Resumo_Mensal_Combustiveis.asp

Tabela 3.12. Valor do IPI % (imposto sobre produtos industrializados) carros de 1.000 cc¹⁹

	Jan	Fev	Mar	Abril	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2002	10	10	10	10	10	10	10	9	9	9	10	10
2003	10	10	10	10	10	10	10	5	5	5	5	6
2004	6	6	6	6	6	6	6	6	6	6	6	6
2005	6	6	6	6	6	6	6	6	6	6	6	6
2006	6	6	6	6	6	6	6	6	6	6	6	6
2007	6	6	6	6	6	6	6	6	6	6	6	6
2008	6	6	6	6	6	6	6	6	6	6	6	6
2009	6	6	6	6	6	6	6	6	6	6	6	0
2010	0	0	0	0	0	0	0	0	0	0	0	0
2011	0	0	0	0	0	0	0	0	0	0	0	0

Tabela 3.13. Valor do IPI % (imposto sobre produtos industrializados) carros de 1.000 cc até 2.000 cc²⁰.

	Jan	Fev	Mar	Abril	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2002	14	14	14	14	14	14	14	14	14	14	13	13
2003	13	13	13	13	13	13	13	11	11	11	12	12
2004	12	12	12	12	12	12	12	12	12	11	11	11
2005	11	11	11	11	11	11	11	11	11	11	11	11
2006	11	11	11	11	11	11	11	11	11	11	11	11
2007	11	11	11	11	11	11	11	11	11	11	11	11
2008	11	11	11	11	11	11	11	11	11	11	11	11
2009	11	11	11	11	11	11	11	11	11	11	11	7
2010	7	7	7	7	7	7	7	7	7	7	7	7
2011	7	7	7	7	7	7	7	7	7	7	7	7

¹⁹ <http://www.receita.fazenda.gov.br/Legislacao/LegisAssunto/impsobproindipi/ImpSobProIndIPI>

²⁰ <http://www.receita.fazenda.gov.br/Legislacao/LegisAssunto/impsobproindipi/ImpSobProIndIPI3.htm>

CAPÍTULO 4 – ANÁLISE DOS RESULTADOS

4.1. A ANÁLISE FATORIAL PARA CONSUMO DE GASOLINA

O método de Análise Fatorial tem sua base nas correlações entre as variáveis, como primeiro passo deve-se analisar a matriz de correlações verificando se existem valores significativos que justifique o uso da técnica. Se as referidas correlações entre todas as variáveis se mostrarem baixas, um forte sinal de que a análise Fatorial não seja a técnica apropriada a ser aplicada (ANDERSON, 2003). A matriz de correlação mede a associação linear entre as variáveis por meio do coeficiente de correlação de Pearson (LUIS PAULO, 2009).

De acordo com Hair, Anderson e Black (2009), se na inspeção visual da matriz de correlações não mostrar um número significativo de valores superiores a 0,30, há fortes indícios de que a utilização da técnica não é apropriada. As variáveis selecionadas foram processadas pelo software aplicativo SPSS for Windows 11.5 (Figura 4.1).

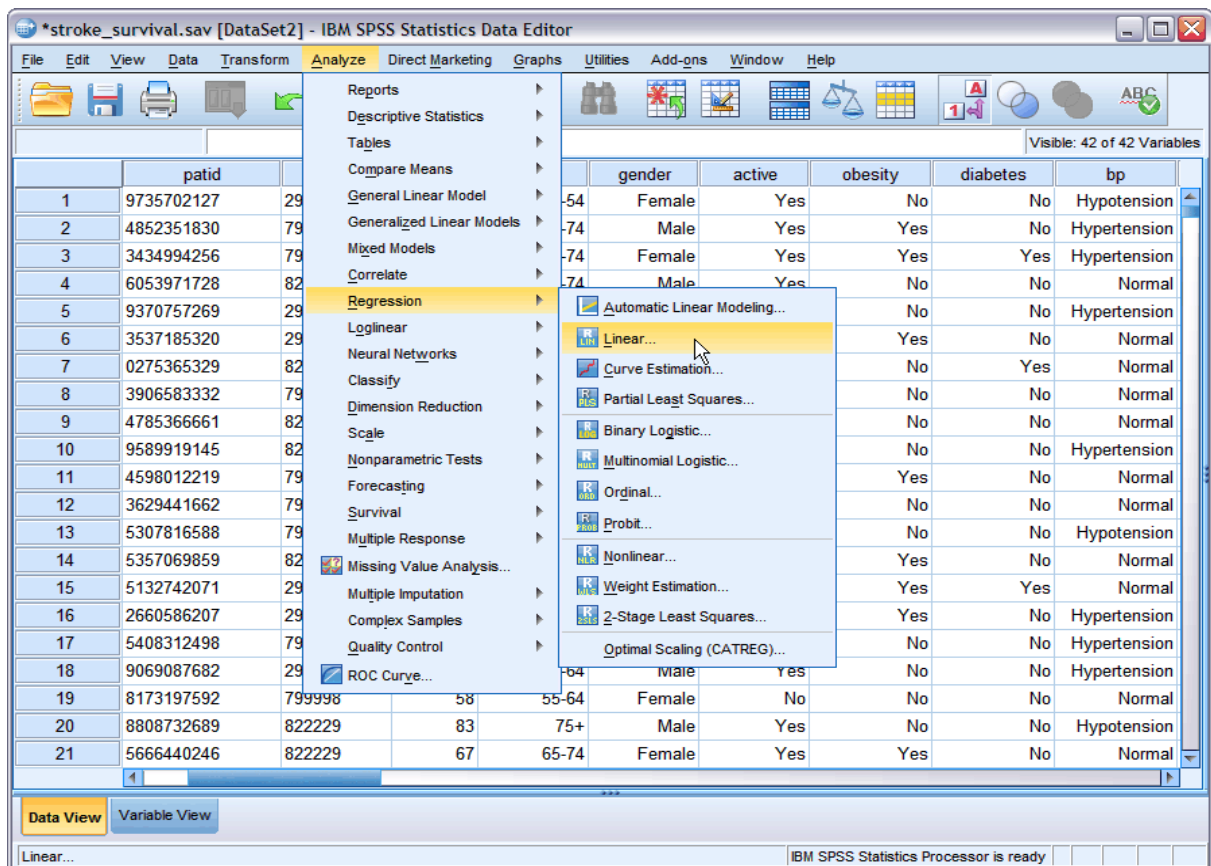


Figura 4.1. Exemplo de tela do aplicativo indicando as variáveis analisadas.

A Tabela 4.1 mostra a matriz de correlação de Pearson. Com base nesta matriz de correlação pode-se afirmar que existem fortes relações entre a população ocupada e o preço da gasolina (0,906), assim como entre preço da gasolina e preço do diesel (0,950), entre financiamento de automóveis e população ocupada (0,932), entre PIB e financiamento de automóveis (0,981), só para ficar em alguns exemplos. A correlação entre IPI e financiamento de automóveis tem valor negativo por variar de maneira inversa, o aumento do imposto reduz o financiamento de automóveis.

Tabela 4.1. Matriz de correlação entre as variáveis.

	ETANOL	DIESELPR	GASOLINA	NAUTOMO	POPOCUPA	PIBREAJU	FINAUTO	PRODGAS	PRODDIES	IMPORGAS	IMPODIES	IPI1000	IPI2000
ETANOLPR	1,000												
DIESELPR	,738	1,000											
GASOLINA	,827	,950	1,000										
NAUTOMOV	,469	,684	,667	1,000									
POPOCUPA	,694	,879	,906	,829	1,000								
PIBREAJU	,666	,765	,787	,852	,945	1,000							
FINAUTO	,704	,777	,786	,844	,932	,981	1,000						
PRODGAS	,015	,028	,046	,068	,074	,098	,094	1,000					
PRODDIES	,043	,162	,139	,168	,199	,199	,195	,463	1,000				
IMPORGAS	,507	,180	,291	,168	,332	,406	,454	,193	,162	1,000			
IMPODIES	,249	,150	,215	,441	,397	,514	,506	,116	,081	,378	1,000		
IPI1000	-,624	-,757	-,708	-,750	-,816	-,847	-,877	-,102	-,240	-,343	-,304	1,000	
IPI2000	-,674	-,787	-,742	-,765	-,830	-,862	-,895	-,069	-,214	-,350	-,309	,988	1,000

A Tabela 4.2 mostra o teste de esfericidade de Bartlett e de precisão de amostragem de Kaiser-Meyer-Olkin (KMO). Ambos indicam a adequação dos dados para a realização da análise fatorial. O teste KMO é uma estatística que indica a proporção da variância dos dados que pode ser considerada comum a todas as variáveis, ou seja, que pode ser atribuída a um fator comum, então quanto mais próximo de 1 (unidade) melhor o resultado, ou seja, mais adequada é a amostra à aplicação da análise fatorial:

Tabela 4.2. Teste de esfericidade de Bartlett e de Kaiser-Meyer-Olkin (KMO).

Precisão de amostragem de acordo com o teste KMO.		,83
Teste de esfericidade de Bartlett	Approx. Chi-Square	2418,39
	df	78
	Sig.	,000

O programa SPSS apresenta o resultado da estatística KMO e do teste de esfericidade de Bartlett. O valor do teste de esfericidade de Bartlett é alto para uma significância de 0.000, o que aponta para uma baixa probabilidade de a matriz de correlação ser uma matriz identidade. Ao rejeitar a hipótese de ser uma matriz identidade, recomenda-se, assim, o uso da análise fatorial. O valor da medida estatística KMO para as variáveis adotadas se classifica como boa (0,811), podendo ser entendido enquanto um indicativo para a utilização da análise fatorial das variáveis.

Tabela 4.3. Matriz de correlação anti-imagem.

	ETAN OLPR	DIESELP R	GASOL INA	NAUTO MOV	POPOCU PA	PIBREA JU	FINAU TO	PRODG AS	PRODDI ES	IMPORG AS	IMPODI ES	IPI100 0	IPI200 0
ETANOLPR	,78	,29	-,72	,178	,541	-,013	-,349	,094	,086	-,215	-,007	-,055	,082
DIESELPR	,29	,86	-,67	,054	,109	,232	-,292	,045	-,070	,305	,131	-,018	,113
GASOLINA	-,72	-,675	,75	-,034	-,711	,003	,412	-,142	,040	-,088	-,007	-,180	,130
NAUTOMOV	,17	,054	-,03	,957(a)	-,085	-,036	-,166	-,046	,018	,292	-,114	-,122	,145
POPOCUPA	,54	,109	-,71	-,085	,829(a)	-,359	-,325	,161	-,052	-,010	,044	,298	-,320
PIBREAJU	-,01	,232	,00	-,036	-,359	,907(a)	-,636	-,045	-,019	,181	-,100	-,013	,011
FINAUTO	-,34	-,292	,41	-,166	-,325	-,636	,846(a)	-,053	,056	-,324	-,140	-,175	,276
PRODGAS	,09	,045	-,14	-,046	,161	-,045	-,053	,482(a)	-,423	-,108	-,039	,200	-,198
PRODDIES	,08	-,070	,04	,018	-,052	-,019	,056	-,423	,717(a)	-,119	,019	,030	,006
IMPORGAS	-,21	,305	-,08	,292	-,010	,181	-,324	-,108	-,119	,763(a)	-,094	,066	-,068
IMPODIES	-,00	,131	-,00	-,114	,044	-,100	-,140	-,039	,019	-,094	,946(a)	-,005	-,070
IPI1000	-,05	-,018	-,18	-,122	,298	-,013	-,175	,200	,030	,066	-,005	,829(a)	-,953
IPI2000	,082	,113	,13	,145	-,320	,011	,276	-,198	,006	-,068	-,070	-,953	,827(a)

Na matriz de correlações anti-imagem da Tabela 4.3, os valores da diagonal principal representam uma medida de adequação dos dados, conhecida por MSA (medida de adequação da amostra), para cada uma das variáveis em análise. Caso algum valor esteja abaixo de 0,5, tal fato indica que esta variável específica pode não se ajustar à estrutura definida pelas outras variáveis e, portanto, merece eventualmente ser eliminada (MAROCO, 2007).

A Tabela 4.4 apresenta as comunalidades das variáveis analisadas, qual seja a quantia total de variância que uma variável original compartilha com todas as outras variáveis. As comunalidades iniciais são iguais a um e, após a extração, varia, de 0 a 1, próximo de zero quando os fatores comuns apresentam baixa ou nenhuma variância da variável, e 1 quando toda a variância é explicada por todos os fatores.

Tabela 4.4. Comunalidade inicial e extração das variáveis analisadas.

	Inicial	Extração
ETANOLPR	1,000	,638
DIESELPR	1,000	,901
GASOLINA	1,000	,854
NAUTOMOV	1,000	,708
POPOCUPA	1,000	,930
PIBREAJU	1,000	,927
FINAUTO	1,000	,953
PRODGAS	1,000	,726
PRODDIES	1,000	,753
IMPORGAS	1,000	,626
IMPODIES	1,000	,724
IPI1000	1,000	,826
IPI2000	1,000	,860

Tabela 4.5. Relação de parâmetros associados à variância total.

Componente	Autovalores iniciais			Peso da extração dos mínimos quadrados			Peso da extração dos mínimos quadrados		
	Total	Variância (%)	Acumulada (%)	Total	Variância (%)	Acumulada (%)	Total	Variância (%)	Acumulada (%)
1	7,772	59,782	59,782	7,772	59,782	59,782	7,082	54,479	54,479
2	1,517	11,670	71,452	1,517	11,670	71,452	1,826	14,050	68,528
3	1,138	8,756	80,208	1,138	8,756	80,208	1,518	11,680	80,208
4	,930	7,156	87,364						
5	,554	4,264	91,628						
6	,474	3,647	95,275						
7	,267	2,056	97,331						
8	,161	1,236	98,567						
9	,116	,890	99,457						
10	,040	,309	99,766						
11	,013	,102	99,868						
12	,011	,084	99,952						
13	,006	,048	100,000						

A Tabela 4.5 mostra a relação de parâmetros associados à variância total. A matriz dos componentes apresenta as cargas que correlacionam as variáveis com os fatores depois da rotação, permite verificar qual Fator melhor explica cada uma das variáveis originais.

Tabela4.6. Rotated Component Matrix Method Varimax.

	Componentes		
	1	2	3
ETANOLPR	,750	,264	-,076
DIESELPR	,946	-,075	,036
GASOLINA	,923	,043	,015
NAUTOMOV	,809	,218	,071
POPOCUPA	,937	,218	,071
PIBREAJU	,879	,383	,083
FINAUTO	,887	,401	,078
PRODGAS	-,025	,151	,838
PRODDIES	,149	-,006	,855
IMPORGAS	,201	,752	,143
IMPODIES	,199	,827	,014
IPI1000	-,876	-,191	-,148
IPI2000	-,901	-,194	-,106

Tabela 4.7.Component Score Coefficient Matrix.

	Component		
	1	2	3
ETANOLPR	,098	,064	-,106
DIESELPR	,200	-,248	,001
GASOLINA	,173	-,151	-,025
NAUTOMOV	,114	,002	,000
POPOCUPA	,139	-,024	-,004
PIBREAJU	,095	,114	-,011
FINAUTO	,093	,126	-,016
PRODGAS	-,063	,035	,569
PRODDIES	,002	-,122	,592
IMPORGAS	-,114	,526	,015
IMPODIES	-,125	,598	-,083
IPI1000	-,130	,041	-,054
IPI2000	-,136	,039	-,024

Com base nasTabela 4.6e 4.7 pode-se calcular os fatores para cada observação da amostra, levando-se em consideração as variáveis selecionadas e padronizadas. Quando se seleciona a opção para salvar os escores fatoriais no aplicativo IBM-SPSS, será inserido na base de dados três novas variáveis (FAC1_1, FAC2_1 e FAC3_1), que representam os fatores 1,2 e 3.

Tabela 4.8. Valores associados ao Fator 1 (gasolina).

	JAN	FEV	MAR	ABRI	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
2002	-0,85	0,33	0,51	0,95	1,02	2,23	1,63	1,78	2,45	2,74	2,70	2,86
2003	-0,45	1,29	0,84	1,72	1,41	1,82	2,05	2,42	2,05	2,91	3,34	3,61
2004	-0,34	1,47	0,69	1,58	2,17	2,66	2,14	2,90	2,30	2,76	3,21	3,69
2005	-0,90	1,36	0,89	0,99	0,66	2,19	1,54	1,43	2,63	2,10	2,79	2,69
2006	-1,10	0,55	-0,02	0,55	0,97	1,24	1,23	1,31	2,41	2,51	2,74	1,87
2007	-0,66	0,58	-0,42	0,87	0,59	0,83	1,11	1,64	2,36	2,21	2,51	2,12
2008	-1,24	0,13	0,13	1,01	1,08	1,05	0,81	1,84	2,40	2,54	2,89	1,97
2009	-0,42	0,72	0,07	1,18	1,32	1,38	1,12	1,34	2,58	2,76	3,09	2,27
2010	-1,42	-0,16	0,30	0,68	1,30	1,15	0,67	1,39	2,17	1,67	1,94	0,47
2011	-1,97	-0,33	-0,78	-0,04	-0,68	0,84	1,13	0,83	0,89	1,10	1,68	0,84

Tabela 4.9. Valores associados ao Fator 2 (gasolina).

	JAN	FEV	MAR	ABRI	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
2002	1,39	-0,50	-0,49	-2,35	-2,40	-4,27	-4,09	-2,85	-2,60	-3,95	-3,03	-2,70
2003	1,84	0,69	-0,28	-2,62	-3,19	-4,49	-4,11	-3,63	-3,02	-3,84	-3,21	-3,11
2004	1,68	-2,24	-1,58	-2,90	-3,43	-4,98	-4,15	-4,29	-3,68	-3,98	-3,37	-5,64
2005	1,35	-2,20	-1,88	-1,60	-1,67	-4,25	-3,36	-3,78	-4,27	-3,53	-2,33	-3,58
2006	1,03	-2,49	-1,49	-1,91	-3,61	-4,57	-4,05	-2,98	-4,53	-4,07	-2,28	-3,06
2007	0,92	-1,87	-1,68	-3,21	-3,97	-4,55	-3,64	-3,62	-4,52	-3,86	-2,69	-1,37
2008	0,93	-1,47	-1,46	-3,34	-4,54	-4,04	-3,56	-3,28	-4,29	-3,25	-2,60	-0,62
2009	0,64	-1,27	-1,76	-3,50	-4,43	-4,27	-3,44	-3,82	-4,52	-3,50	-4,39	-2,09
2010	1,57	-1,24	-2,18	-3,12	-3,85	-3,42	-2,84	-3,79	-4,47	-2,87	-4,61	-0,43
2011	1,60	-1,28	-2,57	-3,13	-4,20	-4,00	-3,89	-3,05	-4,12	-2,89	-4,03	-1,39

Importante ressaltamos que a correlação de Pearson entre Fator 1 e Fator 2 é zero, uma vez que o método escolhido de rotação é o Varimax que gera Fatores ortogonais entre si. Estes fatores criados pela Análise Fatorial podem ser utilizados em substituição às variáveis originais em outras técnicas multivariadas.

4.2. A ANÁLISE FATORIAL PARA CONSUMO DE ÓLEO DIESEL

Selecionou-se as variáveis que tem interferência no consumo de óleo diesel e aplicamos a técnica Análise Fatorial. As tabelas produzidas pelo programa aparecem na língua inglesa, apresenta-se conforme saída do programa.

Tabela 4.10. Matriz de correlação entre as variáveis.

	ETANO LPR	DIESE LPR	GASO LINA	NCAM INHO	POPO CUPA	PIBRE AJU	FINAU TO	PROD GAS	PROD DIES	IMPO RGAS	IMPO DIES
ETANOLPR	1,00										
DIESELPR	,738	1,000									
GASOLINA	,827	,950	1,000								
NCAMINHA	,582	,624	,650	1,000							
POPOCUP	,694	,879	,906	,835	1,000						
PIBREAJU	,666	,765	,787	,897	,945	1,000					
FINAUTO	,704	,777	,786	,892	,932	,981	1,000				
PRODGAS	,015	,028	,046	,089	,074	,098	,094	1,000			
PRODDIES	,043	,162	,139	,184	,199	,199	,195	,463	1,000		
IMPORGA	,507	,180	,291	,429	,332	,406	,454	,193	,162	1,000	
IMPODIES	,249	,150	,215	,604	,397	,514	,506	,116	,081	,378	1,000

A Tabela 4.11 mostra o teste de esfericidade de Bartlett e de precisão de amostragem de Kaiser-Meyer-Olkin (KMO) para o diesel. Ambos indicam a adequação dos dados para a realização da análise fatorial. O teste KMO é uma estatística que indica a proporção da variância dos dados que pode ser considerada comum a todas as variáveis, ou seja, que pode ser atribuída a um fator comum, então quanto mais próximo de 1 (unidade) melhor o resultado, ou seja, mais adequada é a amostra à aplicação da análise fatorial:

Tabela 4.11. Teste de esfericidade de Bartlett e de Kaiser-Meyer-Olkin (KMO) para o diesel.

Precisão de amostragem de acordo com o teste KMO.	,792	
Teste de esfericidade de Bartlett	Approx. Chi-Square	1768,332
	df	55
	Sig.	,000

Tabela 4.12. Matriz de correlação anti-imagem para o diesel.

	ETANO LPR	DIESELP R	GASOLIN A	NCAMINH O	POPOCU PA	PIBREAJ U	FINAUT O	PRODGA S	PRODDI ES	IMPORG AS	IMPODI ES
ETANOLPR	,715(a)	,274	-,756	-,142	,630	,011	-,396	,110	,080	-,272	,077
DIESELPR	,274	,790(a)	-,680	-,021	,170	,248	-,475	,061	-,111	,339	,224
GASOLINA	-,756	-,680	,693(a)	,182	-,743	-,025	,464	-,109	,057	-,074	-,102
NCAMINHO	-,142	-,021	,182	,947(a)	-,222	-,126	-,057	,027	-,028	-,025	-,316
POPOCUPA	,630	,170	-,743	-,222	,798(a)	-,337	-,264	,097	-,043	-,016	,084
PIBREAJU	,011	,248	-,025	-,126	-,337	,865(a)	-,711	-,047	-,014	,205	-,063
FINAUTO	-,396	-,475	,464	-,057	-,264	-,711	,794(a)	-,017	,028	-,279	-,063
PRODGAS	,110	,061	-,109	,027	,097	-,047	-,017	,545(a)	-,438	-,124	-,055
PRODDIES	,080	-,111	,057	-,028	-,043	-,014	,028	-,438	,653(a)	-,129	,056
IMPORGAS	-,272	,339	-,074	-,025	-,016	,205	-,279	-,124	-,129	,780(a)	-,064
IMPODIES	,077	,224	-,102	-,316	,084	-,063	-,063	-,055	,056	-,064	,874(a)

Tabela 4.13. Relação de parâmetros associados à variância total.

Componente	Autovalores iniciais			Peso da extração dos mínimos quadrados			Peso da extração dos mínimos quadrados		
	Total	Variância (%)	Acumulada (%)	Total	Variância (%)	Acumulada (%)	Total	Variância (%)	Acumulada (%)
1	6,283	57,116	57,116	6,283	57,116	57,116	5,245	47,686	47,686
2	1,521	13,824	70,940	1,521	13,824	70,940	2,232	20,289	67,974
3	1,175	10,686	81,626	1,175	10,686	81,626	1,502	13,651	81,626
4	,815	7,411	89,036						
5	,526	4,781	93,817						
6	,329	2,993	96,811						
7	,170	1,542	98,353						
8	,103	,933	99,286						
9	,056	,506	99,791						
10	,013	,122	99,913						
11	,010	,087	100,000						

Tabela 4.14. Matriz de componentes rotacionados.

	Componentes		
	1	2	3
ETANOLPR	,797	,227	-,032
DIESELPR	,967	-,024	,061
GASOLINA	,965	,067	,050
NCAMINHO	,679	,616	,061
POPOCUPA	,919	,303	,084
PIBREAJU	,825	,482	,084
FINAUTO	,828	,491	,081
PRODGAS	-,037	,131	,846
PRODDIES	,128	,030	,851
IMPORGAS	,199	,664	,174
IMPODIES	,124	,878	,001

Com a utilização da técnica de Análise Fatorial pode-se substituir as variáveis originais por um número menor de variáveis inteiramente novas, criado a partir de escores fatoriais (Tabelas 4.15, 4.16 e 4.17). As variáveis geradas pela Análise Fatorial não são correlacionadas entre si e serão usadas na técnica de dependência - Análise de Regressão.

Tabela 4.15.Component Score Coefficient Matrix.

	Componente		
	1	2	3
ETANOLPR	,172	-,038	-,063
DIESELPR	,280	-,262	,025
GASOLINA	,257	-,198	,008
NCAMINHO	,040	,247	-,036
POPOCUPA	,185	-,028	,008
PIBREAJU	,114	,117	-,009
FINAUTO	,113	,123	-,012
PRODGAS	-,061	,011	,578
PRODDIES	,011	-,100	,589
IMPORGAS	-,111	,386	,049
IMPODIES	-,182	,569	-,093

Tabela 4.16.Valores associados ao Fator 1 para o diesel.

	Jan	Fev	Mar	Abri	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2002	-1,60	-0,72	-0,56	-0,09	0,31	0,80	0,92	1,00	0,97	1,20	1,20	1,85
2003	-1,65	-0,49	-0,67	0,08	0,25	1,23	1,05	0,85	1,02	1,37	1,41	1,81
2004	-1,63	-0,03	-0,72	0,27	0,42	1,23	1,09	1,21	1,11	1,36	1,51	2,13
2005	-1,61	0,01	-0,36	0,35	0,54	1,33	1,13	1,41	1,33	1,44	1,61	3,16
2006	-1,53	0,03	-0,23	-0,00	-0,03	1,01	1,07	1,16	1,49	1,40	1,31	1,89
2007	-1,45	-0,15	-0,47	0,04	0,55	1,04	1,08	0,94	1,54	1,67	1,34	1,79
2008	-1,51	-0,32	-0,48	0,46	0,72	1,06	0,90	1,24	1,48	1,65	1,52	1,47
2009	-1,36	-0,37	-0,51	0,47	0,92	0,92	0,80	1,17	1,40	1,64	1,57	1,37
2010	-1,70	-0,50	-0,45	0,52	1,02	0,99	0,81	1,23	1,45	1,69	2,16	1,69
2011	-1,70	-0,72	-0,14	0,36	0,90	0,76	0,60	1,20	1,42	1,36	2,04	1,25

Tabela 4.17.Valores associados ao Fator 2 para o diesel.

	Jan	Fev	Mar	Abri	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2002	-1,10	-2,29	-2,19	-2,36	-2,74	-3,02	-2,36	-2,49	-1,82	-1,99	-0,88	-0,44
2003	-0,93	-1,98	-2,39	-1,10	-1,05	-2,47	-1,75	-2,02	-2,29	-1,40	-0,92	-0,08
2004	-0,73	-2,33	-2,09	-1,57	-2,39	-2,75	-1,10	-1,16	-2,73	-1,83	0,02	-2,01
2005	-0,50	-1,91	-2,20	-2,48	-2,59	-2,70	-0,93	-1,61	-2,70	-1,49	0,02	-0,38
2006	-0,81	-1,74	-2,01	-2,64	-3,06	-2,24	-0,73	-1,30	-2,34	-1,29	-0,28	0,00
2007	-0,99	-1,89	-2,27	-2,76	-2,98	-2,41	-0,50	-1,74	-2,60	-1,58	-0,01	1,44
2008	-0,80	-2,05	-2,28	-2,36	-2,36	-1,64	-0,81	-1,63	-2,64	-0,56	-1,47	1,82
2009	-1,05	-1,98	-2,40	-2,37	-2,77	-2,22	-0,99	-1,11	-2,22	-0,49	-2,24	0,65
2010	-0,18	-1,05	-1,99	-1,72	-2,42	-2,47	-0,80	-0,66	-1,97	-0,67	-1,77	1,82
2011	-0,11	-0,85	-2,10	-2,48	-2,71	-2,50	-1,05	-1,02	-1,90	-1,99	-0,88	1,10

4.3. ANÁLISE REGRESSÃO

As funções que se encaixam nas observações das variáveis existentes apresentam a seguinte forma:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e \quad (4.1)$$

O objetivo é obter estimativas para os parâmetros $\beta_1, \beta_2, \dots, \beta_n$ que ofereçam um melhor ajuste para os dados disponíveis.

O procedimento utilizado neste trabalho foi o chamado STEPWISE que avalia a significância estatística dos parâmetros de determinadas variáveis explicativas e vai incluindo passo a passo apenas aquelas que se mostraram relevantes a determinado nível de confiança.

4.4. PREVISÃO DO CONSUMO DE GASOLINA

Utilizou-se o aplicativo SPSS para aplicação da modelagem de Regressão Múltipla do problema em questão, buscando uma forma de previsão de demanda, gasolina e óleo diesel. Na Análise Fatorial resumiu as variáveis originais em três variáveis escore fatorial sem perda relevante de informações. Estas variáveis foram o ponto de partida de aplicação da técnica nos dois casos, gasolina e óleo diesel. Pela Tabela 4.18 é possível verificar que o conjunto de variáveis explicativas explica 78,4% da variância de Y (consumo de gasolina)

Tabela 4.18. Resumo do modelo.

Modelo	R	R quadrático	R quadráticoajust.	Erropadrãoestimado
1	,657(a)	,431	,426	273693,193
2	,885(b)	,784	,780	169423,650

a Predictors: (Constant), REGR factor score 1 for analysis 1

b Predictors: (Constant), REGR factor score 1 for analysis 1, REGR factor score 2 for analysis 1

c Dependent Variable: CONSUGAS

Tabela 4.19. Resultados estatísticos referentes ao método ANOVA.

Modelo		Soma dos quadrados	df	Médiaquadrática	F	Sig.
1	Regressão	6703428255118,660	1	6703428255118,660	89,489	,000(a)
	Resíduos	8839139764785,330	118	74907964108,350		
	Total	15542568019903,990	119			
2	Regressão	12184156362397,240	2	6092078181198,620	212,235	,000(b)
	Resíduos	3358411657506,746	117	28704373141,083		
	Total	15542568019903,990	119			

a Predictors: (Constant), REGR factor score 1 for analysis 1

b Predictors: (Constant), REGR factor score 1 for analysis 1, REGR factor score 2 for analysis 1

c Dependent Variable: CONSUGAS

Pode-se observar que todas as variáveis explicativas apresentam sig $t < 0,05$. Conforme pode-se observar que as variáveis apresentam valores de tolerância altos e valores baixos de VIF (*Variance Inflation Factor*).

Tabela 4.20. Coeficientes usados no modelo.

Modelo		Coeficientes não-padronizados		Coeficientes padronizados	t	Sig.	Estatísticas colineares	
		B	Erro padrão	Beta			Tolerância	VIF
1	Constante	2128167,908	24984,656		85,179	,000		
	REGR factor score 1 for analysis 1	237342,221	25089,414	,657	9,460	,000	1,000	1,000
2	Constante	2128167,908	15466,192		137,601	,000		
	REGR factor score 1 for analysis 1	237342,221	15531,041	,657	15,282	,000	1,000	1,000
	REGR factor score 2 for analysis 1	214607,872	15531,041	,594	13,818	,000	1,000	1,000

a Dependent Variable: CONSUGAS

Conforme a aplicação da técnica de Análise Fatorial, pode-se chamar o primeiro fator de desenvolvimento econômico ou de e o segundo fator de importação de gasolina e óleo diesel ou imp, para montarmos o modelo final: Pode-se observar que utiliza os dois primeiros fatpres para a construção da equação de regressão, baseado no caráter das variáveis.

$$\text{Consumo de gasolina} = 2.128.167,91 + 237.342,22de + 214.607,87imp + u \quad (4.2)$$

A partir desse modelo (Equação 4.2) poderão ser elaboradas previsões de consumo de gasolina em função dos possíveis valores das variáveis (fatores), desenvolvimento econômico e importação de gasolina e óleo diesel.

O gráfico a seguir apresenta a evolução do consumo de gasolina durante o período estudado (2002 até 2011)

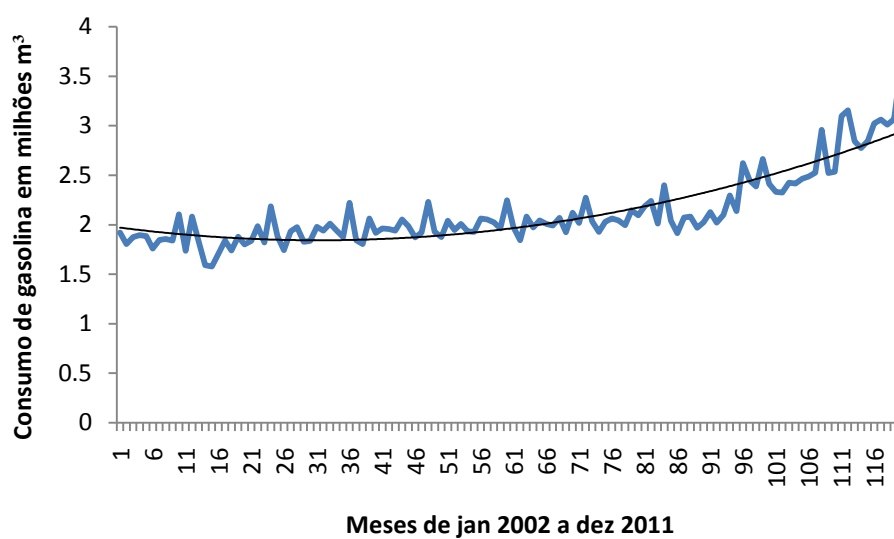


Figura 4.2. Consumo de gasolina entre janeiro de 2002 e dezembro 2011.

A Tabela 4.21 mostra uma estimativa de consumo de gasolina até o ano 2021 comosendo uma estimativa pelo valor esperado dos modelos.

Tabela 4.21. Projeção de consumo de gasolina em 1.000 m³

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Jan	2339	2392	2445	2499	2552	2605	2659	2712	2766	2819
Fev	2456	2519	2582	2646	2646	2709	2772	2835	2898	2961
Mar	2832	2926	3020	3114	3209	3303	3397	3491	3586	3680
Abr	2929	3026	3123	3220	3317	3414	3512	3609	3706	3803
Mai	2554	2622	2689	2757	2825	2893	2961	3028	3096	3164
Jun	2701	2784	2866	2949	3031	3114	3196	3279	3361	3444
Jul	2784	2870	2956	3041	3127	3213	3299	3384	3470	3556
Ago	3115	3234	3353	3473	3592	3711	3830	3949	4069	4188
Set	3236	3355	3473	3592	3711	3830	3949	4067	4186	4305
Out	2967	3058	3150	3242	3334	3426	3517	3609	3701	3793
Nov	3264	3389	3514	3639	3764	3889	4014	4139	4264	4389
Dez	3097	3205	3312	3419	3526	3634	3741	3848	3955	4062

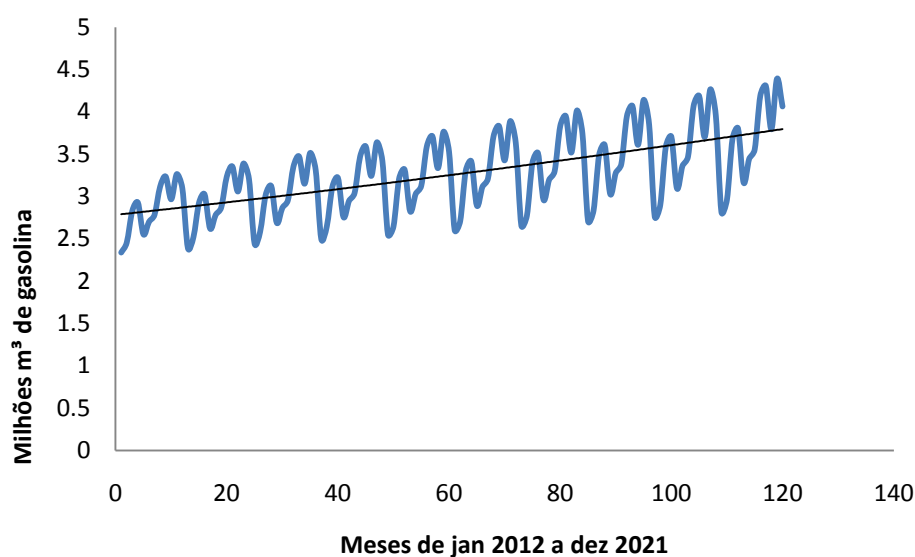


Figura 4.3. Projeção do consumo de gasolina entre 2012 e 2021.

4.5. PREVISÃO DO CONSUMO DE ÓLEO DIESEL

Tabela 4.22. Resumo estatístico do modelo.

Modelo	R	R quadrático	R quadráticoajust.	Erropadrão	Durbin-Watson
1	,665(a)	,442	,437	369881,259	
2	,896(b)	,802	,799	221070,130	1,298

a Predictors: (Constant), REGR factor score 2 for analysis 2

b Predictors: (Constant), REGR factor score 2 for analysis 2, REGR factor score 1 for analysis 2

c DependentVariable: CONSUDIE

Tabela 4.23. Dados estatísticos do modelo ANOVA

Modelo		Soma dos quadrados	df	Médiaquadrática	F	Sig.
1	Regressão	12785106656016,580	1	12785106656016,580	93,45	,000(a)
	Resíduos	16143833218034,870	118	136812145915,55		
	Total	28928939874051,460	119			
2	Regressão	23210915599802,400	2	11605457799901,20	237,46	,000(b)
	Resíduos	5718024274249,050	117	48872002344,01		
	Total	28928939874051,460	119			

a Predictors: (Constant), REGR factor score 2 for analysis 2

b Predictors: (Constant), REGR factor score 2 for analysis 2, REGR factor score 1 for analysis 2

c DependentVariable: CONSUDIE

Tabela 4.24. Coeficientes usados no modelo.

Modelo		Coeficientes não-padronizados		Coeficientes padronizados	t	Sig.	Estatísticas colineares	
		B	Erro padrão	Beta			Tolerância	VIF
1	Constante	3533719,767	33765,385		104,655	,000		
	REGR factor score 2 for analysis 2	327777,167	33906,959	,665	9,667	,000	1,000	1,000
2	Constante	3533719,767	20180,849		175,103	,000		
	REGR factor score 2 for analysis 2	327777,167	20265,466	,665	16,174	,000	1,000	1,000
	REGR factor score 1 for analysis 2	295992,973	20265,466	,600	14,606	,000	1,000	1,000

a Dependent Variable: CONSUDIE

Conforme a aplicação da técnica Análise Fatorial, pode chamar o primeiro fator de desenvolvimento econômico ou de e o segundo fator de importação de gasolina e óleo diesel ou imp, para montarmos o modelo final:

$$\text{Consumo de óleo diesel} = 3.533.719,77 + 295.992,97\text{de} + 327.777,17\text{imp} + \mathbf{u} \quad (4.3)$$

A partir deste modelo (Equação 4.3) poderão ser elaboradas previsões de consumo de gasolina em função dos possíveis valores das variáveis (fatores), desenvolvimento econômico e importação de gasolina e óleo diesel. A Figura 4.4 mostra a evolução do consumo de óleo diesel no período estudado.

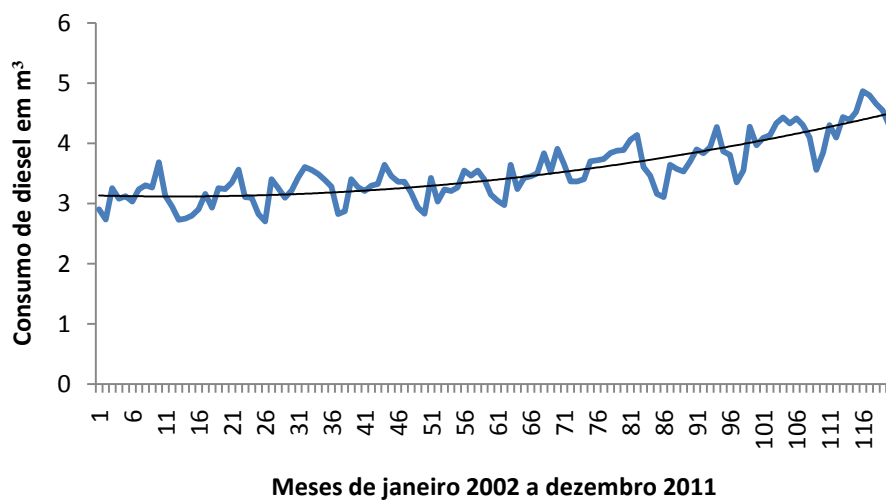


Figura 4.4. Consumo de óleo diesel entre 2002 e 2011.

A Tabela 4.25 mostra uma estimativa de consumo de óleo diesel até o ano 2021 prevista pelo modelo. A tendência da curva na Figura 4.5 mostra um aumento expressivo do consumo de diesel para a próxima década.

Tabela 4.25. Projeção de consumo de gasolina em 1.000 m³

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Jan	3802	3.873	3944	4.015	4086	4157	4228	4299	4370	4441
Fev	3906	3137	3291	3.063	3187	3268	3438	3617	3490	3921
Mar	4364	4475	4587	4.699	4810	4922	5034	5145	5257	5368
Abr	4492	4717	4842	4.967	5093	5218	5343	5468	5593	5718
Mai	4239	4339	4439	4.539	4639	4739	4839	4939	5039	5139
Jun	4417	4536	4656	4.775	4895	5014	5134	5253	5373	5492
Jul	4519	4643	4768	4.893	5017	5142	5267	5391	5516	5641
Ago	4939	5105	5271	5.438	5604	5771	5937	6104	6270	6437
Set	5024	5181	5338	5.495	5652	5809	5966	6123	6280	6437
Out	4702	4827	4953	5.078	5204	5329	5455	5580	5706	5831
Nov	4996	5159	5323	5.487	5651	5815	5978	6142	6306	6470
Dez	4849	4998	5147	5.297	5446	5595	5745	5894	6043	6193

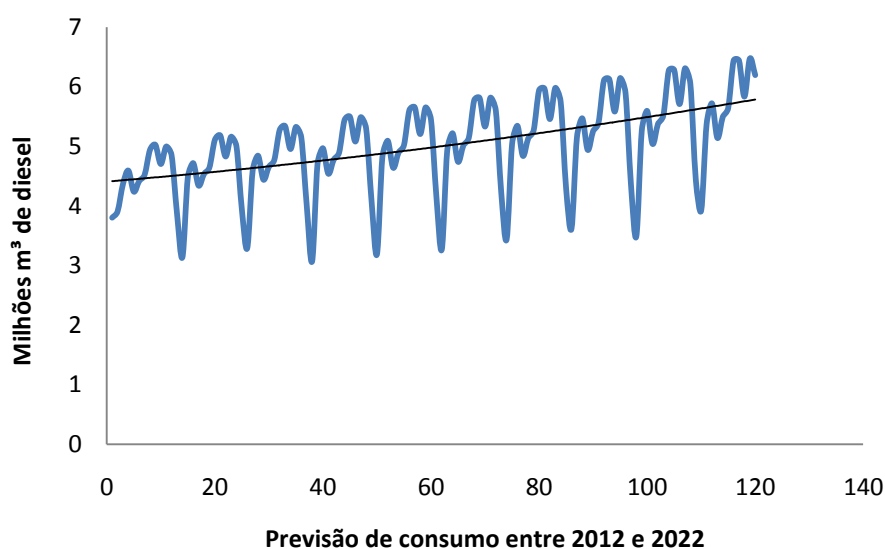


Figura 4.5. Previsão do consumo de diesel e gasolina entre 2012 e 2022.

CAPÍTULO 5 – CONCLUSÃO

Partir de qualquer técnica ou modelo de previsão para inferir situações futuras remete a uma reflexão a respeito da efetividade dos resultados. O aumento da tecnologia e o aprimoramento das técnicas de cálculo e análise de cenários associadas ao avanço tecnológico propicia uma previsão considerável dos acontecimentos do mercado e da economia em geral.

No caso da gasolina e óleo diesel a previsão está sujeita a eventos tecnológicos, econômicos e políticos. No campo tecnológico pode-se imaginar facilmente o impacto no consumo de gasolina se os carros elétricos se viabilizarem comercialmente nos próximos anos. O consumo sob a perspectiva da economia, mostra a dissertação em tela, a forte relação entre crescimento do PIB, aumento da frota de carros e caminhões, financiamentos, aumento da mão de obra ocupada e incentivo fiscal. Qualquer variação das variáveis econômicas listada tem impacto no consumo, tanto de gasolina quanto de óleo diesel.

O Brasil quebrou em parte o monopólio da prospecção de petróleo, o monopólio da distribuição, mas, ainda tem o monopólio do refino nas mãos da Petrobras uma sociedade de economia mista conforme definida pelo Art. 73 da Constituição Federal. A priorização de investimento nas refinarias hoje tem influência de uma lógica política maior, disputa recursos de investimento com o pré sal por exemplo. O investimento forte na infraestrutura, passa também, por termos um parque de refino a altura dos desafios de consumo do país nos próximos anos.

A própria prospecção advinda do pré-sal se transforma em uma pressão significativa pelo refino no Brasil.

Mostrou a dissertação em tela, confirmando de certa forma a intuição, que o crescimento econômico e suas variáveis, formam o principal fator de consumo de combustíveis. O fator chamado neste trabalho de desenvolvimento econômico tem suas maiores cargas nas variáveis originais PIB, número de automóveis, população ocupada e montante de financiamento disponibilizado pelas financeiras.

Observa-se que, tanto no gráfico que demonstra o consumo no período 2002 a 2011, quanto no gráfico de projeção de consumo para o período 2012 a 2021, que existe um certo padrão de consumo que varia dentro do ano. Tem períodos do ano que o consumo se eleva e tem períodos que o consumo se arrefece. Este padrão foi conservado para as projeções de consumo conforme pode-se perceber nos gráficos.

REFERENCIAS BIBLIOGRÁFICAS

- ALBUQUERQUE, J. P. A.; FORTES, J. M. P., FINAMORE, W. A.** Probabilidade, variáveis aleatórias e processos estocásticos. Rio de Janeiro: Interciencia, 2008.
- ANDERSON, D. R.; SWEENEY, D. J.; WILLIAMS, T. A.** Estatística aplicada à administração e economia. 2. Ed. São Paulo: Thomson Pioneira, 2007.
- ANDERSON, T. W.** An introduction to multivariate statistical analysis. 3. Ed. Hoboken: John Wiley & Sons, 2003.
- BARRETO, A. S.** Previsão de comportamento e classificação de contribuintes tributários: uma abordagem por modelos lineares generalizados hierárquicos. Florianópolis. Doutorado (Engenharia de Produção) – Departamento de Engenharia de Produção e Sistemas da Universidade Federal de Santa Catarina, 2005.
- BRUNI, A. L.** Estatística aplicada à gestão empresarial. São Paulo: Atlas, 2007.
- CARVALHO, H.** Análise multivariada de dados qualitativos: utilização do SPSS. Lisboa: Edições Sílabo, 2004.
- COOPER, D. R.; SCHINDLER, P. S.** Métodos de pesquisa em administração. 7. Ed. Porto Alegre: Bookman, 2003.
- CORRAR, L. J.; PAULO, E.; FILHO, J. M. D.** Análise multivariada para administração, Ciências contábeis e economia. São Paulo: Atlas, 2007.
- FÁVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. L.** Modelagem multivariada para tomada de decisões – Dados. São Paulo: Elsevier, 2009.
- ENDERS, W.** Applied econometric time series. 2. Ed. New York: John Wiley & Sons, 2003.
- FERRANDO, P. J.** Introducción al análisis factorial. Barcelona: PPU, 1993.
- HAIR jr., J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L.** Análise Multivariada de dados. 5. Ed. Porto Alegre: Bookman, 2009.
- JOHNSON, R. A.; WICHERN, D. W.** Applied multivariate statistical analysis. New Jersey: Prentice – Hall, 1999.
- MANLY, B. F. J.** Multivariate statistical methods. A primer. 2. Ed. London: Chapman & Hall, 1997.
- MC CLARE, J. T.; BENSON, P. G.; SINCICH, T.** Estatística para Administração e Economia. São Paulo: Pearson, 2009.
- MARTINS, G. A.** Estatística geral e aplicada. 3. Ed. São Paulo: Atlas, 2005.
- MONTGOMERY, D. C.; RUNGER, G. C.** Estatística aplicada e probabilidade para

MONTGOMERY, D. C.; RUNGER, G. C. Estatística aplicada e probabilidade para engenheiros,. 2. Ed. Rio de Janeiro: LTC Editora, 2003.

PEREIRA, J. C. R. Análise de dados qualitativos: estratégias metodológicas para a ciência da saúde, humanas e sociais. 3. Ed. São Paulo: Edusp, 2004.

PESTANA, M. H.; GAGEIRO, J. N. Análise de dados para ciências sociais: a complementaridade do SPSS. 4. Ed. Lisboa: Edições Sílabo, 2005.

SZKLO, ALEXANDRE. Fundamentos do Refino de Petróleo – tecnologia e economia. Rio de Janeiro: Interciencia, 2005.