

**UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
ENGENHARIA DE PRODUÇÃO**

**MODELOS DE PREVISÃO DE RECURSOS PARA  
ANTIMICROBIANOS NO HOSPITAL UNIVERSITÁRIO  
DE SANTA MARIA**

**DISSERTAÇÃO DE MESTRADO**

**Claudio Bastos**

**Santa Maria, RS, Brasil**

**2009**

**MODELOS DE PREVISÃO DE RECURSOS PARA  
ANTIMICROBIANOS NO HOSPITAL UNIVERSITÁRIO DE  
SANTA MARIA**

**por**

**Claudio Bastos**

Dissertação apresentada ao Programa de Pós Graduação em Engenharia de Produção, Área de Concentração em Gerência da Produção da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Mestre em Engenharia de Produção**

**Orientador: Prof. Dr. Adriano Mendonça Souza**

**Santa Maria, RS, Brasil**

**2009**

**Universidade Federal de Santa Maria  
Centro de Tecnologia  
Programa de Pós graduação em Engenharia de Produção**

A Comissão Examinadora, abaixo assinada,  
aprova a Dissertação de Mestrado

**MODELOS DE PREVISÃO DE RECURSOS PARA  
ANTIMICROBIANOS NO HOSPITAL UNIVERSITÁRIO DE SANTA  
MARIA**

elaborada por  
**Claudio Bastos**

como requisito parcial para obtenção do grau de  
**Mestre em Engenharia de Produção**

**COMISSÃO EXAMINADORA:**

---

**Adriano Mendonça Souza, Dr.**  
(Presidente/Orientador)

---

**Roselaine Ruviaro Zanini, Dr<sup>a</sup>.**(UFSM)

---

**Wesley Vieira da Silva, Dr.**(PUC-PR)

Santa Maria, 04 de setembro de 2009.

Aos meus pais Carlos e Zélia,  
por me proporcionarem a existência e educação.  
À Denise, o meu reconhecimento e amor,  
pelo companheirismo e incentivo.  
A Tatiana e a Mayara,  
peço desculpas por não ter dado toda a atenção que precisavam.

## **AGRADECIMENTOS**

Ao Professor Adriano M. Souza, pela orientação na elaboração deste trabalho, pelos conhecimentos repassados, pela paciência e disponibilidade em esclarecer todas as dúvidas ao longo destes dois anos.

Ao Hospital Universitário de Santa Maria, colegas e chefes pelo incentivo e pela concessão da redução parcial no horário de trabalho, fundamental para o desenvolvimento deste trabalho.

A todos os colegas do curso, pelo companheirismo e aprendizado em equipe.

Aos meus pais especialmente à minha mãe, Maria Zélia, exemplo de dedicação, de carinho, de caráter e persistência, a qual sempre me estimulou a progredir e seguir em frente na busca de meus ideais.

À Denise, pelo amor, carinho e incentivo, não só neste trabalho, mas em todos os momentos de nossa caminhada.

## RESUMO

Dissertação de Mestrado  
Programa de Pós Graduação em Engenharia de Produção  
Universidade Federal de Santa Maria

### **MODELOS DE PREVISÃO DE RECURSOS PARA ANTIMICROBIANOS NO HOSPITAL UNIVERSITÁRIO DE SANTA MARIA**

AUTOR: CLAUDIO BASTOS

ORIENTADOR: DR. ADRIANO MENDONÇA SOUZA

Data e Local da Defesa: Santa Maria, 04 de setembro de 2009.

Os escassos recursos da saúde pública impõem ao administrador gerenciar a destinação dos recursos buscando racionalizar e otimizar sua alocação, permitindo, desta forma, melhorar o atendimento aos pacientes, pois o hospital, sendo uma entidade pública, não tem por objetivo o lucro, mas sim promover o bem estar da comunidade. Com isso, a infecção hospitalar que é adquirida após a internação do paciente e se manifesta durante a internação ou mesmo após a alta, podendo ser relacionada com a internação ou procedimentos hospitalares, deve ser evitada. Uma vez que sua total erradicação não é possível, se faz necessário analisar e controlar o custo mensal dos principais antibióticos utilizados no seu tratamento a fim de se ter embasamento suficiente para prever a alocação de recursos para sua aquisição. Nesse contexto, o principal objetivo desta pesquisa é realizar a previsão do custo mensal e de alocação de recursos necessários para aquisição de medicamentos utilizados no tratamento de infecções hospitalares no Hospital Universitário de Santa Maria. Para isso, utilizou-se a metodologia de previsão por regressão linear múltipla e de regressão dinâmica combinada com a técnica multivariada de componentes principais que foi utilizada para eliminar a multicolinearidade existente entre as variáveis originais. Com isso, as componentes principais resultantes foram utilizadas como variáveis independentes na construção do modelo de regressão linear múltipla e de regressão dinâmica. Portanto, essas metodologias são aplicadas a um estudo de caso na saúde pública, a fim de fazer previsões e tirar conclusões a respeito de qual modelo é mais adequado para realizar a previsão do custo mensal dos antibióticos em infecções hospitalares. Os resultados obtidos nos dois modelos foram considerados satisfatórios, mas foi escolhido, como modelo mais adequado para realizar as previsões, o modelo de regressão dinâmica, porque apresentou o menor erro percentual absoluto médio (MAPE). Por fim, as previsões encontradas, podem se constituir em uma ferramenta gerencial para a administração hospitalar ao fornecer subsídios para o planejamento orçamentário e financeiro dos recursos, especialmente em uma época em que há escassez de recursos em escala global, com reflexos muito intensos nos custos da saúde.

**Palavras-chave:** Regressão linear múltipla, regressão dinâmica, análise de componentes principais, custos em saúde.

## **ABSTRACT**

Mastership Dissertation  
Post-graduation Program in Production Engineering  
Federal University of Santa Maria

### **RESOURCE COLLECTION FOR ANTI-MICROBIAL AT THE UNIVERSITY HOSPITAL OF SANTA MARIA BY MEANS OF FORECASTS**

AUTHOR: CLAUDIO BASTOS  
SUPERVISOR: DR. ADRIANO MENDONÇA SOUZA  
Date and Local of Defense: Santa Maria, September 4th, 2009.

The scarce resources of public health makes the administrator manage the destination of resources, aiming to rationalize and optimize its collection, in order to improve the assistance to patients because the hospital is a public institution and does not get profits but promotes the community well-being. Thus, the hospital infection is acquired after the patient comes to the hospital or after he goes home and might be associated with his staying in hospital or with hospital procedures. This cost must be avoided. Once the complete eradication is not impossible, it is necessary to analyze and to control the monthly cost of the main antibiotics used for its treatment so that there is enough knowledge to foresee the resource collection to buy them. In this context, the main reason of this research is to carry out a forecast of the monthly cost and of the resource collection needed to purchase those medicine used in the treatment of hospital infections at the University Hospital of Santa Maria. To do so, a methodology for forecast by dynamic and multiple linear regressions was used. They were combined with to a multivariate technique by principal components. The technique of principal components was used to eliminate the multiple linearity existing among the original variants so, the resulting principal components were used as variables in the construction of the model of multiple linear regression and of dynamic regression. Therefore, these methodologies are applied to a case study of public health, in order to foresee and to conclude about which model is more suitable to forecast the monthly cost of antibiotics in hospital infections. The results obtained from the two models were considered satisfactory but the model of dynamic regression was chosen to be more suitable because it presented a mean absolute percentage error (MAPE). Finally, the findings might be a managerial tool for hospital administration when they offer subsidies for the budget of planning and of the resource finances, especially in a time when resources are globally scarce, making health even more expensive.

**Key words:** Multiple linear regression, dynamic regression, principal components analysis.

## LISTA DE TABELAS

Tabela 1 – Custo médio unitário anual .....	53
Tabela 2 – Custo mensal por ano, dos antibióticos, custo médio, desvio padrão e coeficiente de variação .....	57
Tabela 3 – Autovalores e variância explicada.....	60
Tabela 4 – Correlação das variáveis (antibióticos) com os fatores (componentes).....	61
Tabela 5 – Modelo de regressão linear múltipla .....	64
Tabela 7 – Previsão do custo mensal por regressão linear múltipla.....	67
Tabela 8 – Estimação dos parâmetros do modelo de regressão linear dinâmica .....	68
Tabela 9 – Previsão do custo mensal por regressão linear dinâmica .....	70
Tabela 10 – Previsão de Alocação dos Recursos Financeiros .....	71

## LISTA DE QUADROS

Quadro 1 – Nome do antibiótico e código adotado para a variável .....	45
Quadro 2 – Resumo dos testes de avaliação dos pressuposto do modelo de regressão linear múltipla.....	67
Quadro 3 – Resumo dos testes de avaliação dos pressuposto do modelo de regressão linear dinâmica.....	70

## LISTA DE ILUSTRAÇÕES

Figura 1 – Fluxograma da elaboração dos modelos de previsão .....	50
Figura 2 – Evolução do custo médio dos antibióticos .....	55
Figura 3 – Custo médio mensal dos antibióticos .....	57
Figura 4 – Demonstração gráfica dos autovalores .....	59
Figura 5 – Círculo de correlação unitário .....	62
Figura 6 – Histograma dos resíduos do modelo de regressão múltipla .....	66
Figura 7 – Histograma dos resíduos da regressão dinâmica.....	69

## **LISTA DE ABREVIATURAS E SIGLAS**

ACP - Análise de Componentes Principais  
ANVISA – Agência Nacional de Vigilância Sanitária  
BG – Breush - Godfrey  
CCIH – Comissão de Controle de Infecção Hospitalar  
CP – Componente Principal  
DIEESE – Departamento Intersindical de Estatística e Estudos Socioeconômicos  
DW – Durbin-Watson  
FGV – Fundação Getúlio Vargas  
FIPE – Fundação Instituto de Pesquisas Econômicas  
HUSM – Hospital Universitário de Santa Maria  
IBGE – Instituto Brasileiro de Geografia e Estatística  
IC - Infecção Comunitária  
IH – Infecção Hospitalar  
IPEAD – Instituto de Pesquisas Econômicas, Administrativas e Contábeis de Minas Gerais  
LM – Multiplicador de Langrange  
LSE – London School of Economics  
MAC - Maximal Allowable Cost  
MAPE - Mean Absolute Percentage Error  
MQO - Mínimos Quadrados Ordinários  
MS - Ministério da Saúde  
PCIH - Programa de Controle de Infecções Hospitalares  
SIE - Sistema de Informações para o Ensino  
SINAIS - Controle de Infecção em Serviços de Saúde  
SUS - Sistema Único de Saúde  
UFSM - Universidade Federal de Santa Maria

## LISTA DE APÊNDICES

Apêndice A – Coordenadas das variáveis e percentual de variância explicada .....	79
Apêndice B – Autovetores para compor a combinação linear .....	80
Apêndice C – Escores das componentes principais .....	81

## SUMÁRIO

LISTA DE TABELAS .....	7
LISTA DE QUADROS .....	8
LISTA DE ILUSTRAÇÕES.....	9
LISTA DE ABREVIATURAS E SIGLAS .....	10
LISTA DE APÊNDICES .....	11
1 INTRODUÇÃO .....	14
1.1 Tema da pesquisa.....	16
1.2 Justificativa e importância do trabalho .....	16
1.3 Problema de pesquisa.....	17
1.4 Objetivos .....	17
1.4.1 Objetivo geral.....	17
1.4.2 Objetivos específicos.....	17
1.5 Delimitação do tema.....	18
1.6 Estrutura do trabalho .....	18
2 REFERENCIAL TEÓRICO.....	19
2.1 Infecção hospitalar .....	19
2.2 Análise de regressão linear múltipla .....	24
2.3 Análise de regressão linear dinâmica.....	28
2.4 Método de seleção de variáveis .....	30
2.5 Pressupostos do modelo de regressão.....	31
2.6 Medidas de erro de previsão .....	37
2.7 Análise de componentes principais.....	38
3 METODOLOGIA.....	44
3.1 Caracterização da pesquisa .....	44
3.2 Identificação das variáveis envolvidas no estudo.....	44
3.3 Coleta dos dados.....	45
3.4 Tratamento dos dados .....	46
3.5 Métricas de análise.....	47
3.5.1 Estatísticas descritivas .....	47
3.5.2 Aplicação de modelos de regressão linear múltipla.....	47
3.5.3 Análise de componentes principais .....	48
3.5.4 Análise dos resíduos dos modelos estimados.....	48
3.6 Elaboração e análise das previsões .....	49
3.7 Fluxograma do desenvolvimento da metodologia .....	49
4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS .....	51
4.1 O Hospital Universitário de Santa Maria.....	51

<b>4.2 Mensuração do custo dos antibióticos .....</b>	<b>53</b>
<b>4.3 Análise das estatísticas descritivas.....</b>	<b>55</b>
<b>4.4 Elaboração do modelo de regressão linear múltipla.....</b>	<b>58</b>
<b>4.5 Análise de componentes principais.....</b>	<b>58</b>
<b>4.6 Estimação da regressão linear com as componentes principais.....</b>	<b>63</b>
<b>4.7 Estimação da regressão linear dinâmica .....</b>	<b>68</b>
<b>4.8 Escolha do modelo a partir do MAPE .....</b>	<b>71</b>
<b>4.9 Estimativas das previsões.....</b>	<b>71</b>
<b>5 CONSIDERAÇÕES FINAIS, LIMITAÇÕES E RECOMENDAÇÕES .....</b>	<b>73</b>
<b>REFERÊNCIAS.....</b>	<b>75</b>

# 1 INTRODUÇÃO

As infecções fazem parte da história da civilização, e o seu aparecimento é tão antigo quanto o surgimento dos hospitais, ocorrido por volta do ano 394 a.C. na periferia de Roma (OLIVEIRA, MARUYAMA, 2008).

Os primeiros indícios de infecção em nível “médico-paciente” remetem às sociedades do Antigo Oriente, mais precisamente no Egito, por volta do século IV a. C., onde os Sacerdotes eram orientados a banhar-se várias vezes ao dia, manter-se limpos e usar trajes brancos. Até meados do século XVIII, o hospital não era um local para o doente se curar, e sim, um local para assistência aos pobres que estavam morrendo (FOUCAULT, 1999 *apud* OLIVEIRA, MARUYAMA, 2008). A precariedade do atendimento desses hospitais, com muitas mortes por infecções adquiridas após a internação, perdurou até meados do século XVIII, período no qual se constatou que as pessoas doentes assistidas nos hospitais passavam a desenvolver outras doenças em função da hospitalização.

A infecção surge nos hospitais como uma consequência das precárias condições em que as pessoas eram dispostas e atendidas naqueles ambientes.

A partir do final do século XVIII, os hospitais iniciam sua gradual reestruturação, passando de um local de caridade para um local de cura, de observação, de saberes e de disciplina. Com a inserção dos médicos, as condições de atendimento melhoraram, tornando-os mais funcionais, internando doentes somente com indicação médica, para evitar a superlotação, contribuindo, ainda, para o ensino.

De acordo com Fernandes (2000), Semmelweis instituiu a rotina de higiene de mãos com solução clorada, conseguindo, com esse simples ato, reduzir as taxas de infecção de 11,4% para 1,3% em um período de sete meses. Também William S. Halsted (1852 – 1922) foi um grande inovador no combate às infecções da ferida operatória, introduzindo o uso de luvas como medida preventiva de infecção tanto do médico quanto do paciente.

De acordo com Melo (1986), em meados do século XIX, destaca-se a figura da enfermeira Florence Nightingale (1820-1910), na Inglaterra, que teve uma significativa importância histórica por meio do seu trabalho para melhorar a infraestrutura hospitalar, implantando medidas para o controle das infecções hospitalares, preocupação voltada para os cuidados de higienização, isolamento dos

enfermos, atendimento individual, utilização controlada da dieta e redução de leitos no mesmo ambiente. Foi quem primeiramente instituiu medidas de organização, sistematização do atendimento e treinamento de pessoal, especialmente práticas higiênicas sanitárias que colaboraram para a redução das taxas de mortalidade hospitalar da época.

O século XX trouxe consigo inúmeros avanços no campo da medicina, a exemplo de Alexandre Fleming (1881 – 1955), com a descoberta da penicilina, que iniciou a era antimicrobiana. Na década de 50, surgiram na Inglaterra as primeiras enfermeiras responsáveis exclusivamente por técnicas de controle de infecção hospitalar (*Infection Control Sister*). Em 1958, a *American Hospital Association* recomenda a criação de Comissões de Controle de Infecção Hospitalar (CCIH), com o objetivo de prover os hospitais americanos de um sistema que lhes permitisse apurar as causas das IH e dotá-los de instrumentos necessários contra possíveis ações legais movidas pela clientela.

A infecção hospitalar (IH), segundo Hoefel (2009), “é aquela que não estava presente e nem em incubação (se desenvolvendo sem se manifestar, em "silêncio") no momento em que o paciente internou no hospital”. Há também a Infecção Comunitária (IC), assim denominada por ser aquela constatada ou em incubação no ato da admissão do paciente, desde que não relacionada com internação anterior no mesmo hospital.

Portanto, se considera infecção hospitalar a adquirida após a internação do paciente, que se manifesta durante a internação ou mesmo após a alta, quando puder ser relacionada com a internação ou procedimentos hospitalares

A IH é um custo a ser evitado. Uma vez que sua total erradicação não é possível, se faz necessário analisar e controlar o custo mensal dos principais antibióticos utilizados no seu tratamento a fim de se ter embasamento suficiente para prever a alocação de recursos necessários a aquisição dos mesmos.

A aplicação deste estudo ocorre no Hospital Universitário de Santa Maria (HUSM) que, por ser um hospital de ensino, recebe os pacientes mais graves e, portanto, mais suscetíveis a desenvolver infecções hospitalares, ocasionando uma demanda considerável de antibióticos.

Além disso, a saúde pública no Rio Grande do Sul é um setor complexo e de grande importância que se encontra carente de ferramentas de gestão que

viabilizem novas formas de gerenciamento de alocação de recursos financeiros, pois a tecnologia cada vez mais presente na medicina impõe uma visão empresarial aos serviços de saúde, exige uma administração atuante, eficiente e comprometida com a melhoria da qualidade como um todo.

Com base em previsões, é elaborado o orçamento, que é uma ferramenta imprescindível para o gerenciamento de qualquer empreendimento e uma exigência legal em se tratando de órgãos públicos. Esse orçamento se constitui no elemento de referência para o planejamento e controle da aplicação dos recursos tanto públicos como privados e deve ser elaborado de forma ampla e detalhada, pois quanto mais precisas forem as previsões, maiores serão as possibilidades de realizar um planejamento adequado às necessidades da organização. Assim, oportuniza-se uma maneira de gerenciar os recursos previstos, avaliar necessidades de possíveis alterações e até, se necessário, remanejar os recursos financeiros para outros setores da instituição.

### **1.1 Tema da pesquisa**

A presente pesquisa trata da aplicação da análise de previsão do custo mensal e da previsão de alocação de recursos orçamentários no setor hospitalar para a aquisição de antimicrobianos (antibióticos), usados no combate às infecções hospitalares no Hospital Universitário de Santa Maria; utilizando-se modelos de regressão linear e dinâmica, aplicados às componentes principais.

### **1.2 Justificativa e importância do trabalho**

O Serviço de Contabilidade e Orçamento do HUSM tem como uma de suas funções prever e distribuir os recursos orçamentários e financeiros para que o hospital seja atendido em sua totalidade. Para isso, precisa dispor de informações e controles de como estão sendo utilizados esses recursos. Nesse contexto, o controle análise e previsão dos custos são imperiosos tanto para subsidiar a tomada de decisões pela administração do hospital quanto para atender às solicitações de informações de instâncias superiores da administração federal, bem como prestar contas à sociedade sobre atendimentos e aplicação dos recursos públicos.

A partir da necessidade de previsão dos recursos financeiros para a manutenção das atividades hospitalares e de reduzir gastos surgiu, a proposição de construir um modelo de previsão do custo mensal de antibióticos para o tratamento de infecções hospitalares.

Em função da importância e do papel desempenhado pelo hospital, pretende-se desenvolver modelos de regressão com a finalidade de fazer previsões desses custos.

### **1.3 Problema de pesquisa**

Como prever de forma eficiente os recursos para antimicrobianos no Hospital Universitário de Santa Maria por meio de projeções?

### **1.4 Objetivos**

Neste item descreve-se o objetivo geral e os objetivos específicos que nortearão esta pesquisa.

#### **1.4.1 Objetivo geral**

Realizar a previsão do custo mensal e de alocação de recursos necessários para aquisição de medicamentos utilizados no tratamento de infecções hospitalares, de forma a manter o fluxo de recursos orçamentários no setor de compras do Hospital Universitário de Santa Maria.

#### **1.4.2 Objetivos específicos**

- Analisar o custo dos principais antibióticos utilizados em infecções hospitalares;
- Prever a alocação de recursos orçamentários para aquisição de antibióticos para o tratamento de infecções hospitalares;
- Aplicar os modelos de regressão aos componentes principais oriundos dos dados originais;

- Determinar um critério de alocação de recursos baseados em previsões.

### **1.5 Delimitação do tema**

A pesquisa compreende a análise e previsão do custo mensal dos antibióticos utilizados no tratamento das infecções hospitalares, no período compreendido entre os anos de 2003 e 2008 do HUSM.

A análise fica restrita apenas aos custos dos antibióticos, sem investigar as taxas ou causas de ocorrências nem fazer comparações com outras instituições hospitalares.

### **1.6 Estrutura do trabalho**

Esta pesquisa foi estruturada em cinco capítulos. O primeiro traz a introdução, o tema da pesquisa, a justificativa e importância do trabalho, o problema de pesquisa, o objetivo geral e os específicos, a delimitação do tema e a estrutura do trabalho.

No capítulo dois, encontra-se a revisão de literatura, em que se apresenta uma revisão de alguns conceitos sobre infecção hospitalar, regressão múltipla, regressão dinâmica, método de seleção de variáveis, os pressupostos do modelo de regressão, as medidas de erro de previsão e sobre análise de componentes principais, pois esses serão os tópicos necessários para o desenvolvimento da pesquisa.

Na Metodologia, no capítulo três, são apresentadas as etapas necessárias para a coleta, tabulação e modelagem das variáveis e dos dados em estudo, representados pelos antibióticos mais utilizados no tratamento das infecções hospitalares no HUSM.

Já, no capítulo quatro, são demonstrados os resultados da análise do custo dos antibióticos utilizados no combate às infecções hospitalares e a previsão de alocação dos recursos orçamentários para sua aquisição para três períodos à frente.

As considerações finais e as sugestões constam no capítulo cinco e, após esse, estão relacionadas as referências bibliográficas e os apêndices.

## **2 REFERENCIAL TEÓRICO**

Neste capítulo apresenta-se a revisão de alguns conceitos sobre infecções hospitalares, regressão linear múltipla, regressão dinâmica, método de seleção de variáveis, pressupostos do modelo de regressão, medidas de erro de previsão e sobre análise de componentes principais, pois esses serão os tópicos necessários para o desenvolvimento desta pesquisa.

### **2.1 Infecção hospitalar**

Considera-se infecção hospitalar a adquirida após a internação do paciente, que se manifesta durante a internação ou mesmo após a alta, quando puder ser relacionada com a internação ou procedimentos hospitalares. Segundo Hoefel (2009), a “infecção hospitalar é aquela que não estava presente e nem em incubação (se desenvolvendo sem se manifestar, em "silêncio") no momento em que o paciente internou no hospital”. Há também a Infecção Comunitária (IC), assim denominada por ser aquela constatada ou em incubação no ato da admissão do paciente, desde que não relacionada com a internação anterior no mesmo hospital.

Conforme Hoefel (2009), alguns grupos de pacientes possuem maior risco de contrair infecções hospitalares do que outros, por exemplo, quem está nos extremos das faixas etárias, ou seja, idosos e recém nascidos: os idosos ,em função de que os diversos sistemas do organismo, aos poucos, vão reduzindo sua perfeita capacidade funcional, e os recém-nascidos, por sua imunidade ainda não completamente desenvolvida. Enquadram-se no grupo de maior risco de contrair infecções hospitalares as pessoas diabéticas, leucêmicas, portadoras de câncer, usando quimioterápicos e corticóides; pacientes sob “stress”, com baixa imunidade ou com problemas neurológicos.

As infecções hospitalares são transmitidas de um paciente a outro, principalmente a partir do contato de sujeiras, secreções e eliminações de outras pessoas ou por contato com materiais sujos ou não esterilizados utilizados na manipulação destas secreções e eliminações. De acordo com Hoefel (2009), além do contato, as infecções podem se transmitir dentro do hospital pelo ar circulante, ou, dependendo do tipo de doença, a transmissão pode ocorrer pela respiração, mais raro de acontecer, ou por espirro ou tosse.

Além disso, a higienização, como um todo, é fundamental, pois locais sujos também podem ser atrativos de insetos e roedores, causando a transmissão de doenças por meio destes vetores de infecção e, por último, há a transmissão por fonte comum, que ocorre quando diversos pacientes se submetem ao mesmo tipo de tratamento ou a alimentos contaminados.

Portanto, devido às inúmeras possibilidades de se contrair uma infecção hospitalar é que todos os hospitais tem, por determinação legal, um Programa de Controle de Infecção Hospitalar.

No Brasil, nas últimas décadas do século XIX, os índices de infecção verificados entre os pacientes submetidos a procedimentos cirúrgicos chegavam a até 90%, e suas causas se relacionavam à falta de regras básicas de higiene e de isolamento dos doentes portadores de doenças infecciosas, causando o que se denominava na época de “gangrenas dos hospitais”, representando a principal causa de mortalidade nos hospitais brasileiros (FERNANDES, 2000).

A carência de normas e protocolos para o tratamento das IH, nos hospitais brasileiros, motivou a criação da primeira Comissão de Controle de Infecção Hospitalar (CCIH) do Brasil, criada no ano de 1963, no Hospital Ernesto Dornelles, no Estado do Rio Grande do Sul. Essa comissão foi um marco precursor do controle de infecção hospitalar no país, levando, em 1976, o governo federal a determinar a necessidade de criação de CCIH's nos hospitais próprios da previdência.

Infelizmente essa medida não causa impacto pela falta de fiscalização e, conforme relatos da ANVISA (2005), somente em 1983 foi editada a Portaria nº. 196 do Ministério da Saúde (MS) determinando a criação e a normatização das Comissões de Controle de Infecções Hospitalares. Dois anos mais tarde, em 1985, a morte do recém-eleito Presidente da República Tancredo Neves por septicemia, devido a uma infecção pós-cirúrgica, causou grande repercussão em nível nacional, contribuindo para que o Ministério da Saúde publicasse o “Manual de Controle de Infecção Hospitalar” com a finalidade de delinear as medidas básicas da prevenção e controle das infecções nos hospitais brasileiros.

Passados dois anos desta publicação, em 1987, o Ministério da Saúde criou o Programa Nacional de Controle de Infecção Hospitalar para posteriormente, em 1992, determinar, através da portaria nº. 930, que todos os hospitais do país

deveriam manter um Programa de Controle de Infecção Hospitalar, independente da natureza da entidade mantenedora.

Além disso, a deficiência de indicadores de infecções hospitalares, segundo Prade (2001), levou a ANVISA a desenvolver o Sistema de Informações para o Controle de Infecção em Serviços de Saúde (SINAIS), cujo objetivo é conhecer o perfil epidemiológico e as taxas de infecções hospitalares nos hospitais, buscando uniformizar e padronizar os indicadores com possibilidade de acompanhamento, além de servir como instrumento de orientação para implantação das ações que visam a diminuir sua incidência e gravidade nos serviços de saúde, medir sua eficácia e monitorar a qualidade da assistência hospitalar e riscos.

Com isso, é possível perceber as inúmeras dificuldades e as necessidades de investimento em controle de infecção hospitalar permitindo um uso mais racional e menos oneroso de recursos materiais e humanos.

Já a Comissão de Controle de Infecção Hospitalar é um órgão de assessoria diretamente vinculado à autoridade máxima da instituição, encarregada da elaboração, implantação e avaliação do Programa de Controle de Infecções Hospitalares (PCIH) adequado às características e necessidades da Unidade Hospitalar. Segundo a Agência Nacional de Vigilância Sanitária, o PCIH é um conjunto de ações desenvolvidas deliberada e sistematicamente, com vistas à redução máxima possível da incidência e da gravidade das infecções hospitalares.

O objetivo da Comissão de Controle de Infecção Hospitalar (CCIH) não é somente de prevenir e combater a infecção hospitalar, beneficiando toda a população assistida, mas também proteger o hospital e o corpo clínico. Na busca desse objetivo, desenvolve atividades de orientação quanto à técnica de lavagem das mãos, considerada a medida mais eficaz no controle de infecção hospitalar, e vigilância epidemiológica, através de busca ativa, em que são coletados dados de todos os pacientes internados no hospital com o objetivo de identificar aqueles que adquiriram infecção hospitalar.

Além disso, a CCIH realiza investigação epidemiológica de casos e surtos, sempre que indicado, e implanta as medidas imediatas de controle, elabora relatórios e comunica periodicamente à autoridade máxima da instituição e às chefias de todos os setores do hospital a situação do controle das infecções hospitalares, promovendo seu amplo debate na comunidade hospitalar. Sendo

também de sua competência elaborar, implantar e supervisionar a aplicação de normas e rotinas técnico-operacionais, visando limitar a disseminação de agentes presentes nas infecções em curso no hospital por meio de medidas de precaução e de isolamento.

Também cabe à CCIH notificar, ao Serviço de Vigilância Epidemiológica e Sanitária do organismo de gestão do SUS, os casos e surtos diagnosticados ou suspeitos de infecção associados à utilização de insumos e/ou produtos industrializados.

Em relação aos cuidados de prevenção, a Sociedade Brasileira de Infectologia (2001) e Hoefel (2009) destacam que o principal modo de evitar a transmissão das infecções hospitalares é lavar as mãos antes e após o contato com os pacientes ou antes ou após realizar atividades que possam resultar em contaminações. A limpeza do ambiente e dos materiais que o paciente irá utilizar deve ser efetiva para eliminar patógenos que por ventura ali se alojem, assim como é obrigatória a utilização de luvas estéreis e não estéreis pelos profissionais que mantêm contato com os pacientes.

Além disso, é necessário o controle de qualidade de todos os processos envolvidos que possam suscitar o aparecimento ou favorecimento de IH's, a fim de que não ocorram falhas expondo o paciente, pois o risco de contrair infecção hospitalar existe para todo e qualquer paciente internado.

#### - Custos e infecções hospitalares

As políticas de antimicrobianos visam basicamente ao controle de custos e da resistência microbiana. Entretanto, implantá-las sem engajamento da equipe de saúde ou sem o respaldo de uma criteriosa análise da literatura médica pode, ao contrário, elevar as despesas e selecionar cepas resistentes. O custo anual associado à resistência antimicrobiana nos Estados Unidos é aproximadamente US\$ 47 bilhões. De acordo com Welch et al. (1999), em um estudo realizado em hospitais de Nova York, Colorado e Ohio, pacientes com bacteremia causada por *Staphylococcus aureus meticillino - resistente* ficaram internados em média 2,7 dias a mais do que pacientes com cepas sensíveis e representaram um custo adicional de US\$ 3.500,00 por caso.

O controle de infecção deve ser à base de atuação de todos os profissionais de saúde. Nos Estados Unidos, o sistema pré-pago de reembolso das despesas de saúde torna fundamental esta abordagem.

Conforme estudos de Frank et al. (1997), nos Estados Unidos uma primeira tentativa para controlar custos foi o projeto MAC (*Maximal Allowable Cost*) ou Custo Máximo Permissível, vinculado ao *Medicare*. Sua implantação relacionou-se a numerosos problemas clínicos e burocráticos, surgindo formulários específicos para determinados medicamentos, necessitando de aprovação posterior.

No Brasil, a tendência crescente da elaboração de pacotes assistenciais em substituição ao tradicional “paga-se o que gastou”, também pressiona economicamente nesse sentido. Dada a estrutura de reembolso que prevalece hoje, é provável que os hospitais com taxas altas de infecções hospitalares por germes multirresistentes percam dinheiro.

Durante décadas, mais especificamente até a implantação do plano Real, o Brasil conviveu com o aumento descontrolado do nível de preços, com taxas inflacionárias elevadas, a exemplo do mês de março de 1989, quando a inflação chegou a 81%. Hoje, a situação econômica apresenta-se mais estável, a inflação não deixou de existir, porém se encontra em níveis mais toleráveis, ocorrendo algumas vezes o inverso da inflação, denominada de deflação. Isso ocorre normalmente quando o consumo diminui e as indústrias e fornecedores reduzem os preços com a esperança de estimular o consumo.

Entretanto, em períodos de inflação ou deflação, o valor original de aquisição de um produto ou serviço não é um bom padrão de avaliação para efeitos de análise, devendo ser adotado um critério que reflita o valor atual desses produtos ou serviços, de modo a minimizar os efeitos da inflação ou deflação ocorrida no período estudado.

De acordo com Pereira e Ramalho (1998), o processo de transformar valores nominais ou correntes em valores constantes ou reais, denomina-se deflacionamento. Por meio desse processo, é possível equiparar valores de diversas datas ao valor de uma data-base fixada de acordo com critérios pré-estabelecidos. Este método consiste basicamente em dividir os valores nominais de diversas épocas por um deflator (que nada mais é do que um índice de preços com base

fixa). A dificuldade, portanto, não reside no processo de deflacionamento, mas sim na escolha de qual, dentre diversos índices, será o melhor deflator.

Lopes (1985) afirma que todos os países possuem organismos que mantêm de forma regular os vários índices de preços, com o objetivo de mensurar da melhor forma possível as tensões inflacionárias. São calculados diversos tipos de índices de custo de vida no Brasil, por vários organismos, dentre os quais se destacam a Fundação Getúlio Vargas (FGV), o Instituto Brasileiro de Geografia e Estatística (IBGE), o Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE), a Fundação Instituto de Pesquisas Econômicas (FIPE), o Instituto de Pesquisas Econômicas, Administrativas e Contábeis de Minas Gerais (IPEAD) dentre outros.

Essa diversidade de índices disponíveis dificulta a escolha de um que reflita a realidade do custo mensal dos antibióticos dos anos anteriores a 2008.

## **2.2 Análise de regressão linear múltipla**

O termo regressão foi introduzido no final do século XIX, por Francis Galton, cujos estudos mostraram que, embora houvesse uma tendência de pais altos terem filhos altos e de pais baixos terem filhos baixos, a altura média dos filhos de pais de uma dada altura tendia a se deslocar ou “regredir” até a altura média da população como um todo. Passados alguns anos, a lei de regressão universal de Galton foi confirmada por seu amigo Karl Pearson, com uma amostra de mais de mil registros das alturas dos membros de grupos familiares (GUJARATI, 2006).

Segundo Anderson, Sweeney e Williams (2005), análise de regressão “é o estudo que analisa como uma variável dependente é relacionada com duas ou mais variáveis independentes”.

Conforme explica Gujarati (2006), como o método se prestava à estimação de relações, passou a ser utilizado para o estudo da dependência de uma variável, em relação a uma ou mais variáveis, com o objetivo de estimar e/ou prever média (da população) ou o valor médio da variável endógena, também denominada de dependente, em termos dos valores conhecidos ou fixos (em amostragem repetida) das exógenas, ou variáveis explicativas. Landim (2000) diz que “uma das mais

importantes aplicações da análise de regressão múltipla é a escolha, entre diversas variáveis independentes, daquelas mais úteis na previsão de  $Y$ ".

A interpretação da análise de regressão tem o objetivo de descrever, a partir de um modelo matemático, a relação existente entre uma variável dependente e uma ou mais variáveis independentes. O emprego dos termos "variáveis dependentes" e "variáveis independentes" decorre da relação matemática entre elas, não implicando necessariamente em uma ser a causa da outra.

A operacionalização do modelo se dá por meio do cálculo dos valores das constantes desconhecidas da equação matemática que relaciona as variáveis, denominadas parâmetros da equação ( $\beta_j$ ) e, então, projeta estimativas de valores futuros para a série, através de modelos matemáticos. Considerando que um fenômeno  $Y$  seja caracterizado perfeitamente por todas as seguintes variáveis  $x_i$ , tem-se:

$$Y = h(x_1, x_2, \dots, x_m, x_{m+1}, \dots, x_{m+n}) \quad (1)$$

Dessa forma, admitindo que se possa dividir  $Y = h(x)$  nas seguintes partes:

$$Y = f(x_1, x_2, \dots, x_m) + g(x_{m+1}, x_{m+2}, \dots, x_{m+n}) \quad (2)$$

Supondo que  $g(x)$  assume valores aleatórios, não facilmente determináveis e insignificantes frente à  $f(x)$ . Então, é possível inferir que o fenômeno  $Y$  é razoavelmente bem explicado apenas pela componente  $f$  e que a componente  $g$ , considerada aleatória e irrelevante, será o erro ( $\varepsilon$ ) ou resíduo da equação, deste modo:

$$Y = f(x_1, x_2, \dots, x_m) + \varepsilon \quad (3)$$

A escolha do melhor modelo para a equação de regressão vai depender da relação entre as variáveis dependente e as independentes. Segundo Milone & Angelini (1995), isto é feito pela análise do diagrama de dispersão, através do erro padrão ou mesmo pelos conhecimentos *a priori* do pesquisador.

Neste estudo utiliza-se a análise de regressão múltipla linear, cuja equação matemática é expressa por:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \hat{\beta}_2 X_{2,i} + \dots + \hat{\beta}_m X_{m,i} + \varepsilon_i \quad (4)$$

Nesta equação os  $\hat{\beta}_j$  são estimativas para os valores reais dos parâmetros  $\beta_j$ .

Por extensão, a equação (4), considerando as  $n$  observações, pode ser escrita da seguinte forma:

$$\begin{aligned} Y_1 &= \hat{\beta}_0 + \hat{\beta}_1 X_{1,1} + \hat{\beta}_2 X_{2,1} + \dots + \hat{\beta}_m X_{m,1} + \varepsilon_1 \\ Y_2 &= \hat{\beta}_0 + \hat{\beta}_1 X_{1,2} + \hat{\beta}_2 X_{2,2} + \dots + \hat{\beta}_m X_{m,2} + \varepsilon_2 \\ &\dots \\ Y_n &= \hat{\beta}_0 + \hat{\beta}_1 X_{1,n} + \hat{\beta}_2 X_{2,n} + \dots + \hat{\beta}_m X_{m,n} + \varepsilon_n \end{aligned} \quad (5)$$

Estas  $n$  equações podem ser escritas de uma forma matricial, mais compacta.

$$Y = X\beta + \varepsilon \quad (6)$$

Onde,

$$Y = [Y_1 \ Y_2 \ Y_3 \ \dots \ Y_m]^T \quad (7)$$

é a matriz com as  $n$  observações de  $Y$ , e

$$X = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} & \dots & X_{m,1} \\ 1 & X_{1,2} & X_{2,2} & \dots & X_{m,2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{1,n} & X_{2,n} & \dots & X_{m,n} \end{bmatrix} \quad (8)$$

é a matriz das  $n$  observações das  $m$  variáveis independentes.

Os vetores  $\beta$  e  $\varepsilon$  são:

$$\beta = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \dots \quad \beta_m]^t \quad (9)$$

$$\varepsilon = [\varepsilon_1 \quad \varepsilon_2 \quad \varepsilon_3 \quad \dots \quad \varepsilon_n]^t \quad (10)$$

Após a especificação do modelo, deve-se estimar o valor dos diversos parâmetros  $\hat{\beta}_j$  da equação de regressão. Com esta estimação, deseja-se que a curva  $\hat{Y}$  seja tão próxima quanto possível do conjunto de pontos reais ( $Y$ ), isto é, deseja-se minimizar a discrepância total entre os pontos marcados e a curva estimada.

As estimativas para os parâmetros da regressão linear são obtidas pelo método dos Mínimos Quadrados Ordinários (*MQO*). De acordo com Gujarati (2006), o método dos *MQO* possui algumas propriedades estatísticas interessantes, que o fizeram um dos métodos mais eficientes e usados na análise de regressão.

Esse método de estimação minimiza a soma dos quadrados das distâncias da curva  $\hat{Y}$  aos pontos experimentais  $Y_i$ , ou seja, deve-se encontrar uma curva que minimize a soma dos quadrados dos resíduos.

Na análise de regressão, o objetivo não é somente obter a estimação dos parâmetros, mas, também, fazer inferências aos seus verdadeiros valores, portanto, de acordo com Landim (2000), alguns cuidados devem ser tomados quando da utilização da análise de regressão como demonstrado a seguir:

- As relações entre as variáveis devem ser lineares, ou seja, os acréscimos em  $X$  produzem acréscimos proporcionais em  $Y$ , e a razão de crescimento é constante;
- Evitar um número inferior de casos em relação ao número de variáveis consideradas; é recomendado que tal relação seja da ordem de 10 a 20 vezes superior;
- Evitar variáveis independentes redundantes, isto é, que tenham um alto coeficiente de correlação entre si;
  - Verificar, utilizando resíduos, a presença de valores anômalos ou de viés;
  - Os resíduos não devem estar correlacionados;

- Os valores de  $X_i$  são fixados, isto é,  $X_i$  não é uma variável aleatória;
- A média dos resíduos deve ser nula, isto é,  $E(\varepsilon_i) = 0$ .

Além dos cuidados citados anteriormente, devem ser verificados os pressupostos do modelo de regressão.

### 2.3 Análise de regressão linear dinâmica

Denomina-se modelo de regressão dinâmica o modelo matemático que inclui não somente valores correntes, mas também os defasados (passados) das variáveis dependentes.

De acordo com Gujarati (2006), os modelos dinâmicos são considerados muito importantes na análise econométrica, porque consideram o comportamento de uma variável no decorrer do tempo.

Esta metodologia se desenvolveu a partir das pesquisas de Sargan, na década de 60, Davidson e Mizon e por vários ensaios publicados por Hendry a partir dos anos 70. Consiste em uma abordagem baseada num processo de redução sucessiva de um modelo econométrico geral, começando de um modelo estatístico dinâmico, que captura as características essenciais do conjunto de dados. O ponto de partida deve ser um modelo dinâmico com uma ordem de defasagem grande o suficiente para incluir a defasagem verdadeira das variáveis. É importante se tentar evitar, durante as reduções, a perda de informações importantes. Para isso, são usados testes para reduzir a complexidade do modelo, eliminando-se as variáveis não significativas estatisticamente, verificando-se a redução em cada estágio para garantir a validade do modelo a ser escolhido. Até chegar ao melhor modelo são, efetuadas várias séries de transformações e reduções.

A avaliação dessas regressões pode ser feita considerando a exigência de algumas restrições lineares no modelo e, gradativamente, à luz dos testes *t-Student*, determina-se um modelo parcimonioso (simples) e bem especificado a ser usado para fins de previsão.

Neste estudo a estimação de todos os modelos obedece à metodologia do geral para o específico *the General to Specific* da Escola de Economia de Londres (*London School of Economics*), ou metodologia *Hendry Approach*, cuja terminologia

ficou popularizada em função do grande esforço de convencimento de seu maior defensor na academia, o econometrista escocês David Hendry (SOUZA *et al.*, 2004).

Fildes (1985), recomenda a utilização de modelos econométricos para a elaboração de previsões mais ajustadas. Tais modelos normalmente compreendem sistemas de relacionamento entre variáveis de interesse, em que os relacionamentos são estimados a partir dos dados disponíveis.

Considerando o comportamento, no decorrer do tempo, dos valores defasados das variáveis, utiliza-se a seguinte equação dinâmica:

$$\begin{aligned} \hat{Y}_t = & \hat{\beta}_0 + \hat{\gamma}_1 Y_{t-1} + \dots + \hat{\gamma}_k Y_{t-k} + \hat{\beta}_{1,t} X_{1,t} + \hat{\beta}_{1,t-1} X_{1,t-1} + \\ & \dots + \hat{\beta}_{1,t-k} X_{1,t-k} + \hat{\beta}_{2,t} X_{2,t} + \hat{\beta}_{2,t-1} X_{2,t-1} + \dots + \varepsilon_t \end{aligned} \quad (11)$$

Onde os índices  $t-i$  são as variáveis e parâmetros com  $i$  defasagens.

A estimação dos parâmetros da equação (11) pode ser feita utilizando-se o Método dos Mínimos Quadrados Ordinários (MQO) apresentado anteriormente, desde que o termo estocástico  $\varepsilon_t$  tenha as propriedades de: média zero ( $E(\varepsilon_t) = 0$ ), homocedasticidade ( $Var(\varepsilon_t) = \sigma^2$ ) e não-correlação entre os erros ( $E(\varepsilon_i, \varepsilon_j) = 0$ , se  $i \neq j$ ).

Para testar cada parâmetro, são formuladas duas hipóteses:

- $H_0 : \beta = 0$ , não existe regressão (parâmetro não é significativo)
- $H_1 : \beta \neq 0$ , existe regressão (parâmetro é significativo)

Fixando  $\alpha = 5\%$ , e, neste caso, utilizando  $t$  com  $n - 2$  graus de liberdade (pois se tem dois parâmetros), onde a estatística do teste de hipóteses é dada por:

$$t_c = \frac{(b - \beta)}{\left( \frac{s}{\sqrt{S_{xx}}} \right)} \quad (12)$$

Onde:

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} \quad (13)$$

Desta forma, conclui-se que, se  $-t_{\alpha/2} < t_c < t_{\alpha/2}$ , então, aceita-se  $H_0$  com risco  $\alpha$  de que o parâmetro não é significativo.

Os parâmetros não significativos vão sendo paulatinamente retirados do modelo, ficando apenas os parâmetros significativos, o que torna o modelo, parcimonioso.

Neste caso, parte-se de um modelo matemático que leva em conta uma elevada estrutura de defasagens em todas as variáveis, na qual, utilizando a metodologia do geral para o específico, encontra-se um modelo de regressão dinâmica, apenas com parâmetros significativos.

## 2.4 Método de seleção de variáveis

Geralmente os *softwares* estatísticos oferecem mais de um método de seleção de variáveis para construir um modelo de regressão linear múltipla, onde uma das opções é o método de estimação *stepwise*, também chamado de passo a passo ou por etapas.

De acordo com Corrar, Paulo e Dias Filho (2007), este método “possibilita examinar a contribuição adicional de cada variável independentemente ao modelo, pois cada variável é considerada para inclusão antes do desenvolvimento da equação”.

Segundo Abbad (2002), regressão *stepwise* geralmente é a opção escolhida para estudos exploratórios, principalmente quando o pesquisador pretende descrever relacionamentos pouco conhecidos entre as variáveis.

Esse tipo de regressão começa com um modelo de regressão simples, em que a primeira variável preditora selecionada é aquela que apresentar o maior coeficiente de correlação com a variável dependente, ou seja, a sequência de entrada das variáveis independentes é feita estatisticamente; eliminando a necessidade de elaborar um modelo teórico a ser seguido (CORRAR, PAULO e DIAS FILHO, 2007, ABBAD, 2002).

A seleção das próximas variáveis preditoras é efetuada com base na correlação parcial com a equação de regressão, sendo testada a significância com as demais variáveis já incluídas no modelo pelo teste F de *Fischer*. Caso não exista

significância, a estimação *stepwise* permite que as variáveis que já estão no modelo sejam eliminadas.

Neste trabalho utiliza-se a opção de estimação *stepwise* disponível no *software* estatístico.

## 2.5 Pressupostos do modelo de regressão

Nos subitens a seguir, serão apresentados os pressupostos e os testes para validação dos modelos de regressão linear múltipla e dinâmica.

### - Homocedasticidade dos resíduos

Segundo Sartoris (2003), a homocedasticidade é uma propriedade importante e deve ser garantida, sob pena de invalidar toda a análise estatística. Ocorre quando se tem variância constante dos resíduos. Os erros devem ser aleatórios, ou seja, não devem ser relacionados. A ausência de homocedasticidade origina a heterocedasticidade. As consequências da heterocedasticidade são que as estimativas dos parâmetros da regressão ( $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$ ) não são tendenciosas, mas são ineficientes, e as estimativas das variâncias, por sua vez, são tendenciosas. Os testes “*t*” de *Student* e *F* de *Fischer* tendem a dar resultados incorretos. Neste caso, os resultados não são confiáveis, ou seja, o modelo pode parecer bom, mas ele não é adequado ao conjunto de dados.

Para detectar a homocedasticidade dos resíduos, pode-se usar o teste geral de heterocedasticidade de White, em que de acordo com Gujarati (2006), é feita uma regressão com os resíduos ao quadrado da regressão original sobre as variáveis originais (*X*) ou regressores, seus valores elevados ao quadrado e seus produtos cruzados, conforme a seguir:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i \quad (14)$$

A seguir, deve-se obter o  $R^2$  desta regressão auxiliar e testar as seguintes hipóteses:

- $H_0$  = os resíduos são homocedásticos;

- $H_1$  = os resíduos não são homocedásticos.

Então, sob a hipótese nula de homocedasticidade, pode-se mostrar que o tamanho da amostra ( $n$ ) multiplicado pelo ( $R^2$ ), obtido da regressão auxiliar, segue uma distribuição qui-quadrado com número de graus de liberdade igual ao número de regressores (excluindo a constante) na regressão auxiliar, definido pela seguinte expressão:

$$n.R^2 \cong \chi^2_{gl} \quad (15)$$

Na interpretação do teste de hipóteses, se o valor de qui-quadrado calculado em (15) for maior do que o valor qui-quadrado tabelado, conclui-se que há heterocedasticidade. Se for menor, não há heterocedasticidade e, sim, homocedasticidade dos resíduos.

Em alguns pacotes de *softwares* estatísticos, este teste é informado como *hetero test*.

#### - Normalidade dos resíduos

Em sua obra, Levine (2005) ensina que a Análise de Regressão, baseia-se na hipótese de que os erros seguem uma distribuição normal (distribuição de Gauss). Em presença de falta de Gaussianidade, os estimadores são não tendenciosos, mas os testes não têm validade, principalmente em amostras pequenas.

A hipótese da normalidade é justificada teoricamente pelo Teorema Central do Limite da Estatística. Este teorema diz que, se houver um grande número de variáveis aleatórias independentes e identicamente distribuídas (representadas pelo erro  $\varepsilon$  no modelo), então, a distribuição de sua soma tenderá para uma distribuição normal à medida que o número dessas variáveis for aumentando indefinidamente.

A causa da não-normalidade dos resíduos é possível de se dar pela heterocedasticidade destes, ou pela escolha de um modelo incorreto para a equação.

A normalidade pode ser verificada pelo histograma dos resíduos, pois é um dispositivo gráfico que permite verificar a forma da função, densidade de probabilidade dos resíduos. Quanto mais próximo à curva teórica normal estiver a curva dos resíduos, mais normais serão estes.

Também pode ser aplicado o teste de normalidade Jarque-Bera, que consiste numa estatística que segue a distribuição qui-quadrado com 2 graus de liberdade, definida pela expressão (16)

$$JB = n \left[ \frac{S^2}{6} + \frac{(k-3)^2}{24} \right] \quad (16)$$

Onde:

$S$  = coeficiente de assimetria;

$K$  = coeficiente de curtose;

$n$  = tamanho da amostra

As hipóteses a serem formuladas são:

- $H_0$ : Os resíduos se distribuem normalmente;
- $H_1$ : Os resíduos não se distribuem normalmente.

O teste Jarque-Bera é utilizado em alguns *softwares* estatísticos para testar a normalidade dos resíduos sendo denominado como *Normality test*.

- Autocorrelação serial dos resíduos

De acordo com Gujarati (2006), o modelo clássico de regressão pressupõe que o termo de erro relacionado a qualquer uma das observações não é influenciado pelo termo de erro de qualquer outra observação.

O fenômeno da autocorrelação ocorre quando os erros são correlacionados com os valores anteriores ou posteriores na série, podendo ocorrer em função da especificação incorreta do modelo da regressão por causa de erros na forma do modelo ou, ainda, por exclusão de variáveis independentes importantes. Existindo autocorrelação residual, os estimadores ordinários de mínimos quadrados não são mais os melhores estimadores lineares não tendenciosos. Neste caso, existirão outros métodos que produzem menor variância amostral nos estimadores. Além disso, em presença de correlação serial, os testes de significância ( $t$  e  $F$ ) e de construção de intervalos de confiança dos coeficientes da regressão também oferecem conclusões incorretas, isto é, as regiões de aceitação e os intervalos de confiança podem ser mais largos ou mais estreitos do que os calculados, dependendo da tendência ser positiva ou negativa.

A autocorrelação pode ser verificada pela aplicação do teste de Durbin-Watson (DW) que, de acordo com Gujarati (2006) e Silva (2005), é considerado o mais comum para o diagnóstico da autocorrelação serial de primeira ordem. O teste não inclui variáveis dependentes defasadas, e é necessário incluir a constante na regressão. Se não estiver presente, deve-se refazer a regressão para incluí-la.

A estatística para o teste (DW) pode ser definida pela seguinte expressão matemática:

$$DW = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2} \quad (17)$$

Conforme Corrar (2007), a estatística DW é tabelada para valores críticos segundo o nível de confiança escolhido, sendo formuladas as seguintes hipóteses:

- $H_0$ : Não existe correlação serial dos resíduos;
- $H_1$ : Existe correlação serial dos resíduos.

Considerando o tamanho da amostra e o número de variáveis independentes, são estabelecidos os valores críticos inferior ( $d_L$ ) e crítico superior ( $d_U$ ). Então, se  $0 < d < 2$ , existe algum grau de autocorrelação positiva, sendo mais forte à medida que ( $d$ ) se aproxima de zero; e se  $2 < d < 4$ , a autocorrelação serial será negativa, ficando mais elevada à medida que ( $d$ ) se aproxima de 4.

Corrar (2007) sugere uma “regra de bolso”, a qual diz que valores de estatística DW próximos a 2 atendem ao pressuposto.

Em modelos de regressão dinâmicos, com valores defasados do regressando, deve ser usado outro teste como o teste de Breush-Godfrey (BG), também conhecido como teste de Multiplicador de Lagrange (LM) e, em alguns *softwares* estatísticos, como teste ARCH. O teste de hipóteses inicia pelo cálculo do  $R^2$  da seguinte regressão:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 \varepsilon_{t-1} + \beta_3 \varepsilon_{t-2} + \dots + \beta_k \varepsilon_{t-m} + \mu_t \quad (18)$$

As hipóteses a serem formuladas são:

- $H_0$ :  $n \cdot R^2 = 0$  Não existe autocorrelação serial dos resíduos;

- $H_1: n \cdot R^2 > 0$  Existe autocorrelação serial dos resíduos

A estatística de teste do teste LM é avaliada pela seguinte equação:

$$LM = n \cdot R^2 \cong \chi^2_{k-1} \quad (19)$$

Onde  $k$  é o número de regressores na regressão formulada.

#### - Multicolinearidade

Segundo Malhotra (2001), a multicolinearidade ocorre quando há relações lineares entre as variáveis independentes, de tais formas correlacionadas umas com as outras, tornando difícil ou até impossível isolar suas influências e obter uma estimativa precisa de seus efeitos relativos. Na maioria das vezes, não se encontram variáveis independentes perfeitamente relacionadas. Esse fato não traz problemas, pois é facilmente detectado e pode ser resolvido a partir da eliminação de uma ou mais variáveis independentes do modelo. O maior problema no que tange à multicolinearidade está nos casos em que ela ocorre com alto grau, isto é, quando duas variáveis independentes estão significativamente correlacionadas, ou em muitos casos quando há uma combinação linear entre um conjunto de variáveis independentes.

De acordo com Moreira (2006), identificar a presença de multicolinearidade em pouco contribui no sentido de saber se isso representa um problema ou não para o modelo. É preciso verificar a magnitude da multicolinearidade para saber se ela é severa, o que exigirá tratamento; em caso contrário, se for de valor reduzido, não se exigirá esse procedimento.

O fato de muitas funções e regressões diferentes proporcionarem bons ajustes para um mesmo conjunto de dados ocorre porque os coeficientes de regressão atendem várias amostras em que as variáveis independentes são altamente correlacionadas. Quando isso acontece as variáveis independentes estão altamente correlacionadas, os coeficientes de regressão estimados variam de uma amostra para outra, levando as informações a serem imprecisas a respeito dos coeficientes verdadeiros.

Geralmente a multicolinearidade é causada pela própria natureza dos dados, principalmente nas áreas de economia com variáveis referentes a valores mercadológicos. Em Análise de Regressão Linear Múltipla, existe normalmente um interesse com relação à natureza e significância das relações entre as variáveis independentes e a variável dependente. Em muitas aplicações de administração e economia, frequentemente se encontram variáveis independentes que estão correlacionadas entre si e, também, com outras variáveis que não estão incluídas no modelo, mas estão relacionadas à variável dependente.

Moreira (2006), em seu artigo, afirma que a presença de multicolinearidade severa faz com que os coeficientes não reflitam nenhum efeito particular da variável explicativa na variável dependente, pois o coeficiente estará contaminado pelo efeito de outras variáveis e apenas um efeito parcial lhe poderá ser atribuído.

A multicolinearidade pode ser detectada e calculada pela maioria dos *softwares* que fazem regressão múltipla e, de acordo com Gujarati (2006), é muitas vezes identificada pelo alto valor do coeficiente de determinação  $R^2$  da regressão. Em geral se verifica multicolinearidade quando a estatística  $F$  da regressão é altamente significativa, mas o valor da estatística  $t$  não aponta algum  $\hat{\beta}_j$  significativo.

Isso ocorre porque  $F$  mede a relação de regressão do conjunto de preditores contra a variável dependente, ao passo que  $t$  mede a contribuição de uma única variável explicativa (MCCLAVE; SINCICH, 2000, *apud* MOREIRA, 2006).

A alta correlação linear simples entre as variáveis é condição suficiente, mas não necessária, para que exista multicolinearidade, portanto, sua presença não é indício infalível de multicolinearidade, salvo no caso de apenas duas variáveis preditoras. O exame das correlações parciais pode, também, ser útil (GUJARATI, 2006).

Ainda de acordo com Moreira (2006), como a multicolinearidade se manifesta em função dos vetores das variáveis explicativas sendo linearmente dependentes, um meio de descobrir qual variável  $X$  se relaciona com outras é regressar  $X_k$  sobre as demais e calcular o  $R^2$  para cada regressão. Se o valor da estatística  $F$  calculado for significativo, pode-se presumir a existência de multicolinearidade.

De acordo com Gujarati (2006) e Moreira (2006), pode-se usar a regra prática formulada por Klein, na qual a multicolinearidade só é preocupante se o  $R^2$  calculado

na regressão auxiliar for maior que o  $R^2$  global. Essa regra evita o teste de cada regressão auxiliar.

A multicolinearidade pode ser identificada, também, pela matriz de correlação entre as variáveis explicativas. Se as correlações entre as variáveis forem elevadas, pode haver um indício de problemas de colinearidade entre elas.

Uma maneira de se eliminar a multicolinearidade do sistema, segundo Alexander (2005), é a aplicação da técnica de Análise de Componentes Principais para reduzir as dimensões de modo que sejam usadas apenas as fontes de informações mais importantes. Dessa forma, as cargas fatoriais resultantes da aplicação da ACP podem ser usadas como as variáveis do sistema original.

## 2.6 Medidas de erro de previsão

O termo previsão tem origem no latim, *praevidere*, que significa ver com antecipação, antever algo antes que exista no futuro. Em alguns casos, em se tratando de receitas ou custos, utiliza-se o termo “projeção”.

As previsões se constituem em um meio de fornecer informações para uma consequente tomada de decisões e ações futuras. As previsões dos mais variados tipos de situações são elaboradas utilizando-se técnicas quantitativas e qualitativas ou, ainda, uma mistura de ambas. Segundo Pellegrini e Fogliatto (2000), métodos quantitativos utilizam dados históricos para prever a demanda em períodos futuros.

Para se escolher entre um método de previsão ou outro é importante utilizar uma medida de erro com a finalidade de se encontrar aquele método que melhor atenda aos resultados esperados. A escolha do critério a ser utilizado para medir a acurácia dos modelos não é arbitrária e deve ser feita baseada nas características do problema em estudo. A palavra acurácia refere-se “ao quão bem ajustado” é o modelo, isto é, “quanto” o modelo é capaz de reproduzir os dados já conhecidos.

Existem na literatura diversas medidas de erro disponíveis, dentre as quais se podem citar a comparação gráfica, diagramas de dispersão, coeficiente de correlação, Erro Percentual Absoluto Médio (MAPE), dentre outros. Neste trabalho, em específico, são utilizados o Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error* - MAPE), o coeficiente de determinação múltiplo ( $R^2$ ) e o coeficiente de determinação ajustado ( $R^2$  ajustado), para avaliar qual modelo de

regressão está mais bem ajustado. De acordo com Gujarati (2006), além do emprego do  $R^2$  e do  $R^2$  ajustado, são utilizados outros critérios para avaliar a adequação de um modelo de regressão.

O Erro Percentual Absoluto Médio: (*Mean Absolute Percentage Error - MAPE*) é um valor absoluto médio em percentual, usado para verificar a margem de acerto em comparação com o valor previsto e é calculado a partir das previsões um passo à frente, gerado por cada modelo estimado.

O valor do MAPE é encontrado a partir da expressão (20).

$$MAPE(\%) = \frac{\sum \left( \left| \frac{y_t - \hat{y}_t}{y_t} \right| \right)}{n} \times 100 \quad (20)$$

Onde:

$n$  é o número de observações;

$y_t$  é o valor observado no instante  $t$ ;

$\hat{y}_t$  é o valor previsto no instante  $t$ .

Quanto menor for o valor apresentado, melhor é o ajuste do modelo.

## 2.7 Análise de componentes principais

Nesta seção apresenta-se a fundamentação teórica a respeito da técnica de Análise de Componentes Principais, abordando-se aspectos como a geração, a seleção e a interpretação das componentes investigadas. Ainda, pretende-se determinar as variáveis de maior influência na formação de cada componente com o objetivo de redução do número de variáveis a serem utilizadas no modelo.

A Análise de Componentes Principais (ACP), de acordo com Karlsson (1992), é uma técnica estatística multivariada que possibilita, em investigações com um grande número de dados disponíveis, a identificação das medidas responsáveis pelas maiores variações entre os resultados sem perdas significativas de informações. Segundo Souza (2000), a idéia matemática do método de Análise de Componentes Principais é conhecida há muito tempo, apesar do cálculo das

matrizes dos autovalores e autovetores não ter sido possível até o advento dos computadores.

Em alguns casos, em especial nesta pesquisa, o objetivo da Análise de Componentes Principais é a redução do número de variáveis a serem avaliadas e a interpretação das combinações lineares, uma vez que a informação contida nas  $p$ -variáveis originais é substituída pela informação contida em  $k$  ( $k < p$ ) componentes principais (MINGOTI, 2005). Essa análise é utilizada na redução das variáveis e para identificar as variáveis que possuem maior influência no sistema, obtendo a participação de cada uma na construção de um índice ou no alcance de resultados.

De acordo com Mingoti (2005), essa técnica foi introduzida por Karl Pearson em 1901 e tem como objetivo principal explicar a estrutura de variância e covariância de um vetor aleatório, composto de  $p$ -variáveis aleatórias, a partir do cálculo de combinações lineares das variáveis originais, resultando, então, no que se denomina de componentes principais. Estas, por sua vez, são não correlacionadas entre si.

Outra aplicação muito importante da análise de componentes principais é a construção de grandes matrizes de covariância positivas definidas. Neste caso, a vantagem da ACP não se encontra tanto na redução da dimensionalidade, mas na ortogonalização das variáveis

Dessa forma, o objetivo da ACP é a redução de dimensão, utilizando-se apenas os primeiros componentes principais calculados. Essa redução se torna útil em sistemas com altas correlações, porque há apenas algumas fontes independentes de variação, com grande parte dessa variabilidade explicada por algumas componentes principais, em número bem inferior ao das variáveis. Se aplicadas às transformações para componentes principais, as cargas fatoriais podem ser utilizadas para relacionar essas transformações com o sistema original.

Diz-se que duas variáveis são ortogonais quando elas têm correlação não condicional nula. Uma vez que os componentes principais são ortogonais, então sua matriz de covariância não condicional é diagonal. As variâncias das componentes principais podem ser rapidamente transformadas na matriz de covariância do sistema original, usando-se as cargas fatoriais. Somente as variâncias dos  $m$ -componentes precisam ser calculadas, ao invés das variâncias dos  $k(k+1)/2$  diferentes elementos da matriz de covariância  $k \times k$  do sistema original (ALEXANDER, 2005).

O objetivo desta análise é verificar se os primeiros componentes são responsáveis pela maior parte da variação nos dados originais. Se o são, então se argumenta que a dimensionalidade eficiente do problema é menor do que  $p$ . Neste caso, espera-se que os primeiros componentes sejam intuitivamente significativos, podendo-se trabalhar com as primeiras componentes, sendo possível operar com um número menor de variáveis.

Dessa forma, os objetivos mais importantes da ACP são:

- Gerar novas variáveis que podem expressar a informação contida em um conjunto original de dados;
- Reduzir o conjunto de dados a ser avaliado;
- Interpretar as combinações lineares das variáveis;
- Eliminar algumas das variáveis originais que possuem pouca informação para o sistema.

- Geração dos componentes principais

Segundo Souza (2000), Morrison (1976), Seber (1984), Reinsel (1993), Jackson (1980, 1981) e Johnson & Wichern (1992), seja  $B$  uma matriz quadrada de dimensão  $(p \times p)$ , é possível encontrar um escalar  $\lambda$  e um vetor  $X$  de dimensão  $(p \times 1)$ , não nulo tal que:

$$B X = \lambda X \quad (21)$$

$$B X - \lambda X = 0 \quad (22)$$

$$(B - \lambda I) X = 0 \quad (23)$$

Para que a operação matricial seja possível, deve-se encontrar outra equação que permita completar o sistema, determinando-se a condição em que os vetores próprios estejam normalizados. Isto equivale a dizer, em termos algébricos, que a soma dos quadrados dos elementos do vetor deve ser igual a 1.

Considera-se o vetor aleatório  $X' = [X_1, X_2, \dots, X_p]$ , encontra-se os autovalores  $\lambda$  que são úteis nas interpretações em termos da densidade constante dos elipsóides, em que se tem a matriz de covariância  $\Sigma$ , com os seguintes autovalores:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0 \quad (24)$$

Considera-se a combinação linear, baseada nos autovalores ( $\lambda$ ), que deram origem aos autovetores ( $\ell$ ):

$$\begin{aligned} Y_1 &= \ell'_1 X = \ell_{11} X_1 + \ell_{21} X_2 + \dots + \ell_{p1} X_p \\ Y_2 &= \ell'_2 X = \ell_{12} X_1 + \ell_{22} X_2 + \dots + \ell_{p2} X_p \\ &\dots \\ Y_p &= \ell'_p X = \ell_{1p} X_1 + \ell_{2p} X_2 + \dots + \ell_{pp} X_p \end{aligned} \quad (25)$$

Sabe-se que:

$$Var(Y_i) = \ell'_i \Sigma \ell_i \quad i = 1, 2, \dots, p \quad (26)$$

$$Cov(Y_i, Y_k) = \ell'_i \Sigma \ell_k \quad i, k = 1, 2, \dots, p \quad (27)$$

As combinações lineares são não correlacionadas de  $Y_1, Y_2, \dots, Y_p$ , cuja variância é a maior possível.

O primeiro componente é a combinação linear com máxima variância, isto é,  $Var(Y_1) = \ell'_1 \Sigma \ell_1$ . Para eliminar esta indeterminação, é conveniente restringir o vetor de coeficientes à unidade. Então se pode definir:

CP<sub>1</sub> = Combinação linear  $\ell'_1 X$  que maximiza  $Var(\ell'_1 X)$ , sujeito à restrição de  $\ell'_1 \ell_1 = 1$ ;

CP<sub>2</sub> = Combinação linear  $\ell'_2 X$  que maximiza  $Var(\ell'_2 X)$ , sujeito às restrições de  $\ell'_2 \ell_2 = 1$  e  $Cov(\ell'_2 X, \ell'_1 X) = 0$ ;

CP<sub>(i-ésimo)</sub> = Combinação linear  $\ell'_i X$  que maximiza  $Var(\ell'_i X)$  sujeito às restrições de  $\ell'_i \ell_i = 1$  e  $Cov(\ell'_i X, \ell'_k X) = 0$  para  $k < i$ .

Seja  $\Sigma$  a matriz de variância do vetor aleatório  $X' = [X_1, X_2, \dots, X_p]$ . Se  $\Sigma$  possui o par de autovalores e autovetores  $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ , onde  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Então  $Y_1 = e'_1 X, Y_2 = e'_2 X, \dots, Y_p = e'_p X$  os componentes:

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i) \quad (28)$$

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \text{tr}(\Sigma) \quad (29)$$

A proporção explicada pelo  $k$ -ésimo componente é fornecida por:

$$\left( \begin{array}{l} \text{Proporção da variância total} \\ \text{explicada pelo } k\text{-ésimo} \\ \text{Componente Pincipal} \end{array} \right) = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad k = 1, 2, \dots, p \quad (30)$$

Para a determinação das componentes principais, é necessário calcular a matriz de variância-covariância ( $\Sigma$ ) ou a matriz de correlação ( $\mathbf{R}$ ), encontrar os autovalores e os autovetores e, por fim, escrever as combinações lineares que serão as novas variáveis, denominadas de componentes principais.

Os autovalores, muitas vezes, são denominados de raízes características ou raízes latentes, e os autovetores são denominados de vetores característicos ou ainda vetores próprios.

Procurando-se eliminar a influência que uma variável possa causar sobre a outra na formação da componente, utilizar-se-á as componentes principais derivadas

de variáveis padronizadas. A padronização é feita por meio da relação  $Z_p = \frac{X_p - \mu_p}{\sqrt{\sigma_{pp}}}$ , onde a representação matricial será igual a  $\mathbf{Z} = (\mathbf{V}^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu})$ , onde  $E(\mathbf{Z}) = \mathbf{0}$  e a covariância será  $\text{Cov}(\mathbf{Z}) = (\mathbf{V}^{1/2})^{-1}\Sigma(\mathbf{V}^{1/2})^{-1} = \mathbf{R}$ .

A definição do número de componentes a serem utilizadas é feita por meio de dois critérios: o primeiro, denominado de método gráfico, representa graficamente a porcentagem de variação explicada pela componente nas ordenadas e os autovalores em ordem decrescente nas abscissas. Este critério, que considera as componentes anteriores ao ponto de inflexão da curva, foi sugerido por Cattell (1966) e utilizado por Lírio (2005) em sua pesquisa; o segundo critério de seleção consiste em incluir somente aquelas componentes cujos valores próprios sejam superiores a

1. Este critério é sugerido por Kaiser (1960 *apud* MARDIA, KENTE; BIBBY, 1979). Ele tende a incluir poucas componentes quando o número de variáveis originais é inferior a vinte e, em geral, utilizam-se as componentes que conseguem sintetizar uma variância acumulada em torno de 70%.

A determinação da variável que possui maior influência na combinação linear será encontrada pela correlação mostrada entre as variáveis originais e as componentes encontradas. Aquelas que possuírem maior correlação são as de maior influência.

O sinal da correlação indica o modo dessa influência (JOHNSON; WICHERN, 1992), sendo, assim, identificada a que deve ser monitorada para manter o sistema estável.

### **3 METODOLOGIA**

Neste capítulo são apresentadas as etapas necessárias para a coleta, tabulação e modelagem das variáveis e dos dados em estudo, representados pelos dezoito antibióticos mais utilizados no tratamento das infecções hospitalares no HUSM.

#### **3.1 Caracterização da pesquisa**

Este trabalho caracteriza-se por ser uma pesquisa aplicada e quantitativa. Quanto aos objetivos, esta pesquisa é classificada como descritivo-explicativa, e, quanto aos procedimentos, o trabalho pode ser caracterizado como uma pesquisa *ex-post facto*. Cabe ressaltar que, frente à importância das técnicas estatísticas, em especial os modelos de previsão, desenvolve-se nesta pesquisa um arcabouço a partir de técnicas de previsão, por meio de regressão linear múltipla e de regressão dinâmica, aplicadas à técnica multivariada de componentes principais.

#### **3.2 Identificação das variáveis envolvidas no estudo**

Para identificar quais são os antibióticos mais utilizados no tratamento de infecções hospitalares pelo HUSM, efetuou-se a pesquisa junto ao Serviço de Farmácia e à Comissão de Controle de Infecção Hospitalar. A relação de antibióticos foi comparada com informações disponíveis em sítios sobre infecções hospitalares disponíveis na rede mundial de computadores, a *internet*, onde se constatou a quase total igualdade na utilização dos antibióticos pelo hospital em relação aos demais hospitais no Brasil. Neste estudo foram selecionados, como variáveis explicativas, dezoito antibióticos utilizados no tratamento das infecções hospitalares, cujos nomes e codificações adotadas constam do quadro 1.

Ordem	Nome do Antibiótico	Código da Variável
1	AMICACINA, 500mg	AMICA500A.
2	AMOXICILINA TRIIDRAT 1g + CLAVULANATO POTÁSSIO 200mg	ACPOT001A.
3	AMPICILINA SÓDICA, 1g	AMPIC001A.
4	ANFOTERICINA B, 50mg	ANFOT050A.
5	CEFAZOLINA SÓDICA, 1g	CEFAZ001A.
6	CEFEPIME, 2g	CEFEP002A.
7	CEFTAZIDIMA PENTAIDRATADA, 1g	CEFTA001A.
8	CEFTRIAXONA DISSÓDICA, 1g	CEFTR001A.
9	CIPROFLOXACINO 200mg	CIPRO200FR
10	CLINDAMICINA, 600mg	CLIND600A.
11	GENTAMICINA, 80mg	GENTA080A.
12	IMIPENEM 500mg/CILASTATINA SÓDICA 500mg,	ICSMO500FR
13	MEROPENEM, 1g	MEROP001FR
14	METRONIDAZOL, 500mg,	METRO100FR
15	OXACILINA SÓDICA, 500mg	OXACI500A.
16	TEICOPLANINA, 400mg	TEICO400A.
17	VANCOMICINA, 1g	VANCO001A.
18	VANCOMICINA, 500mg	VANCO500A.

**Quadro 1 – Nome do antibiótico e código adotado para a variável**

### 3.3 Coleta dos dados

Os dados foram coletados junto ao banco de dados do Sistema de Informações para o Ensino, (SIE) da Universidade Federal de Santa Maria (UFSM), disponibilizados pelo relatório de Consumo Mensal do Almojarifado na opção Almojarifado da Farmácia do HUSM.

O SIE disponibiliza inicialmente as informações para serem visualizadas na tela, podendo ser impressas na forma de relatório, ou armazenadas em meio magnético na estação de trabalho do usuário. Optou-se por armazenar os dados com as quantidades de antibióticos consumidas anualmente em planilhas eletrônicas a fim de facilitar a importação dos dados pelos programas estatísticos utilizados.

Em função da legislação vigente, que regulamenta e normatiza os processos de compra dos órgãos da administração pública federal, não é permitida a aquisição de medicamentos pelo seu nome comercial. Em todas as aquisições, deve obrigatoriamente constar o princípio ativo do medicamento e, por isso, foi adotado o mesmo critério nesta pesquisa.

### 3.4 Tratamento dos dados

Para atender ao objetivo principal desta pesquisa, que consiste em prever as necessidades orçamentárias a serem alocadas na aquisição de antibióticos para debelar infecções hospitalares a partir de modelos de regressão, é necessário ajustar uma variável dependente representativa do custo total mensal, expressa em Reais.

O cálculo dessa variável exigirá a reestruturação de um novo banco de dados com a finalidade de converter as unidades de antibióticos utilizadas em unidades monetárias. Desta forma, esse banco de dados auxiliar é constituído de 19 variáveis com 75 observações cada uma, num total de 1.425 registros, e é estruturado com dezoito variáveis, formadas pela multiplicação das quantidades mensais pelo seu custo médio e uma variável denominada Custo Total Mensal, calculada a partir da soma algébrica do custo mensal de cada uma das outras 18 variáveis. Esses dados serão calculados inicialmente com base nos valores originais do custo médio apresentado por cada um dos antibióticos.

No caso dos preços médios originais de cada ano apresentarem uma variação significativa, será necessário utilizar um indexador para estabilizar o custo e expurgar os efeitos da ocorrência das oscilações dos preços e manter um padrão uniforme de valor de moeda ao longo do período estudado.

Desta forma, é necessário ajustar as variáveis novamente para estabilizar a variável custo total mensal. Pretende-se utilizar, como indexador para o cálculo do custo mensal, o custo médio dos antibióticos do ano de 2008. Assim, com o auxílio de uma planilha eletrônica, serão obtidas 75 observações ajustadas da variável custo total mensal, a qual será utilizada na estimação do modelo de regressão apropriado para se realizar as previsões necessárias.

### 3.5 Métricas de análise

#### 3.5.1 Estatísticas descritivas

Serão calculados a média, o desvio padrão e o coeficiente de variação mensal, anual e um cálculo geral das três estatísticas descritivas, abrangendo todo o período estudado com a finalidade de analisar o comportamento do custo mensal.

#### 3.5.2 Aplicação de modelos de regressão linear múltipla

A análise de regressão linear múltipla é utilizada com o objetivo de analisar e estimar o custo mensal dos antibióticos e as necessidades orçamentárias para sua aquisição três períodos à frente.

Pode-se utilizar a quantidade e posteriormente ponderar pelo valor indexado do medicamento para encontrar o valor total, ou estimar diretamente o custo da compra destes medicamentos. Nesta pesquisa será utilizada a que fornecer os melhores valores previstos, a serem determinados durante o estudo.

Na etapa de estimação da equação de regressão linear múltipla, estima-se uma regressão estática pelo método *stepwise* com o objetivo de verificar o impacto de cada antibiótico na formação do custo mensal.

Com o objetivo de captar o efeito temporal do custo dos antibióticos na variável custo total mensal, é estimada uma regressão dinâmica, pautando-se na metodologia geral para o específico, seguindo a metodologia *Hendry Approach*, na qual a variável que permanecerá no modelo será decidida com base em um teste de hipótese para validade do parâmetro. Neste sentido, pretende-se captar o efeito retroativo do custo dos medicamentos, se existirem, na composição do custo mensal.

Verificada a existência de multicolinearidade, será aplicada a Análise de Componentes Principais com o objetivo de formar um novo conjunto de variáveis independentes para se construir o modelo de regressão. Desta forma, procura-se solucionar o problema da multicolinearidade entre as variáveis originais do sistema.

### 3.5.3 Análise de componentes principais

Solucionado o caso da multicolinearidade, a variável custo total mensal tentará ser explicada em função das variáveis independentes, que serão as componentes principais selecionadas pelos métodos citados na seção 2.9 do referencial teórico.

Também será construído um modelo de regressão linear dinâmica, utilizando-se as componentes principais selecionadas e obedecendo aos critérios das regressões definidos nas itens 2.3 a 2.5 respectivamente.

Será testado um modelo de regressão com as CP, com autovalor menor que um, ou seja, as não selecionadas, com a finalidade de verificar se é possível fazer previsões com as CP com reduzido autovalor.

De acordo com Alexander (2005), a determinação das componentes principais selecionadas seguirá as seguintes etapas:

- As variáveis explicativas serão normalizadas de forma que tenham média zero e variância igual a 1. Assim  $X_i^* = (X_i - \mu_i)/\sigma_i$  em que  $\sigma_i$  e  $\mu_i$  são o desvio padrão e a média de  $X_i$  para  $i = 1, \dots, K$ ;
- As novas variáveis  $X_1^*, \dots, X_k^*$  pela ACP para obter todos os  $K$  componentes principais  $P_1, \dots, P_k$ , com a matriz de cargas fatoriais.

Assim, ao invés de regredir a variável custo mensal contra as variáveis originais correlacionadas, a regressão será realizada contra as componentes principais não correlacionadas.

### 3.5.4 Análise dos resíduos dos modelos estimados

Para validação dos modelos, são efetuados testes de validação dos resíduos para que se possam realizar previsões de curto prazo, conforme descrito na seção 2.5, da revisão de literatura, intitulada pressupostos do modelo de regressão. Estando válido, o modelo será utilizado para a previsão.

### 3.6 Elaboração e análise das previsões

As previsões serão efetuadas pelos melhores modelos, previsões com três passos à frente, o que corresponde a um horizonte de compra para três meses. Também serão comparados os valores previstos com os valores reais, utilizando também critérios para validação das previsões.

São comparados os valores previstos oriundos das equações de regressão estática e dinâmica, verificando-se, assim, qual o modelo mais adequado para este tipo de estudo.

Para esta comparação, serão deixadas de fora da amostra as três últimas observações correspondentes aos meses de janeiro, fevereiro e março de 2009. Nesta etapa serão utilizados, como critério de comparação, o Erro Percentual Absoluto Médio – MAPE, o coeficiente de determinação múltiplo ( $R^2$ ) e o coeficiente de determinação ajustado ( $R^2$  ajustado) para avaliar qual modelo de regressão está mais bem ajustado.

Com base no melhor modelo, que fornece a melhor previsão, serão alocados os recursos orçamentários do hospital para a aquisição dos antibióticos, de acordo com os valores previsto obtidos no modelo.

### 3.7 Fluxograma do desenvolvimento da metodologia

Por meio da Figura 1, pode-se visualizar os passos adotados para a elaboração dos modelos de regressão linear múltipla e regressão linear dinâmica.

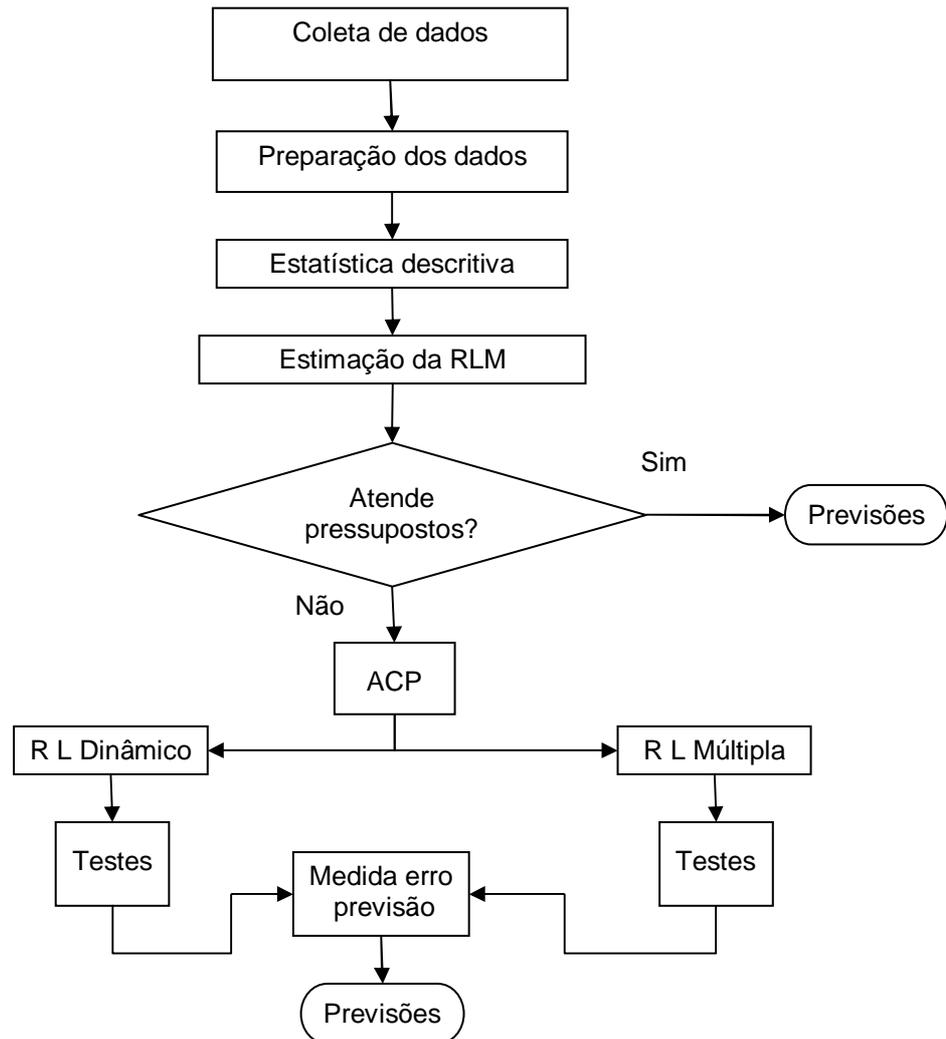


Figura 1 – Fluxograma da elaboração dos modelos de previsão

Pelo fluxograma da Figura 1, é possível visualizar resumidamente as etapas da metodologia adotada para a realização da presente pesquisa.

## **4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS**

Neste capítulo são apresentados os resultados da análise do custo dos antibióticos utilizados no combate às infecções hospitalares e a previsão de alocação dos recursos orçamentários para sua aquisição para três períodos à frente. Dessa forma, os resultados obtidos contribuirão para revelar à administração do hospital, além das necessidades de alocação dos recursos, pela previsão, o comportamento dos custos dos medicamentos sob diversos aspectos ao longo do período estudado.

As técnicas estatísticas univariadas, multivariadas, regressão linear múltipla e regressão dinâmica forneceram o suporte técnico necessário para as análises efetuadas, revelando fatores e correlações, permitindo efetuar previsões compatíveis com a realidade pesquisada.

Assim, é possível fornecer diretrizes visando subsidiar o planejamento orçamentário e financeiro do HUSM, permitindo a alocação racional e otimizada, de forma a maximizar os recursos que, em um hospital público, são bastante escassos.

### **4.1 O Hospital Universitário de Santa Maria**

O Hospital Universitário de Santa Maria constitui-se no único hospital público da região, para onde são canalizados pacientes terciários, portadores de complicações clínicas ou cirúrgicas, ocasionando internações de alto custo e alto índice de permanência no hospital.

Tem como missão primordial recuperar, manter e incrementar os padrões de saúde de sua clientela. Para cumprir sua missão, necessita distribuir de maneira eficiente os recursos de que dispõe, sejam eles materiais, humanos ou financeiros. Ao contrário dos hospitais privados onde o objetivo da gestão é maximizar os lucros, nos hospitais públicos, como o HUSM, que não pode recusar atendimento porque tem como finalidade promover o bem-estar social da população atendida, os gestores possuem o objetivo principal de atender o maior número possível de pacientes com o menor custo possível.

Em se tratando de um hospital público com atendimento exclusivo pelo Sistema Único de Saúde (SUS), que é a sua principal fonte de recursos financeiros,

é imprescindível que haja um planejamento orçamentário adequado a fim de otimizar a aplicação dos escassos recursos, visando a um melhor atendimento do paciente.

O Hospital Universitário de Santa Maria iniciou como Hospital Regional de Tuberculose, em 1959, e foi transformado em Hospital Universitário com o início de suas atividades em 1970, com a finalidade de desenvolver um sistema de ensino, pesquisa e extensão por meio da assistência à comunidade na área da saúde.

Constitui-se no único hospital público da região, para onde são canalizados pacientes terciários, portadores de complicações clínicas ou cirúrgicas, ocasionando internações de alto custo e alto índice de permanência no hospital.

O hospital atende mais de 33 municípios da região central do Rio Grande do Sul, oferecendo 28 especialidades médicas; Anestesia, Angiologia e Cirurgia Vascular, Cardiologia, Cirurgia Cardíaca, Cirurgia de Cabeça e Pescoço, Cirurgia Geral, Cirurgia Torácica, Dermatologia, Endocrinologia e Nutrição, Gastroenterologia, Ginecologia, Hemato – Oncologia, Hemodinâmica, Infectologia, Nefrologia, Neurologia, Obstetrícia, Oftalmologia, Otorrinolaringologia, Pediatria, Pneumologia, Proctologia, Quimioterapia, Radioterapia, Reumatologia, Traumatologia – Ortopedia, Urologia, Serviços Unidade de Terapia Intensiva para Adultos, Recém Nascidos, Pediatria e Hematologia.

Além disso, o hospital constitui-se em centro de ensino e pesquisa no âmbito das ciências da saúde, centro de programação e manutenção de ações voltadas à saúde das comunidades local e regional, desenvolve programas específicos à comunidade, devidamente integrado à rede regional de saúde.

Atualmente conta com 305 leitos, com 1386 servidores, 6 salas do Centro Cirúrgico em funcionamento, 2 salas de parto e 1 de cesariana em funcionamento, 64 consultórios, 94 professores que atuam no hospital, 104 médicos residentes e 146 médicos no quadro funcional.

Segundo os dados mensais do Serviço de Estatística do HUSM, em 2008 foram realizados em média 914 internações, 900 procedimentos cirúrgicos, 166 partos, 10.332 consultas ambulatoriais, 4.285 consultas no Pronto Atendimento, 1.144 seções de fisioterapia, 63.808 exames e 700 seções de quimioterapia.

Atender essa demanda expressiva de clientes, muitas vezes portadores de moléstias graves, favorece ao aparecimento de infecções hospitalares e encarece o tratamento.

De acordo com os registros da Comissão de Controle de Infecção Hospitalar (CCIH) do HUSM, que monitora as infecções hospitalares, ocorreram 346 casos de IH em 2008, 386 no ano de 2007, 434 casos em 2006, 344 em 2005, 357 no ano de 2004 e 355 em 2003.

#### 4.2 Mensuração do custo dos antibióticos

Para a mensuração do custo mensal, foi utilizado o valor do custo médio anual de cada variável, multiplicado pelas quantidades consumidas mensalmente, permitindo, dessa forma quantificar os custos incorridos e efetuar análises sobre seu comportamento ao longo do período estudado. A Tabela 1 mostra a evolução do custo médio unitário para cada antibiótico durante o período compreendido entre os anos de 2003 e 2008.

**Tabela 1 – Custo médio unitário anual**

Nome do antibiótico	Custo médio unitário (R\$)					
	2003	2004	2005	2006	2007	2008
AMICACINA, 500mg	11,15	10,09	12,20	7,98	4,96	3,69
AMOXICILINA TRIIDRATADA 1g + CLAVULANATO POTÁSSIO 200mg	0,64	0,82	0,81	0,85	0,72	0,72
AMPICILINA SÓDICA ,1g	0,86	0,99	0,98	0,78	0,69	0,68
ANFOTERICINA B,50mg	8,78	8,86	10,72	10,73	10,68	11,11
CEFAZOLINA SÓDICA,1g	2,06	1,98	1,93	1,35	1,11	0,93
CEFEPIME, 2g	43,75	39,14	32,16	14,54	9,19	6,86
CEFTAZIDIMA PENTAIDRATADA, 1g	4,14	5,76	4,92	3,18	2,48	2,32
CEFTRIAXONA DISSÓDICA, 1g	2,37	2,51	2,30	1,36	1,11	1,00
CIPROFLOXACINO 200mg	3,86	3,33	2,91	3,67	2,94	4,14
CLINDAMICINA, 600mg	1,95	2,08	1,77	1,54	1,14	1,04
GENTAMICINA, 80mg	0,22	0,24	0,30	0,30	0,29	0,29
IMIPENEM 500mg/CILASTATINA SÓDICA 500mg,	41,32	52,14	57,03	44,02	25,81	23,54
MEROPENEM, 1g	93,33	104,24	105,47	86,84	57,90	45,77
METRONIDAZOL, 500mg,	0,66	0,64	0,66	0,75	1,85	3,24
OXACILINA SÓDICA, 500mg	0,62	0,74	0,88	0,75	0,62	0,59
TEICOPLANINA, 400mg	111,00	127,92	153,36	176,23	100,61	64,29
VANCOMICINA, 1g	15,75	18,37	13,50	17,23	18,59	18,60
VANCOMICINA, 500mg	6,68	9,50	7,67	8,25	3,99	3,68

Observa-se, na Tabela 1, um fato atípico, pois mesmo numa economia estável, com reduzida taxa inflacionária, era de se esperar um aumento no preço desses antibióticos, o que não ocorreu. Uma análise horizontal da Tabela 1 indica

que a maioria dos antibióticos teve seu custo reduzido em relação aos anos anteriores a 2008.

A causa desta redução é, em parte, devido às quebras de patentes de medicamentos promovidas pelo Governo Federal, pois a patente é um dos mecanismos legais de proteção à propriedade intelectual e visa a garantir ao seu detentor os direitos de produção e comercialização desse produto (CUNHA, 2001). Mas, no caso do Brasil e de outros países menos desenvolvidos, fatores sociais têm prevalecido sobre o aspecto econômico em função da grande defasagem tecnológica desses países, cuja escassez de recursos impede a aquisição dos produtos de última geração fabricados pelos grandes centros econômicos.

Assim, a política brasileira de incentivo à produção de medicamentos genéricos traz uma redução substancial no custo de todos os medicamentos e beneficia o próprio governo, uma vez que os hospitais públicos e privados necessitam destinar um volume menor de recursos financeiros para aquisição de medicamentos.

Outro fator que contribui para essa redução decorre da compra dos medicamentos serem obrigatoriamente efetuada por meio de pregão eletrônico, uma modalidade de compra semelhante a um leilão de maneira inversa, ou seja, os lances vão diminuindo de valor e se consagra vencedor aquele que ofertar o menor preço. O pregão eletrônico é efetuado pela rede mundial de computadores, a *internet*, e garante acesso a um grande número de fornecedores, forçando uma concorrência acirrada entre eles.

Na Figura 2, demonstra-se o comportamento do custo médio dos antibióticos e observa-se a ocorrência da redução dos preços. Optou-se por demonstrar os cinco medicamentos que apresentaram as maiores oscilações, porque a plotagem de todos resulta em um aglomerado de linhas e pontos que não permite interpretar adequadamente o gráfico.

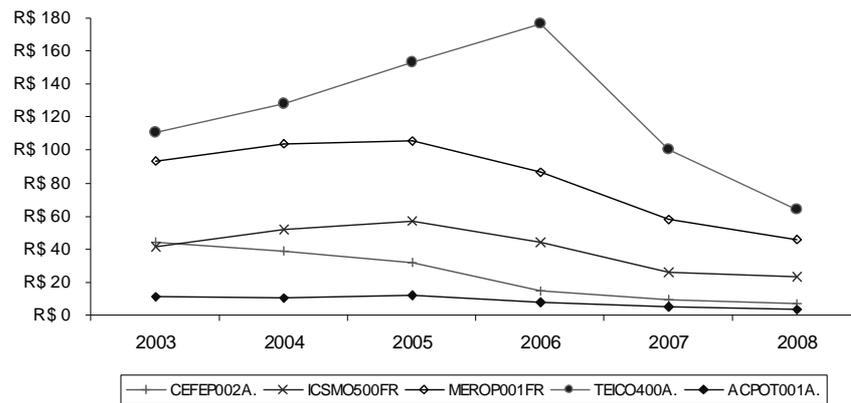


Figura 2 – Evolução do custo médio dos antibióticos

O custo médio do medicamento Teicoplanina, 400 mg, representado pela variável TEICO 400A, apresentou uma curva ascendente de preço até o ano de 2006, conforme mostra a Figura 2. No ano de 2007, o Ministério da Saúde liberou a comercialização desse medicamento para outros laboratórios, motivo pelo qual houve uma redução de preços em torno de 64% em relação a 2006.

Em função da deflação apresentada pelos custos dos antibióticos, foi necessário adotar um indexador dos valores, de forma a minimizar as variações ocorridas. Existem muitos índices disponíveis, o que torna difícil escolher o mais adequado, em função dos critérios adotados para o cálculo de cada um depender de metodologia própria. Dessa forma, optou-se por utilizar a mesma sistemática adotada pelo HUSM, ou seja, utilizar como fator de atualização/deflação, o custo médio ponderado de cada antibiótico.

### 4.3 Análise das estatísticas descritivas

A análise do custo mensal dos antibióticos a partir da estatística descritiva oportunizou uma visão geral do comportamento destas observações. Foi possível, assim, tecer comentários sobre quais os meses e os anos que apresentaram maior dispêndio com antibióticos, suas médias e a variabilidade ocorrida em relação a elas.

Em razão de diversos fatores e particularidades, as infecções hospitalares não ocorrem de forma constante e permanente, estando, portanto, sujeitas aos períodos de altas e baixas incidências, esses os mais desejados; porque a administração hospitalar busca a redução dos custos, tanto de antibióticos como dos demais serviços hospitalares. Nesse ambiente sujeito às variações, a análise

estatística desempenha papel fundamental para avaliar e interpretar o quanto as observações refletem a realidade pesquisada. Neste sentido, a análise do custo mensal, a partir da estatística descritiva, forneceu os anos com maiores médias de consumo.

Na Tabela 2, com a utilização da estatística descritiva, demonstra-se a evolução dos custos mensais dos antibióticos nos anos de 2003 a 2008, em que se calculou o custo mensal e o custo médio mensal dos seis anos estudados. Esta tabela permite ainda visualizar o comportamento dos custos mensais incorridos em cada ano, evidenciando uma visão global dos recursos despendidos.

Como a ocorrência das infecções hospitalares não é constante, observam-se variações no custo mensal dos antibióticos em relação aos outros meses do ano observado e, em relação ao mesmo mês, nos demais anos. Em função dessa oscilação no custo, considerada normal em se tratando de IH, foi calculada a média para cada mês, em relação a todos os anos objeto do estudo. Com base nessas médias, foi possível inferir que a média mensal de consumo dos antibióticos oscila entre R\$ 52.400,00 e R\$ 68.700,00, o que já pode indicar, embora de forma pouco precisa, parâmetros para alguma previsão de alocação de recursos.

**Tabela 2 – Custo mensal por ano, dos antibióticos, custo médio, desvio padrão e coeficiente de variação**

	2003	2004	2005	2006	2007	2008	Custo médio	D. Pad. meses	C V mês %
jan	53.406,82	77.937,65	63.806,08	78.951,90	56.463,06	54.461,04	<b>64.171,09</b>	4.752,87	18,14
fev	36.835,32	48.234,90	74.643,45	64.637,64	71.412,24	69.691,98	<b>60.909,26</b>	6.137,78	24,68
mar	33.751,16	65.511,67	57.436,99	81.232,14	75.328,12	79.727,95	<b>65.498,01</b>	7.347,39	27,48
abr	48.759,60	49.824,42	55.378,11	56.890,94	60.444,40	62.344,40	<b>55.606,98</b>	2.242,16	9,88
mai	52.560,70	50.139,92	76.325,49	61.576,34	72.990,84	75.999,84	<b>64.932,19</b>	4.831,73	18,23
jun	59.995,45	67.269,30	70.331,35	88.257,90	65.983,30	60.406,60	<b>68.707,32</b>	4.239,29	15,11
jul	57.253,15	49.457,60	73.243,93	65.315,90	45.342,25	45.332,40	<b>55.990,87</b>	4.675,81	20,46
ago	44.596,50	58.672,04	61.175,41	97.478,65	62.668,28	65.414,48	<b>65.000,89</b>	7.144,79	26,92
set	52.949,10	74.309,10	77.270,50	59.276,60	34.320,32	63.613,12	<b>60.289,79</b>	6.390,76	25,96
out	31.739,71	49.460,34	71.440,62	62.848,53	37.086,27	61.756,30	<b>52.388,63</b>	6.401,18	29,93
nov	46.997,64	54.423,14	76.451,51	78.366,00	38.781,50	49.491,68	<b>57.418,58</b>	6.655,30	28,39
dez	41.474,95	65.938,36	81.007,75	74.683,68	28.511,74	34.095,68	<b>54.285,36</b>	9.132,32	41,21
C. médio Anual	46.693,34	59.264,87	69.875,93	72.459,69	54.111,03	60.194,62	<b>60.433,25</b>	***	***
Desvio Padrão	9.201,63	10.586,37	8.407,88	12.698,20	17.076,78	12.852,52	14.683,53	***	***
CV %	19,70	17,90	12,0	17,50	31,56	21,35	24,30	***	***

Na figura 3, plotou-se o custo médio mensal com a finalidade de se observar de forma gráfica o comportamento do valor médio e verificar a amplitude entre o custo médio mensal mínimo e o máximo de aproximadamente R\$ 16.300,00. O desvio padrão foi de 8.793,72, e o coeficiente de variação foi de 14,55%.

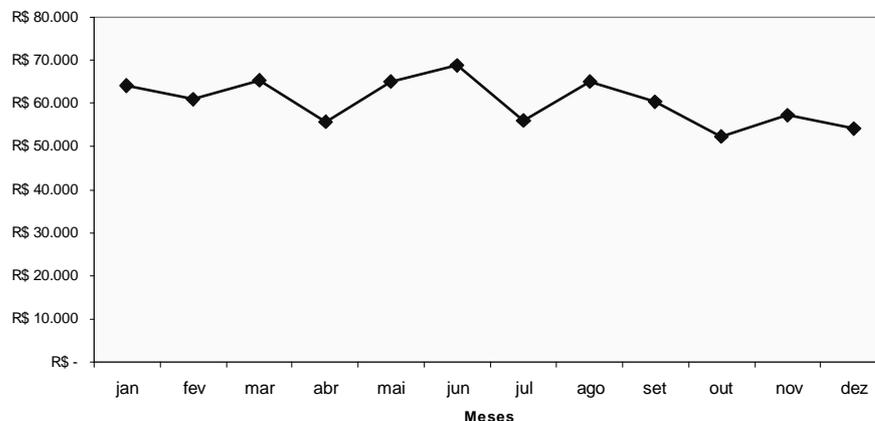


Figura 3 – Custo médio mensal dos antibióticos

O custo médio do período, referente aos 72 meses estudados, foi de R\$ 60.433,25, com um desvio padrão de 14.683,53 e coeficiente de variação de 24,30%, o qual, por ser inferior a 30%, é considerado representativo.

#### **4.4 Elaboração do modelo de regressão linear múltipla**

Inicialmente partiu-se para a construção de um modelo de regressão linear múltipla pelo método *stepwise*, utilizando-se as quantidades de antibióticos dispensadas em cada mês como variáveis preditoras, com a finalidade de prever o custo mensal dos antibióticos.

Como são dezoito as variáveis independentes consideradas, e, em alguns casos, a administração de um tipo de medicamento é acompanhada de outro em um período subsequente ou muitas vezes concomitantemente, podem ocorrer problemas de multicolinearidade entre as variáveis preditoras.

Neste estudo não foi encontrado um modelo adequado em função de haver multicolinearidade entre as variáveis preditoras. A presença de multicolinearidade entre os regressores do modelo foi detectada por meio da matriz de correlação das variáveis independentes e pelo coeficiente de determinação  $R^2$  muito alto, com valor igual a um. A existência de multicolinearidade, tendo sido detectada, é considerada prejudicial, indicando que o pesquisador deve procurar soluções para suavizar seus efeitos nocivos.

Por essa razão, decidiu-se utilizar a técnica de estatística multivariada denominada de Análise de Componentes Principais.

#### **4.5 Análise de componentes principais**

Foi efetuada com o objetivo de reduzir o número de variáveis a serem avaliadas, uma vez que a informação contida nas  $p$ -variáveis originais é substituída pela informação contida em  $k$  ( $k < p$ ) componentes principais. Essa redução se torna útil em sistemas com elevadas correlações, tal como encontra-se neste estudo, em função de haver apenas algumas fontes independentes de variação, com grande parte dessa variação explicada por algumas componentes principais, em número bem inferior ao das variáveis.

A análise de componentes principais foi realizada utilizando-se as variáveis com as quantidades consumidas em cada mês e extraídas as componentes principais a serem utilizadas como variáveis para determinação do modelo de regressão linear múltipla e do modelo de regressão dinâmica.

A técnica de componentes principais apresenta tantos componentes principais quantos forem o número de variáveis que estão sendo analisadas, porém, com a propriedade de que, em ordem decrescente, os primeiros componentes principais retêm a maior parte da variância dos dados originais.

Para seleção dos autovalores foi inicialmente examinado por meio do gráfico *screeplot*, conforme mostra a Figura 4, que indicou a existência de três fatores, os quais respondem por apenas 47,77% da explicação da variância total. Este percentual de variância explicada é considerado pequeno, o que impossibilita o uso desse critério.

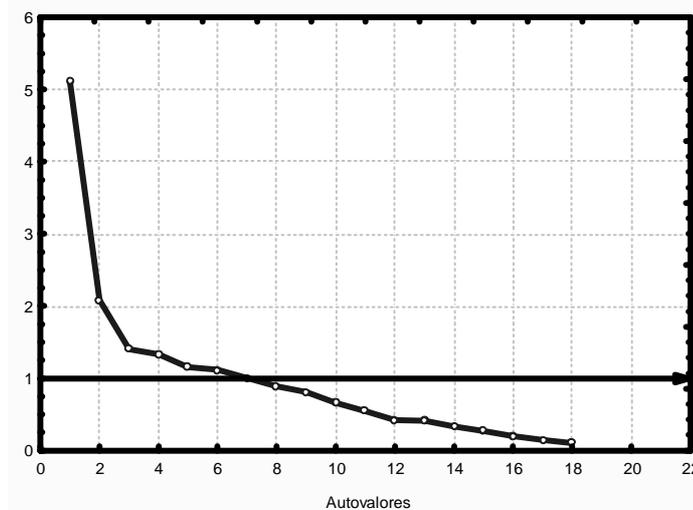


Figura 4 – Demonstração gráfica dos autovalores

Dessa maneira, aplicou-se o critério de Kaiser, que consiste em selecionar os autovalores maiores que a unidade (1), juntamente com o critério do percentual de variância explicada maior do que 70%. Na Tabela 3, estão demonstrados os fatores representados pelos autovalores resultantes da matriz de correlação, o percentual da variância explicado por cada fator e o percentual total de explicação da variância do conjunto de fatores.

**Tabela 3 – Autovalores e variância explicada**

Autovalores				
Extração dos componentes principais				
Fator	Autovalores	% da variância explicada	Autovalores acumulados	% da variância acumulada
1	5,108	28,38	5,11	28,38
2	2,082	11,57	7,19	39,95
3	1,408	7,82	8,60	47,77
4	1,330	7,39	9,93	55,16
5	1,159	6,44	11,09	61,59
6	1,117	6,20	12,20	67,80
7	1,003	5,57	13,21	<b>73,37</b>

Por meio da ACP, utilizando-se o *software Statistica for windows* aplicado ao conjunto das variáveis, foram selecionados sete CP ou novas variáveis, que representam 73,37% das variáveis originais.

Constata-se que a primeira componente principal (CP1) representa 28,38%, a segunda (CP2) 11,57%, a terceira (CP 3) 7,82%, a quarta (CP 4), 7,39%, a quinta (CP 5), 6,44%, a sexta (CP 6) 6,20% e a sétima (CP 7) representa 5,57% da variabilidade das variáveis pesquisadas. Verifica-se, ainda, que são sete os autovalores acima de um e representam uma variância explicada acumulada de 73,37%.

A partir dessa análise, ao invés de utilizar as dezoito variáveis, passam a ser utilizados apenas sete fatores, representados pelos seus autovalores superiores a 1, com uma perda de informação de aproximadamente 26% das variáveis originais.

Na Tabela 4, estão demonstradas, por ordem de importância, as correlações existentes entre os fatores e as variáveis originais, refletindo quais componentes principais (os fatores) que estão mais ou menos correlacionados com as variáveis, (os antibióticos).

**Tabela 4 – Correlação das variáveis (antibióticos) com os fatores (componentes)**

Variáveis	CP1	CP2	CP3	CP4	CP5	CP6	CP7
VANCO500A.	0,89						
CEFTA001A.	0,82						
CEFAZ001A.	0,81						
VANCO001A.	-0,75						
CEFEP002A.	-0,68						
OXACI500A.	0,59						
AMICA500A.	0,55						
GENTA080A.	0,53						
ACPOT001A.		0,62					
ANFOT050A.		-0,58					
CIPRO200FR		-0,54					
CEFTR001A.			-0,50				
TEICO400A.			0,45				
ICSMO500FR					0,49		
METRO100FR					0,46		
CLIND600A.						-0,63	
AMPIC001A.						-0,49	
MEROP001FR				0,54			-0,62

Assim, com base no poder de explicação dos autovalores, foram selecionadas as sete componentes principais utilizadas para determinação do modelo de previsão.

Nesta análise utilizaram-se as sete CP, que representam 73,37% das variáveis explicadas. Pode-se afirmar que a CP 1 é a mais importante para o estudo, pois ela é derivada do maior autovalor, explica 28,38%, e representa o maior número de variáveis, oito das dezoito variáveis.

A variável com maior correlação com esta CP VANCO500A representa o antibiótico vancomicina, 500mg, seguida das variáveis CEFTA001A, - ceftazidima pentaidratada1g, CEFAZ001A - cefazolina sódica1g, VANCO001A - vancomicina 1g, CEFEP002A - cefepime 2g, OXACI500A - oxacilina sódica 500mg, AMICA500A - amicacina 500mg e GENTA080A - gentamicina 80mg.

A CP 2 explica 11,57% da variância e representa três antibióticos: amoxicilina triidratada 1g + clavulanato potássio 200mg (ACPOT001A), anfotericina B 50mg (ANFOT050A) e ciprofloxacino 200mg (CIPRO200FR).

A terceira CP, que representa 7,39% de variância explicada, representa duas variáveis representadas pelos antibióticos ceftriaxona dissódica1g (CEFTR001A) e teicoplanina, 400mg (TEICO400A).

A CP 4 e a CP 7 representam 7,39% e 5,57% da variância explicada e estão mais fortemente associados à variável MEROP001FR - meropenem1g. Os CP 5 e 6 explicam 6,44% e 6,20% da variância respectivamente e representam as variáveis correspondentes aos seguintes antibióticos: ICSMO500FR - imipenem 500mg/cilastatina sódica 500mg, METRO100FR - metronidazol 500mg, CLIND600A - clindamicina 600mg e AMPIC001A - ampicilina sódica1g.

As correlações negativas de algumas variáveis em relação aos fatores mostram que as variáveis se relacionam de forma oposta, isto é, quando a utilização de um antibiótico aumenta o outro diminui.

A seguir, tem-se a análise dos componentes principais, com uma abordagem das correlações das variáveis com as Cps que permitiram reduzir de 18 para 7 as variáveis a serem usadas nos modelos de regressão linear e dinâmica.

Na Figura 5, mostra-se o círculo de correlação unitário, com a nuvem de variáveis distribuídas no plano fatorial, cujas variáveis mais próximas ao círculo de correlação são altamente representativas.

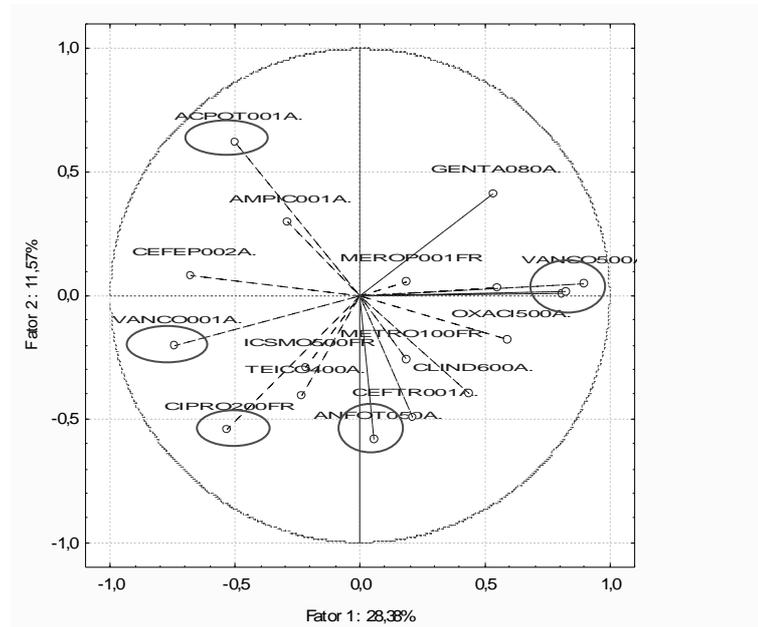


Figura 5 – Círculo de correlação unitário

Na Figura 5, observa-se as variáveis que estão mais correlacionadas e assim melhor explicam a CP 1 e a CP 2, pois uma das principais utilizações do círculo unitário de correlação é realizar a sua sobreposição sobre o plano fatorial;

possibilitando, com isso, identificar visualmente as variáveis relacionadas com os fatores em estudo.

As coordenadas das variáveis com as sete CPs no plano espacial e o percentual de variabilidade explicado por cada uma das CPs estão demonstradas no Apêndice A.

Para a determinação das componentes principais, foram calculados os autovetores para compor a combinação linear que dará origem as novas variáveis, cuja matriz dos autovetores, obtida a partir das variáveis originais, encontra-se demonstrada no apêndice B.

Como foi mostrado, a técnica de componentes principais permite reduzir o número de variáveis originais, correlacionadas em variáveis não correlacionadas em uma mesma quantidade de variáveis, agora transformada em componentes ortogonais, ou seja, nas componentes principais. Cada uma dessas é uma combinação linear de todas as variáveis originais, independentes entre si e escolhidas por ordem decrescente dos autovalores com o máximo de informação, em termos de variação total contida nos dados iniciais, permitindo, com isso, reduzir a dimensão do conjunto original de variáveis e permitir construir os modelos de regressão.

#### **4.6 Estimação da regressão linear com as componentes principais**

A partir do novo conjunto de variáveis formado pelas componentes principais, partiu-se para a construção do modelo de regressão linear, utilizando-as como variáveis independentes para efetuar a previsão do custo mensal dos antibióticos e das necessidades de alocação dos recursos orçamentários para sua aquisição. O modelo foi construído utilizando-se o método *stepwise* do *software Statistica for windows 7.0*.

Testou-se um modelo de regressão com as componentes principais de baixo valor de autovalor, isto é, com aquelas cujo autovalor foi inferior a um, mas este não se mostrou eficiente, razão pela qual se manteve a utilização das sete primeiras componentes principais.

Os escores fornecidos pelas sete componentes, que passam a ser os valores das variáveis explicativas consideradas para o ajuste do modelo de regressão linear, estão demonstrados no Apêndice C.

Na Tabela 5, tem-se o resumo dos resultados estatísticos dos modelos, na qual verifica-se, com o teste “t”, que todas as estimativas dos parâmetros são significativas, indicando que as variáveis transformadas em componentes principais contribuíram satisfatoriamente para a construção do modelo. Esse fato é reforçado pelo coeficiente de determinação ( $R^2 = 0,76$ ) e pelo coeficiente de determinação ajustado ( $R^2$  ajustado = 0,74), mostrando quanto da variação dos valores da variável dependente custo mensal é explicada pela regressão sobre as variáveis independentes representadas pelos componentes principais.

O coeficiente de correlação múltipla  $R^2$ , que mede a qualidade do ajuste do modelo, cujo valor encontrado foi de  $R^2 = 0,76$ , explica cerca de 76% da variabilidade do custo mensal dos antibióticos.

**Tabela 5 – Modelo de regressão linear múltipla**

Variáveis	$\beta$	t (Calculado)	valor de p
Intersecção	60.433,25	67,90877	0,000000
CP <sub>1</sub>	-2.316,64	-2,58506	0,012025
CP <sub>2</sub>	-6.632,01	-7,40044	0,000000
CP <sub>3</sub>	3.491,12	3,89563	0,000237
CP <sub>4</sub>	4.329,50	4,83115	0,000009
CP <sub>5</sub>	4.796,58	5,35235	0,000001
CP <sub>6</sub>	-6.870,69	-7,66678	0,000000
CP <sub>7</sub>	-3.702,79	-4,13183	0,000107

Pelos valores da estatística  $F$ , em um nível de confiança de 95%, a hipótese nula de que a previsão do custo mensal não pode ser explicada por um modelo de regressão linear múltipla foi rejeitada, indicando que as componentes principais podem ser utilizadas para efetuar as previsões neste estudo.

**Tabela 6 – Análise de variância da regressão**

Fonte	Soma dos quadrados	Grau de liberdade	Quadrado médio	F	valor-p
Modelo	1,17E10	7	1,67E+09	29,20918	0,000000
Resíduos	3,65E09	64	5,70E+07		
Total	1,53E10				

Assim, a equação de regressão linear múltipla que descreve a relação entre o custo mensal e as sete componentes principais é dada pela equação (31).

$$\text{Custo mensal} = 60.433,25 - 2.316,64\text{CP1} - 6.632,01\text{CP2} + 3.491,12\text{CP3} + 4.329,50\text{CP4} + 4.796,58\text{CP5} - 6.870,69\text{CP6} - 3.702,79\text{CP7} \quad (31)$$

A análise de variância, contida na Tabela 6, mostra que se deve rejeitar a hipótese nula de não haver regressão, ou seja, o modelo estimado anteriormente é, de fato, significativo, uma vez que forneceu valor-p, de  $p < 0,001$ .

Para verificar o ajuste e a significância do modelo encontrado, foram aplicados testes para avaliar os pressupostos da regressão. Inicialmente testou-se a multicolinearidade, analisando-se o coeficiente de determinação  $R^2$  que apresentou o valor de 0,76 e pela matriz de correlação das componentes principais. A ausência de multicolinearidade foi atendida, pois uma vez que as CP são ortogonais entre si, não há correlação entre elas.

Além disso, o valor da estatística *t-student* aponta todos os  $\hat{\beta}_j$  como sendo significantes.

A avaliação da independência entre os erros – ausência de correlação serial – foi efetuada a partir do teste de DURBIN-WATSON, que se baseia em cálculo da medida conhecida como Estatística DW, cujo valor calculado foi  $DW = 2,042$ . De acordo com a “regra de bolso” descrita na seção 2.5, o valor próximo de 2 atende ao pressuposto da regressão.

Dessa forma, foi aceita a hipótese nula ( $H_0$ ), avaliada a partir do teste DW, de que não existe autocorrelação serial entre resíduos.

O pressuposto da normalidade dos resíduos foi verificado pelo histograma dos resíduos e pelo teste de Jarque-Bera, em que se testaram as seguintes hipóteses:

- $H_0$ : Os resíduos se distribuem normalmente;
- $H_1$ : Os resíduos não se distribuem normalmente.

Esse teste de hipóteses apresentou uma estatística calculada de 0,173 com  $p$ -valor = 0,2721, aceitando a hipótese nula ( $H_0$ ) de que os resíduos se distribuem normalmente.

A normalidade dos resíduos também foi aceita pelo método gráfico, a partir do exame da curva teórica normal e da curva dos resíduos do modelo, na qual verifica-se que esta segue o padrão da normalidade, conforme demonstrado na Figura 6.

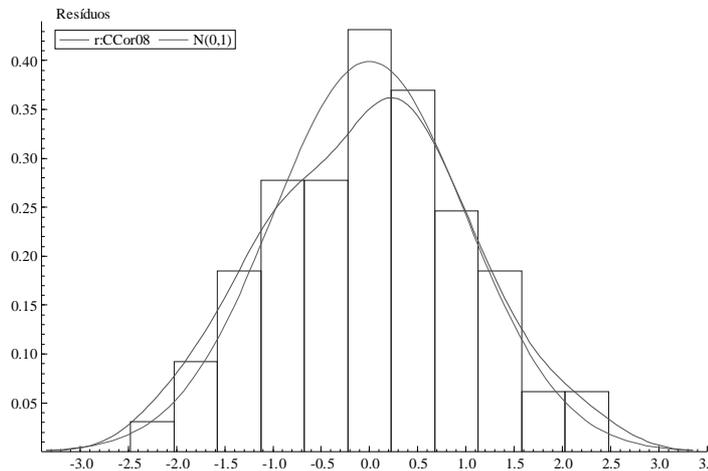


Figura 6 – Histograma dos resíduos do modelo de regressão múltipla

A homocedasticidade dos resíduos foi verificada pelo teste de hipóteses de White, que apresentou um valor calculado igual a 1,172 e um  $p$ -valor = 0,3257, valores estes que permitem aceitar a hipótese de que os resíduos são homocedásticos.

Os testes, os resultados e a interpretação da avaliação dos pressupostos do modelo de análise de regressão linear múltipla estão demonstrados de forma resumida no Quadro 2.

Pressuposto	Teste	Estatística	Interpretação
Homocedasticidade	<ul style="list-style-type: none"> <li>• <b>WHITE</b> H<sub>0</sub>=homocedasticidade H<sub>1</sub>=Heterocedasticidade</li> </ul>	1,172 <i>p-valor = 0,3257</i>	Aceita-se H <sub>0</sub>
Normalidade	<ul style="list-style-type: none"> <li>• <b>Jarque-Bera</b> H<sub>0</sub>: normal; H<sub>1</sub>: não normal</li> <li>• <b>Histograma resíduos</b></li> </ul>	0,173 <i>p-valor = 0,2721</i>	Aceita-se H <sub>0</sub> Curva se aproxima da normal
Autocorrelação serial	<ul style="list-style-type: none"> <li>• <b>DW</b> H<sub>0</sub>=Não correlação H<sub>1</sub>=há correlação</li> </ul>	DW = 2,042	Aceita-se H <sub>0</sub>
Multicolinearidade	<ul style="list-style-type: none"> <li>• R<sup>2</sup></li> <li>• Matriz correlação</li> </ul>	0,76 Correlação = 0	Não há multicolinearidade

## Quadro 2 – Resumo dos testes de avaliação dos pressuposto do modelo de regressão linear múltipla

Após a avaliação dos pressupostos da análise de regressão para o modelo encontrado, efetuou-se a previsão para três períodos à frente, para os meses de janeiro a março do ano de 2009. Os valores preditos pela Equação (31) ajustada e os valores observados, bem como os erros da predição, são apresentados na Tabela 7.

**Tabela 7 – Previsão do custo mensal por regressão linear múltipla**

Período	Custo Real	Previsão	Erro
Janeiro/09	50.288,27	49.694,48	593,79
Fevereiro/09	34.741,63	65.057,88	-30.316,25
Março/09	45.684,50	46.905,08	-1.220,58

O modelo apresentou um erro percentual absoluto médio (MAPE) de 11,24%, valor esse considerado aceitável para efetuar as previsões.

#### 4.7 Estimação da regressão linear dinâmica

Estimou-se um modelo dinâmico para captar os efeitos dos meses passados, utilizando-se a abordagem geral para o específico, no qual inicialmente foram consideradas todas as variáveis com três defasagens cada uma, em função dos pedidos do setor efetuarem-se de três em três meses.

As variáveis foram retiradas paulatinamente durante o processo de estimação, testando-se quais parâmetros eram estatisticamente significativos no modelo e quais não eram; passando a excluir os não significativos e reduzindo o modelo até se chegar ao resultado apresentado na Tabela 8.

**Tabela 8 – Estimação dos parâmetros do modelo de regressão linear dinâmica**

Variáveis	Coeficiente	Erro Padrão	Estatística t	t-prob	Part. R <sup>2</sup>
Cmensal <sub>1</sub>	0,467219	0,07855	5,95	0,000	0,3830
Cmensal <sub>2</sub>	0,526659	0,07927	6,64	0,000	0,4364
CP <sub>2</sub>	-5.270,40	1.136	-4,64	0,000	0,2742
CP <sub>2_1</sub>	5.675,58	1.254	4,53	0,000	0,2644
CP <sub>3</sub>	-2.833,39	1.148	-2,47	0,017	0,0966
CP <sub>3_1</sub>	3.351,58	1.086	3,09	0,003	0,1432
CP <sub>4_2</sub>	-2.499,31	1.045	-2,39	0,020	0,0912
CP <sub>5</sub>	-3.302,71	1.098	-3,01	0,004	0,1371
CP <sub>5_2</sub>	3.164,61	1.100	2,88	0,006	0,1268
CP <sub>6</sub>	7.218,78	1.109	6,51	0,000	0,4262
CP <sub>6_1</sub>	-4.265,17	1.154	-3,70	0,000	0,1934
CP <sub>6_2</sub>	-3.658,47	1.213	-3,02	0,004	0,1376
CP <sub>7</sub>	-4.136,40	1.105	-3,74	0,000	0,1973

Observando-se a Tabela 8, nota-se que todas as estimativas para os parâmetros são estatisticamente significativas em nível de 5%, indicando que no modelo as componentes principais utilizadas em substituição às variáveis originais explicam 74,06 ( $R^2 = 0,7406$ ) das variações do custo mensal dos antibióticos utilizados em IH.

Para verificar o ajuste e a significância do modelo encontrado, foram aplicados testes para avaliar os pressupostos da regressão dinâmica, em que

inicialmente se analisou o coeficiente de determinação  $R^2$ , que apresentou o valor de 0,74, e a matriz de correlação das componentes principais.

Como já foram detalhadas anteriormente, as CP são ortogonais entre si, não havendo a ocorrência de multicolinearidade entre elas.

O teste para ausência de autocorrelação serial dos resíduos foi efetuado pelo teste de Breush-Godfrey (BG), que apresentou o valor calculado igual a 0,4343 e um  $p$ -valor = 0,8223, indicando que se deve aceitar a hipótese nula ( $H_0$ ), de que não existe autocorrelação serial nos resíduos.

A normalidade dos resíduos deste modelo foi verificada por meio do histograma dos resíduos e pelo teste de hipóteses de Jarque-Bera, que apresentou um valor calculado de 0,7982 e um  $p$ -valor de 0,6709. Esses valores indicam que se deve aceitar a hipótese nula de que os resíduos se distribuem normalmente.

A normalidade dos resíduos pode ser percebida pela análise da curva dos resíduos que segue o padrão da curva teórica normal, conforme demonstrado na Figura 7.

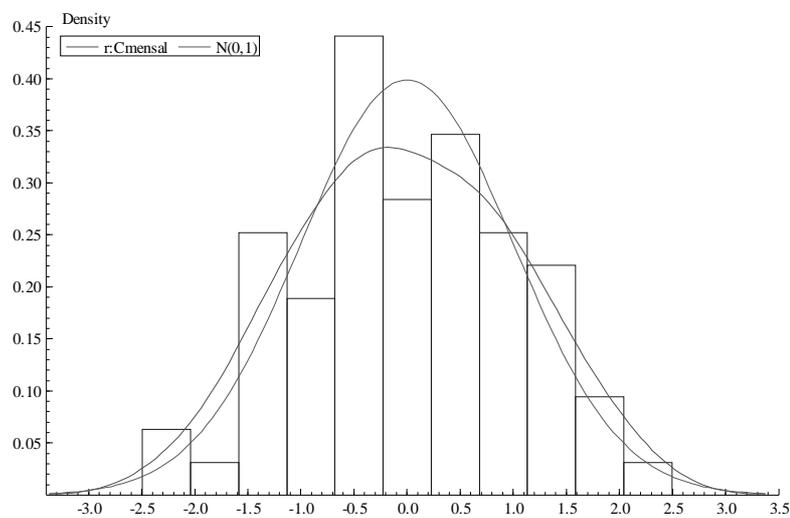


Figura 7 – Histograma dos resíduos da regressão dinâmica

A homocedasticidade dos resíduos foi verificada pelo teste de White, que apresentou um valor calculado igual a 1,256 e um  $p$ -valor = 0,2721, indicando, assim, que deve ser aceita a hipótese nula de que os resíduos são homocedásticos.

Os testes, os resultados estatísticos e a interpretação da avaliação dos pressupostos do modelo de análise de regressão linear dinâmica estão demonstrados de forma sintética no Quadro 3.

Pressuposto	Teste	Estatística	Interpretação
Homocedasticidade	<ul style="list-style-type: none"> <li>• <b>WHITE</b> H<sub>0</sub>=homocedasticidade H<sub>1</sub>=Heterocedasticidade</li> </ul>	1,256 <i>p-valor</i> = 0,2721	Aceita-se H <sub>0</sub>
Normalidade	<ul style="list-style-type: none"> <li>• <b>Jarque-Bera</b> H<sub>0</sub>: normal; H<sub>1</sub>: não normal</li> <li>• <b>Histograma resíduos</b></li> </ul>	0,7982 <i>p-valor</i> = 0,6709	Aceita-se H <sub>0</sub> Curva se aproxima da normal
Autocorrelação serial	<ul style="list-style-type: none"> <li>• <b>Breush-Godfrey BG</b> H<sub>0</sub>=Não autocorrelação H<sub>1</sub>=há autocorrelação</li> </ul>	0,4343 <i>p-valor</i> = 0,8223	Aceita-se H <sub>0</sub>
Multicolinearidade	<ul style="list-style-type: none"> <li>• R<sup>2</sup></li> <li>• Matriz correlação</li> </ul>	0,74 Correlação = 0	Não há multicolinearidade

### Quadro 3 – Resumo dos testes de avaliação dos pressuposto do modelo de regressão linear dinâmica

Analisando-se os testes de especificação do modelo, conclui-se que o modelo se presta para realizar as previsões do custo mensal dos antibióticos utilizados em infecções hospitalares no HUSM.

Os resultados obtidos pelo modelo de regressão linear dinâmica podem ser visualizados na Tabela 9.

**Tabela 9 – Previsão do custo mensal por regressão linear dinâmica**

Período	Previsão	Erro Padrão	Custo real	Erro
Janeiro/09	56.422,13	8.002,00	50.288,27	-6.133,86
Fevereiro/09	35.423,67	8.833,00	34.741,63	-682,04
Março/09	48.525,40	1.066,00	45.684,50	-2.840,89

Os resultados auferidos para as previsões mostram que este modelo apresentou um erro percentual absoluto médio (MAPE) de 6,79%, índice considerado aceitável para efetuar as previsões.

#### 4.8 Escolha do modelo a partir do MAPE

A escolha do melhor modelo de previsão foi efetuada a partir dos resultados encontrados para o MAPE, por ele ser uma medida da precisão (ou da incerteza) do método. Quanto menor o valor do MAPE, melhor será a previsão.

O modelo de regressão linear múltipla apresentou um valor de MAPE de 11,24% e, para o modelo de regressão linear dinâmica, o MAPE resultou em um valor de 6,79%

Baseando-se, então, especificamente no MAPE, o modelo mais adequado para se fazer a previsão do custo mensal de antibióticos é o de regressão dinâmica.

#### 4.9 Estimativas das previsões

As previsões de alocação de recursos orçamentários para aquisição dos antibióticos para os meses de janeiro a março do ano de 2009 estão demonstradas na tabela 10.

**Tabela 10 – Previsão de Alocação dos Recursos Financeiros**

Período	Previsão de Alocação dos Recursos
Janeiro/09	56.422,13
Fevereiro/09	35.423,67
Março/09	48.525,40

Calculou-se o percentual de erro de previsão em relação ao erro absoluto total para os meses de janeiro, fevereiro e março de 2009. Os resultados encontrados indicam que a previsão do custo mensal dos antibióticos do mês de janeiro/2009 apresentou um erro absoluto de 1,44%, o mês de fevereiro, 0,16%, e a previsão do custo mensal para o mês de março de 2009 resultou em erro absoluto de 0,66%.

Estes percentuais indicam que o modelo apresenta previsões com elevado índice de exatidão e, portanto, se presta para realizar as previsões do custo mensal dos antibióticos. Espera-se, assim, que este trabalho auxilie no gerenciamento da alocação de recursos financeiros do Hospital Universitário de Santa Maria.

## 5 CONSIDERAÇÕES FINAIS, LIMITAÇÕES E RECOMENDAÇÕES

Esta pesquisa teve por propósito formular modelos de regressão visando a prever o custo mensal de antibióticos e as necessidades de alocação de recursos financeiros para a sua aquisição no Hospital Universitário de Santa Maria.

O estudo teve início na determinação do custo mensal dos antibióticos, em que se observou seu comportamento inconstante, uma vez que a ocorrência das IH e de outras doenças não obedecem a um padrão pré-definido.

Em relação ao custo médio mensal, foi possível concluir que, em média, não ocorreram grandes oscilações, sendo que, observando-se a evolução nos anos de 2003 a 2008, o mês que apresentou o menor custo médio foi outubro, e junho, o maior. Tal fato se deve às doenças de inverno, muito frequentes nesta época do ano.

O estudo possibilitou afirmar que a redução da ocorrência de infecções hospitalares diminuirá o custo dos antibióticos.

Na estimação dos modelos de regressão, foi detectada a presença de multicolinearidade nas variáveis preditoras, o que fez com que fosse aplicada a metodologia de Análise de Componentes Principais. Esta técnica possibilitou formar um novo conjunto de variáveis que, por serem ortogonais, eliminou a multicolinearidade existente entre as variáveis originais, possibilitando, então, construir os modelos de regressão linear e dinâmica.

Com a utilização da metodologia “geral para o específico” da *London School of Economics*, foi possível capturar, a partir de defasagens, um modelo parcimonioso capaz de fornecer previsões mais precisas, com menor erro percentual, em relação ao modelo de regressão linear múltipla.

Na análise e comparação dos modelos, pode-se concluir que ambos os modelos são potencialmente aceitáveis, pois satisfizeram os testes de hipóteses aplicados aos resíduos e foram capazes de prever valores próximos do real para o custo mensal dos antibióticos.

Foi escolhido como o modelo mais adequado para realizar as previsões, o modelo da regressão dinâmica, por ter apresentado o menor erro percentual absoluto médio (MAPE) de 6,79%, contra um MAPE de 11,24% do modelo de regressão linear.

As previsões encontradas podem se constituir em uma ferramenta gerencial para a administração hospitalar ao fornecer subsídios para o planejamento orçamentário e financeiro dos recursos, especialmente em uma época em que há uma escassez de recursos em escala global, com reflexos muito intensos nos custos da saúde; especialmente no caso do Hospital Universitário de Santa Maria, que por ser um hospital de ensino e prestar atendimento gratuito a mais de 33 municípios em 28 especialidades médicas, necessita de ferramentas gerenciais eficazes de apoio à tomada de decisões.

Em se tratando de gerenciamento de estoques, as previsões são importantes tanto nos processos de compras, saber o quanto e quando comprar; como no dimensionamento dos estoques, evitando a falta ou o excesso.

É de fundamental importância para os órgãos de planejamento contar com previsões lastreadas em ferramentas estatísticas que permitam destinar os recursos com segurança, pois, inúmeras vezes, o administrador se depara com a difícil decisão de reduzir gastos para não extrapolar o orçamento.

Dessa forma, com este trabalho, pretende-se fornecer ao hospital os meios para melhor gerir seus custos com antibióticos, aperfeiçoar o planejamento e a execução dos recursos financeiros disponíveis.

Dentre as limitações desta pesquisa, estão as interações com outros tipos de infecções que não as hospitalares, as quais não foram objeto do presente estudo.

Outra limitação encontrada foi que, ao utilizar-se a técnica de componentes principais para a realização das previsões, há uma perda de explicação da variabilidade dos dados. No entanto, esta técnica foi utilizada por ser o melhor caminho para eliminar a multicolinearidade existente entre as variáveis originais.

Seria interessante para pesquisas futuras estender as previsões para outros tipos de medicamentos e materiais de alto custo, de maneira a se dispor de modelos de previsão que permitam uma melhor alocação dos recursos financeiros, tanto no HUSM quanto em outras instituições.

## REFERÊNCIAS

ABBAD, G.; TORRES, C. V. **Regressão múltipla stepwise e hierárquica em psicologia organizacional**: aplicações, problemas e soluções. 2002. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413294X200200030004&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413294X200200030004&lng=en&nrm=iso)>. Acesso em: 05 out. 2008.

ALEXANDER, C. **Modelos de mercado**: um guia para análise de informações financeiras. Tradução: José Carlos de Souza Santos. São Paulo: Bolsa de Mercadorias & Futuros, 2005.

ANDERSON, D. R.; SWEENEY, D. J.; WILLIAMS, T. A. **Estatística aplicada à administração e a economia**. São Paulo: Pioneira Thompson Learning, 2005.

ANVISA. **Breve histórico do controle de infecção no país**. Brasília, 2005. Disponível em: <[http://anvisa.gov.br/divulga/noticias/2005/130505\\_4.htm](http://anvisa.gov.br/divulga/noticias/2005/130505_4.htm)>. Acesso em: 3 mai. 2009.

CATTEL, R. B. The screen test for the number of factors. **Multivariate Behavior Research**, v.1, p. 245-276, 1966.

CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. (Coords.). **Análise multivariada**: para os cursos de administração, ciências contábeis e economia. São Paulo: Atlas, 2007.

CUNHA, Rodrigo. **A quebra de patente de medicamentos anti-Aids**: benefícios sociais e econômicos para países periféricos. 2001. Disponível em: <<http://www.comciencia.br/especial/pataids/pataids01.htm>>. Acesso em: 25 fev. 2009.

FERNANDES, A.T. **Infecção hospitalar e suas interfaces na área da saúde**. São Paulo: Atheneu, 2000.

FILDES, R. Quantitative forecasting the state of the art: econometric models. **Journal of the Operational Research Society**, v. 36, n. 7, 549-580, 1985.

FRANK, M. O. et al. **Decrease in expenditures and selected nosocomial infections following implementation of an antimicrobial-prescribing improvement program**. Clin Perform Qual Health Care. 1997.

GUJARATI, D.N. **Econometria básica**. 4. ed. Rio de Janeiro: Elsevier, 2006.

HOEFEL, H. H. K. **Controle e prevenção de infecções humanas**. 2009. Disponível em: <<http://www.cih.com.br/>>. Acesso em: 30 mai. 2009.

JACKSON, J.E. **Principal components and factor analysis**: Part I – principal components. Journal of Quality Technology, October. v.12, n.4, pp.201 – 213, 1981

- JOHNSON, R. A & WICHERN, D. N. **Applied multivariate statistical analysis**. 3 ed. New Jersey: Prentice-Hall, 1992.
- KARLSSON, A. The use of principal component analysis (PCA) for evaluating results from pig meat quality measurement. **Meat Science**, Essex, v.31, n.4, p.423-34, 1992.
- LANDIM, P.M.B. **Análise estatística de dados geológicos multivariados**. Disponível em <<http://www.rc.unesp.br/igce/aplicada/textodi.html>>. Acesso em: 22 mar. 2009.
- LEVINE, D. M.; BERENSON, M. L.; STEPHAN, D. **Estatística: teoria e aplicações usando Microsoft Excel em português**. Rio de Janeiro: Livros Técnicos e Científicos, 2005.
- LÍRIO, G. S. W., Multivariate methods: a methodology to assessing the client's satisfaction about RBS-TV in the northwest region of Rio Grande do Sul - Brazil. In: THE 9TH WORLD MULTI - CONFERENCE ON SYSTEMICS, CYBERNETICS AND INFORMATICS, 1., 2005, Orlando. **Anais...** Orlando: International Institute of Informatics and Systemics, 2005.293-297.
- LOPES, Francisco L. **A medida da inflação no Brasil**. PUC: Departamento de Economia, 1985.
- MALHOTRA, N. K. **Pesquisa de marketing: uma orientação aplicada**. 3 ed. Porto Alegre: Boockman, 2001.
- MARDIA, K. V.; KENTE, J. T. & BIBBY, J. M. **Multivariate analysis**. London: Academic, 1979.
- MELO, C. M. M. **Divisão social do trabalho e enfermagem**. São Paulo: Cortez, 1986.
- MILONE, G.; ANGELINI, F. **Estatística aplicada**. São Paulo: Atlas, 1995.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Ed. UFMG, 2005.
- MOREIRA, L. F. **Multicolinearidade em análise de regressão**. Disponível em: <[http://www.ead.fea.usp.br/Semead/9semead/resultado\\_semead/trabalhosPDF/455.pd](http://www.ead.fea.usp.br/Semead/9semead/resultado_semead/trabalhosPDF/455.pd)>. Acesso em: 12 jan. 2009.
- MORRISON, D.F. **Multivariate statistical methods**. 2. Ed., New York, NY. Mc Graw Hill. (1976).
- OLIVEIRA, R., MARUYAMA, S. A. T. **Controle de infecção hospitalar: histórico e papel do estado**. Disponível em: <<http://www.fen.ufsg.br/revista/v10/n3/v10n3a23.htm>>. Acesso em: 22 jan. 2009.

PELLEGRINI, F. R.; FOGLIATTO, F. Estudo comparativo entre modelos de Winters e de Box & Jenkins para a previsão da demanda sazonal. **Revista Produto e Produção**, v. 4, número especial, p. 72-85, abr. 2000.

PEREIRA, Ana Lúcia A. & RAMALHO, Wanderley. **Números índices: conceitos e aplicações**. UFMG: Departamento de Estatística, 1998.

PRADE, S. S.; VASCONCELLOS, M. M. **Identificando fontes de dados nos sistemas de informações do SUS para a vigilância epidemiológica das IH**. Informe Epidemiológico do SUS, 2001.

REINSEL, G. C. **Elements of multivariate time series analysis**. Springer- Verlag. New York, 1993

SARTORIS, A. **Estatística e introdução à econometria**. São Paulo: Saraiva, 2003.

SEBER, G.A.F. **Multivariate observation**. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc. NY, 1984

SILVA, W. V., **Curso de análise de regressão linear: teoria e aplicações**. Santa Maria: Universidade Federal de Santa Maria/Departamento de Estatística, 2005. Apresentação em PowerPoint.

SOCIEDADE BRASILEIRA DE INFECTOLOGIA. **Projeto diretrizes**. 2001. Disponível em: <[http://www.projetodiretrizes.org.br/projeto\\_diretrizes/065.pdf](http://www.projetodiretrizes.org.br/projeto_diretrizes/065.pdf)>. Acesso em: 3 dez. 2008.

SOUZA, A. M. et al. **Modelo de previsão de curto prazo para o consumo de combustível de uma empresa de coleta de resíduos sólidos domésticos**. Disponível em: <<http://www.posgraduacao.ufla.br/gauss/congresso/49rbras/pub/t029.pdf>>. Acesso em: 14 out.2008.

\_\_\_\_\_. **Monitoração e ajuste de realinhamento em processos produtivos multivariados**. 2000, 166f. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2000.

WELCH, K. E. et al. **A multi-center economic analysis of bacteremia caused by methicillin-resistant *Staphylococcus aureus***. Paper presented at 39th Interscience Conference on Antimicrobial Agents and Chemotherapy. San Francisco, CA; 1999

## **APÊNDICES**

## Apêndice A – Coordenadas das variáveis e percentual de variância explicada

Variáveis	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7
ACPOT001A.	-0,50	0,62	-0,08	-0,08	-0,12	-0,12	0,01
AMICA500A.	0,55	0,04	0,32	-0,01	-0,34	-0,06	-0,08
AMPIC001A.	-0,29	0,30	0,12	0,40	-0,23	-0,49	0,33
ANFOT050A.	0,06	-0,58	0,38	0,39	-0,16	0,10	-0,16
CEFAZ001A.	0,81	0,01	-0,15	-0,01	-0,16	-0,07	-0,01
CEFEP002A.	-0,68	0,08	0,44	0,14	0,00	-0,12	0,15
CEFTA001A.	0,82	0,02	0,04	0,05	-0,02	-0,12	0,01
CEFTR001A.	0,21	-0,49	-0,50	-0,11	-0,45	-0,06	0,03
CIPRO200FR	-0,53	-0,54	-0,17	0,06	0,12	-0,25	-0,27
CLIND600A.	0,44	-0,40	0,04	-0,27	-0,09	-0,63	0,07
GENTA080A.	0,53	0,42	0,04	0,22	-0,06	-0,29	0,09
ICSMO500FR	-0,22	-0,29	0,40	-0,47	0,49	-0,31	0,10
MEROP001FR	0,19	0,06	0,15	0,54	0,18	-0,17	-0,62
METRO100FR	0,19	-0,26	-0,37	0,43	0,46	-0,04	0,37
OXACI500A.	0,59	-0,18	0,06	0,30	0,23	0,19	0,38
TEICO400A.	-0,23	-0,40	0,45	0,06	-0,36	0,25	0,28
VANCO001A.	-0,75	-0,20	-0,28	0,21	-0,16	-0,22	0,04
VANCO500A.	0,89	0,05	0,17	-0,13	0,11	-0,05	-0,02
% Var Explicada	28,38%	11,57%	7,82%	7,39%	6,44%	6,20%	5,57%

## Apêndice B – Autovetores para compor a combinação linear

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
ACPOT001A.	-0,221	0,433	-0,065	-0,073	-0,110	-0,116	0,011
AMICA500A.	0,242	0,024	0,274	-0,013	-0,314	-0,054	-0,081
AMPIC001A.	-0,130	0,207	0,098	0,350	-0,211	-0,459	0,328
ANFOT050A.	0,026	-0,405	0,323	0,335	-0,146	0,098	-0,162
CEFAZ001A.	0,356	0,008	-0,125	-0,007	-0,151	-0,067	-0,006
CEFEP002A.	-0,301	0,054	0,374	0,124	-0,003	-0,114	0,154
CEFTA001A.	0,364	0,013	0,036	0,046	-0,018	-0,113	0,005
CEFTR001A.	0,093	-0,341	-0,418	-0,093	-0,423	-0,061	0,031
CIPRO200FR	-0,237	-0,376	-0,139	0,055	0,108	-0,239	-0,265
CLIND600A.	0,193	-0,275	0,037	-0,234	-0,087	-0,592	0,065
GENTA080A.	0,236	0,289	0,033	0,193	-0,055	-0,273	0,089
ICSMO500FR	-0,098	-0,199	0,338	-0,406	0,452	-0,298	0,097
MEROP001FR	0,083	0,039	0,124	0,467	0,171	-0,159	-0,616
METRO100FR	0,084	-0,179	-0,314	0,374	0,425	-0,037	0,368
OXACI500A.	0,260	-0,123	0,054	0,259	0,211	0,180	0,382
TEICO400A.	-0,103	-0,278	0,383	0,051	-0,331	0,233	0,282
VANCO001A.	-0,330	-0,140	-0,235	0,183	-0,146	-0,205	0,044
VANCO500A.	0,396	0,033	0,140	-0,111	0,098	-0,050	-0,024

## Apêndice C – Escores das componentes principais

Mês/ano	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
jan/03	1,010438	-0,06428	1,051418	-0,26071	-1,11248	-1,6758	1,970425
fev/03	0,536481	-0,53533	0,549587	-1,20666	1,306109	-1,25304	-1,76628
mar/03	1,368522	1,108847	0,654529	-2,68376	-0,99647	-2,18694	-0,90411
abr/03	1,87998	0,39323	0,561734	-0,13068	-0,01628	0,381366	1,232181
mai/03	1,085073	1,78149	0,628095	-0,37551	1,130802	0,157565	-1,58491
jun/03	2,040454	-0,13813	-1,38243	1,920718	-1,94466	-0,79517	0,715635
jul/03	1,366599	0,146572	-0,15596	0,778808	1,607632	0,681942	-0,64943
ago/03	1,113102	1,317644	0,979234	-0,95809	-0,85316	-1,12457	-0,23188
set/03	1,217589	0,853034	-0,78286	-0,03908	-0,75814	-1,1932	-0,39976
out/03	1,612605	0,375176	0,886594	-0,38157	2,311405	-0,62438	1,703938
nov/03	1,482173	0,194209	0,984273	-0,29694	1,701021	-0,46056	-1,40509
dez/03	0,79888	0,92181	-0,19554	0,326243	1,17043	-0,54831	-0,28744
jan/04	2,904326	0,174208	0,710506	1,094662	-2,62011	1,352332	1,746037
fev/04	0,907395	1,1132	0,261233	-0,07418	-1,89017	-0,78722	-0,93158
mar/04	1,326125	1,284389	0,284492	1,038117	-1,6987	0,076735	-0,85058
abr/04	1,624543	-0,36403	0,602487	1,664483	2,775413	1,217048	1,795899
mai/04	0,885875	0,359846	0,852777	-0,28777	0,038572	0,695859	0,521981
jun/04	0,606433	-0,05401	-0,48067	0,286592	0,072807	0,30794	0,503917
jul/04	0,680483	-0,95213	-2,02886	-1,14545	0,518735	-1,283	1,288648
ago/04	1,04759	-0,04825	-1,21345	-1,47385	-0,32477	-0,1103	0,575196
set/04	0,807307	-0,96715	-2,33959	0,261386	1,493575	1,495297	0,497533
out/04	0,379711	-0,22402	0,062169	-0,14656	0,792741	0,820238	0,401525
nov/04	0,239536	0,092247	1,005638	-0,60147	-0,18716	0,130617	0,078794
dez/04	-0,03034	0,895454	-0,61704	-0,14054	1,229707	1,094154	-1,25614
jan/05	1,292121	-0,93782	0,520375	-0,63822	0,888808	0,80538	0,071647
fev/05	0,754943	-2,25755	-2,32797	1,167725	-0,42356	-0,33064	-0,85245
mar/05	0,486117	-0,69945	-1,0268	1,013256	0,320875	0,134768	-0,81434
abr/05	0,500811	-0,06226	0,086082	-0,59737	0,688912	-0,18384	-2,62859
mai/05	0,253649	-1,52524	0,18296	-0,05704	0,942616	0,176585	-1,48681
jun/05	0,839843	-0,71432	-0,60294	-2,73543	-0,88049	2,038915	1,141732
jul/05	0,52551	-0,70299	-1,33148	-2,6435	-0,6307	1,790955	-0,08706
ago/05	0,497235	-0,59127	0,679779	0,239875	0,857607	0,624873	-0,51352
set/05	0,630553	0,622471	0,328035	1,251331	-0,77835	0,940391	-2,00024
out/05	0,460365	-0,11568	-0,20327	0,923818	0,17984	-0,859	-0,57349
nov/05	-0,41798	-0,51511	0,858565	-0,08152	0,002638	0,828287	0,008289
dez/05	-0,34667	0,432623	-0,00354	0,816172	-1,18565	-0,10111	-1,05917
jan/06	-0,57872	-1,93022	1,332915	-0,7134	-1,52282	0,290288	-0,5242
fev/06	-0,48293	-0,06664	-0,07839	-0,01196	-1,39308	-0,18529	-0,22485
mar/06	-0,47015	-1,62173	-0,21015	0,758627	-0,0929	-0,06711	0,737898
abr/06	-0,43874	-0,69183	2,006358	-0,1958	1,104767	-0,11245	0,425951
mai/06	-0,5964	-0,99812	0,566777	0,551766	0,346515	0,098213	0,002715
jun/06	-0,94655	-0,81622	-0,97693	-0,13296	-0,82333	1,275629	-0,28534
jul/06	-0,56818	-1,11697	1,124294	0,421861	-0,64768	0,215617	-0,8536
ago/06	-0,60438	-2,55472	0,215974	0,459974	-1,41918	0,073937	0,752623
set/06	-0,85916	-1,55583	-0,67615	0,497367	1,155813	-1,70774	-0,53982
out/06	-0,59758	-1,81247	1,435017	0,151447	-0,44675	-1,33532	0,279357
nov/06	-0,71566	-1,95227	0,263318	1,303531	-0,33522	-1,94131	0,396587

Mês/ano	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
dez/06	-0,95775	-0,69422	-0,70511	0,80861	0,365029	0,962921	-0,39559
jan/07	-0,8069	-1,40915	1,596433	-1,12076	-0,0323	-0,69154	-0,70638
fev/07	-0,64983	0,813649	-1,94281	-0,06468	-0,71731	0,003526	-1,77812
mar/07	-0,94705	0,84072	1,251011	1,019219	0,278762	2,06426	0,055296
abr/07	-0,34853	0,232914	-0,32144	1,552574	0,116633	-0,33935	0,176679
mai/07	-1,22069	-0,16316	-0,70682	-1,2109	-0,77183	1,706192	0,682507
jun/07	-1,04193	0,476899	0,141402	-0,82219	0,136543	0,227213	-0,21294
jul/07	-1,22511	1,437892	-0,01152	-0,0466	-0,26056	-0,06559	1,020542
ago/07	-1,06895	0,565701	0,895753	0,453719	-0,09886	0,072085	-0,19693
set/07	-1,28608	1,227087	0,341589	0,045693	0,177496	-0,39426	1,04462
out/07	-0,75022	0,658902	0,832451	0,211119	-0,08716	-0,60883	2,360067
nov/07	-0,85274	0,278388	-1,50066	0,198544	0,457208	-2,22839	1,335916
dez/07	-1,11385	1,061299	0,000328	-1,50317	0,447561	-0,91031	0,824299
jan/08	-0,8069	-1,40915	1,596433	-1,12076	-0,0323	-0,69154	-0,70638
fev/08	-0,64983	0,813649	-1,94281	-0,06468	-0,71731	0,003526	-1,77812
mar/08	-0,94705	0,84072	1,251011	1,019219	0,278762	2,06426	0,055296
abr/08	-0,34853	0,232914	-0,32144	1,552574	0,116633	-0,33935	0,176679
mai/08	-1,22069	-0,16316	-0,70682	-1,2109	-0,77183	1,706192	0,682507
jun/08	-1,04193	0,476899	0,141402	-0,82219	0,136543	0,227213	-0,21294
jul/08	-1,22511	1,437892	-0,01152	-0,0466	-0,26056	-0,06559	1,020542
ago/08	-1,05613	0,571466	0,909595	0,579137	-0,16141	0,161046	-0,37724
set/08	-1,14931	1,288577	0,489244	1,383491	-0,48975	0,554661	-0,87873
out/08	-0,63503	0,710688	0,956804	1,337796	-0,64911	0,190347	0,740248
nov/08	-0,80273	0,30087	-1,44668	0,687677	0,213245	-1,88143	0,632692
dez/08	-1,08778	1,07302	0,028475	-1,24815	0,320367	-0,72942	0,457661
jan/09	-0,36616	1,355335	-1,29525	0,320461	-0,60079	0,347086	0,417191
fev/09	-1,10988	1,367384	-1,20558	0,691212	1,80707	-0,79955	0,492345
mar/09	-0,79217	0,295557	-1,35464	-1,49721	1,143704	0,615999	0,930457