

JAQUELINE PIVETA

**MÓDULO PARA ENSINO-APRENDIZAGEM EM
ANÁLISE DE CORRESPONDÊNCIA MÚLTIPLA
INTEGRADO AO SESTAT.NET**

FLORIANÓPOLIS – SC

MARÇO DE 2003

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM
CIÊNCIA DA COMPUTAÇÃO**

Jaqueline Piveta

**MÓDULO PARA ENSINO-APRENDIZAGEM EM
ANÁLISE DE CORRESPONDÊNCIA MÚLTIPLA
INTEGRADO AO SESTAT.NET**

Dissertação submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Mestre em Ciência da Computação.

Silvia Modesto Nassar,
Orientadora

Florianópolis, Março de 2003

MÓDULO PARA ENSINO-APRENDIZAGEM EM ANÁLISE DE CORRESPONDÊNCIA MÚLTIPLA INTEGRADO AO SESTAT.NET

Jaqueline Piveta

Esta Dissertação foi julgada adequada para obtenção do título de Mestre em Ciência da Computação, Área de Concentração Sistemas de Conhecimento e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Prof. Fernando A. Ostuni Gauthier, Dr. (coordenador)

Banca Examinadora

Profa. Silvia Modesto Nassar, Dra. (orientadora)

Profa. Araci Hack Catapan, Dra.

Prof. Pedro Alberto Barbetta, Dr.

Prof. Marcelo Menezes Reis, Dr.

Agradecimentos

À professora Silvia Modesto Nassar, pela sua orientação, confiança e amizade.

A todos os professores e funcionários do Programa de Pós-graduação em Ciência da Computação, pelas contribuições fornecidas a esta pesquisa.

À toda minha família, pelo amor, apoio e paciência.

A todos que contribuíram direta ou indiretamente para a realização desta pesquisa.

SUMÁRIO

LISTA DE FIGURAS.....	iii
LISTA DE QUADROS.....	iv
LISTA DE TABELAS	v
LISTA DE SIGLAS	vi
RESUMO.....	vii
ABSTRACT	viii
1. INTRODUÇÃO	1
1.1. Justificativa.....	2
1.2. Objetivo Geral	3
1.3. Objetivos Específicos	3
1.4. Estrutura do Trabalho	4
2. EDUCAÇÃO E TECNOLOGIA.....	5
2.1. Tecnologia Educacional	8
2.1.1. Ensino mediado por computador.....	9
2.1.2. Educação a distância mediada por computador.....	12
2.2. Considerações Finais	16
3. ENSINO DE ESTATÍSTICA	18
3.1. SEstat.....	20
3.1.1. A primeira versão - o SEstat.....	20
3.1.2. A segunda versão - o Edust	23
3.1.3. A terceira versão - o SEstat.Net	24
3.2. Considerações Finais	25
4. ANÁLISE ESTATÍSTICA DE DADOS.....	26
4.1. Tabela de Frequência Observada ou Tabela de Contingência.....	27
4.2. Tabela de Frequência Esperada	27
4.3. Contribuição do Qui-quadrado	28
4.4. Análise Estatística Multivariada.....	29
4.4.1. Representação dos dados.....	30
4.4.2. Técnicas de análise multivariada.....	31
4.5. Análise de Correspondência Múltipla	32
4.5.1. Tabela de Burt	33
4.5.2. Autovalores e autovetores	34
4.5.3. Número máximo de dimensões	36
4.5.4. Inércia	36
4.6. Considerações Finais	36
5. INTELIGÊNCIA ARTIFICIAL	38

5.1. Inteligência Artificial na Educação	38
5.2. Sistemas Especialistas	39
5.2.1. Características dos sistemas especialistas	39
5.2.2. Vantagens dos sistemas especialistas	40
5.2.3. Problemas enfrentados pelos sistemas especialistas.....	41
5.2.4. Componentes de um SE.....	42
5.2.5. Ciclo de Vida para o Desenvolvimento de Sistemas Especialistas	44
5.3. Considerações Finais	47
6. DESENVOLVIMENTO DO MÓDULO MAC.....	48
6.1. Análise.....	48
6.2. Especificação.....	49
6.3. Desenvolvimento do módulo MAC.....	51
6.3.1. Tecnologias utilizadas para o desenvolvimento	51
6.3.1.1. Java Server Pages (JSP).....	52
6.3.1.2. JavaBeans	53
6.3.1.3. JDBC	54
6.3.1.4. Java Servlet.....	54
6.3.2. Modelagem do módulo MAC.....	55
6.3.2.1. Classe administrador	56
6.3.2.2. Classe estatística	57
6.3.2.3. Classe base de dados	59
6.3.2.4. Classe leitor DBF	59
6.3.2.5. Classe gráfico	60
6.3.3. Concepção da interface.....	60
6.3.3.1. Interfaces do módulo MAC	61
6.3.4. Avaliação e Validação dos Algoritmos	77
6.4. Disseminação.....	77
6.5. Considerações Finais	78
7. CONCLUSÃO.....	79
7.1. Trabalhos Futuros	80
REFERÊNCIAS	81
ANEXOS	

LISTA DE FIGURAS

Figura 2.1 – Relação Tripolar num ambiente pedagógico informatizado, adaptado de (QUEIROZ , 2001).....	7
Figura 3.1 – Arquitetura do Edust, adaptado de (DIAS, 2000).....	23
Figura 4.1 – Representação matricial dos dados	30
Figura 5.1 – Componentes de um Sistema Especialista, adaptado de (STAIR, 1998).....	42
Figura 6.1 – Modelo de Classes do SEstat.Net, adaptado de (NAKAZAWA & MARAFON, 2003).....	56
Figura 6.2 – Interface de entrada do sistema.....	61
Figura 6.3 – Interface de escolha da base de dados.....	62
Figura 6.4 – Interface de escolha do método.....	63
Figura 6.5 – Interface de escolha das variáveis.....	64
Figura 6.6 – Interface de tipificação das variáveis	65
Figura 6.7 – Interface de Aviso sobre a escolha de variáveis quantitativas e qualitativas	66
Figura 6.8 – Interface de categorização da variável quantitativa	67
Figura 6.9 – Interface de alerta sobre o número de categorias.....	68
Figura 6.10 – Interface de criação das categorias.....	69
Figura 6.11 – Escolha da variável suplementar.....	70
Figura 6.12 – Interface do resultado: Tabela de Frequência por Variável	71
Figura 6.13 – Interface de reagrupamento de variáveis.....	72
Figura 6.14 – Interface de resultados obtidos.....	73
Figura 6.15 – Interface de resultado: Tabela de Frequência Observada	74
Figura 6.16 – Interface de resultado: Tabela de Frequência Esperada.....	74
Figura 6.17 – Interface de resultado: Tabela de Contribuição do Qui-quadrado	75
Figura 6.18 – Interface de resultado: Tabela de Autovalores e Inércia.....	75
Figura 6.19 – Interface de resultado: Gráfico de Autovalores	76
Figura 6.20 – Interface de resultado: Gráfico de dimensão.....	76

LISTA DE QUADROS

QUADRO I – Comparação entre as vantagens para o aluno, professor e objetivo no uso da Tecnologia Educacional, adaptado de (NISKIER, 1993)	9
QUADRO II – Ciclo de vida de um Sistema Especialista, adaptado de (SCOTT et al, 1991).....	45
QUADRO III – Resumo dos métodos implementados na classe Estatística.....	58

LISTA DE TABELAS

Tabela 4.1 – Tabela de frequência observada.....	27
Tabela 4.2 – Tabela de frequência esperada.....	27
Tabela 4.3 – Tabela de contribuições do qui-quadrado.....	28
Tabela 4.4 – Tabela de Burt de frequência observada.....	34

LISTA DE SIGLAS

ACM	Análise de Correspondência Múltipla
API	<i>Application Programming Interface</i>
CASE	<i>Computer Assisted Software Engineering</i>
DBF	<i>Database File</i>
HTML	<i>Hypertext Markup Language</i>
HTTP	<i>Hipertext Transfer Protocol</i>
IA	Inteligência Artificial
INE	Departamento de Informática e de Estatística
J2EE	<i>Java 2 Enterprise Edition</i>
JDBC	<i>Java DataBase Connectivity</i>
JSP	<i>Java Server Pages</i>
LEA	Laboratório de Estatística Aplicada
MAC	Módulo de Análise de Correspondência Múltipla
PAP	Programa de Alimentação Popular
SE	Sistema Especialista
SEstat	Sistema Especialista para Apoio ao Ensino de Estatística
SETip	Sistema Especialista para Tipificar Dados de uma Pesquisa: Variáveis Qualitativas e Variáveis Quantitativas
UFSC	Universidade Federal de Santa Catarina
WEB	<i>World Wide Web</i>

RESUMO

Nesta pesquisa foi desenvolvido um módulo inteligente para apoio ao ensino-aprendizagem de estatística, presencial ou a distância. Este módulo, a partir de uma base de dados fornecida pelo aluno, realiza a análise de correspondência múltipla e será integrado a um ambiente de apoio ao ensino-aprendizagem de estatística, o SEstat.Net.

A análise de correspondência é um procedimento de análise estatística multivariada que busca identificar padrões de associação em um conjunto de variáveis categóricas. Os resultados estatísticos desta técnica são as tabelas de frequência de Burt, autovalores e autovetores, inércia e gráficos que permitem verificar exploratoriamente a existência de associação entre as categorias das variáveis.

O módulo foi desenvolvido utilizando a plataforma Java 2, mais especificamente a tecnologia J2EE, com os componentes JSP, JavaBeans, JDBC e Java Servlets, que foram as mesmas tecnologias usadas para desenvolver o SEstat.Net.

Com o desenvolvimento deste módulo foi possível ampliar a base de conhecimento do SEstat.Net, permitindo que sejam identificados padrões de associação em um conjunto de variáveis qualitativas, e também disponibilizar para os alunos de graduação e pós-graduação uma técnica complexa de análise de dados.

Palavras-Chave: Análise de Correspondência, Sistemas Especialistas, Ensino de Estatística.

ABSTRACT

In this research an intelligent module was developed to support the teaching of statistics presential or in a distance. This module, starting from a database supplied by the student, accomplishes the multiple correspondence analysis on a group of qualitative variables observed in a quantitative survey. This module will be integrated in a support environment for statistics teaching, SEstat.Net.

The correspondence analysis is a procedure of multivariated statistic analysis that aims to identify association patterns in a group of categorical variables. The statistic results of this technique are the frequency tables of Burt, self-values and self-vectors, inertia and graphs allowing to explore and evaluate the existence among the categories of the variables.

The module was developed by Java 2 platform, specifically the technology J2EE, with the components: JSP, JavaBeans, JDBC and Java Servlets, that were the same technologies used to develop the SEstat.Net.

With the development of this module it was possible to enlarge the base of knowledge of Sestat.Net, allowing us to identify association patterns in a group of qualitative variables, providing the complex technique of analysis and data to undergraduate and master degree students.

Keywords: Correspondence Analysis; Specialist Systems; Statistics teaching.

1. INTRODUÇÃO

O conhecimento não é transferido ou depositado, nem é inventado, mas é construído na relação com os outros e com o mundo. Dessa forma, o conteúdo apresentado pelo professor precisa ser trabalhado, refletido e reelaborado, pelo aluno, para se constituir em conhecimento. Caso contrário, o aluno não aprende, podendo, quando muito, apresentar um comportamento condicionado, baseado na memorização (VASCONCELLOS, 1997).

De acordo com o mesmo autor, o processo de construção do conhecimento por parte dos alunos pode ser mediado pelo professor e isto exige dele a mobilização para o conhecimento. Cabe ao professor não apenas apresentar os elementos a serem conhecidos, mas despertar, como frequentemente é necessário, e acompanhar o interesse dos alunos pelo conhecimento.

Não se pode desconsiderar que as novas tecnologias trouxeram diferentes maneiras de ensinar e de aprender. Ao rejeitar as facilidades que elas trazem ao sistema escolar, principalmente no que se refere às maiores escalas de disseminação do conhecimento, acabará por limitar a educação às paredes do passado, o que não representaria a realidade, pois vive-se em um mundo globalizado onde o conhecimento se multiplica a cada instante e enquanto instituição responsável pelo ensino do saber historicamente acumulado, deve-se conhecer e fazer uso das tecnologias disponíveis (UNOPAR, 2002).

E essas novas tecnologias, com qualidades inovadoras, não são apenas uma nova maneira de disseminar o conhecimento disponível, que pode facilitar a vida do professor e de seus alunos, mas sim uma dinâmica diferente para que o aprendizado ocorra dentro de novas condições (UNOPAR, 2002).

O importante é analisar sem preconceitos as vantagens reais das novas tecnologias, aceitando uma postura educacional crítica, reflexiva e dinâmica que serve a qualquer sistema, promovendo um processo de ensino-aprendizagem bilateral no qual professores e alunos possam aprender e ensinar juntos (UNOPAR, 2002).

Partindo desses pressupostos, para que o aluno construa o conhecimento e sua autonomia, é necessário estar inserido em um ambiente onde haja intervenções pedagógicas. Deve-se valorizar o processo de aprender a aprender, a formação das capacidades, o desenvolvimento da criatividade pessoal e do reconhecimento do outro como sujeito, a criação de atividades que privilegiem o conhecimento e, por fim, a possibilidade de verificar o desempenho dos alunos nas diversas práticas escolares, para encadear sempre a correção de rumos e o replanejar (SHUDO, 2001).

Segundo SAMPAIO & LEITE (1999), a utilização das tecnologias só auxiliará o desenvolvimento de uma educação transformadora se for baseada em conhecimento que permita ao professor interpretar, refletir e dominar criticamente a tecnologia.

Os *softwares* educacionais encontram-se ainda em um estágio bastante rudimentar quanto à qualidade e concepção, pois muitos deles funcionam apenas como uma ferramenta de transmissão de conteúdo, o que o faz remeter ao ensino tradicional onde o aluno assume um papel passivo.

Baseado nessas teorias é que o Departamento de Informática e de Estatística (INE) da Universidade Federal de Santa Catarina (UFSC), por meio de seu Laboratório de Estatística Aplicada (LEA), iniciou o projeto de um software de apoio ao ensino-aprendizagem de análise estatística de dados, o SEstat, e é dando continuidade a este trabalho que surgiu a motivação de integrar ao mesmo um novo módulo inteligente, com o intuito de proporcionar o apoio ao ensino de análise estatística multivariada, em especial da análise de correspondência múltipla.

1.1. Justificativa

A análise multivariada é utilizada quando se deseja estudar um fenômeno complexo, onde se tenha que lidar com múltiplas variáveis para conhecer suas relações e interações. Existem várias técnicas que podem ser utilizadas para diversas finalidades específicas, todas elas com elevado grau de complexidade e exigindo, portanto, uma matemática relativamente sofisticada.

Segundo PEREIRA (1999), com o advento das facilidades de computação providas pela revolução tecnológica dos últimos 30 anos, a aplicação de análises multivariadas cresce expressivamente. É a tecnologia sendo incorporada para a viabilização da análise de dados.

Embora o conhecimento das modernas técnicas de análise tenha precedido a evolução tecnológica, foi apenas graças a ela que o conhecimento encontrou possibilidade real de aplicação (PEREIRA, 1999).

A integração de dois novos módulos para análise multivariada ao SEstat.Net, um de análise de correspondência múltipla, a presente pesquisa, e outro de análise fatorial por componentes principais (PAVAN, 2003) tanto contribuirá para o aumento do conhecimento do SEstat.Net como também facilitará o ensino-aprendizagem destas técnicas estatísticas tão complexas.

1.2. Objetivo Geral

Desenvolver um módulo inteligente para apoio ao ensino de análise de correspondência múltipla e integrá-lo ao ambiente de educação a distância SEstat.Net.

1.3. Objetivos Específicos

- Investigar o uso de tecnologias educacionais no processo de ensino-aprendizagem, buscando compor um fundamento teórico para o desenvolvimento do módulo inteligente de apoio ao ensino;
- Explorar o funcionamento do SEstat.Net para possibilitar a integração do módulo MAC com o mesmo;
- Investigar as tecnologias para desenvolvimento de sistemas para *Web*;
- Adotar um modelo de ciclo de vida para o desenvolvimento do módulo.

- Implementar algoritmos computacionais para o desenvolvimento da análise de correspondência múltipla;
- Implementar o módulo MAC proposto de acordo com a concepção de ensino pesquisada;
- Validar o módulo com a aplicação de vários testes com bases de dados diferentes.

1.4. Estrutura do Trabalho

Esta pesquisa está estruturada em 7 capítulos da seguinte maneira:

O capítulo 2 apresenta um estudo sobre a utilização da informática no processo ensino-aprendizagem, salientando o uso das tecnologias educacionais como o computador e da educação a distância mediada pelo mesmo.

O capítulo 3 relata a situação do ensino de estatística nos últimos tempos e a proposta do SEstat como um sistema de apoio a este ensino, bem como sua evolução desde a primeira versão até o SEstat.Net com ênfase na educação a distância.

O capítulo 4 mostra os conceitos estatísticos envolvidos na análise multivariada com uma ênfase maior em análise de correspondência múltipla.

O capítulo 5 traz a fundamentação teórica sobre inteligência artificial e sistemas especialistas, na qual foi baseado o desenvolvimento do módulo proposto.

O capítulo 6 apresenta a modelagem do módulo, as tecnologias empregadas no desenvolvimento e a concepção de interface.

As conclusões e trabalhos futuros desta pesquisa são apresentados no capítulo 7.

2. EDUCAÇÃO E TECNOLOGIA

Uma volta ao passado é necessária para se entender a atual situação da Educação no Brasil. São fatos marcantes: a política de educação implantada pelos jesuítas e desmantelada pelo Marquês de Pombal, a criação de escolas durante o reinado de Dom João VI, os direitos adquiridos pelas constituições, e o desenvolvimento da industrialização. A partir desse momento, a grande massa começou a tomar consciência da educação como meio de ascensão social e as comunicações mais rápidas começaram a favorecer a expansão da idéia do direito à educação e da igualdade de oportunidades (NISKIER, 1993).

A sociedade de hoje não é mais a sociedade do saber fechado como há pouco tempo atrás. A pós-modernidade, com o avanço das tecnologias, tem como característica a instabilidade, as mudanças rápidas, seja no nível dos conhecimentos, das técnicas ou das atitudes. O saber hoje é aberto, reorganizável e construído.

Dessa forma, a introdução de novas tecnologias no ambiente educacional requer uma reflexão sobre a maneira de ensinar e aprender, uma vez que estas não podem ser consideradas apenas como meros instrumentos, mas esta inclusão deve ser feita com objetivos pedagógicos claros (OLIVEIRA, 2001).

Este novo contexto já provocou dúvidas aos professores e traz algumas preocupações no âmbito pedagógico. Os mais radicais resistem ao uso dos meios tecnológicos em sala de aula, como se estes fossem inverter os papéis, desapropriando o professor da detenção do conhecimento, para assumir um papel degradante de auxiliar de máquinas (OLIVEIRA, 2001).

Em outras palavras, professores e escolas começam a tomar decisões lentas, em relação à utilização desses recursos, mais especificamente o computador, no processo de ensino. As dificuldades vão desde o número de equipamentos e sua qualidade até a qualificação do professor.

LEVY (1994) aponta que a escola é uma instituição que há cinco mil anos se baseia no falar/ditar do mestre, na escrita manuscrita do aluno e, há quatro séculos, em

um uso moderado da impressão. Dessa forma, a tradição da escola é muito forte, dificultando a utilização de outros elementos que não sejam giz, cadernos e livros.

Uma adequada integração da informática supõe, portanto, o abandono de um hábito milenar, o que não pode ser feito em alguns anos. De acordo com LIBÂNEO (2000), novas exigências educacionais pedem às escolas um novo professor, capaz de ajustar sua didática às novas realidades da sociedade, do conhecimento, do aluno, dos meios de comunicação.

A interação entre aluno e professor passa a ter função extremamente importante no processo de ensino. Para QUEIROZ (2001), por interação, em termos genéricos, entende-se a influência ou ação recíproca. Este é um dos principais eixos da teoria vygotskyana para se entender o desenvolvimento humano, pois o homem, como sujeito histórico e social, constitui-se nas interações estabelecidas com o meio, apropriando-se das experiências vividas pela humanidade.

Para VYGOTSKY (1989), as funções especificamente humanas têm um caráter mediatizado, isto é, primeiro se formam no meio social para depois se tornarem individual, através dos instrumentos e dos signos. Enquanto os instrumentos orientam a ação do sujeito externamente, os signos são orientados para o seu interior. A relação se dá através da mediação da trilogia: estímulo, resposta e elemento mediador, onde o sujeito categoriza as informações recebidas e as transforma em conhecimento, através dos elementos mediadores (ferramentas, instrumentos, signos, interventores, escola, por exemplo), sendo a escola imprescindível aos que vivem em uma sociedade letrada, por lhes propiciar condições para a formação dos conceitos científicos e introduzir novas formas de interpretar os aspectos do mundo real.

Diferentemente dos conhecimentos espontâneos adquiridos via interação direta com seu meio, o conhecimento fornecido pela escola é desde o início planejado, sistematizado e orientado para o desenvolvimento do sujeito (QUEIROZ, 2001).

De acordo com VYGOTSKY (1989), o ensino escolarizado deve: se adiantar ao desenvolvimento; trabalhar com as potencialidades; considerar o aluno como sujeito ativo e fornecer o conhecimento científico para que os conceitos espontâneos se elevem

ao nível dos científicos e estes favoreçam a tomada de consciência dos conceitos espontâneos trazidos pelos sujeitos.

Segundo QUEIROZ (2001), as considerações de Vygotsky, acerca do ensino escolarizado, devem ser entendidas como uma intervenção pedagógica interativa que lhe permite internalizar os conhecimentos externos existentes no âmbito da cultura, o que só será possível, a partir de interações de qualidade, onde o aluno desempenhe um papel ativo, caso contrário, o professor, ao invés de interventor, será um repetidor de informações que não serão internalizadas de fato pelo sujeito receptor.

Na concepção de VYGOTSKY (1989), o desenvolvimento humano só pode ser entendido a partir do desenvolvimento da sociedade. Isto significa dizer que na sociedade da informação, cada vez menos o sujeito precisará memorizar informações e cada vez mais precisará ter autonomia de pensamento, ou seja, será necessário aprender e dominar as formas de acesso à tecnologia, apropriar-se delas e das inúmeras informações existentes ao seu redor.

Segundo QUEIROZ (2001), um ambiente pedagógico informatizado que queira explorar as potencialidades dos alunos necessita, além do computador, instrumento mediador (software e sujeito), que também o professor faça a mediação entre o sujeito e a máquina (hardware e software), exercendo ainda o papel de interventor pedagógico, de forma que se estabeleça uma relação tripolar, ilustrada na Fig. 2.1.

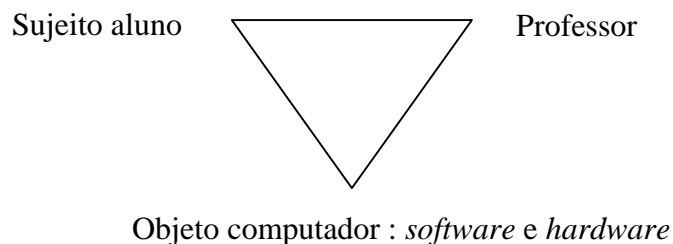


Figura 2.1 – Relação Tripolar num ambiente pedagógico informatizado, adaptado de (QUEIROZ , 2001)

Isto porque, na concepção de VYGOTSKY (1989), a preocupação de todo professor deve ser com o futuro, ou seja, aquilo que o sujeito pode ainda desenvolver e não apenas o que já desenvolveu até o momento. Para ele, as leis do desenvolvimento são as mesmas para todos e a educação tem um papel imprescindível, pois é ela quem vai auxiliá-lo na estrutura e na organização do conhecimento e da experiência acumulada pela história da humanidade.

2.1. Tecnologia Educacional

“A Tecnologia educacional é, fundamentalmente, a relação entre a Tecnologia e a Educação, que se concretiza em conjunto dinâmico e aberto de princípios e processos de ação educativa, resultantes da aplicação do conhecimento científico e organização à solução ou encaminhamento de soluções para problemas educacionais”.

Francisco Silveira Lobo Neto

Muitas têm sido as abordagens sobre o conceito de tecnologia educacional, segundo NISKIER (1993):

- Um exercício crítico com utilização de instrumentos a serviço de um projeto pedagógico;
- Uma linha de prestação de serviços à educação;
- Uma conjugação de recursos humanos e não-humanos para atingir um rendimento ótimo (qualitativo e quantitativo), com utilização de meios tecnológicos e tecnologias educacionais, com inúmeras possibilidades para solução de problemas;
- Uma mediação do encontro entre Ciência, Técnicas e Pedagogia.

O mesmo autor pontua algumas das vantagens da tecnologia educacional, que podem ser vistas no Quadro I, a seguir.

QUADRO I – Comparação entre as vantagens para o aluno, professor e objetivo no uso da Tecnologia Educacional, adaptado de (NISKIER, 1993)

Quanto ao aluno:	Quanto ao professor:	Quanto ao objetivo:
<ul style="list-style-type: none"> • Conhece os objetivos a serem atingidos; • É responsável pela escolha de seus próprios objetivos e dos meios para alcançá-los; • Compromete-se individualmente; • Torna-se capaz de auto-avaliar-se, de tomar decisões e determinar seu próprio tempo. 	<ul style="list-style-type: none"> • Não passa informações; coordena as informações, diagnostica os alunos, manipula os meios a seu dispor; • Torna-se capaz de especialização com o domínio do processo ensino-aprendizagem; • Pode introduzir novos conteúdos ou novos meios para atingir novos objetivos ou melhorar os objetivos já determinados. 	<ul style="list-style-type: none"> • São selecionados de acordo com os conhecimentos do aluno; • Permitem ao aluno comparar-se a si mesmo e não ao grupo; • São expressos de forma mensurável ou de comportamentos observáveis; • Permitem variedade de medições de acordo com sua natureza; • Podem ser modificados e redirecionados.

As tecnologias educacionais envolvem recursos, como cinema educativo, ensino por correspondência, videotexto e o computador. Destas, o computador tem apresentado maior crescimento, com possibilidades de uso para a educação a distância com o desenvolvimento da internet. Por isso, será dedicado um item à parte para explicitá-los e fundamentar a presente pesquisa.

2.1.1. Ensino mediado por computador

Conforme MARQUES et al (1995), a invenção do primeiro computador pode ser considerada um evento relativamente recente, levando em conta o período de tempo em que o homem vem desenvolvendo esforços no sentido de processar informações de

forma mais eficiente. Do primeiro computador até os atuais, ocorreu uma enorme evolução em relação à versatilidade, sofisticação e rapidez de processamento, bem como compactação das dimensões físicas.

A necessidade de realizar cálculos e processar informações condicionou o desenvolvimento da informática, sendo assim, ela está intimamente associada à evolução da vida econômica e ao desenvolvimento das atividades das pessoas. Os computadores tornaram-se parte de suas vidas, elas visitam lojas e escritórios desenhados com a ajuda deles, lêem revistas e pagam contas preparadas por eles. Os benefícios da tecnologia também estão sendo explorados no ensino. O uso do computador está se integrando ao sistema educacional, mas com atraso se comparado com outros setores da sociedade.

Pode-se destacar duas posições diferenciadas sobre a forma mais adequada de utilizar o computador em educação: uma das posições afirma que o computador deve ser usado só como um recurso de aprendizagem e nesta abordagem, dispondo do instrumental necessário, em geral linguagens de programação, o aluno dirige seu próprio aprendizado; a outra posição defende o uso do computador como instrumento didático, fornecendo ao aluno programas educativos estruturados que visam cumprir um determinado objetivo, vinculado ou não ao currículo (MARQUES et al, 1995).

De acordo com MARQUES et al (1995), as duas formas de utilização não são excludentes, porém não se pode negar que a utilização do computador apenas como recurso de aprendizagem pode significar uma subutilização de um equipamento versátil. Por outro lado, o uso do computador como instrumento de ensino possibilita a introdução de novos “aspectos” a qualquer momento do processo ensino-aprendizagem.

MARQUES et al (1995) apontam vantagens que o computador apresenta em relação aos outros instrumentos por possuir as seguintes características:

- Interatividade: é um recurso audiovisual superior aos demais por permitir interação entre o aluno e a máquina;
- Individualidade: o computador possui a vantagem de poder obedecer ao ritmo próprio de cada aluno;

- Disponibilidade: a flexibilidade de horários. O aluno faz o seu próprio tempo;
- Prontidão: o aluno recebe o feedback às suas intervenções de forma instantânea.

Segundo o mesmo autor, estas características podem talvez ser responsáveis pelo alto grau de motivação, por parte dos alunos, em usar o instrumento sempre que possível. Isto por que, mesmo já tendo tido algum contato com o computador, os alunos continuam predispostos a novos contatos. A motivação é importante para qualquer aprendizagem, pois sem ela, é pouco provável que a atenção do indivíduo esteja voltada para o que deve aprender. Nesse sentido, a motivação, aliada a outros pontos positivos do computador, pode contribuir significativamente para o ensino-aprendizagem.

Para NISKIER (1993), ao se planejar o ensino-aprendizagem através do computador, tem que se ter em vista as seguintes dificuldades:

- A área educacional não possui o mesmo atrativo da área comercial;
- A área educacional é muito mais diversificada, o que exige programas mais flexíveis e diversidade;
- As mudanças de conteúdo e de filosofia educacional;
- A adaptação de textos para o computador;
- A falta de pesquisas amplas de como ele poderá ser utilizado;
- A subutilização, como substituição do material impresso;
- A falta de especialistas em conteúdos e no domínio do instrumento;
- A possibilidade de obsolescência rápida do equipamento.

Ainda segundo NISKIER (1993), é recomendável que para a implantação do ensino mediado por computador se tenha em mente:

- Tipo de programa (tutoria, simulação, relação com o ensino tradicional);

- A substituição de esforços rotineiros, reservando as pessoas para funções mais dignificadoras;
- Conhecimento de como será usado, para quê, por quem;
- A flexibilidade para atualizações posteriores;
- A natureza do que se deseja transmitir;
- A preparação do horário para trabalhar com a mente.

Ou seja, a adequação do uso da informática como instrumento auxiliar ao ensino-aprendizagem depende de diversos fatores, tais como: forma de aplicação, intensidade de uso, nível de ensino ao qual se destina, idade ou maturidade dos alunos expostos ao uso do computador. Por isso, é necessário que sejam alocados recursos adequados e regulares para a pesquisa e desenvolvimento de protótipos de sistemas de computação (*hardware* e *software*) voltados para ambientes educacionais.

2.1.2. Educação a distância mediada por computador

O presidente da república, em 10 de fevereiro de 1998 decreta: “Educação a distância é uma forma de ensino que possibilita a auto-aprendizagem, com a mediação de recursos didáticos sistematicamente organizados, apresentados em diferentes suportes de informação, utilizados isoladamente ou combinados, e veiculados pelos diversos meios de comunicação” (UNOPAR, 2002).

A expansão da internet permite eliminar as barreiras características do ensino tradicional / presencial, constituindo-se assim num meio alternativo no desenvolvimento de aplicações para a educação. Estas aplicações correspondem a diferentes produtos para utilização em aula ou para auto-aprendizagem, como, por exemplo, permitindo a possibilidade de comunicação por correio eletrônico, grupos de discussão, *chats* e ainda o *download* e *upload* de documentos (SÁ, 2001).

A educação a distância pela Internet possui características particulares que o distingue do ensino tradicional. Possibilita um processo de aprendizagem contínuo

através dos métodos alternativos de distribuição da informação proporcionados pela Internet.

A educação a distância é uma nova modalidade educativa, uma alternativa pedagógica que não vem substituir a educação presencial, mas é fruto de uma série de determinações presentes no atual estágio de desenvolvimento científico-tecnológico, econômico, político, cultural, midiático e educacional (SÁ, 2001).

Condicional a ela, há um novo paradigma educativo, na medida em que o docente não está mais, necessariamente, permanentemente presente face a face com seu aluno, no entanto, esta relação continua existindo num novo patamar, alicerçando-se na mediação estabelecida pelos meios didático-pedagógicos contemporâneos (SÁ, 2001).

As novas mídias em pleno desenvolvimento neste final de século, aportadas na informática e na comunicação têm possibilitado novas formas de intercomunicação que vêm garantindo ou possibilitando que a relação presencial / dialógica possa ser, também, realizada sobre outro paradigma que envolve as dimensões do tempo / espaço através de uma relação dialógica virtual (SÁ, 2001).

O processo de aprendizagem passa a ter ênfase na ação discente. O aluno tem um papel preponderante na construção e apropriação do conhecimento, se apoiando na mediação desenvolvida entre o material didático-pedagógico produzido pelo professor, o qual deveria possibilitar-lhe uma aprendizagem mais autônoma e segura (SÁ, 2001).

Por aprendizagem autônoma entende-se um processo de ensino e aprendizagem centrado no aluno, cujas experiências são aproveitadas como recurso, e no qual o professor deve assumir-se como recurso do aluno, considerado como um ser autônomo, gestor de seu processo de aprendizagem, capaz de auto-dirigir e auto-regular este processo. Este modelo de aprendizagem é apropriado a adultos com maturidade e motivação necessárias à auto-aprendizagem (BELLONI, 1999).

O controle direcionado pelo aluno é centrado em problemas, é experimental, realístico, envolve a realização de atividades práticas, de pesquisas e solicita, continuamente, o desenvolvimento de operações de pensamento superior, como a tomada de posição, comprovada por forte argumentação. Estimula a criatividade do

aluno, na geração de alternativas de situações e de respostas não programadas (WICKERT, 1999)¹

O docente nesta prática educativa procurará garantir o aprendizado do seu aluno à distância. Torna-se muito mais complexo seu papel na medida em que precisa passar a lecionar através de outras linguagens suportadas por meios informacionais e de comunicação, que atingem dezenas de pessoas, numa relação de tempo e espaço diferentes da atividade presencial que se conhece (SÁ, 2001).

Existem diferentes mecanismos de aprendizado mediados por computador, cada qual suportado de maneira diferente. Segundo MAGALHÃES (2002), a atual tecnologia base da *Web* provê suporte para criação de materiais que representam um relativo avanço sobre os convencionais livros e textos em papel, mas não é suficiente para o estabelecimento de um ambiente computacional para o apoio à aprendizagem. Entre os requisitos de um ambiente computacional para apoio à autoria de materiais educacionais podem ser citados:

- a escolha de um ou mais modelos cognitivos, que estabeleçam, entre outras coisas, se a aprendizagem deve ser alcançada individual ou cooperativamente;
- estabelecimento de mecanismos e metáforas de cooperação, conforme o caso;
- estabelecimento de modelos de avaliação adequados;
- a adaptação tanto do conteúdo do material educacional como de sua seqüência de apresentação ao perfil do aprendiz;
- modos mais avançados para a interação entre aluno e documentos dinâmicos, que auxiliem o mesmo na exploração do material didático e dos conceitos nele contidos;

¹ WICKERT, Prof^a Maria Lucia Scarpini. Palestra apresentada na mesa redonda: **O futuro da educação a distância no Brasil**. Promovida pelo Centro de Educação Aberta, Continuada / CEAD, da Universidade de Brasília em 05/04/1999.

- a natural dificuldade da criação de materiais educacionais digitais, considerando a hipermídia.

Para o ensino mediado por computador, correntes de pensamento contrastantes que vão do objetivismo ao construtivismo estabelecem extremos de um contínuo. Os teóricos objetivistas baseiam-se na idéia de que os sujeitos aprendem porque lhes falam e que o conhecimento é objetivo e existe independentemente daquele que aprende. Como conseqüência, o material usado para aquisição de conhecimento, que resulta desse tipo de pensamento, tende a ser mais explícito e declarativo. Por outro lado, os construtivistas presumem que os indivíduos aprendem fazendo, ou seja, constroem o conhecimento situando experiências cognitivas em atividades autênticas. Conseqüentemente, o conhecimento emergente se consolida nas raízes da compreensão que se tem do mundo e dos contatos que são estabelecidos com ele (MAGALHÃES, 2002).

SCHÄR & KRUEGER (2000) sugerem quatro diferentes tipos de sistemas de ensino-aprendizagem, que, conseqüentemente, requerem diferentes ferramentas e atividades de autoria:

- Baseados em simulação: aprende-se fazendo, através de sistemas que podem demonstrar as condições de ações e eventos no mundo real (construtivista).
- Hipertexto e hipermídia: organização do material didático em nós, em texto ou outras mídias, conectadas por ligações, que permite o acesso não estruturado através desses nós (construtivista).
- Sistemas tutores: aprendizado guiado por objetivos pré-programados, com a aquisição de conhecimentos sendo verificada através de testes, correções e sugestões para repetição de atividades (objetivista).
- Prática por repetição: tipo de aquisição estabelecida na forma de perguntas e respostas; o aprendizado se dá com recompensa a respostas corretas (objetivista).

- Representação abstrata: modelos gráficos, bastante comuns, para a percepção de representações químicas, físicas e matemáticas, que tornam atrativo o ensino de matérias básicas (objetivista).

Será cada vez mais imprescindível, intransferível e relevante, a responsabilidade do professor na busca da excelência do processo educacional. O professor necessita, mais do que nunca, exercer seu papel de coordenador das ações da equipe multidisciplinar que concebe, planeja e produz materiais educativos. Esta equipe reúne psicólogos, especialistas em informática, em comunicação, em Tecnologia Educacional e em cada meio tecnológico, sendo a criatividade e a flexibilidade atributos essenciais a todos os componentes (WICKERT, 1999)².

2.2. Considerações Finais

Como foi visto ao longo deste capítulo, há uma proposta de mudança no processo de ensino-aprendizagem, ou seja, de um modelo tradicional onde o professor é o único detentor do conhecimento, aquele que leva a informação para a sala de aula, para um modelo construtivista onde é o aluno quem constrói o seu conhecimento, participando ativamente deste processo.

Aliadas a essa forma de ensino-aprendizagem, inserem-se as tecnologias educacionais, que propõem meios para que o aluno construa seu conhecimento. Entre esses meios pode-se destacar o uso do computador e os *softwares* educacionais como ferramentas de apoio nesse processo de aprendizagem.

O papel do software educacional não deve ser apenas de substituir os livros e textos em papel pela transmissão da informação em mídia, mas sim prover meios onde o aluno possa interagir com o ambiente, ter autonomia para explorar o mesmo e desenvolver seu raciocínio.

² WICKERT, Prof^a Maria Lucia Scarpini. Palestra apresentada na mesa redonda: **O futuro da educação a distância no Brasil**. Promovida pelo Centro de Educação Aberta, Continuada / CEAD, da Universidade de Brasília em 05/04/1999.

Esta é a proposta da pesquisa, criar um módulo onde o aluno possa se apoiar para aprender análise estatística multivariada utilizando análise de correspondência múltipla, fornecendo ao mesmo, mecanismos de interação com o software e total liberdade para traçar o caminho que desejar na construção de seu conhecimento nesta técnica estatística.

3. ENSINO DE ESTATÍSTICA

POTTER (1995) destaca que o objetivo mais importante de um curso de Estatística é encorajar os estudantes a serem praticantes deste instrumental. O conhecimento estatístico definitivamente nada significa se ele não se relaciona a questões e problemas reais.

ROITER & PETROCZ (1996) identificam quatro abordagens principais para os cursos de Estatística, com seus respectivos conteúdos:

- Estatística como ramo da matemática: combinatórias, teoria da probabilidade e variáveis aleatórias, com atividades voltadas para demonstrações e derivações;
- Estatística como análise de dados: análise exploratória de dados, testes de hipótese, regressão e correlação e atividades voltadas para a coleta, investigação e análise de dados e confirmação de hipótese;
- Estatística como delineamento experimental: análise do efeito de variáveis sobre uma resposta, regressão e atividades voltadas para delineamento de experimentos, coleta de dados e interpretação de resultados.
- Estatística como resolução de problemas: análise exploratória de dados, delineamento de experimento e atividades voltadas para a resolução de problemas nas áreas de atuação.

O ensino de estatística tem sido prescritivo, impessoal, não dando lugar para as opiniões dos alunos e para a criatividade. As aulas são reduzidas a séries de pequenas e triviais atividades, nas quais são lecionados alguns conhecimentos, em que o livro e o professor são as únicas fontes de informação válidas.

Este tipo de ensino vai inibindo o desenvolvimento das personalidades dos alunos e das suas aspirações. Portanto, uma das tarefas dos professores é tornar significativa para os alunos a aprendizagem de estatística.

Problemas tais como, desinteresse dos alunos, dificuldades dos mesmos no trato com o conteúdo e na vinculação do conhecimento estatístico oferecido nas aulas com a

realidade dos campos de conhecimento, têm se tornado questões emergentes da docência. Muitas vezes, os alunos não conseguem visualizar como a metodologia estatística será aplicada na sua futura prática profissional e terminam o curso sem a instrumentação necessária para a utilização da Estatística na solução de problemas em sua área de atuação (OLIVEIRA & GRACIO, 1999).

O ensino de qualquer disciplina científica deve pretender obter, mais que um aluno que possua uma grande quantidade de conhecimento, desenvolver no indivíduo uma maneira de identificar problemas, pensar como resolvê-los utilizando os métodos mais adequados, controlando certas variáveis, observando e medindo determinados parâmetros, para finalmente verificar uma hipótese de trabalho (CAPRILE, 1991).

Dessa forma, o ensino e a aprendizagem da Estatística não podem limitar-se ao estudo de técnicas formais nem à aplicação rotineira de procedimentos descontextualizados e deverá em vez disso, permitir interpretar, analisar e criticar. De preferência, o aluno deveria poder trabalhar dados reais, começando por uma análise exploratória da informação, para em seguida, e utilizando a calculadora ou um software adequado, descobrir por si próprio as potencialidades e as desvantagens do método em questão.

Um outro problema encontrado no ensino de estatística são as ferramentas disponíveis para auxiliar o aluno e o professor nesse processo de ensino-aprendizagem. Existem, atualmente, muitos *softwares* estatísticos no mercado, porém, a maioria deles é de difícil utilização por serem voltados a um tipo de usuário que já tenha um conhecimento estatístico razoável, servindo apenas como um facilitador do seu trabalho. No caso do aluno, ele está numa posição de aprendiz, ou seja, ele precisa de uma ferramenta que o ensine a fazer o seu trabalho, que digam quais os procedimentos escolher para resolver um determinado problema e como interpretar os resultados, por exemplo.

Com o intuito de tentar minimizar os problemas no ensino de estatística, surgiu a idéia de criar um *software* que além de servir como uma ferramenta na qual o aluno teria um suporte na aprendizagem, também daria total liberdade ao mesmo de trabalhar com os dados, ou seja, o problema, elaborado por ele, de seu domínio, e não problemas

descontextualizados com sua realidade. Dessa idéia surgiu o SEstat, o qual será discorrido nas próximas seções.

3.1. SEstat

O SEstat (Sistema Especialista de Apoio ao Ensino de Estatística) é um projeto que vem sendo desenvolvido pelo Departamento de Informática e Estatística (INE) da Universidade Federal de Santa Catarina (UFSC), através do seu Laboratório de Estatística Aplicada (LEA), com o objetivo de ser usado como ferramenta de apoio ao ensino-aprendizagem da Análise Estatística de Dados, nas disciplinas de estatística.

3.1.1. A primeira versão - o SEstat

O SEstat foi desenvolvido pela primeira vez, em 1997 por dois alunos do curso de Ciência da Computação da UFSC, Cristian Cechinel e Lidiane Amorim. Para CECHINEL (1999), a utilização do SEstat, como integrante de um ambiente de ensino de Estatística, fez com que o usuário/aluno passasse a ser *sujeito* do processo de aprendizagem. A intenção era que o usuário/aluno tivesse uma atitude de buscar ativamente a construção do seu conhecimento, saindo da tradicional postura passiva de puro receptor. De acordo com essa abordagem, durante o processo de aprendizagem, quando sentisse necessidade de apoio, o aluno demandava o auxílio ou intervenção do(s) professor(es), com o intuito de obter uma orientação para o progresso de seu aprendizado: onde ele estivesse naquele momento, e como prosseguir para atingir os objetivos levando em conta seus conhecimentos prévios.

Para o desenvolvimento do SEstat, foram usados paradigmas de Inteligência Artificial (IA), pois segundo CECHINEL (1999), um dos campos no qual a IA se destaca é o que trata de Sistemas Especialistas (SE), onde se desenvolve um sistema que possui a habilidade de tratar os problemas relacionados a um domínio, de forma semelhante a um especialista naquele domínio. Dotados de grande interatividade, os SEs proporcionam aos seus usuários uma espécie de diálogo que prende a atenção, e

acaba por transmitir a eles algum conhecimento sobre o assunto abordado. Essas características motivaram a utilização dos paradigmas de inteligência artificial direcionados à área de sistemas especialistas, na construção do *software* de apoio ao ensino de estatística, que se decidiu chamar de SEstat – Sistema Especialista de Apoio ao Ensino de Estatística.

Conforme CECHINEL (1999), o sistema funcionava da seguinte maneira: a partir de uma Base de Dados (em formato *.dbf*) fornecida pelo usuário, o SEstat faz um conjunto de perguntas a ele, e fornece ao mesmo suporte para que essas perguntas sejam respondidas corretamente. Ao longo da interação do usuário com o sistema, o caminho a ser percorrido nessa interação é escolhido através das respostas dadas pelo usuário às perguntas referentes à análise estatística que se está realizando. No final da interação, o sistema escolhe e aplica uma técnica de análise estatística para os dados fornecidos pelo usuário, e explica ao mesmo por que aquela técnica foi escolhida, bem como apresenta explicações que possibilitam a interpretação dos resultados estatísticos encontrados.

O SEstat foi estruturado em diversos módulos, onde cada um possuía uma característica, uma tarefa específica e a capacidade de se comunicar entre eles para decidir qual o módulo capaz de executar determinada tarefa. Os módulos eram os seguintes:

- *Gerenciador*: decide o módulo que irá executar a tarefa.
- *Estratégia*: a partir das informações fornecidas pelo Gerenciador, procura a estratégia adequada para resolver o problema.
- *Processamento*: responsável pela comunicação com o *software Statistica 5.0*.
- *Software Statistica 5.0*: realiza a análise de dados.
- *Mecanismo de Explicação*: módulo de ajuda do sistema, cuja finalidade é explicar ao usuário os conceitos estatísticos, além de mostrar o caminho que foi percorrido para se chegar às decisões tomadas.

- *Interface*: recebe do usuário as informações que serão processadas pelo sistema e mostra ao usuário sua evolução, assim como as soluções obtidas e as respostas às perguntas feitas pelo usuário.

A partir de 1999, o SEstat passou a ser usado como uma ferramenta de auxílio no ensino da disciplina de Métodos Estatísticos para cursos de graduação da UFSC e teve um grande êxito, pois os alunos, antes passivos diante do processo de aprendizagem, viraram, então, sujeitos fundamentais desse processo.

Mas alguns problemas surgiram com a sua utilização: devido ao fato do SEstat trabalhar com bases de dados flexíveis, o sistema não possuía conhecimento sobre as características das variáveis que estavam sendo analisadas. Dessa forma, o SEstat permitia que o aluno respondesse a perguntas de forma incorreta, ocasionando uma análise estatística incorreta. Em momento algum o aluno era avisado que havia errado e que essa análise estava incorreta.

Tentando solucionar este problema, foi desenvolvido por um aluno da pós-graduação em Ciência da Computação, José Gonçalo dos Santos, um sistema que, integrado ao SEstat, pudesse classificar as variáveis em qualitativas ou quantitativas, pois segundo SANTOS (2001), a importância dessa identificação é devido ao fato de que dados categorizados têm um tratamento diferenciado de dados quantitativos em análise estatística. E, se não for feita esta distinção, os resultados da pesquisa podem ficar seriamente comprometidos. A este sistema foi dado o nome de SETip – Sistema Especialista para Tipificar dados de uma Pesquisa: Variáveis Qualitativas e Quantitativas.

Um outro problema encontrado foi a fragilidade na comunicação entre o SEstat e o *software Statistica*, pois, devido ao fato de o canal de comunicação entre os dois ser basicamente unidirecional, as requisições chegavam até o *Statistica* com certa facilidade, mas esse não oferecia respostas quanto ao sucesso ou insucesso da operação. Pode-se destacar também problemas na execução devido a mudanças na configuração do *software*.

Devido a esses problemas, surgiu a idéia de criar um novo sistema independente do *Statistica*, onde os cálculos seriam executados dentro do próprio sistema, o que acabaria com os problemas citados anteriormente.

3.1.2. A segunda versão - o Edust

O Edust, como foi chamada a nova versão do sistema, foi desenvolvido pelo aluno da graduação, Kirliam Maciel Dias. Ele herdou da primeira versão do SEstat apenas a semântica educacional e o seu conhecimento.

Segundo DIAS (2000), O Edust é composto de três unidades principais: a Interface, o Gerente e a unidade Funcional. Cada uma dessas unidades é formada por um ou mais módulos, sendo que cada módulo é um conjunto de classes³ que operam em conjunto para executar uma determinada tarefa. A arquitetura do Edust é representada na Fig. 3.1

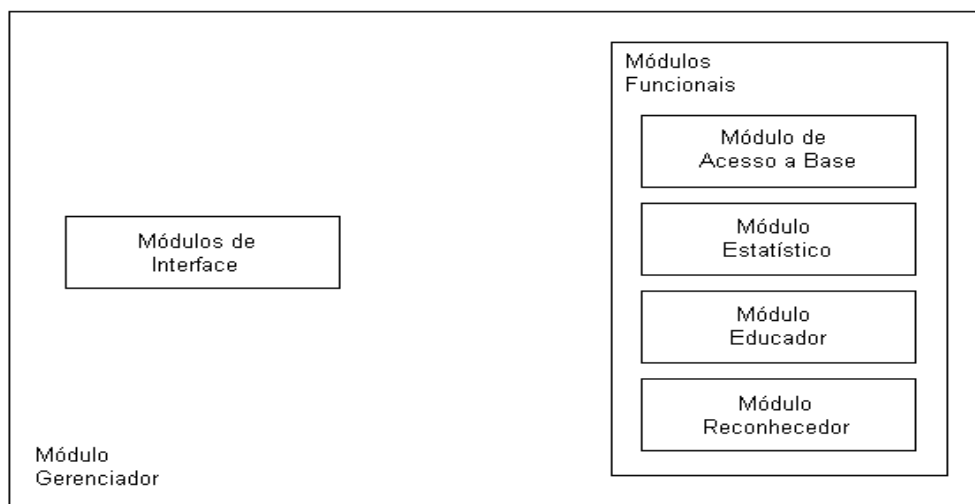


Figura 3.1 – Arquitetura do Edust, adaptado de (DIAS, 2000)

³ Uma classe é uma descrição de um conjunto de objetos que compartilham os mesmos atributos, operações, relacionamentos e semântica. Uma classe implementa uma ou mais interfaces (BOOCH et al, 2000).

A unidade Funcional, assim chamada porque é a que exerce todas as funções do sistema, é composta de diversos módulos independentes. São módulos independentes porque funcionam como componentes, isto é, são classes reutilizáveis que podem, eventualmente, serem utilizadas por qualquer outro sistema que venha a ser desenvolvido. Há quatro módulos nessa unidade: o módulo de Acesso à Base, o Estatístico, o Professor e o Reconhecedor. Com exceção do último, todos os outros três foram todos frutos do presente trabalho (DIAS, 2000).

Já as unidades Interface e Gerente são específicas para o sistema desenvolvido, não podendo dessa forma ser reutilizadas por outra aplicação. A Interface é responsável pela interação com o aluno. Os módulos de Interface possuem pouquíssimo código, a fim de que futuras modificações na Interface não causem grandes impactos no sistema. Já o Gerente é responsável por receber as informações do aluno através da Interface; consultar os módulos da unidade Funcional e transmitir as respostas recebidas desses módulos à Interface (DIAS, 2000).

Com o uso do SEstat e depois do Edust, pode-se perceber que os alunos usavam o sistema criando seus próprios caminhos, ou seja, não havendo a necessidade de estarem reunidos num mesmo local para a aprendizagem. A partir de então surgiu a idéia de criar um novo sistema com os mesmos princípios pedagógicos, mas abordando conceitos de educação a distância, para que pudesse contribuir nas experiências de educação a distância da UFSC.

3.1.3. A terceira versão - o SEstat.Net

O SEstat.Net vem sendo implementado com o apoio dos alunos da graduação, Carlos Alberto Nakazawa e Márcio Juliandrei Marafon (NAKAZAWA & MARAFON, 2003). Seguindo a mesma concepção pedagógica, o objetivo do sistema é oferecer páginas Web como interface com o aluno que se tornem dinâmicas dependendo da interação escolhida, como também os conceitos estatísticos na forma de páginas HTML e os resultados numéricos e gráficos aplicados aos dados de cada aluno conforme o

objetivo da análise. O SEtat.Net trabalhará simultaneamente com diferentes alunos, que possuem diferentes bases de dados e realizam diferentes análises estatísticas.

Para que o SEtat.Net pudesse ser usado num ambiente de educação a distância suportada por computador, o aluno da pós-graduação, Renato Bica Noal, desenvolveu, em sua pesquisa, um ambiente colaborativo para o ensino de estatística com o SEtat, utilizando para tanto, sistemas distribuídos.

De acordo com NOAL (2002), a forma de comunicação do ambiente SEtat com o cliente é através de um arquivo criado sempre que o aluno termina uma interação de aprendizagem. Nesse arquivo é incluído todo o caminho da aprendizagem e todas as soluções certas são enviadas para o servidor, donde o professor pode analisar o andamento dos estudos de cada aluno, tornando possível determinar em qual parte da disciplina os alunos encontram mais dificuldades.

3.2. Considerações Finais

O SEtat como uma ferramenta de apoio ao ensino de estatística contribuiu de uma forma bastante significativa na redução dos problemas levantados para o processo de ensino-aprendizagem de estatística. O aluno passou a ter um maior interesse na aprendizagem e passou também a ser o construtor de seu próprio conhecimento.

Com o intuito de aumentar a base de conhecimento do SEtat, ou seja, incluir novos métodos de análise, surgiram dois novos projetos que foram desenvolvidos: a presente pesquisa e a do aluno da pós-graduação, Marcus Vinicius Pavan (PAVAN, 2003), abordando a análise estatística multivariada com os métodos de análise de correspondência múltipla e análise fatorial por componentes principais, respectivamente. Os dois módulos serão integrados ao SEtat.Net e poderão, também, serem usados no ensino de estatística, para alunos dos cursos de graduação e pós-graduação.

4. ANÁLISE ESTATÍSTICA DE DADOS

Na busca de soluções para um problema usando a análise estatística de dados, o estatístico deve conhecer as especificidades do problema, os objetivos (gerais e específicos) para que possam auxiliá-lo na identificação das variáveis e na sua classificação (quantitativa ou qualitativa), além de extrair o processo de medição dessas variáveis. Com esses conhecimentos, pode-se, então, aplicar a técnica estatística apropriada, ou seja, escolher a estratégia de análise estatística.

As técnicas estatísticas de análise estão agrupadas em duas grandes áreas: a estatística não-paramétrica e a estatística paramétrica. De forma geral, a partir do tipo de variável, são selecionadas as técnicas para realizar a análise estatística. Se as variáveis forem qualitativas, devem ser buscadas as não-paramétricas; se forem quantitativas, buscam-se técnicas paramétricas (NASSAR, 1995).

Sobre a teoria estatística, NASSAR (1995) diz que dispõe de um grande número de técnicas para o planejamento de experimentos e, principalmente, para análise de dados. Este fato, provavelmente leva a pensar de forma limitada que Estatística é justamente a fase de análise de dados.

O módulo desenvolvido nesta pesquisa é capaz de realizar a análise de dados referentes a um conjunto de variáveis qualitativas, utilizando o método Análise de Correspondência Múltipla (ACM). Este método é uma das técnicas da análise estatística multivariada, pois analisa duas ou mais variáveis simultaneamente.

Serão apresentados a seguir, alguns conceitos da Análise Estatística Básica que serão úteis para entender Análise de Correspondência Múltipla. Será colocada, também, uma breve explanação sobre a Análise Estatística Multivariada e suas principais técnicas, além de um estudo mais aprofundado sobre a técnica de Análise de Correspondência Múltipla, por ser de fundamental importância para o desenvolvimento do módulo proposto.

4.1. Tabela de Frequência Observada ou Tabela de Contingência

A tabela de frequência observada pode ser obtida através da organização dos dados de acordo com as ocorrências dos diferentes resultados observados. Para construir uma distribuição de frequências com dados de uma variável qualitativa, basta contar a quantidade de resultados observados em cada categoria BARBETTA (1999). A tabela 4.1 mostra um exemplo de tabela de frequência observada.

Tabela 4.1 – Tabela de frequência observada do grau de instrução do chefe da casa, segundo a localidade da residência. Amostra de 120 famílias do Bairro Saco Grande II, Florianópolis, SC, 1988, adaptado de (BARBETTA, 1999).

Grau de Instrução	Localidade			Total
	Monte Verde	Pq. Da Figueira	Encosta do Morro	
Nenhum	6	14	18	38
Primeiro grau	11	14	13	38
Segundo grau	23	15	6	44
Total	40	43	37	120

4.2. Tabela de Frequência Esperada

De acordo com BARBETTA (1999), para se obter a tabela de frequência esperada, o cálculo pode ser simplificado com a aplicação da fórmula a seguir, aplicada a cada casela da tabela de frequência observada. A tabela 4.2 ilustra um exemplo de tabela de frequência esperada.

$$E = \frac{(\text{total da linha}) \times (\text{total da coluna})}{(\text{total geral})}$$

Tabela 4.2 – Tabela de frequência esperada do grau de instrução do chefe da casa, segundo a localidade da residência. Amostra de 120 famílias do Bairro Saco Grande II, Florianópolis, SC, 1988, adaptado de (BARBETTA, 1999).

Grau de Instrução	Localidade			Total
	Monte Verde	Pq. Da Figueira	Encosta do Morro	
Nenhum	12,67	13,62	11,72	38,00
Primeiro grau	12,67	13,62	11,72	38,00
Segundo grau	14,67	15,77	13,57	44,00
Total	40,00	43,00	37,00	120,00

4.3. Contribuição do Qui-quadrado

O teste de associação qui-quadrado é um dos testes estatísticos mais antigos e um dos mais usados em pesquisa social. É um método que permite testar a significância da associação entre duas variáveis qualitativas. Pode ser usado para comparar duas ou mais amostras, quando os resultados da variável resposta estão dispostos em categorias (BARBETTA, 1999).

Ainda segundo BARBETTA (1999), a estatística do teste, que pode ser designado por χ^2 , é uma espécie de medida entre as frequências observadas, O, e as frequências que se espera encontrar em cada casela, E, na suposição das variáveis serem independentes.

A estatística do teste qui-quadrado, χ^2 , é definida por:

Contribuição em cada casela: $\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

Para calcular o qui-quadrado total: $\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

Um exemplo da tabela de contribuições do qui-quadrado pode ser observado na tabela 4.3.

Tabela 4.3 – Tabela das contribuições do qui-quadrado do grau de instrução do chefe da casa, segundo a localidade da residência. Amostra de 120 famílias do Bairro Saco Grande II, Florianópolis, SC, 1988, adaptado de (BARBETTA, 1999)

Grau de Instrução	Localidade			Total
	Monte Verde	Pq. Da Figueira	Encosta do Morro	
Nenhum	3,509	0,011	3,369	6,889
Primeiro grau	0,219	0,011	0,140	0,371
Segundo grau	4,735	0,037	4,220	8,992
Total	8,463	0,059	7,730	16,252

4.4. Análise Estatística Multivariada

A análise multivariada de dados refere-se a todos os métodos estatísticos que simultaneamente analisam múltiplas medidas de cada indivíduo ou objeto sob investigação. Qualquer análise simultânea de mais de duas variáveis pode ser considerada análise multivariada. Para serem consideradas verdadeiramente multivariadas, todas as variáveis devem ser aleatórias e inter-relacionadas de tal forma que seus diferentes efeitos não possam ser significativamente interpretáveis separadamente. O propósito da análise multivariada é medir, explicar e prever o grau de relacionamento entre variáveis (CARVALHO, 2001).

Segundo JOHNSON (1998), os objetivos das investigações científicas a que os métodos multivariados se aplicam são:

1. Redução de dados ou simplificação estrutural. O fenômeno que está sendo estudado é representado tão simples quanto possível sem sacrificar a importância da informação. Espera-se que este faça a interpretação mais fácil.

2. Classificação e agrupamento. Grupos de objetos “similares” ou variáveis são criados, baseados em características mensuráveis. Alternativamente, regras para classificar objetos em grupos bem-definidos podem ser obtidas.

3. Investigação da dependência entre variáveis. A natureza dos relacionamentos entre variáveis é do interesse. São todas as variáveis mutuamente independentes ou são uma ou mais variáveis dependentes umas das outras? Se forem, como?

4. Predição. Os relacionamentos entre variáveis devem ser determinados com a finalidade de prever os valores de uma ou mais variáveis na base de observações nas outras variáveis.

5. Construção e teste de hipótese. As hipóteses estatísticas específicas, formuladas nos termos dos parâmetros de populações multivariadas, são testadas. Isto pode ser feito para validar suposições ou para reforçar convicções prévias.

O método proposto por esta pesquisa, Análise de Correspondência Múltipla (ACM), tem o objetivo de investigar a associação entre variáveis qualitativas, ou seja, identificar padrões de associação em um conjunto de variáveis categóricas.

4.4.1. Representação dos dados

A estrutura dos dados num problema multivariado sempre poderá ser representada na forma matricial, onde nas linhas são representados os indivíduos sobre os quais foram coletados os dados e nas colunas, as variáveis medidas sobre esses indivíduos. A célula mais simples de representação multivariada acontece quando se mede p variáveis em n indivíduos, conservando a identidade do indivíduo, de tal forma a representá-lo na correspondência da medida das p variáveis (CRIVISQUI, 1999).

Uma representação da matriz de dados é ilustrada na Fig. 4.1

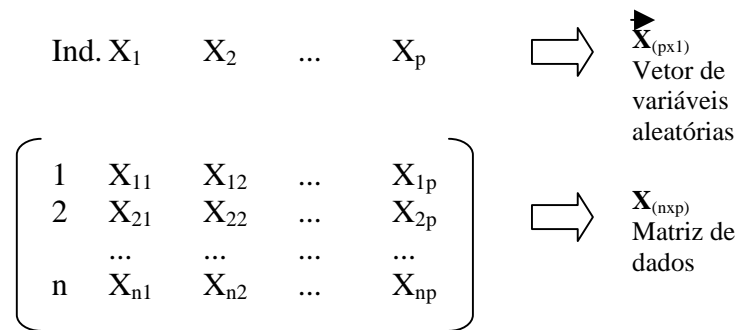


Figura 4.1 – Representação matricial dos dados

A complexidade de representação dos problemas multivariados passa por entender o que acontece com indivíduos e com as variáveis. É a entender as sobreestruturas que esses acontecimentos impõem sobre as linhas (indivíduos) ou sobre as colunas (variáveis) da matriz. Sobreestrutura é uma estrutura que vai além do caso uma população/uma amostra. O conjunto dessas p variáveis não é visto como função homogênea de mesma importância no problema. As p variáveis se dividem em dois ou mais grupos com funções diferenciadas. Essas funções podem ser pelo caráter exploratório ou pelo caráter de resposta, ou ainda por agrupamento natural. A maneira pela qual as técnicas multivariadas são abordadas tem, portanto, a ver com o tipo de representação (CRIVISQUI, 1999).

4.4.2. Técnicas de análise multivariada

A análise multivariada apresenta um conjunto de técnicas para análise dos dados. Seguem abaixo as principais delas:

Componentes Principais: essa técnica permite a visualização dos n indivíduos observados levando em conta medidas simultâneas de p variáveis em subespaços que serão propriamente determinados segundo critério a ser definido, de tal forma que se possa ver o padrão de dispersão desses indivíduos nas suas direções mais favoráveis. Na determinação dessas novas direções, necessárias na presença de três ou mais variáveis, a dispersão dos pontos vai aparecer sobre variáveis que não são observáveis, as chamadas componentes principais, que poderão ou não ter interpretabilidade e permitirão conhecer o padrão de dispersão dos indivíduos revelado na composição das componentes principais, portanto através dos coeficientes de suas combinações lineares, a estrutura de covariância ou a estrutura de correlação do conjunto de dados estudados (CRIVISQUI, 1999).

Análise Fatorial: seu interesse é determinar o menor número de variáveis, construtos ou variáveis latentes que em geral tem uma justificativa no campo substantivo da técnica, ou seja, são conceitos não diretamente mensuráveis, mas já parte do campo substantivo tais como inteligência, atitude, posição política, por exemplo. A análise fatorial cria novas variáveis que, mimetizando o sistema, reproduzem as inter-relações entre as variáveis originais. Esta é uma das poucas técnicas que passa pela construção de modelos, ou seja, a explicação do vetor de variáveis originais através de um vetor de variáveis que não são observáveis. (CRIVISQUI, 1999)

Análise de Correlação Canônica: pressupõe um agrupamento, uma sobreestruturação, sobre as colunas da matriz de dados, ou seja, as operações são realizadas levando-se em conta que as variáveis tem um agrupamento natural em dois conjuntos de variáveis que tem o mesmo tipo de função no problema. O que se busca é entender as inter-relações entre esses dois conjuntos. (CRIVISQUI, 1999).

Análise de Discriminante: é uma técnica que trata de obter uma combinação linear de duas ou mais variáveis independentes que melhor irá discriminar grupos

definidos *a priori*. A variável dependente é do tipo categórica ou nominal e as variáveis independentes são métricas. A técnica é baseada na maximização da variância entre grupos, com relação à variância dentro dos grupos. (ALMEIDA, 2002)

Análise de Agrupamentos: é uma técnica para agrupar observações, indivíduos ou objetos em grupos, de modo que as observações de um mesmo grupo pareçam-se mais entre si do que com as observações de outros grupos. A análise de agrupamento recebe diversos nomes em outras áreas de conhecimento como: Análise do tipo Q, Tipologia, Análise de Classificação ou Taxonomia Numérica. É uma técnica exploratória, descritiva, uma vez que não determina testes de significância. As variáveis envolvidas na análise podem ser métricas ou categóricas ou combinações delas (HAIR, 1995).

Uma outra técnica da análise estatística multivariada é a análise de correspondência múltipla que, por ser a técnica utilizada para o desenvolvimento desta pesquisa, será descrita a seguir com maiores detalhes.

4.5. Análise de Correspondência Múltipla

Desenvolvida por franceses, a análise de correspondência é uma técnica recente da análise multivariada, popular na França e no Japão. De acordo com SAS (1989), na França, a análise de correspondência foi desenvolvida sob a forte influência de Jean-Paul Benzécri e, no Japão, sob Chikio Hayashi.

Para PEREIRA (1999), a análise de correspondência é usada “... para examinar relações geométricas do cruzamento, ou contingenciamento, de variáveis categóricas”. Ela busca identificar padrões de associação em um conjunto de variáveis qualitativas.

Segundo BARBETTA (1999), uma variável é considerada qualitativa “... quando os possíveis resultados são atributos ou qualidades”. As variáveis qualitativas podem ser classificadas em categórica nominal e em categórica ordinal, onde, na nominal, cada categoria é independente, sem relação com as outras; e na ordinal, cada categoria mantém uma relação de ordem com as outras que pode ser ou não regular.

Uma outra definição de análise de correspondência pode ser encontrada em STATSOFT (2002) que diz ser uma técnica descritiva ou exploratória, construída para analisar tabelas de dupla entrada ou de múltiplas entradas, contendo alguma medida de correspondência entre linhas e colunas. Os resultados são semelhantes aos da Análise Fatorial e permitem a exploração de variáveis categóricas incluídas na análise.

A análise de correspondência baseia-se em informações contidas em tabelas de contingência. Assim, segundo JOHNSON (1998), se a tabela de contingência tiver I linhas e J colunas, o assinalamento produzido pela análise de correspondência contém dois grupos de pontos: um grupo de I pontos correspondendo às linhas e um grupo de J pontos correspondendo às colunas. As posições dos pontos refletem associações.

4.5.1. Tabela de Burt

Conforme STATSOFT (2002), a tabela de Burt tem uma estrutura claramente definida e é uma matriz simétrica. No exemplo de 3 variáveis categóricas, a matriz dos dados consiste em $3 \times 3 = 9$ divisórias, criadas por cada variável tabulada em relação às categorias de todas as outras variáveis. Um exemplo da tabela é ilustrado na tabela 4.4.

Tabela 4.4 – Tabela de Burt de Frequência Observada, do grau de instrução do chefe da casa, localidade da residência e uso de programas de alimentação popular (PAP). Amostra de 120 famílias do Bairro Saco Grande II, Florianópolis, SC, 1988.

		Localidade			PAP		Grau Instrução			Total
		Monte verde	Pq. Da Figueira	Encosta do Morro	Sim	Não	Nenhum	Primeiro grau	Segundo grau	
Localidade	Monte Verde	40	0	0	18	22	6	11	23	129
	Pq. Da Figueira	0	43	0	12	31	14	14	15	129
	Encosta do morro	0	0	37	12	25	18	13	6	111
PAP	Sim	18	12	12	42	0	7	16	19	126
	Não	22	31	25	0	78	31	22	25	234
Grau Instrução	Nenhum	6	14	18	7	31	38	0	0	114
	primeiro grau	11	14	13	16	22	0	38	0	114
	Segundo grau	23	15	6	19	25	0	0	44	132
Total		120	129	111	126	234	114	114	132	1080

4.5.2. Autovalores e autovetores

Os autovalores de uma matriz real ou complexa de ordem $n \times n$ são números reais ou complexos não nulos. Segue abaixo a determinação de autovalores e autovetores conforme STEINBRUCH & WINTERLE (1987).

Por exemplo, dada a matriz A de 3 linhas por 3 colunas:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Se v e λ são, respectivamente, autovetor e o correspondente autovalor, tem-se:

$$A \cdot v = \lambda v \quad (v \text{ é matriz-coluna } 3 \times 1) \quad \text{ou} \quad Av - \lambda v = 0$$

Tendo em vista que $v = Iv$ (I é a matriz-identidade), pode-se escrever:

$$Av - \lambda Iv = 0 \quad \text{ou} \quad (A - \lambda I)v = 0$$

Para que esse sistema homogêneo admita soluções não-nulas, isto é:

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

deve-se ter:

$$\det(A - \lambda I) = 0$$

ou:

$$\det \left(\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \right) = 0$$

ou, ainda:

$$\det \begin{pmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{pmatrix} = 0$$

A equação $\det(A - \lambda I) = 0$ é denominada equação característica da matriz A , e suas raízes são os autovalores da matriz A . O determinante $\det(A - \lambda I)$ é um polinômio em λ denominado polinômio característico.

A substituição de λ pelos seus valores no sistema homogêneo de equações lineares permite determinar os vetores próprios associados.

4.5.3. Número máximo de dimensões

De acordo com JOHNSON (1998), o número máximo de dimensões ou autovalores que pode ser extraído de uma tabela de dupla entrada é igual ao número mínimo de colunas (colunas – 1) ou de (linhas – 1).

4.5.4. Inércia

Segundo STATSOFT (2002), o termo inércia na análise de correspondência é usado por analogia com a definição na matemática aplicada de “momento de inércia”. É definida como o total do qui-quadrado de Pearson dividido pelo total geral de frequências observadas.

A análise de correspondência pode ser vista como um método de decomposição da estatística do qui-quadrado total da tabela ($\text{Inércia} = \chi^2/N$, onde N =Total de frequências observadas), por identificação de um número menor de dimensões, com os quais os desvios entre as frequências observadas e esperadas podem ser representados (STATSOFT, 2002).

A qualidade de um ponto representa a proporção da contribuição do ponto para a inércia total (qui-quadrado) que pode ser explicada pelo número de dimensões escolhido. Contudo, ele não indica se, e em que extensão, o respectivo ponto contribui de fato para a inércia total. A inércia relativa representa a proporção da inércia total explicada pelo respectivo ponto, e é independente do número de dimensões que foi escolhido pelo pesquisador. A inércia relativa para cada dimensão contém a contribuição relativa do respectivo ponto (linha) para a inércia explicada por essa dimensão (STATSOFT, 2002).

4.6. Considerações Finais

Para esta pesquisa, foi utilizado o método ACM para ser implementado e integrado ao Sestat.Net como mais um conhecimento para o sistema.

Através do módulo MAC, o usuário poderá ter acesso a todos os conceitos envolvidos em análise multivariada e análise de correspondência múltipla e interagir com o módulo para que o mesmo possa realizar a análise e mostrar os resultados obtidos.

Dentre esses resultados, estarão disponíveis: as tabelas de frequência observada de cada variável da base de dados informada pelo aluno; as tabelas de Burt de frequência observada, esperada e da contribuição do qui-quadrado; os autovalores e os gráficos gerados.

5. INTELIGÊNCIA ARTIFICIAL

RICH & KNIGHT (1993) definem IA como o estudo de como fazer os computadores realizarem coisas que, no momento, as pessoas fazem melhor.

Para BITTENCOURT (1998), o objetivo central da IA é simultaneamente teórico – a criação de teorias e modelos para a capacidade cognitiva – e prático – a implementação de sistemas computacionais baseados nestes modelos. Nesse sentido, a IA tem uma relação com seu objeto de estudo semelhante à da psicologia, mas com uma importante diferença: os modelos e teorias da IA são implementadas em um computador, o que os torna de certa forma autônomos.

Neste capítulo, serão abordados o uso da inteligência artificial na educação e um estudo mais amplo sobre sistemas especialistas, por ser o componente da IA usado para o desenvolvimento do módulo MAC.

5.1. Inteligência Artificial na Educação

CURILEM (1998) pontua vários motivos para a aplicação da Inteligência Artificial nas áreas de educação e treinamento. Entre eles:

- busca de ferramentas mais poderosas para a construção de sistemas educacionais;
- oportunidade de desenvolver e testar novas técnicas e novos modelos, aplicando ferramentas de ensino inovadoras que cumprem funções de explicação, diagnóstico, entre outras, esperadas em um sistema de instrução;
- capacidade da Inteligência Artificial de modelar conhecimento, o sistema pode resolver problemas que o próprio aluno tem que resolver (o sistema consegue “entrar” no problema junto com o aluno, discutindo os passos intermediários) e baseia-se no conhecimento do domínio a ser ensinado;

- possibilidade de produção de interações bidirecionais entre o usuário e o sistema, que são especialmente importantes na obtenção de habilidades para a resolução de problemas complexos.

5.2. Sistemas Especialistas

Os sistemas especialistas solucionam problemas que normalmente são solucionados por “especialistas” humanos. Para solucionar tais problemas, os SE precisam acessar uma substancial base de conhecimento do domínio da aplicação, que precisa ser criada do modo mais eficiente possível. Eles também precisam explorar um ou mais mecanismos de raciocínio, para aplicar seu conhecimento aos problemas que têm diante de si. Depois eles precisam de um mecanismo para explicar o que fizeram aos usuários que deles dependem (RICH & KNIGHT, 1993).

5.2.1. Características dos sistemas especialistas

Conforme STAIR (1998), os sistemas especialistas possuem algumas características e recursos. Entre eles:

- *Explicar seu raciocínio ou as decisões sugeridas:* a capacidade de explicar os processos de raciocínio pode ser o recurso mais valioso dos sistemas especialistas computadorizados. O usuário do SE, dessa forma, tem acesso ao raciocínio por trás da conclusão.
- *Apresentar comportamento inteligente:* considerando um conjunto de dados, um SE pode propor novas idéias ou abordagens para a solução de problemas.
- *Manipular informações simbólicas e tirar conclusões:* as pessoas estão acostumadas a trabalhar com imagens visuais; algum dia os sistemas especialistas serão capazes de imitar esse comportamento.

- *Tirar conclusões de relacionamentos complexos:* os SE podem avaliar dados relacionados de modos altamente complexos para tirar conclusões e encontrar soluções para problemas.
- *Proporcionar conhecimento portátil:* podem ser usados para absorver o conhecimento humano que, de outra forma, poderia ser perdido.
- *Poder lidar com a incerteza:* capacidade de lidar com conhecimento incompleto ou não inteiramente preciso. Esses sistemas lidam com esses problemas por meio da satisfação e da heurística.

Ainda segundo STAIR (1998), algumas dessas características limitam a utilidade dos SE. Muitas dessas características limitadoras têm a ver com custo, controle e complexidade. Eis algumas dessas características: não amplamente utilizados ou testados; difíceis de serem usados; limitados a problemas relativamente simples; não podem lidar prontamente com conhecimento “mesclado”; possibilidade de erro quando o especialista passa o conhecimento incorretamente ou incompleto, ou ainda erros que envolvem a programação; não podem refinar o próprio banco de conhecimentos; dificuldade de manutenção; podem ter elevados custos de desenvolvimento e ainda levantam questões legais e éticas pelo fato de não se saber quem é o responsável quando os SE são utilizados para tomar decisões ou ajudar no processo decisório.

5.2.2. Vantagens dos sistemas especialistas

Os sistemas especialistas são dotados de inteligência e conhecimento e, por isso, dispõem de vantagens diferentes dos sistemas tradicionais. MENDES (2000) pontua algumas dessas vantagens:

- Um sistema especialista é capaz de estender as facilidades de tomada de decisão para muitas pessoas. O conhecimento dos especialistas pode ser distribuído, de forma que possa ser utilizado por um grande número de pessoas;

- Um sistema especialista pode melhorar a produtividade e desempenho de seus usuários, considerando que o provê de um vasto conhecimento, que, certamente, em condições normais, demandaria mais tempo para assimilá-lo e, conseqüentemente, utilizá-lo em suas tomadas de decisão;
- Sistemas especialistas reduzem o grau de dependência que as organizações mantêm quando se vêem em situações críticas, inevitáveis, como, por exemplo, a falta de um especialista;
- Sistemas especialistas são ferramentas adequadas para serem utilizadas em treinamento de grupos de pessoas de forma rápida e agradável, podendo servir, após o treinamento, como instrumento para coleta de informações, obtendo subsídios para reformulação das lições para a obtenção de melhor desempenho, além de prestar suporte imediato para os treinandos durante a utilização dos conhecimentos na realização de suas tarefas diárias.

5.2.3. Problemas enfrentados pelos sistemas especialistas

Segundo RICH & KNIGHT (1993), os sistemas especialistas enfrentam alguns problemas que são:

- *Fragilidade*: como os sistemas especialistas só têm acesso a conhecimentos altamente específicos do domínio, eles não podem contar com conhecimentos mais genéricos quando a necessidade surge.
- *Falta de Metaconhecimento*: os sistemas especialistas não têm conhecimentos muito sofisticados sobre sua própria operação. Eles normalmente não conseguem raciocinar sobre seu próprio escopo e restrições, dificultando ainda mais a tarefa de lidar com a sua fragilidade.
- *Aquisição de Conhecimento*: a aquisição é um dos maiores obstáculos à aplicação da tecnologia dos sistemas especialistas a novos domínios.

- *Validação*: medir o desempenho de um sistema especialista é difícil, porque não se sabe como quantificar o uso do conhecimento.

Sobre a aquisição de conhecimento, BITTENCOURT (1998) também salienta ser a parte mais sensível no desenvolvimento de um SE. Esta não pode limitar-se à adição de novos elementos de conhecimento à base de conhecimentos; é necessário integrar o novo conhecimento ao conhecimento já disponível, através da definição de relações entre os elementos que constituem o novo conhecimento e os elementos já armazenados na base.

5.2.4. Componentes de um SE

A arquitetura de um sistema especialista é uma estrutura que representa o conjunto dos componentes do sistema e a forma como eles estão integrados e relacionados.

Para STAIR (1998), este conjunto consiste de um banco de conhecimentos, um motor de inferência, um recurso de explicação, um recurso de aquisição de conhecimento e uma interface com o usuário. Uma arquitetura típica de sistema especialista é mostrada na Fig. 5.1.

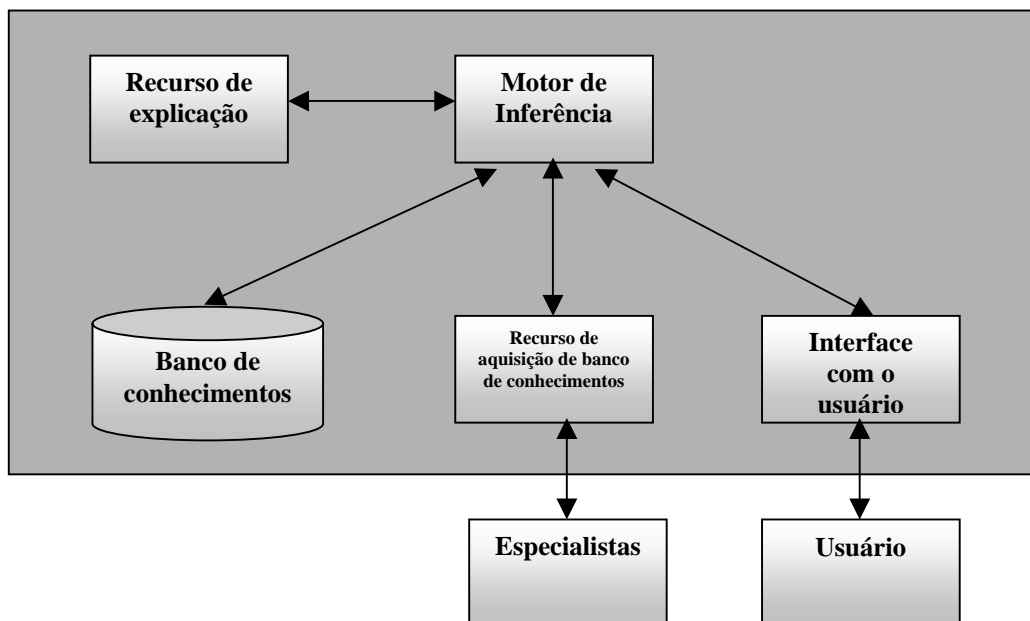


Figura 5.1 – Componentes de um Sistema Especialista, adaptado de (STAIR, 1998).

Banco de conhecimentos

O banco de conhecimentos é o local onde estão armazenadas todas as informações, dados, regras, casos e relacionamentos, ou seja, a soma do conhecimento e da experiência do especialista que passarão a ser utilizados pelo sistema. Uma base completa só tende a melhorar o desempenho do sistema.

Motor de inferência

O motor de inferência é o componente do sistema especialista responsável pela busca das informações e relacionamentos armazenados no banco de conhecimentos, ou seja, ele precisa encontrar os fatos, interpretações e regras e construir um sentido lógico das informações.

Recurso de explicação

Este recurso possibilita ao usuário entender como o sistema chegou a um determinado resultado, pois se o mesmo não tiver convencido da precisão do processo de raciocínio que o resultado produziu, ele não aceitará.

Recurso de aquisição de conhecimento

O objetivo geral desse recurso é fornecer um meio conveniente e eficiente para capturar e armazenar todos os componentes da base de conhecimentos. (STAIR, 1998)

Interface com o usuário

A interface com o usuário é o componente que faz a comunicação entre o usuário e o sistema se tornar mais fácil e assim como em qualquer outro *software*, é fundamental para o sucesso do mesmo.

Para a implementação de uma interface, deve-se levar em conta os critérios ergonômicos e a facilidade no uso, além do grau de familiarização do usuário com o domínio de trabalho do sistema.

5.2.5. Ciclo de Vida para o Desenvolvimento de Sistemas Especialistas

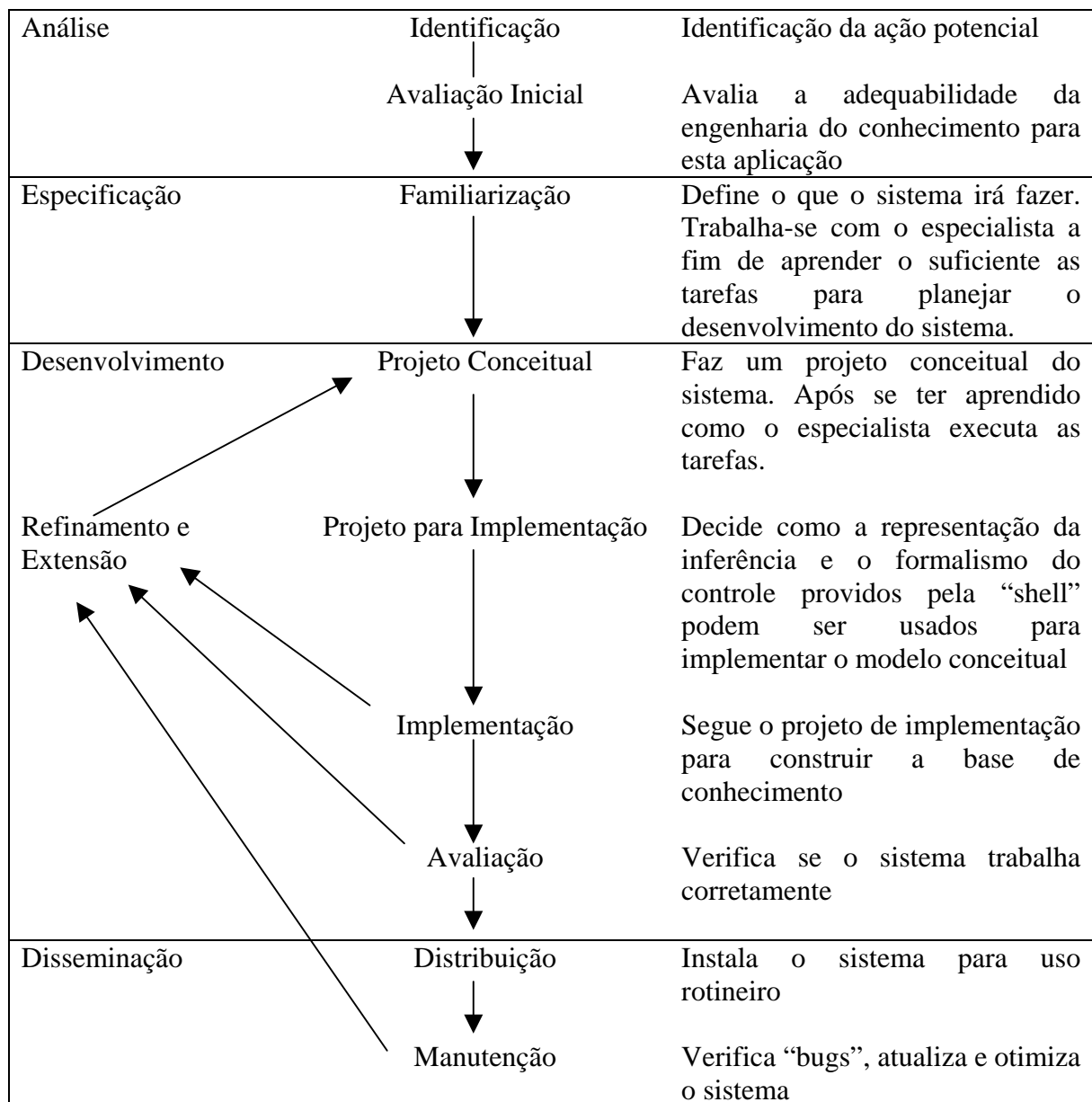
De acordo com SCOTT et al (1991), o desenvolvimento de um sistema especialista pode ser dividido em quatro fases: análise, especificação, desenvolvimento e disseminação.

Na fase de análise, todas as possibilidades de desenvolvimento de um sistema especialista são analisadas pelos interessados. As bases para o desenvolvimento do sistema são estabelecidas durante a fase de especificação. O projeto e a implementação do sistema ocorrem na fase de desenvolvimento. Na fase de disseminação, o sistema estará disponível para uso, ocorrendo em seguida a verificação e correção de problemas e a otimização do sistema.

O ciclo pode ser interrompido inúmeras vezes e em qualquer época, a partir da fase de desenvolvimento. Esta interrupção é feita para retornar ao projeto conceitual. Este processo é chamado de refinamento e extensão.

O Quadro II a seguir representa o ciclo de vida para o desenvolvimento de um sistema especialista proposto por Scott. Cada fase e sub-fase do ciclo são bem detalhadas.

QUADRO II – Ciclo de vida de um Sistema Especialista, adaptado de (SCOTT et al, 1991)



- **Identificação:** verifica-se a aplicação. As regras a seguir identificam bons candidatos a um SE:
 - a) Tarefas em que o desempenho aumenta consideravelmente com a experiência;
 - b) Existem poucos especialistas para a tarefa

- c) Um importante especialista está se retirando de atividade
- d) Tarefas que requerem freqüentes referências a material com pouca disponibilidade;
- **Avaliação inicial:** verifica-se a viabilidade da construção de uma aplicação baseado em conhecimento. Esta fase deve considerar três aspectos: técnico, econômico e prático;
 - **Familiarização:** é necessário definir o que o sistema proposto fará. Além da definição, será observada a forma como o especialista executa suas tarefas e a divisão da funcionalidade do sistema, em passos, para o desenvolvimento incremental;
 - **Projeto conceitual:** após estar clara a definição do sistema, deve-se ter um modelo conceitual, especificando a seqüência de passos que o sistema realiza ao realizar uma tarefa, a busca pelas regras e as informações que serão utilizadas;
 - **Projeto para implementação:** será especificado como o sistema irá realizar as tarefas definidas no modelo conceitual, baseando-se em uma ferramenta computacional que utiliza técnicas de Inteligência Artificial para geração automática de sistemas especialistas;
 - **Implementação:** dá-se a construção do sistema, de acordo com o projeto para implementação. Nesta fase pode haver inconstâncias, solucionadas com a integração do sistema com o seu ambiente de trabalho;
 - **Avaliação:** verifica-se se todas as tarefas estão sendo executadas adequadamente pelo sistema. É nesta fase que será definido se o sistema pode ser distribuído ou necessita de um refinamento;
 - **Distribuição:** ocorre a integração do sistema com outros programas, redes, equipamentos e treinamentos de usuários;
 - **Manutenção:** é necessária, pois inclui outras versões atualizadas do sistema, resolve problemas com os usuários e outros problemas deste nível.

5.3. Considerações Finais

O uso de paradigmas de Inteligência Artificial para o desenvolvimento de sistemas na área da educação pode garantir a interatividade, ou seja, o diálogo entre o sistema e o aluno, permitindo assim a transmissão do conhecimento.

Para o desenvolvimento do módulo proposto nesta pesquisa foi utilizado o componente da IA que é o sistema especialista. Como visto anteriormente, os sistemas especialistas possuem características, vantagens e problemas específicos.

Para a construção de um sistema especialista é necessária a contribuição de um ou vários especialistas humano no domínio e, para o desenvolvimento desta pesquisa, participaram os professores de estatística do INE.

Como características do módulo MAC, podem ser destacadas: ele apresenta comportamento inteligente, pois a partir de uma base de dados fornecida pelo aluno, ele escolhe o melhor método estatístico para ser aplicado, realiza a análise de dados seguindo um raciocínio estatístico e utilizando-se de algoritmos de cálculo e, por fim, mostra os resultados visualmente na tela do computador, além de disponibilizar para o aluno todos os conceitos estatísticos envolvidos. Uma outra característica importante é a explicação do raciocínio seguido, ou seja, a cada interação, o aluno tem detalhado o caminho percorrido até então.

Dentre as vantagens que o módulo MAC fornece, é um módulo que terá embutido nele o conhecimento de uma equipe de especialistas (professores de estatística) e esse conhecimento poderá ser estendido para várias pessoas, ou seja, ficará disponível para todos que quiserem aprender. Dessa maneira, o aluno fica livre para buscar seu conhecimento no momento em que desejar. Uma outra vantagem é poder fornecer o raciocínio que o módulo seguiu pra chegar no resultado, permitindo assim, que o aluno aprenda o porque da aplicação do procedimento estatístico em questão.

O SEstat.Net é aberto para o aluno modelar seu conhecimento, ou seja, ele pode explorar o ambiente livremente e traçar um percurso conforme seu interesse e desenvolvimento do assunto. Não há pressupostos sobre o modelo do estudante encontrado em alguns *softwares* de ensino.

6. DESENVOLVIMENTO DO MÓDULO MAC

Para o desenvolvimento do módulo, foi utilizado o ciclo de vida para desenvolvimento de sistemas especialistas proposto por SCOTT (1991). O ciclo apresenta 4 fases: análise, especificação, desenvolvimento e disseminação (distribuição e manutenção).

A seguir, serão apresentadas todas as atividades realizadas em cada fase.

6.1. Análise

Nesta fase, faz-se uma avaliação para detectar se a aplicação é adequada para ser um sistema especialista, ou seja, se ela é baseada em conhecimento e regras. Com base nestas considerações, verificou-se que o módulo MAC é adequado e, entre os fatores que levaram a tal conclusão, pode-se citar:

- existem especialistas no domínio para reunir o conhecimento e transmiti-lo, contribuindo desta maneira para com o módulo;
- deve ser dotado de conhecimento (procedimentos estatísticos e regras) para que possa apresentar um comportamento inteligente a cada solicitação do usuário;
- pode adquirir conhecimento com a experiência a cada vez que o usuário interage com ele;
- deve mostrar ao usuário o raciocínio seguido (explicação) para alcançar o resultado;
- necessita de uma interface onde o usuário deve interagir com o módulo.

6.2. Especificação

Na fase de especificação devem ser levantados todos os requisitos do sistema, ou seja, definir o que o sistema fará. Para este levantamento, é necessário que se trabalhe com o(s) especialista(s), a fim de conhecer a forma como ele executa as tarefas, neste caso, a análise estatística multivariada utilizando análise de correspondência múltipla e, também muito importante, a sua experiência com o ensino de estatística que é de fundamental importância para descobrir as necessidades do aluno.

Os especialistas envolvidos nesta atividade foram os professores do INE.

Com base no levantamento realizado, foram definidas as funções do módulo MAC, que são:

- Disponibilizar o método de análise multivariada no SEstat.Net;
- Capturar a base de dados informada pelo aluno, através da qual será feita a análise;
- Permitir que o aluno escolha as variáveis que deseja trabalhar na análise;
- Fornecer ao aluno um mecanismo para que o mesmo possa classificar estas variáveis como quantitativas ou qualitativas;
- Utilizar o SETip para verificar o tipo correto da variável e alertar o aluno caso tenha cometido um erro na classificação;
- Propor ao aluno a categorização de variáveis quantitativas e realizar caso o mesmo solicitar, conforme número de categorias informado por ele;
- Verificar o número de categorias informado pelo aluno no caso de categorização e alertar caso esse número seja maior que o número recomendado pelo módulo;
- Permitir que o aluno escolha a variável suplementar para trabalhar na análise;
- Realizar a análise:

- Gerar e mostrar a tabela de frequência observada para cada variável escolhida pelo aluno;
 - Verificar para cada variável, se existe alguma categoria com uma frequência menor ou igual a 10 e, caso tenha, propor o reagrupamento;
 - Gerar e mostrar a matriz de autovalores;
 - Gerar e mostrar a tabela de Burt de frequência observada;
 - Gerar e mostrar a tabela de Burt de frequência esperada;
 - Gerar e mostrar a tabela de Burt das contribuições do qui-quadrado;
 - Gerar e mostrar os autovalores e a inércia.
 - Gerar e mostrar o gráfico de autovalores;
 - Gerar e mostrar os gráficos bidimensionais;
- A cada interface de interação do aluno com o módulo, disponibilizar ajuda com os conceitos estatísticos relativos ao assunto abordado;
 - Mostrar o raciocínio seguido pelo módulo (explicação) a cada interação com o aluno.

Devido à necessidade de todas estas funções estarem integradas ao SEstat.Net, foi realizado um levantamento para verificar como se daria esta integração, e também como seria a relação entre o módulo MAC e o Módulo de Análise por Componentes Principais (PAVAN, 2003), já que os dois módulos devem compartilhar de interações com o aluno bastante semelhantes.

Com base neste levantamento, decidiu-se por usar as mesmas tecnologias utilizadas no desenvolvimento do SEstat.Net (ver seção 6.3.1), e também aproveitar algumas funcionalidades que o mesmo já possui e são necessárias para realizar a análise de correspondência múltipla, como a geração da tabela de frequência para cada variável, por exemplo.

O SEstat.Net oferece, hoje, as opções do aluno realizar uma análise univariada, bivariada ou inferência. Será incluída, então, mais uma opção de análise multivariada onde, dependendo das variáveis escolhidas pelo aluno, qualitativas ou quantitativas, o sistema realizará o método de análise de correspondência múltipla ou análise fatorial, respectivamente.

A modelagem do módulo será melhor explanada a seguir.

6.3. Desenvolvimento do módulo MAC

A fase de desenvolvimento envolve várias atividades: um projeto conceitual do sistema, onde deve ficar especificada a seqüência de passos que o sistema realiza para alcançar o objetivo; um projeto para implementação da base de conhecimento; a implementação do sistema e a avaliação do mesmo com a aplicação de testes.

6.3.1. Tecnologias utilizadas para o desenvolvimento

Para uma perfeita integração com o SEstat.Net, o módulo MAC foi desenvolvido utilizando as mesmas tecnologias, ou seja, a plataforma Java 2, mais especificamente, J2EE.

A tecnologia J2EE tem seu modelo baseado em componentes, simplificando assim seu desenvolvimento. Ela controla a infra-estrutura e a interoperabilidade das aplicações, suportando serviços *Web* com um desenvolvimento seguro e robusto.

Dentre os vários componentes que formam essa tecnologia, foi utilizado apenas JSP, JavaBeans, JDBC e Java Servlets para o desenvolvimento.

6.3.1.1. Java Server Pages (JSP)

De acordo com NAKAZAWA & MARAFON (2003), dentre as diversas tecnologias existentes que permitem a elaboração de *websites* com conteúdo dinâmico (PHP – *Personal Home Page*, ASP – *Active Server Pages*, Delphi, CGI-*Common Gateway Interface*), optou-se pelo componente JSP, pois, segundo NETO (2003), trata-se de uma tecnologia para desenvolvimento de aplicações *Web* que possui grande portabilidade de plataforma, podendo ser executado em diversos sistemas operacionais. Ela permite o desenvolvimento de aplicações que acessem banco de dados, arquivos-texto, capte informações a partir de formulários, informações sobre o visitante e sobre o servidor, a geração de gráficos dinâmicos, o uso de variáveis e *loops* entre outras coisas.

Além disso, essa tecnologia permite separar a programação lógica (parte dinâmica) da programação visual (parte estática), facilitando o desenvolvimento de aplicações mais robustas (NETO, 2003).

JSP é a combinação de HTML (*Hypertext Markup Language*) com Java dentro de uma mesma página. Usando *tags* especiais, pode-se introduzir Java em qualquer parte da página, todo o código fora dos *tags* é, em princípio, HTML puro. Uma página JSP sempre é convertida em programa Java (um *servlet*) antes de entrar em ação (BATISTA, 2003).

Segundo NETO (2003), no processamento da requisição de uma página JSP, o cliente faz a solicitação de um arquivo JSP, com isso, é enviado um *object request* para a *JSP engine*. A *JSP engine* envia a solicitação de um *JavaBean component* especificado no arquivo. O componente controla a requisição possibilitando a obtenção de um gráfico, geração de dados como resultado de uma análise feita ou acesso a um arquivo em banco de dados, em seguida, passa o objeto *response* de volta para a *JSP engine*. A *JSP engine* e o *WEB server* enviam a página JSP revisada de volta para o cliente, onde o usuário pode visualizar os resultados através do *Web browser*.

Segundo NAKAZAWA & MARAFON (2003), todas as páginas do SStat.Net que trocam informações com os alunos na análise estatística dos dados são JSP e realizam a comunicação direta com os *JavaBeans*, atuando como uma interface para os

resultados estatísticos gerados por esses. Algumas páginas fazem uso de *JavaScript*, que é responsável por encaminhar para uma determinada página dependendo da escolha do usuário. A seqüência de aprendizado que o aluno segue se dá por meio de uma série de páginas, onde cada uma é anexada em uma ou mais páginas, formando assim os diversos caminhos de aprendizagem e que são armazenados no servidor.

6.3.1.2. JavaBeans

De acordo com LIMA (2003), JavaBeans (também conhecidos simplesmente como *Beans*) são componentes de software projetados para serem unidades reutilizáveis, onde uma vez criados podem ser reutilizados sem modificação de código.

Um modelo de componente é definido como um conjunto de classes e interfaces na forma de pacotes Java que deve ser usado em uma forma particular para isolar e encapsular um conjunto de funcionalidades (LIMA, 2003).

De acordo com NAKAZAWA E MARAFON (2003), a forma de acesso utilizada pelas páginas JSP às classes Java do SEstat.Net é feita através de *Beans*. As classes “Administrador” e as classes responsáveis pela geração dos dados dos gráficos são instanciadas como *Beans* e são acessadas de forma direta pelas páginas JSP. As classes onde seus objetos são compartilhados entre todos os usuários, como as classes “Gráfico”, “base de dados”, “leitor DBF”, “estatística”, “Setip”, “educador” são instanciadas como *Beans*, porém são encapsuladas dentro do *Bean* “Administrador” pois não necessitam de acesso direto com as páginas JSP.

Os *Beans* são declarados dentro da página JSP (fazendo uso da *tag* `jsp:useBean`) e cada um deve definir qual é o seu escopo. O escopo define o ciclo de vida de um objeto, que pode ser definido como:

- **page:** Objetos definidos com o escopo *page* são vistos apenas pela página onde ele foi criado. As referências para objetos no contexto *page* são armazenados no objeto *pageContext*.
- **request:** Objetos são vistos nas páginas usadas para responder a requisição do usuário. Se uma página é redirecionada para outra os objetos de escopo *request*

são preservados uma vez que fazem parte da mesma requisição. As referências são armazenadas no objeto implícito *request*.

- **session:** Objetos que são acessíveis por páginas que fazem parte da mesma sessão (Uma sessão é iniciada no logon do usuário ao sistema e finalizada em sua saída) . As referências são armazenadas no objeto *session*.
- **application:** Objetos que são acessíveis por toda aplicação JSP em questão e são armazenadas no objeto implícito *application*.

6.3.1.3. JDBC

A tecnologia JDBC é uma API que possibilita qualquer aplicação Java acessar qualquer base de dados (SUN, 2003b)

No projeto SEstat.Net foi fundamental o uso do componente JDBC, já que há a necessidade da utilização de um banco de dados para armazenar os caminhos das bases dos alunos, como também armazenar as senhas e usuários, a fim de distinguir o acesso ao sistema entre alunos e convidados (NAKAZAWA & MARAFON, 2003).

Também sua utilização foi essencial para realizar a conexão e iterações com os arquivos do formato Dbase (DBF), já que as bases dos alunos serão fornecidas, ou armazenadas no servidor, no formato DBF.

6.3.1.4. Java Servlet

A tecnologia Java Servlet fornece um mecanismo simples de estender a funcionalidade de servidores *Web* para um sistema que proporciona um acesso mais rápido. Um servlet possibilita abranger um maior número de tipos de aplicações *Web*, estendendo e realçando assim as funcionalidades do servidor (SUN, 2003a).

Basicamente, Servlets são módulos de código em Java que rodam em uma aplicação no servidor para responder pedidos do cliente. Eles não são direcionados para

um tipo de protocolo cliente-servidor, embora sejam mais comumente usados com HTTP, e a palavra "Servlet" seja frequentemente encontrada com o significado "HTTP Servlet".

Como são escritos em Java, eles criam extensões sofisticadas em um servidor, independentemente do sistema operacional, e são comumente usados para acessar conteúdos dinâmicos, processar e armazenar dados mandados de formulários e gerenciar o estado de informação sobre uma conexão de estado HTTP.

No SEstat.Net, Servlets possuem uma participação marcante. Como existem diversas páginas JSP funcionando como uma interface para com o aluno, e para cada tipo de análise realizada são processadas diversas páginas no servidor, significando que a cada página acessada, sem o uso de servlets, seria preciso compilá-la novamente. Com o uso dos servlets, apenas no primeiro acesso à página será compilada, fazendo com que os próximos acessos se tornem mais rápidos e eficientes (NAKAZAWA & MARAFON, 2003).

6.3.2. Modelagem do módulo MAC

Na concepção do módulo MAC, foi empregado o paradigma orientado a objeto, devido às vantagens que oferece como: reusabilidade, estabilidade, confiabilidade, integridade, por exemplo. A ferramenta CASE utilizada na modelagem do módulo foi a Together 6.0.

O modelo de classes representado na Fig. 6.1 é o modelo do SEstat.Net. Para a implementação do módulo MAC, não foi necessária a criação de classes novas, todos os procedimentos estatísticos envolvidos na análise de correspondência múltipla foram implementados como métodos da classe Estatística e a comunicação entre as páginas (interfaces) e as outras classes foi realizada na classe Administrador. Para um melhor entendimento do modelo, será feita uma descrição das classes mais importantes, conforme NAKAZAWA & MARAFON (2003).

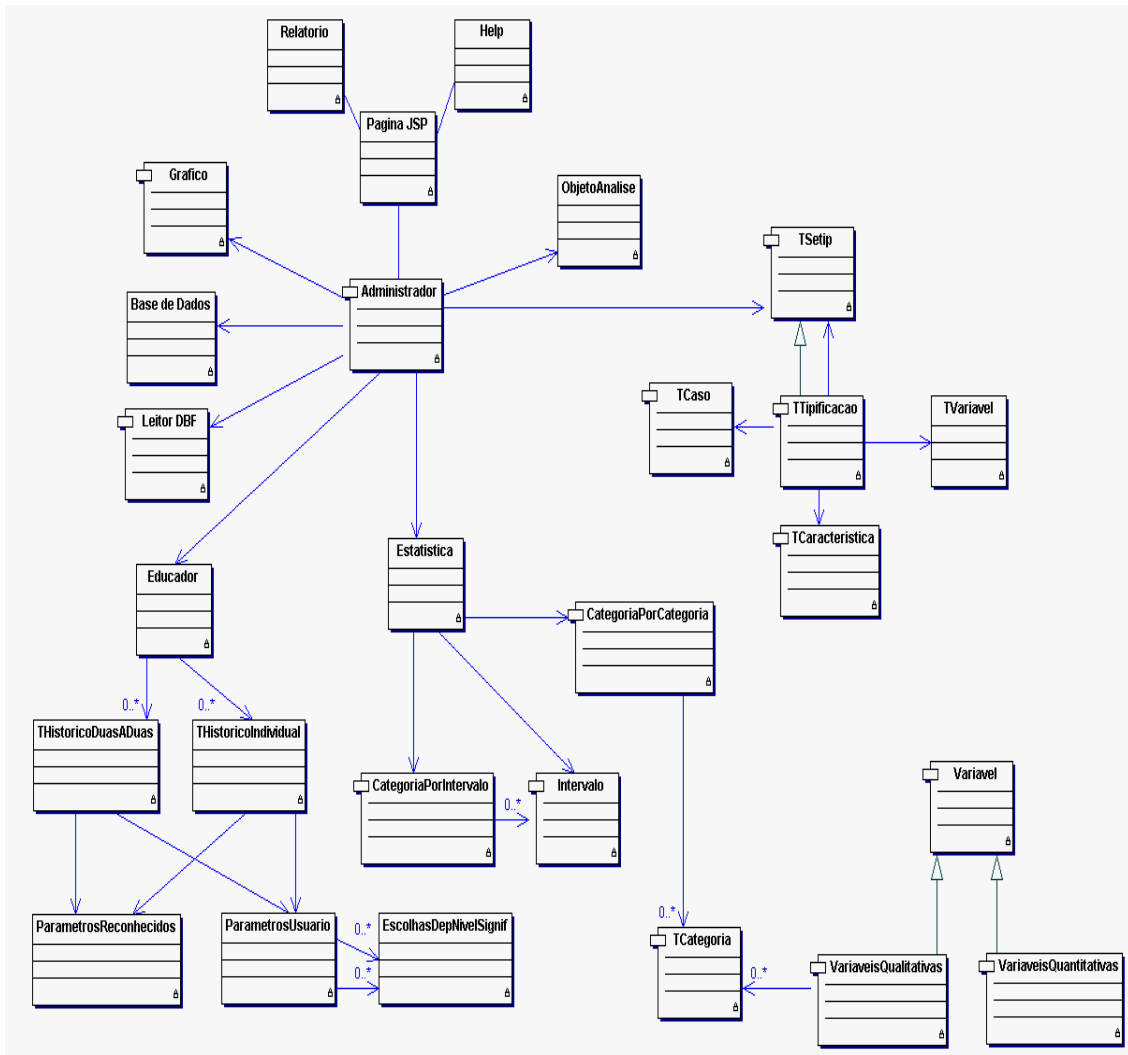


Figura 6.1 – Modelo de Classes do SStat.Net, adaptado de (NAKAZAWA & MARAFON, 2003)

6.3.2.1. Classe administrador

A classe Administrador é a classe responsável pela comunicação entre as páginas e as outras classes do sistema. Cada objeto desta classe é responsável pela gerência dos dados de um aluno específico, e por isso o seu *Bean* é declarado no escopo *session*, fazendo com que cada aluno tenha uma instância particular dessa classe para se relacionar com o sistema.

Todas as requisições das páginas JSP serão feitas diretamente para um *Bean* desta classe, o qual se encarregará de tratar tal requisição e realizar chamadas de funções das outras classes caso sejam necessárias.

Todas as informações passadas pelas páginas são utilizadas apenas para processar os dados e armazená-los nos *Beans*, notando, aqui, que a classe Administrador age como um verdadeiro gerente do sistema. Todas as informações requeridas pelas páginas JSP são encaminhadas, também, somente para essa classe, como a requisição de uma tabela, ou dos dados para geração de um gráfico.

A classe Administrador foi utilizada pelo módulo MAC para permitir a comunicação entre as outras classes do sistema e as novas páginas de interação criadas para a técnica de análise de correspondência múltipla.

6.3.2.2. Classe estatística

Essa classe é responsável, principalmente, pela geração dos resultados estatísticos (gráficos e medidas numéricas) e cálculo de todos os procedimentos estatísticos presentes no sistema.

Um objeto dessa classe é instanciado pela página JSP como sendo um *Bean* com escopo *application*, e encapsulado em seguida pelo *Bean* Administrador. Esse encapsulamento é realizado para poder disponibilizar o uso de uma mesma instância para diversos alunos, a fim de economizar memória e processamento no servidor, ao mesmo tempo tornando a aplicação mais rápida pelo simples fato de não necessitar da geração dessa nova instância.

Para o desenvolvimento do módulo MAC, foram criados novos métodos na classe Estatística a fim de contemplar os procedimentos estatísticos envolvidos na análise de correspondência múltipla. A implementação destes métodos pode ser vista no (ANEXO 2) e um resumo está descrito no Quadro III:

QUADRO III – Resumo dos métodos implementados na classe Estatística

Método	Parâmetros	Descrição
Somatorio	_valor: valor que se deseja realizar o somatório.	Calcula e retorna o somatório de um número.
Sign	_valor: valor que se deseja analisar.	Retorna o sinal de um determinado número.
AcheMenorValor	_valores: valores da variável; _n: retorna o número de casos da variável; _media: retorna a média da variável; _mínimo: retorna o valor mínimo da variável.	Encontra o valor mínimo da variável, além de calcular a média desta variável.
AcheMaiorValor	_valores: valores da variável; _n: retorna o número de casos da variável; _media: retorna a média da variável; _mínimo: retorna o valor máximo da variável.	Acha o valor máximo da variável, além de calcular a média desta variável.
CalculeQ25	_valores: valores da variável; _Q25: retorna o primeiro quartil da variável.	Calcula e retorna o primeiro quartil da variável.
CalculeMediana	_valores: valores da variável; _mediana: retorna a mediana da variável.	Calcula e retorna a mediana da variável.
CalculeQ75	_valores: valores da variável; _Q75: retorna o terceiro quartil da variável.	Calcula e retorna o terceiro quartil da variável.
CategorizeVariavel	_numeroClasses: número de categorias para a categorização; _valores Quanti: valores da variável a ser categorizada; _intervalos: retorna o número de intervalos; _n: retorna o número de casos da variável; _mínimo: retorna o menor valor da variável; _maximo: retorna o maior valor da variável;	Categoriza uma variável quantitativa para qualitativa.

	_media: retorna a média da variável _DP: retorna o desvio-padrão da variável.	
AcheCategorias PorCategorias	_quali1: primeira variável de entrada; _quali2: segunda variável de entrada; _categoriaPorCategoria: objeto onde será colocada a tabela.	Gera a tabela de frequência observada de duas variáveis
CalculeFrequenciasEsperadas	_catPorCat: tabela de frequência observada; _tabContribuicaoQui2: armazenada a tabela de contribuição do qui-quadrado.	Gera a tabela de frequência esperada e a tabela de contribuições do qui-quadrado.

6.3.2.3. Classe base de dados

Essa classe faz uso do Driver “Connector/J” para realizar a conexão do JDBC ao banco de dados MySQL. Ela é responsável por todas as consultas e atualizações feitas ao banco de dados.

Todas as informações dos alunos são armazenadas no banco de dados, como usuário e senha, e o caminho onde se encontram os arquivos DBF gerados por esse aluno, como também algumas informações sobre esses arquivos.

Assim como a classe “estatística”, uma instância dessa classe é gerada na forma de um *Bean* pela página JSP e em seguida é encapsulada pelo *Bean* “administrador”.

6.3.2.4. Classe leitor DBF

Essa classe faz uso do Driver *Type 4 DBF JDBC 2.0* para realizar a conexão do JDBC com o arquivo DBF requerido pelo aluno. Ela é responsável pela obtenção dos dados contidos dentro desses arquivos.

Segundo NAKAZAWA & MARAFON (2003), optou-se pela utilização desse formato de arquivo pela facilidade de criação (pode ser gerado em qualquer programa de planilha eletrônica) e para manter o raciocínio de elaboração e armazenamento dos dados obtidos pelas versões anteriores do SEstat.Net.

Uma instância dessa classe é gerada pela página JSP como sendo um *Bean* com escopo *session* (isso significa que cada usuário terá um *Bean* “leitorDBF” diferente), sendo também encapsulado no *Bean* “administrador”. Isso se faz necessário, pois cada instância dessa classe necessita guardar os dados da base DBF escolhida por cada usuário.

6.3.2.5. Classe gráfico

A classe gráfico é responsável pela geração de gráficos que é uma das partes fundamentais do sistema e tem por finalidade mostrar de uma forma visual as análises estatísticas feitas sobre os dados do aluno.

6.3.3. Concepção da interface

Devido ao fato do módulo MAC estar integrado ao Sestat.Net, foi utilizado o mesmo padrão para a criação das interfaces. A cada interface de interação do aluno com o módulo, é disponibilizado o *Help* dos assuntos abordados, ou seja, os conceitos que o aluno pode querer consultar para aprender ou relembrar. É disponibilizada, também, a visualização da base de dados completa escolhida pelo aluno para a realização da análise.

6.3.3.1. Interfaces do módulo MAC

Segue abaixo uma seqüência das interfaces que o aluno irá interagir caso escolha realizar uma análise multivariada e o sistema decida por aplicar análise de correspondência múltipla.

A primeira interface do sistema (Fig. 6.2) é a entrada, onde o aluno deve informar o seu usuário e sua senha.

Home




help ≡≡≡

base ≡≡≡

docs ≡≡≡

- Artigo

≡≡≡

Entrada no SEstat.Net

Digite seus dados para entrar

Usuário:

Senha:

Entrar

2003 - LEA - Laboratório de Estatística Aplicada

Figura 6.2 – Interface de entrada do sistema

Com a entrada realizada, o sistema busca do Banco de dados no servidor todas as bases de dados .dbf referentes ao aluno e mostra (Fig 6.3) para que o mesmo possa escolher qual a base que ele deseja trabalhar. A base pap.dbf está sempre disponível, independente do aluno.

SEstat.Net

Home

help

- Criando uma base
- Base PAP
- Padrão dBase

base

docs

- Artigo

Escolha da base de dados

Arquivo	Descrição	Criação	Última modificação
<input type="radio"/> pap.dbf	Base de Dados PAP	2003-01-16	2003-01-17
<input type="radio"/> teste.dbf	Base Teste	2003-01-17	2003-01-17

Avançar >

2003 - LEA - Laboratório de Estatística Aplicada

POWERED BY

Figura 6.3 – Interface de escolha da base de dados

Escolhida a base de dados, o aluno deve optar, então, pelo método que deseja aplicar na mesma. A partir desta interface (Fig. 6.4), caso o aluno escolha análise multivariada, é que o módulo MAC começa funcionar.



The screenshot shows the SEstat.Net website interface. At the top, there is a blue header with the text "SEstat.Net" and a Brazilian flag. Below the header, the page is divided into two main sections. On the left, there is a navigation menu with the following items: "Home", "help" (with a hamburger menu icon), a list of links: "- Descrição", "- Descrição Univariada", "- Descrição Bivariada", "- Inferência", and "- Análise Multivariada", "base" (with a hamburger menu icon), "- Visualizar toda base", "docs" (with a hamburger menu icon), and "- Artigo". At the bottom of the left sidebar, there are logos for a penguin, "JAVA", and "POWERED BY MySQL". The main content area on the right is titled "Escolha do método" and contains the question "Qual é o método você deseja realizar?". Below this question, there are four blue, underlined links: "Descrição Univariada", "Descrição Bivariada", "Inferência", and "Análise Multivariada". At the bottom of the main content area, there is a horizontal line and the text "2003 - LEA - Laboratório de Estatística Aplicada".

Figura 6.4 – Interface de escolha do método

Com a escolha do método de Análise multivariada, o próximo passo é selecionar as variáveis que ele deseja trabalhar na análise. O aluno deve escolher três ou mais variáveis e, caso isto não ocorra, o módulo emite uma mensagem. A escolha das variáveis pode ser vista na Fig. 6.5.



The screenshot shows the SEstat.Net web interface. The header features the logo 'SEstat.Net' and a Brazilian flag. The left sidebar contains navigation links: 'Home', 'help' (with sub-links for 'Análise multivariada' and 'Variável'), 'base' (with sub-link for 'Visualizar toda base'), and 'docs' (with sub-link for 'Artigo'). The main content area is titled 'Escolha das variáveis' and asks the user to select variables. It includes instructions: 'Quais são as variáveis que você deseja trabalhar?' and 'Caso se queira trabalhar com uma variável suplementar, esta deve ser selecionada também. (utilize as teclas Ctrl ou Shift para selecionar mais de uma variável)'. A list of variables is shown in a scrollable box: CASO, LOCAL, PAP, INSTR, TAM, RENDA88, RENDA98, ALIM98, APROV88, and APROV98. An 'Avançar >' button is at the bottom. The footer indicates '2003 - LEA - Laboratório de Estatística Aplicada'.

Figura 6.5 – Interface de escolha das variáveis

Escolhidas as variáveis, a próxima etapa é a tipificação das mesmas, que será feita na próxima interface (Fig. 6.6). Neste ponto é utilizado o módulo SETip, que faz a verificação se o aluno informou corretamente o tipo da variável. Caso não tenha informado, o módulo mostra uma mensagem e o aluno tem a opção de prosseguir ou voltar e atribuir o tipo correto.

The screenshot shows the SEstat.Net web application interface. The main heading is "Tipo das variáveis". Below it, the instruction reads "Selecione o tipo de cada variável ativa:". A list of active variables is shown in a box: LOCAL, PAP, INSTR, and APPROV88. To the right of this list are two vertical input fields for selecting the variable type. Below the list is a "Reiniciar Tipificação" button. To the right of the input fields are two buttons: "Qualitativa >" and "Quantitativa >". At the bottom of the main area is an "Avançar >" button. The footer of the page reads "2003 - LEA - Laboratório de Estatística Aplicada".

SEstat.Net

Home

help

- [Análise Multivariada](#)
- [Variável](#)
- [Variável qualitativa](#)
- [Variável quantitativa](#)

base

- [Visualizar variável](#)
- [Visualizar toda base](#)

docs

- [Artigo](#)

Tipo das variáveis

Selecione o tipo de cada variável ativa:

Variáveis ativas:

LOCAL
PAP
INSTR
APROV88

Qualitativa >

Quantitativa >

Reiniciar Tipificação

Avançar >

2003 - LEA - Laboratório de Estatística Aplicada

Figura 6.6 – Interface de tipificação das variáveis

Após a tipificação, o sistema decide qual a técnica a ser aplicada: análise fatorial por componentes principais (módulo MAF) se todas as variáveis forem quantitativas ou análise de correspondência múltipla (módulo MAC) se todas forem qualitativas. Caso o aluno tenha escolhido qualitativas e quantitativas, serão fornecidas ao mesmo as seguintes opções: categorizar a(s) variável(is) quantitativa(s), ou seja, criar categorias para que essa variável se torne qualitativa; descartar a(s) variável(is) quantitativa(s), caso este descarte não resulte na sobra de menos de duas variáveis para a análise; ou ainda prosseguir, mas alertando-o de que os resultados estatísticos podem não ter validade porque será aplicada a ACM em variáveis quantitativas. Esta interface pode ser vista na Fig. 6.7.

SEstat.Net

Home

help ≡≡≡

- [Análise de Correspondência Múltipla](#)
- [Variável](#)
- [Variáveis Ativas](#)
- [Variáveis qualitativas](#)
- [Variáveis quantitativas](#)

base ≡≡≡

- [Visualizar variável](#)
- [Visualizar toda base](#)

docs ≡≡≡

- [Artigo](#)

Categorização

Considerando que seu conjunto de variáveis ativas possui variáveis qualitativas, então será aplicada a técnica de **Análise de Correspondência Múltipla**, mas existem variáveis quantitativas e essa técnica pressupõe que todas as variáveis sejam qualitativas.

Portanto, você tem as seguintes opções:

- Categorizar as variáveis quantitativas
- Descartar as variáveis quantitativas
- Prosseguir com as variáveis quantitativas, porém alertando que os resultados estatísticos podem não ter validade

Avançar >

2003 - LEA - Laboratório de Estatística Aplicada

Figura 6.7 – Interface de Aviso sobre a escolha de variáveis quantitativas e qualitativas

Caso o aluno escolha a opção de categorizar a(s) variável(is), o mesmo será indagado sobre a categorização de cada uma delas, caso seja mais de uma variável. E se ele decidir por categorizar, é mostrado, então, o número total de observações e o valor mínimo e máximo da variável quantitativa. O aluno deverá informar ainda, a quantidade de categorias que ele deseja criar para a variável. Veja Fig. 6.8.



The screenshot shows the SEstat.Net website interface. The header features the logo 'SEstat.Net' and a Brazilian flag. The main content area is titled 'Categorização' and asks 'Você deseja categorizar a variável (TAM)?'. There are two radio button options: 'Sim' (selected) and 'Não'. Below this, there are three input fields for 'Número total de observações' (120), 'Valor mínimo' (1), and 'Valor máximo' (9). A question 'Quantas categorias você deseja trabalhar?' is followed by an empty input field. A large blue button labeled 'Avançar >' is positioned below the input fields. The footer contains the text '2003 - LEA - Laboratório de Estatística Aplicada'. The left sidebar contains navigation links under 'help', 'base', and 'docs', along with logos for Java and MySQL.

SEstat.Net

Home

help

- [Análise de Correspondência Múltipla](#)
- [Variável](#)
- [Número total de observações](#)
- [Categorias](#)
- [Valor mínimo](#)
- [Valor máximo](#)

base

- [Visualizar variável](#)
- [Visualizar toda base](#)

docs

- [Artigo](#)

Categorização

Você deseja categorizar a variável (TAM)?

Sim

Não

Número total de observações:

Valor mínimo:

Valor máximo:

Quantas categorias você deseja trabalhar?

Avançar >

2003 - LEA - Laboratório de Estatística Aplicada

Figura 6.8 – Interface de categorização da variável quantitativa

O módulo MAC calcula o número de categorias recomendado para o caso através da fórmula: $1 + \log_2 N$ (onde N é o total de observações) e compara com o valor informado pelo aluno. Caso seja menor, o módulo mostra o valor recomendado e o especificado e um alerta dizendo que o número de categorias especificado é maior que o recomendado e, portanto, o aluno tem a opção de voltar e informar um novo número ou prosseguir, porém os resultados estatísticos podem não ter validade. Esta interface pode ser vista na Fig. 6.9.

The screenshot shows the SEstat.Net web interface. The header features the site name 'SEstat.Net' and a Brazilian flag. The left sidebar contains navigation links: 'Home', 'help' (with a menu icon), 'Múltipla' (with sub-links for 'Análise de Correspondência', 'Variável', and 'Número de categorias recomendado'), 'base' (with sub-links for 'Visualizar variável' and 'Visualizar toda base'), and 'docs' (with a link for 'Artigo'). The main content area is titled 'Categorização' and displays a warning: 'O número de categorias que você informou é maior que o número de categorias recomendado. Você pode voltar e informar um novo número ou avançar, porém alertando que os resultados estatísticos podem não ter validade.' Below the warning are two input fields: 'Número de categorias recomendado:' with the value '7' and 'Número de categorias especificado:' with the value '10'. A prominent 'Avançar >' button is centered below the fields. The footer contains the text '2003 - LEA - Laboratório de Estatística Aplicada' and logos for Java and MySQL.

Figura 6.9 – Interface de alerta sobre o número de categorias

Tendo informado o número de categorias, o módulo disponibiliza para o aluno a criação das mesmas (Fig. 6.10). São então calculados os intervalos para as categorias com um valor mínimo e um valor máximo, porém o aluno pode alterá-los conforme achar necessário. Os nomes das categorias padronizou-se chamar de Categoria1, Categoria2....CategoriaN.

SEstat.Net

Home

help ≡≡≡

- [Análise de Correspondência](#)

Múltipla

- [Variável](#)
- [Categorias](#)
- [Valor mínimo](#)
- [Valor máximo](#)

base ≡≡≡

- [Visualizar variável](#)
- [Visualizar toda base](#)

docs ≡≡≡

- [Artigo](#)

Categorização

Categorização da variável TAM

Para cada categoria criada, foram sugeridos o valor máximo e o valor mínimo, porém você pode alterar estes valores:

Categoria	Valor Mínimo	Valor Máximo
Categoria 1	<input type="text" value="1"/>	<input type="text" value="2"/>
Categoria 2	<input type="text" value="2"/>	<input type="text" value="3"/>
Categoria 3	<input type="text" value="3"/>	<input type="text" value="5"/>
Categoria 4	<input type="text" value="5"/>	<input type="text" value="6"/>
Categoria 5	<input type="text" value="6"/>	<input type="text" value="7"/>
Categoria 6	<input type="text" value="7"/>	<input type="text" value="8"/>
Categoria 7	<input type="text" value="8"/>	<input type="text" value="9"/>

Avançar >

2003 - LEA - Laboratório de Estatística Aplicada

Figura 6.10 – Interface de criação das categorias

O aluno tem, num próximo passo, a opção de trabalhar ou não com uma variável suplementar (Fig. 6.11). Caso escolha a opção sim, ele deve informar qual será a variável suplementar.



The screenshot shows the SEstat.Net web application interface. The header features the logo 'SEstat.Net' in large blue letters on a dark blue background, with a Brazilian flag graphic to the right. The left sidebar contains a navigation menu with sections: 'Home', 'help' (with a hamburger menu icon), 'base' (with a hamburger menu icon), and 'docs' (with a hamburger menu icon). Under 'help', there are links for 'Análise Multivariada', 'Variável', and 'Variável Suplementar'. Under 'base', there are links for 'Visualizar variável' and 'Visualizar toda base'. Under 'docs', there is a link for 'Artigo'. At the bottom of the sidebar are logos for a cartoon character, 'JAVA', and 'POWERED BY MySQL'. The main content area is titled 'Variável Suplementar' and contains the following text and controls:

Deseja utilizar uma variável suplementar na análise?

Sim
 Não

Qual destas variáveis será utilizada como a variável suplementar da análise?

LOCAL ▾

Avançar >

At the bottom of the page, there is a footer: 2003 - LEA - Laboratório de Estatística Aplicada.

Figura 6.11 – Escolha da variável suplementar

Os primeiros resultados que o módulo MAC mostra são as tabelas de frequência e gráfico de cada variável (Fig. 6.12).

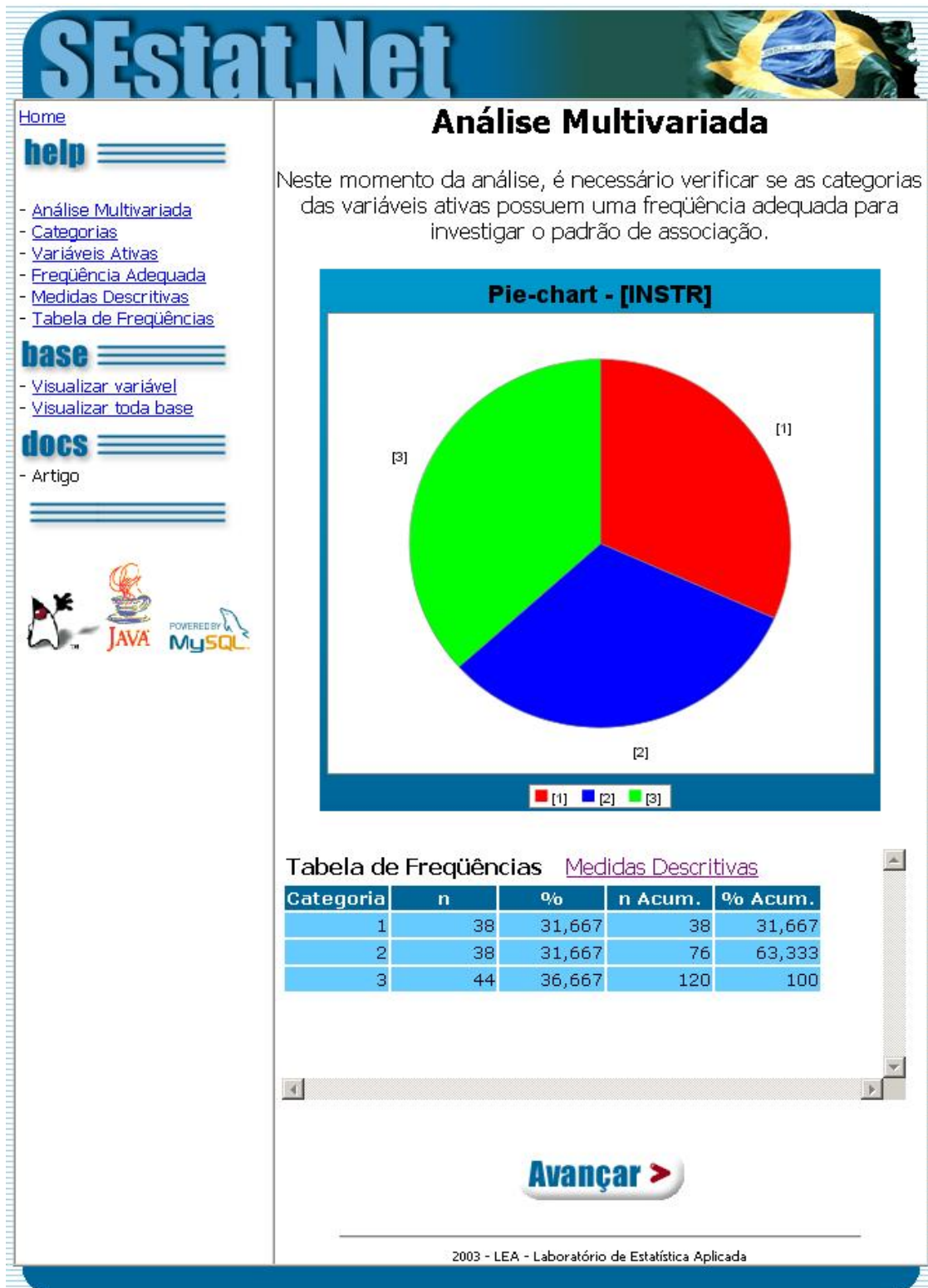


Figura 6.12 – Interface do resultado: Tabela de Frequência por Variável

O módulo verifica se todas as categorias das variáveis apresentam frequência maior que 10, considerada uma frequência adequada para investigar o padrão de associação. Caso exista alguma categoria que não se encaixe nesta regra, o módulo propõe o reagrupamento dos casos em uma outra e o aluno deve escolher em qual categoria ele deseja reagrupar (Fig. 6.13).



The screenshot shows the SEstat.Net web application interface. The header features the logo 'SEstat.Net' in large blue letters and a Brazilian flag graphic. The main content area is titled 'Reagrupamento de variáveis' and contains the following text: 'A categoria 2 da variável **APROV98** possui uma frequência muito baixa, deseja reagrupá-la para qual categoria?'. Below this text is a dropdown menu with the number '3' selected. A prominent blue button with a white arrow and the text 'Avançar >' is positioned below the dropdown. At the bottom of the main content area, there is a footer line that reads '2003 - LEA - Laboratório de Estatística Aplicada'. On the left side of the interface, there is a navigation menu with sections: 'Home', 'help' (with a sub-menu containing 'Reagrupamento de variáveis', 'Categorias', and 'Frequência baixa'), 'base' (with sub-menu 'Visualizar variável', 'Visualizar toda base'), and 'docs' (with sub-menu 'Artigo'). At the bottom left, there are logos for Java and MySQL, with the text 'POWERED BY' above the MySQL logo.

Figura 6.13 – Interface de reagrupamento de variáveis

A última etapa é a visualização dos resultados (Fig. 6.14). Estarão disponíveis para o aluno: Tabelas de Burt (Frequência Observada (Fig. 6.15), Frequência Esperada (Fig. 6.16) e Contribuições do Qui-quadrado (Fig. 6.17)), tabela de autovalores e inércia (Fig. 6.18) e gráficos de autovalores (Fig. 6.19) e de dimensões (Fig.6.20).

SEstat.Net

Home

help ≡

- [Análise de Correspondência Múltipla](#)
- [Tabela de Frequência Observada](#)
- [Tabela de Frequência Esperada](#)
- [Tabela de Contribuições do qui-quadrado](#)
- [Tabela de autovalores e inércia](#)
- [Gráfico de autovalores](#)
- [Gráfico de dimensões](#)
- [Número de categorias recomendado](#)

base ≡

- [Visualizar variável](#)
- [Visualizar toda base](#)

docs ≡

- [Artigo](#)

Análise de Correspondência Múltipla

Estes foram os resultados obtidos pela análise de correspondência múltipla que você pode analisar:

- [Tabela de Frequência Observada](#)
- [Tabela de Frequência Esperada](#)
- [Tabela de Contribuições do qui-quadrado](#)
- [Tabela de autovalores e inércia](#)
- [Gráfico de autovalores](#)
- [Gráfico de dimensões](#)

2003 - LEA - Laboratório de Estatística Aplicada

POWERED BY

Figura 6.14 – Interface de resultados obtidos

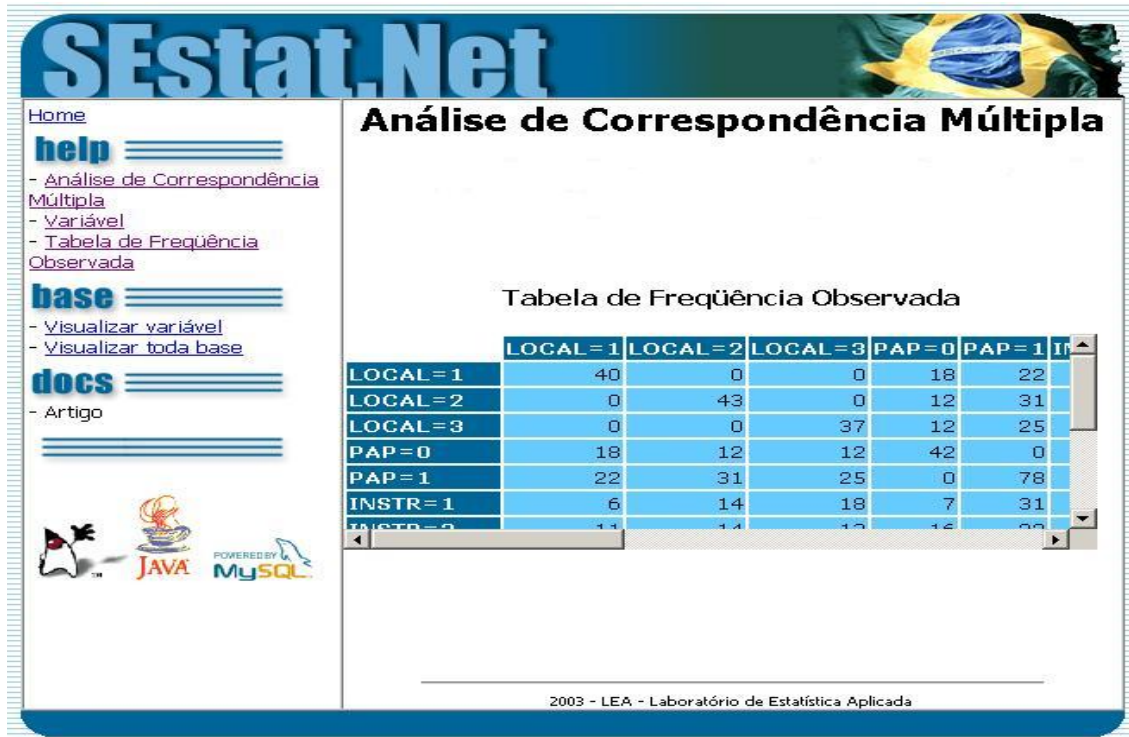


Figura 6.15 – Interface de resultado: Tabela de Frequência Observada



Figura 6.16 – Interface de resultado: Tabela de Frequência Esperada

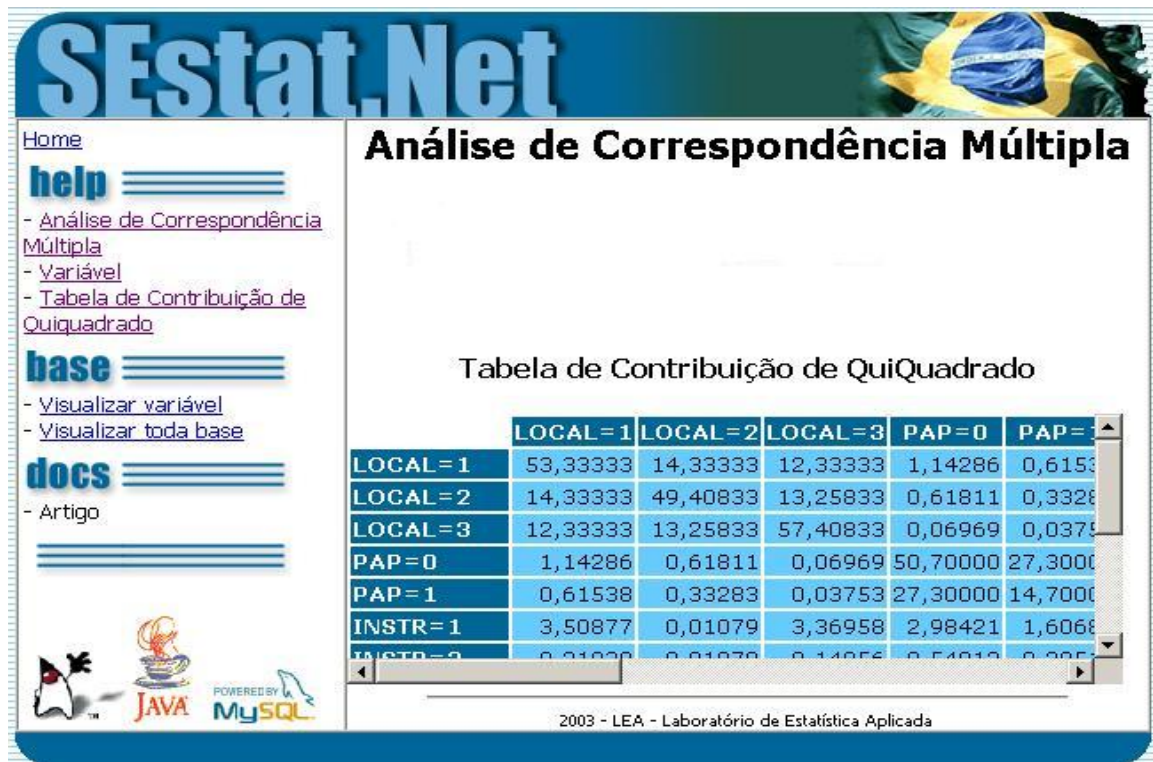


Figura 6.17 – Interface de resultado: Tabela de Contribuição do Qui-quadrado



Figura 6.18 – Interface de resultado: Tabela de Autovalores e Inércia

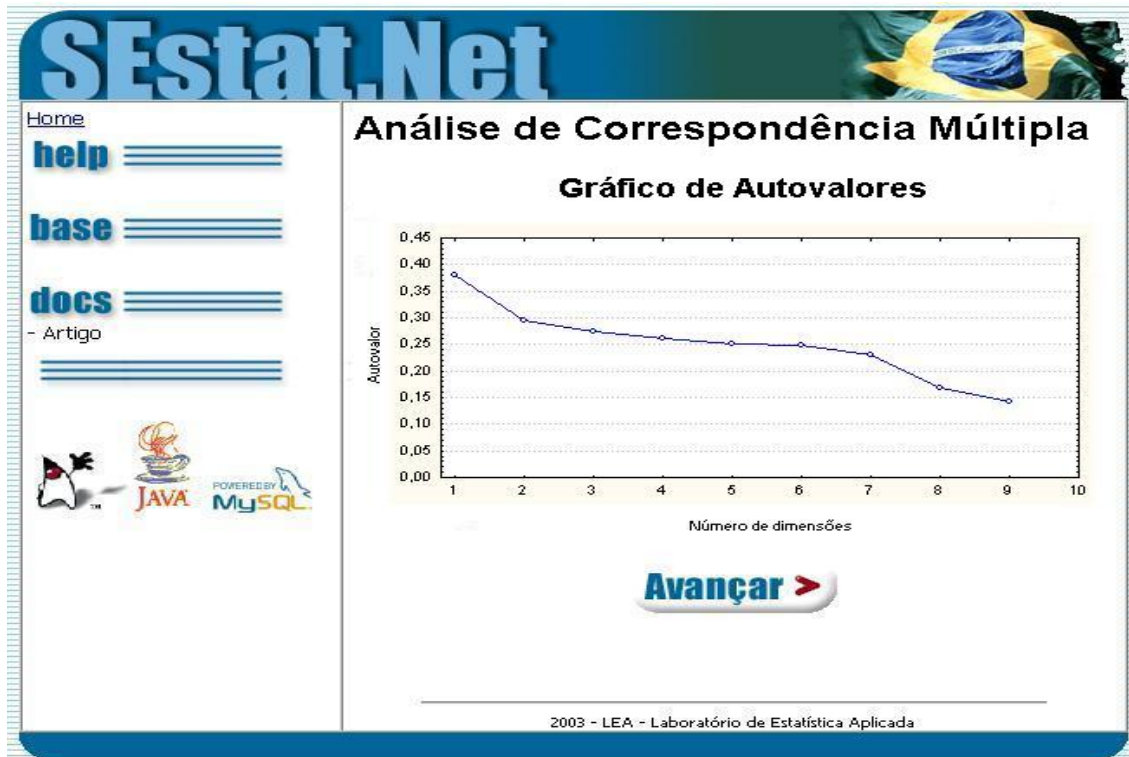


Figura 6.19 – Interface de resultado: Gráfico de Autovalores

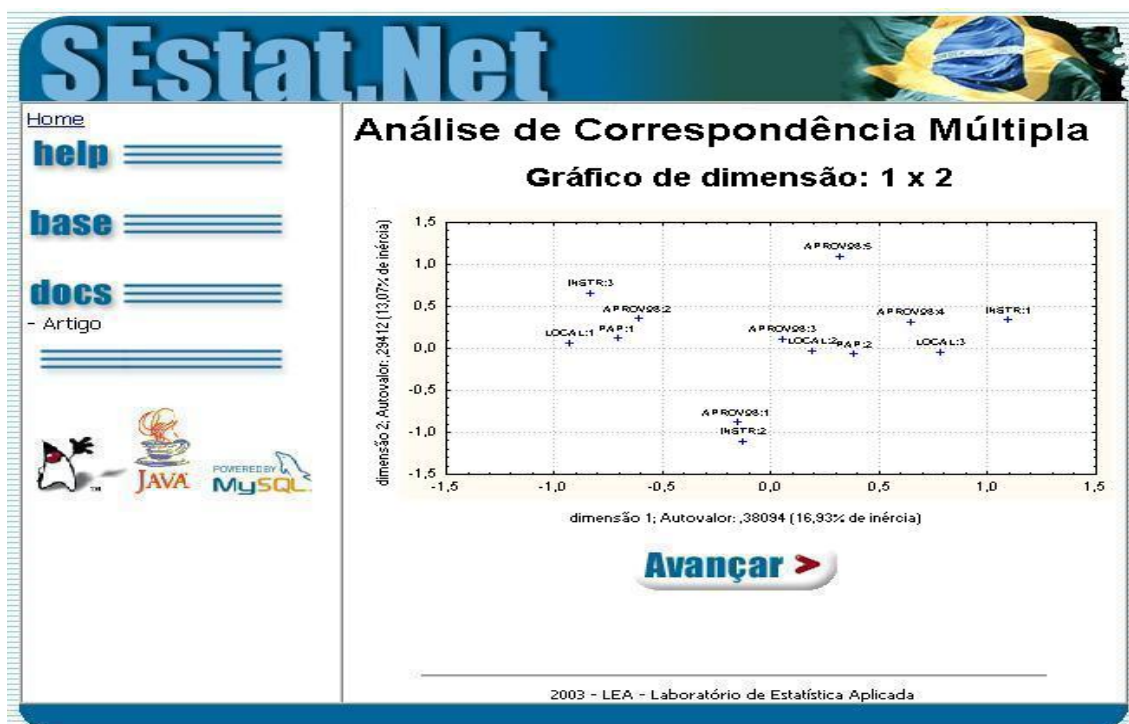


Figura 6.20 – Interface de resultado: Gráfico de dimensão

Para o cálculo dos autovalores, foi realizada uma pesquisa sobre os algoritmos computacionais existentes. Escolheu-se um algoritmo e, com base nele, foi criado um novo para ser implementado no módulo MAC. O algoritmo pode ser visto no (ANEXO 1).

6.3.4. Avaliação e Validação dos Algoritmos

Para a avaliação e validação do módulo, foram realizados vários testes com bases de dados distintas a fim de verificar a precisão dos procedimentos estatísticos implementados. O *software Statistica 5.0* foi usado para realizar a comparação dos resultados obtidos por ele e pelo módulo MAC na aplicação do método em uma mesma base de dados. Os testes apresentaram um bom resultado, pois na comparação realizada, obtiveram-se resultados semelhantes.

O uso efetivo do módulo ocorrerá a partir do primeiro semestre de 2003 em disciplinas de estatística ministradas no CTC/UFSC, onde poderão ser avaliados aspectos da aprendizagem de análise de correspondência múltipla e mudanças necessárias no módulo.

6.4. Disseminação

A última fase do ciclo de vida é a disseminação do sistema, ou seja, sua distribuição e possíveis manutenções para correção de problemas ou otimização. Serão mostrados aqui os recursos necessários para o funcionamento do SEstat.Net, e por consequência, do módulo MAC.

Como o SEstat.Net utiliza tecnologias como JSP, JavaBeans e Servlets é necessário que se tenha um servidor específico que compreenda essas tecnologias utilizadas. Para esse projeto, o servidor escolhido e utilizado foi o RESIN Enterprise 2.1.6, devido a sua ótima indicação por *sites* especializados, simplicidade de

configuração e ótima performance obtida nos testes realizados (NAKAZAWA & MARAFON, 2003).

Também foi necessária a utilização de um servidor de banco de dados. O servidor escolhido foi o MySQL, por se tratar de um banco de dados com performance comprovada para aplicações *Web* e de fácil utilização. Um ponto importante para essa escolha foi por se tratar de um software sem custo e *open-source*.

Além disso, dentro do servidor JSP devem estar instaladas as bibliotecas *Cewolf* e *JFreeChart* e os drivers *Connector/J 2.0* e *Type 4 DBF JDBC 2.0*.

6.5. Considerações Finais

Este capítulo procurou mostrar o processo de desenvolvimento do módulo MAC, ou seja, as fases do ciclo de vida e as atividades realizadas em cada fase. Buscou-se destacar as tecnologias adotadas, a modelagem do módulo e também as interfaces e as regras relacionadas a cada uma.

7. CONCLUSÃO

O processo de ensino-aprendizagem está passando por mudanças e a implantação das novas tecnologias educacionais exige uma fundamentação teórica que enfatize o processo de construção do conhecimento pelo aluno e não deixe que esses meios tecnológicos se tornem simplesmente uma forma de transmissão-recepção da informação.

Há, hoje, uma deficiência muito grande em *softwares* com esta concepção de ensino-aprendizagem, principalmente na área de estatística, e o SEstat.Net foi um passo muito importante na tentativa de suprir essa necessidade. Como complementação, disponibilizar o método de análise multivariada só teve a acrescentar ao software, pois é um método com uma certa complexidade e o módulo MAC ajudará a facilitar a compreensão pelo aluno.

A investigação realizada sobre o uso das tecnologias educacionais no processo de ensino-aprendizagem contribuiu para que o módulo fosse desenvolvido baseado em uma concepção pedagógica, o construtivismo, neste caso.

Com o estudo do SEstat.Net, foi possível a integração com o mesmo, realizada de forma fácil, devido ao uso das mesmas tecnologias e também pelo reaproveitamento de classes e métodos já criados para determinadas funções do módulo.

O desenvolvimento do módulo seguiu um modelo de ciclo de vida, o que facilitou bastante o processo de criação. Foram apresentadas as fases do desenvolvimento e as atividades concluídas em cada fase, proporcionando o seu entendimento.

A implementação dos algoritmos computacionais exigiu um certo grau de complexidade matemática, daí uma das maiores dificuldades nesta pesquisa.

O módulo foi validado com os testes realizados e com as comparações com os resultados obtidos através do *software Statistica*, porém esta validação somente estará completa quando o MAC for utilizado pelos alunos nas aulas de estatística, onde

realmente será possível constatar possíveis problemas, principalmente nas interfaces de interação do módulo durante o processo ensino-aprendizagem.

7.1. Trabalhos Futuros

Como trabalhos futuros, recomenda-se:

- Permitir que o aluno informe o percentual acumulativo de contribuição da inércia e também o número de dimensões que deseja trabalhar;
- A implementação da aquisição de conhecimento pelo módulo fazendo com que o aluno não responda as mesmas perguntas a cada vez que realiza uma nova análise com a mesma base de dados. Por exemplo, a tipificação das variáveis;
- O desenvolvimento de novos módulos para serem integrados ao SStat.Net como: análise de discriminante, regressão múltipla, análise de correlação canônica, análise de agrupamentos, entre outros.

REFERÊNCIAS

ALMEIDA, Prof. Tabajara Lucas de. **Análise de Discriminante**. Departamento de Matemática – FURG. Documento disponível na Internet via WWW: <URL: <http://lula.dmat.furg.br/~taba/posanadis.htm>> out-2002.

BARBETTA, Pedro A. **Estatística Aplicada às Ciências Sociais**. Florianópolis: Editora da UFSC, 1999. 3.ed.

BATISTA, Galileu. **Conteúdo Dinâmico, CGI, Servlets e JSP**. JSPBrasil. Documento disponível na Internet via <URL: <http://jspbrasil.com.br:8081/jspbrasil/forum.jsp?forum=1001&action=10&pageSize=100>>. Fev - 2003.

BELLONI, Maria Luiza. **Educação à Distância**. Campinas: Autores Associados, 1999.

BITTENCOURT, Guilherme. **Inteligência Artificial: ferramentas e teorias**. Florianópolis: Editora da UFSC, 1998.

BOOCH, Grady; RUMBAUGH, James; JACOBSON, Ivar. **UML, Guia do Usuário**. Rio de Janeiro: Campus, 2000.

CAPRILE, Rodolfo H. **Una Propuesta para Evaluar el Trabajo de Laboratorio em Biología y Ciências Naturales**. Revista de Tecnologia Educativa, 1991.

CARVALHO, Luis Alfredo Vidal de. **Datamining: A mineração de dados no marketing, medicina, economia, engenharia e administração**. São Paulo: Érica, 2001.

CECHINEL, Cristian et al. **Concepção e Implementação de um Ambiente de Ensino de Estatística**. In: PRESTA – CONFERÊNCIA INTERNACIONAL: EXPERIÊNCIAS E PERSPECTIVAS DO ENSINO DE ESTATÍSTICA, UFSC: Florianópolis, 1999.

CRIVISQUI, Eduardo. **Análise Estatística Multivariada**. Presta - II Seminário, por videoconferência, de métodos exploratórios de estatística multivariada. Florianópolis: UFSC, 1999.

CURILEM, G. M. **Tutoriais Inteligentes**. Seminário da disciplina de Inteligência Aplicada à Educação. Curso de Pós-graduação em Ciência da Computação da UFSC, Florianópolis, 1998.

DIAS, Kirliam Maciel. **Sistema especialista para auxílio ao ensino de estatística**. Trabalho de Conclusão de Curso da graduação em Ciências da Computação. Florianópolis: UFSC, 2000.

HAIR, Joseph F. **Multivariate Data Analysis**. New Jersey: Prentice-Hall, 1995. 4.ed.

JOHNSON, Richard A & WICHERN, Dean W. **Applied Multivariate Statistical Analysis**. New Jersey – USA: Prentice Hall, 1998. 4 ed.

LEVY, Pierre. **As tecnologias da inteligência**. São Paulo: Editora 34, 1994.

LIBÂNEO, José Carlos. **Adeus professor, adeus professora: novas exigências educacionais e profissão docente**. São Paulo: Cortez, 2000. 4 ed.

LIMA, Gleydson. **Tutoriais: JavaBeans**. JSPBrasil. Documento disponível na Internet via <URL: <http://jspbrasil.com.br:8081/jspbrasil/tutobasico/javabeans.jsp>>. Fev - 2003.

MAGALHÃES, Léo Pini. **Autoria de materiais educacionais**. Relatório final de atividades: Projeto Sapiens. Unicamp/USP. Documento disponível na Internet via WWW: <URL: www.dca.fee.unicamp.br/projects/sapiens/Reports/rf2000/node14.html>. dez-2002

MARQUES, Cristina P.C., MATTOS, M. Isabel L. de, TAILLE, Yves de la. **Computador e Ensino: uma aplicação à língua portuguesa**. São Paulo: Ática, 1995. 2. ed.

MENDES, Raquel Dias. **Sistemas especialistas no gerenciamento da informação**. IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia. Documento disponível na Internet via WWW: <URL: <http://www.ibict.br/cionline/260197/26019705.pdf>> dez-2002.

NAKAZAWA, Carlos Alberto; MARAFON, Márcio Juliandrei. **SEstat.Net – Ensino de estatística mediado por computador**. Trabalho de Conclusão de Curso da graduação em Ciências da Computação. Florianópolis: UFSC, 2003.

NASSAR, Silvia M. **Sistema Estatístico Inteligente para Apoio a Pesquisas Médicas**. Tese de doutorado. Florianópolis: UFSC, Depto. de Engenharia Elétrica, 1995.

NETO, Ulisses Telemaco. **Tutoriais: O que é JSP**. JSPBrasil. Documento disponível na Internet via <URL: <http://jspbrasil.com.br:8081/jspbrasil/tutobasico/oqueehjsp.jsp>>. Fev - 2003.

NISKIER, Arnaldo. **Tecnologia educacional: uma visão política**. Petrópolis: Vozes, 1993.

NOAL, Renato Bica. **Ambiente colaborativo para ensino de estatística com o SEstat**. Dissertação de Mestrado em Ciência da Computação. Florianópolis: CPGCC, UFSC, 2002.

OLIVEIRA, Diene Eire M. B. **Informática educativa: uma reflexão sobre o papel dos professores**. IN: Educação: identidades e contextos. Londrina: UNOPAR, 2001.

OLIVEIRA, Ely F. T. de.; GRACIO, Maria C. Cabrini. **O Ensino de Estatística na Graduação: Ensaio para sua Reformulação.** In: PRESTA – CONFERÊNCIA INTERNACIONAL: EXPERIÊNCIAS E PERSPECTIVAS DO ENSINO DE ESTATÍSTICA, Florianópolis: UFSC, 1999.

PAVAN, Marcus Vinicius. **SEstat.Net: Módulo de Análise por Componentes Principais.** Trabalho Individual. Florianópolis: CPGCC, UFSC, 2003.

PEREIRA, Júlio C. R. **Análise de Dados Qualitativos.** São Paulo: EDUSP-FAPESP, 1999.

POTTER, A.M. **Statistics for Sociologists: Teaching Techniques that Work.** Teaching Sociology, v.23, July, 1995.

QUEIROZ, Vivina Dias Sol. **Uma contribuição da psicologia sócio-histórica para o uso do computador na educação especial.** IN: Associação Brasileira de Tecnologia Educacional. Tecnologia educacional. Rio de Janeiro: Revista Brasileira de Tecnologia Educacional. Ano XXIX. Nº 152/153, 2001.

RICH, Elaine; KNIGHT, Kevin. **Inteligência artificial.** Tradução Maria Cláudia Santos Ribeiro Ratto; revisão técnica Alvaro Antunes. São Paulo: Makron Books, 1993.

ROITER, K., PETROCZ, P. **Introductory statistics courses – A new way of thinking.** Journal of Statistics Education., v.4, n.2, 1996.

SÁ, Ricardo Antunes de. **Educação à Distância: considerações preliminares de um percurso na UFPR.** IN: Associação Brasileira de Tecnologia Educacional. Tecnologia educacional. Rio de Janeiro: Revista Brasileira de Tecnologia Educacional. Ano XXIX. Nº 152/153, 2001.

SAMPAIO, Maria Narciso; LEITE, Lígia Silva. **Alfabetização Tecnológica do professor.** Petrópolis: Vozes, 1999.

SANTOS, José Gonçalo dos. **SETip – Sistema Especialista para Tipificar Dados de uma Pesquisa: Variáveis Qualitativas e Quantitativas.** Dissertação de Mestrado em Ciência da Computação. Florianópolis: CPGCC, UFSC, 2001.

SAS Institute Inc., **SAS/STAT User's Guide.** Version 6, 4.ed. v.1, Cary, NC: SAS Institute Inc., 1989.

SCHÄR, S.G.; KRUEGER, H. **Using new learning technologies with multimedia.** IEEE Multimedia, 2000.

SCOTT, Carlisle; CLAYTON, Jan E.; GIBSON, Elizabeth L. **A Practical Guide to Knowledge Acquisition.** USA: Addison Wesley Publishing Company, 1991.

SHUDO, Regina. **Sala de Aula e Avaliação: Caminhos e Desafios**. In: Nova Escola: a revista do professor. Ano 2001.

STAIR, Ralph M. **Princípios de Sistemas de Informação**. Rio de Janeiro: LTC, 1998. 2.ed.

STATSOFT. **STATISTICA**. [on line] Documento disponível na Internet via WWW: <URL: <http://www.statsoft.com> > nov – 2002

STEINBRUCH, Alfredo; WINTERLE, Paulo. **Álgebra Linear**. São Paulo: McGraw-Hill, 1987. 2.ed.

SUN Microsystems. **JAVA Servlet Technology**. Documento disponível na Internet via <URL: <http://java.sun.com/products/servlet/> >. Fev – 2003a.

SUN Microsystems. **JDBC Data Access API**. Documento disponível na Internet via <URL: <http://java.sun.com/products/jdbc/> >. Fev – 2003b.

UNOPAR – Universidade Norte do Paraná. **Projeto pedagógico do curso normal superior – ensino a distância**. Londrina, 2002.

VASCONCELLOS, Celso dos S. **Metodologia dialética em sala de aula**. In Revista Associação de educadores católicos, 1997.

VYGOTSKY, Lev Semiovith. **A formação social da mente**. São Paulo: Martins Fontes, 1989. 3.ed.

ANEXO 1 – Algoritmo para cálculo de autovalores e autovetores

Geração de AutoValores e AutoVetores

```
inicio
  N <- R
  L <- C
  A <- 0.00000001
  C <- 0
  para I de 2 ate N faça
    para J de 1 até I - 1 faça
      C <- C + 2 * (R2[I, J] ^ 2)
    fim para
  fim para
  Y <- sqrt(C)
  O <- (A / N) * Y
  T <- Y
  D <- 0
  inicio
    T <- T / N
    para Q de 2 até N faça
      para P de 1 até Q - 1 faça
        se ABS(R2[P, Q]) >= T
          inicio
            D <- 1
            V <- R2[P, P]
            Z <- R2[P, Q]
            E <- R2[Q, Q]
            F <- .5 * (V - E)
            se F = 0
              g <- -1
            senao
              G <- -(SGN(F))
            fim se
            G <- G * Z / (sqrt(Z ^ 2 + F ^ 2))
            se G < -1 entao
              G <- -.999999
            fim se
            H <- G / (sqrt(2 * (1 + sqrt(1 - G ^ 2))))
            K <- sqrt(1 - H ^ 2)
            para I de 1 ate N faça
              se (I = P) ou (I = Q) entao
                C <- B2[I, P]
                F <- B2[I, Q]
                B2[I, Q] <- C * H + F * K
                B2[I, P] <- C * K - F * H
              senao
                C <- R2[I, P]
                F <- R2[I, Q]
                R2[Q, I] <- C * H + F * K      -> AUTOVALORES
                R2[I, Q] <- R2[Q, I]         -> AUTOVALORES
                R2[P, I] <- C * K - F * H     -> AUTOVALORES
                R2[I, P] <- R2[P, I]         -> AUTOVALORES
            fim se
          fim para
          R2[P, P] <- V * K ^ 2 + E * H ^ 2 - 2 * Z * H * K -> AUTOVALORES
          R2[Q, Q] <- V * H ^ 2 + E * K ^ 2 + 2 * Z * H * K -> AUTOVALORES
          R2[P, Q] <- (V - E) * H * K + Z * (K ^ 2 - H ^ 2) -> AUTOVALORES
          R2[Q, P] <- R2[P, Q]                -> AUTOVALORES
        fim
      fim para
    fim para
    se T > 0
      d <- 0
    fim se
    D <- 0
  fim
  para I de 1 até N
    Q1[I] <- I
  fim para
  J <- 0
  inicio
```

```

V1 <- 0
J <- J + 1
para I de 1 até N - J
  se R2[I, I] < R2[I + 1, I + 1] entao
    V1 <- 1
    V2 <- R2[I, I]
    R2[I, I] <- R2[I + 1, I + 1]    -> AUTOVALORES
    R2[I + 1, I + 1] <- V2        -> AUTOVALORES
    P <- Q1[I]
    Q1[I] <- Q1[I + 1]
    Q1[I+1] <- P
  fim se
fim para
se V1 = 0
  fim
fim se
fim

Q <- 0
D7 <- 0
para J de 1 até N faça
  S2[J] <- R2[J, J]                -> AUTOVALORES
  se S2[J] > 0 entao
    S3[J] <- sqrt(S2[J])
    D7 <- D7 + (S2[J])
  senao
    S2[J] <- 0.00000001           -> AUTOVALORES
  end
fim para
para J de 1 até N faça
  D8[J] <- 100 * S2[J] / D7
  D9 <- D9 + (S2[J]) / D7
  D6[J] <- 100 * D9
  K <- Q1[J]
  V <- 0
  para I de 1 até N faça
    V <- V + B2[I, K] ^ 2
  fim para
  para I de 1 até N faça
    R2[I, J] <- B2[I, K] * sqrt(1 / V)    -> AUTOVALORES
  fim para
fim para
para I de 1 até N faça
  R2[I, 2] <- R2[I, 2] * (-1)    -> AUTOVALORES
  R2[I, 3] <- R2[I, 3] * (-1)    -> AUTOVALORES
fim para
para I de 1 até N faça
  para J de 1 até N faça
    C1[I, J] <- 0
    B2[I, J] <- 0
  fim para
  C1[I, I] <- S3[I]
fim para
para I de 1 até N faça
  para J de 1 até N faça
    para K de 1 até N faça
      B2[I, J] <- B2[I, J] + (R2[I, K] * C1[K, J])
    fim para
  fim para
fim para
fim para

```

Cálculo dos Scores fatoriais

```
C <- L
para I de 1 até C faça
  para J de 1 até R faça
    X1[I, J] <- 0
  fim para
fim para
para I de 1 até C faça
  para J de 1 até R faça
    para K de 1 até R faça
      X1[I, J] <- X1[I, J] + X2[K, I] * R2[K, J]
    fim para
  fim para
fim para
```

ANEXO 2 – Métodos implementados na classe Estatística

```
/**
// Retorna o somatório de um número.
// Parâmetros:
//   __valor: valor que se deseja realizar o somatório.
/**
public long getSomatorio(int __valor) {
    if (__valor <= 0)
        return 0;
    else
        return __valor + getSomatorio(__valor-1);
}

/**
// Retorna o sinal de um determinado valor.
// Parâmetros:
//   __valor: valor que se deseja analisar.
/**
protected int getSign(double _valor) {
    if (_valor < 0)
        return -1;
    else if (_valor >0)
        return 1;
    else
        return 0;
}

/**
// Acha o valor mínimo da variável, além de calcular a média desta variável.
// Parâmetros:
//   __valores: valores da variável.
//   __n: retorna o número de casos da variável.
//   __media: retorna a média da variável.
//   __mínimo: retorna o valor mínimo da variável.
/**
public int AcheMenorValor(String[] __valores, Objeto __n, Objeto __media, Objeto
__minimo) {
    long i,tamanho;
    double aux2,soma;
    int ret;
    tamanho = __valores.length;
    Objeto aux = new Objeto(0);
    Objeto j = new Objeto(0);
    ret = PrimeiroValor(__valores,aux,j);
    if(ret != 0){
        return(ret);
    }else{
        // aux contém o primeiro (não nulo) valor do array
        __n.setLong(1);
        soma = aux.doubleValue();
        for(i=j.intValue()+1; i<=tamanho-1;i++){
            if(!__valores[(int)i].equals("")){
                try{
                    aux2 = Double.valueOf(__valores[(int)i]).doubleValue();
                    if(aux.doubleValue() > aux2){
                        aux.setDouble(aux2);
                    }
                }
                soma = soma + aux2;
                __n.setLong(__n.longValue() + 1);
            }catch(NumberFormatException e){
                return(3);
            }
        }
    }
}
```

```

    }
    }
    __media.setDouble(soma/__n.doubleValue());
    __minimo.setDouble(aux.doubleValue());
    return(0);
}
}

//*****
// Acha o valor máximo da variável, além de calcular a média desta variável.
// Parâmetros:
//   __valores: valores da variável.
//   __n: retorna o número de casos da variável.
//   __media: retorna a média da variável.
//   __mínimo: retorna o valor máximo da variável.
//*****
public int AcheMaiorValor(String[] __valores, Objeto __n, Objeto __media, Objeto
__maximo) {
    long i,tamanho;
    double aux2,soma;
    int ret;
    tamanho = __valores.length;
    Objeto aux = new Objeto(0);
    Objeto j = new Objeto(0);
    ret = PrimeiroValor(__valores,aux,j);
    if(ret !=0){
        return(ret);
    }else{
        // aux contém o primeiro (não nulo) valor do array
        __n.setLong(1);
        soma = aux.doubleValue();
        for(i=j.intValue()+1; i<=tamanho-1;i++){
            if(!__valores[(int)i].equals("")){
                try{
                    aux2 = Double.valueOf(__valores[(int)i]).doubleValue();
                    if(aux.doubleValue() < aux2){
                        aux.setDouble(aux2);
                    }
                    soma = soma + aux2;
                    __n.setLong(__n.longValue()+1);
                }catch(NumberFormatException e){
                    return(3);
                }
            }
        }
        __media.setDouble(soma/__n.doubleValue());
        __maximo.setDouble(aux.doubleValue());
        return(0);
    }
}

//*****
// Calcula o primeiro quartil da variável
// Parâmetros:
//   __valores: valores da variável.
//   __Q25: retorna o primeiro quartil da variável.
//*****
public int CalculeQ25(double[] __valores, Objeto __Q25){
    long n;
    int ret;
    Objeto _valores = new Objeto();
    _valores.__double = __valores;
    n = _valores.__double.length;
    if(n==0){
        return(8);
    }
}

```

```

    }else{
        ret = OrdeneValores(_valores);
        if(ret != 0){
            return(ret);
        }else{
            __Q25.setDouble(_valores.__double[Math.round((n+1)/4)-1]);
            return(0);
        }
    }
}

}

//*****
// Calcula a mediana da variável
// Parâmetros:
//   __valores: valores da variável.
//   __mediana: retorna a mediana da variável.
//*****
public int CalculeMediana(double[] __valores, Objeto __mediana) {
    long n;
    int ret;
    Objeto _valores = new Objeto();
    _valores.__double = __valores;
    n = _valores.__double.length;
    if(n==0){
        return(8);
    }else{
        ret = OrdeneValores(_valores);
        if(ret != 0){
            return(ret);
        }else{
            if(n%2 == 0){ // É Par
                __mediana.setDouble((_valores.__double[(int)(n/2)-
1]+__valores[(int)(n/2)])/2);
            }else{
                __mediana.setDouble(_valores.__double[(int)((n+1)/2)-1]);
            }
            return(0);
        }
    }
}

//*****
// Calcula o terceiro quartil da variável
// Parâmetros:
//   __valores: valores da variável.
//   __Q75: retorna o terceiro quartil da variável.
//*****
public int CalculeQ75(double[] __valores, Objeto __Q75) {
    long n;
    int ret;
    Objeto _valores = new Objeto();
    _valores.__double = __valores;
    n = _valores.__double.length;
    if(n==0){
        return(8);
    }else{
        ret = OrdeneValores(_valores);
        if(ret != 0){
            return(ret);
        }else{
            __Q75.setDouble(_valores.__double[Math.round(3*(n+1)/4)-1]);
            return(0);
        }
    }
}
}

```

```

//*****
// Função para categorizar uma variável.
// Parâmetros:
//   __numeroClasses: número de categorias para a categorização.
//   __valoresQuanti: valores da variável a ser categorizada.
//   __intervalos: retorna o número de intervalos.
//   __n: retorna o número de casos da variável.
//   __minimo: retorna o menor valor da variável.
//   __maximo: retorna o maior valor da variável.
//   __media: retorna a média da variável.
//   __DP: retorna o desvio-padrão da variável.
//*****

public int CategorizeVariavel(long __numeroClasses, String[] __valoresQuanti,
ObjetoIntervalo __intervalos, Objeto __n, Objeto __minimo, Objeto __maximo, Objeto
__media, Objeto __DP) {
    long NumeroClasses,i;
    NumeroClasses = __numeroClasses;
    double tamanhoIntervalo;
    int ret;
    // Acha o Número de Classes pela regra de Sturges
    Objeto vMin = new Objeto(0); // será passado para referencia.
    ret = AcheMenorValor(__valoresQuanti, __n, __media, vMin);
    if(ret != 0){
        return(ret);
    }else{
        Objeto vMax = new Objeto(0);
        ret = AcheMaiorValor(__valoresQuanti,__media.doubleValue(), __n, __DP, vMax);
        if(ret != 0){
            return(ret);
        }else{
            //NumeroClasses = Math.round(1 + 3.3*(Math.log(__valoresQuanti.length) *
(1/Math.log(10))));
            __intervalos.intervalo = new Intervalo[(int)NumeroClasses];
            for(int p=0; p<(int)NumeroClasses; p++){
                __intervalos.intervalo[p] = new Intervalo();
            }
            for(i = 0; i<= NumeroClasses-1; i++){
                __intervalos.intervalo[(int)i].setQtidade(0);
            }
            tamanhoIntervalo = (vMax.doubleValue()-vMin.doubleValue())/NumeroClasses;
            if(tamanhoIntervalo==0){ // array constante - Minimo e Maximo são iguais
                __intervalos.intervalo = new Intervalo[1];
                __intervalos.intervalo[0] = new Intervalo();
                __intervalos.intervalo[0].setMin(vMin.doubleValue());
                __intervalos.intervalo[0].setMax(vMin.doubleValue());
            }else{
                __intervalos.intervalo[0].setMin(vMin.doubleValue());
                __intervalos.intervalo[0].setMax(vMin.doubleValue()+tamanhoIntervalo);
                for(i=1; i<=NumeroClasses-1; i++){
                    __intervalos.intervalo[(int)i].setMin(__intervalos.intervalo[(int)i-
1].getMax());
                    __intervalos.intervalo[(int)i].setMax(__intervalos.intervalo[(int)i-
1].getMax()+tamanhoIntervalo);
                }
            }
            __intervalos.intervalo[(int)NumeroClasses-1].setMax(vMax.doubleValue());
            ret = ColoqueValoresNosIntervalos(__valoresQuanti,__intervalos);
            if(ret != 0){
                return(ret);
            }else{
                __minimo.setDouble(vMin.doubleValue());
                __maximo.setDouble(vMax.doubleValue());
                return(0);
            }
        }
    }
}

```

```

    }

    /*******
    // Gera a tabela de frequência observada de duas variáveis.
    // Parâmetros:
    //   __quali1: primeira variáveis de entrada.
    //   __quali2: segunda variáveis de entrada.
    //   __categoriaPorCategoria: objeto onde será colocada a tabela; é passada por
    //                               referência
    /*******
    public int AcheCategoriasPorCategorias(VariaveisQualitativas __quali1,
    VariaveisQualitativas __quali2, ObjetoRefCategoriaPorCategoria __categoriaPorCategorias)
    {
        long auxTam,tamanho,i,j,k,l,m;
        int ret;
        ret = VerificaParametrosEntrada(__quali1,__quali2);
        k = 0;
        m = 0;
        if(ret != 0){
            return(ret);
        }else{
            tamanho = __quali1.getValores().length;
            __categoriaPorCategorias.categoriaPorCategoria = new
            CategoriaPorCategoria[(int)__quali1.getNumCategorias()];
            for(int p =0; p< (int)__quali1.getNumCategorias(); p++){
                __categoriaPorCategorias.categoriaPorCategoria[p] = new
            CategoriaPorCategoria();
            }
            for(i=0; i<= __quali1.getNumCategorias()-1; i++){
                __categoriaPorCategorias.categoriaPorCategoria[(int)i].setValor(__quali1.getCategorias()
                [(int)i].getValor());
                __categoriaPorCategorias.categoriaPorCategoria[(int)i].setQtidade(0);
                __categoriaPorCategorias.categoriaPorCategoria[(int)i].Categorias = new
            TCategoria[(int)__quali2.getNumCategorias()];
                for(int p =0; p< (int)__quali2.getNumCategorias(); p++){
                    __categoriaPorCategorias.categoriaPorCategoria[(int)i].Categorias[p] =
            new TCategoria();
                }
                for(j=0; j<= __quali2.getNumCategorias()-1; j++){
                    __categoriaPorCategorias.categoriaPorCategoria[(int)i].Categorias[(int)j].setValor(__qua
                    li2.getCategorias()[(int)j].getValor());

                    __categoriaPorCategorias.categoriaPorCategoria[(int)i].Categorias[(int)j].setQuantidade(
                    0);
                }
            }
            for(i=0; i<= tamanho-1; i++){
                if(("".equals(__quali1.getValores()[(int)i])) &&
                (!"".equals(__quali2.getValores()[(int)i] ))){
                    j=0;
                    auxTam = __categoriaPorCategorias.categoriaPorCategoria.length;
                    while(j<auxTam){

            if(!__categoriaPorCategorias.categoriaPorCategoria[(int)j].getValor().equals(__quali1.ge
            tValores()[(int)i])){
                j=j+1;
            }else{
                k=j;
                j=auxTam+1;
            }
            }
            if(j==auxTam){
                return(11);
            }

            __categoriaPorCategorias.categoriaPorCategoria[(int)k].setQtidade(__categoriaPorCategoria
            s.categoriaPorCategoria[(int)k].getQtidade()+1);
        }
    }

```



```

        l=0;
        auxTam =
__categoriaPorCategorias.categoriaPorCategoria[(int)k].Categorias.length;
        while(l<auxTam){

if(!__quali2.getValores()[ (int)i].equals(__categoriaPorCategorias.categoriaPorCategoria[
(int)k].Categorias[(int)l].getValor())){
            l=l+1;
        }else{
            m=l;
            l=auxTam+1;
        }
        }
        if(l==auxTam){
            return(l1);
        }

__categoriaPorCategorias.categoriaPorCategoria[(int)k].Categorias[(int)m].setQuantidade(
__categoriaPorCategorias.categoriaPorCategoria[(int)k].Categorias[(int)m].getQuantidade(
)+1);
    }
    }
    return(0);
}
}
}

```

```

//*****

```

```

// Gera a tabela de frequência esperada e a tabela de Contribuição de Quiquadrado.

```

```

// Retorno: a tabela de frequência esperada.

```

```

// Parâmetros:

```

```

//     _catPorCat: tabela de frequência observada.

```

```

//     _tabCorrelacaoQui2: objeto onde será armazenada a tabela de contribuição de

```

```

//     Quiquadrado.

```

```

//*****

```

```

public double[][] CalculeFrequenciasEsperadas(CategoriaPorCategoria[] _catPorCat,
double[][] _tabCorrelacaoQui2) {
    int _linhas = _catPorCat.length;
    int _colunas = _catPorCat[0].Categorias.length;
    double[][] _fEsperada = new double[_linhas][_colunas];
    int _somaTotal = 0;
    double[] _somaColunas = new double[_colunas];
    double[] _somaLinhas = new double[_linhas];
    for (int i = 0; i < _linhas; i++) {
        _somaLinhas[i] += _catPorCat[i].getQtdade();
        _somaTotal += _somaLinhas[i];
        for (int j = 0; j < _colunas; j++) {
            _somaColunas[j] += _catPorCat[i].Categorias[j].getQuantidade();
        }
    }
    for (int i = 0; i < _linhas; i++) {
        for (int j = 0; j < _colunas; j++) {
            _fEsperada[i][j] = _somaColunas[j]*_somaLinhas[i]/_somaTotal;
            double _temp = _catPorCat[i].Categorias[j].getQuantidade() -
            _fEsperada[i][j];
            _tabCorrelacaoQui2[i][j] = (_temp * _temp) / _fEsperada[i][j];
        }
    }
    return _fEsperada;
}

```