

Universidade Federal de Santa Catarina
Departamento de Engenharia de Produção e Sistemas
Programa de Pós-Graduação em Engenharia de Produção

**PROPOSTA METODOLÓGICA PARA CONSTRUÇÃO DE INDICADORES PARA
AVALIAÇÃO DA REDE ESCOLAR BRASILEIRA**

Gretel VILLAMONTE

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de
Produção da Universidade Federal de Santa Catarina como requisito parcial
para obtenção do título de Mestre em Engenharia de Produção

Florianópolis

2001



03386851

**PROPOSTA METODOLÓGICA PARA
CONSTRUÇÃO DE INDICADORES PARA AVALIAÇÃO DA REDE ESCOLAR
BRASILEIRA**

Gretel VILLAMONTE


Esta dissertação foi julgada e aprovada para a obtenção do título de Mestre em Engenharia de Produção no Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina.

Florianópolis, 10 de Abril de 2001




Prof. Ricardo M. Barcia, PhD
Coordenador do Curso


Banca Examinadora



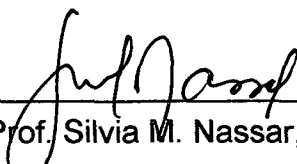
Prof. Jair dos Santos Lapa, PhD.
Orientador



Prof. Carlos M. Batista, Dr.
Membro



Prof. Pedro A. Barbetta, Dr.
Membro



Prof. Silvia M. Nassar, Dra.
Membro

AGRADECIMENTOS

À Universidade Federal de Santa Catarina,
Ao Vice-Reitor da UFSC, Prof. Lúcio José Botelho,
Ao Conselho Nacional de Desenvolvimento
Científico e Tecnológico (CNPq),
Ao orientador, Prof. Jair dos Santos Lapa,
Ao co-orientador, Prof. Eduardo Crivisqui,
Aos colegas do Programme de Recherche et
d'Enserignement en Statistique Appliquée (PRESTA),
Aos professores do INE da UFSC,
Aos meus colegas do Curso de Pós-graduação.

SUMÁRIO

1. INTRODUÇÃO	001
1.1. Antecedentes da pesquisa	001
1.2. O problema da pesquisa	003
1.3. Objetivos da pesquisa	004
1.4. Método e Definições	006
1.5. Estrutura do trabalho	014
2. FUNDAMENTAÇÃO TEÓRICA	015
2.1. Fundamentos e funcionamento do método de Análise de Correspondências Simples	017
2.2. Fundamentos e funcionamento do método de Análise Fatorial de Correspondências Múltiplas (AFCM)	031
2.3. Fundamentos e funcionamento do método de Análise Fatorial Múltipla	043
3. A PROPOSTA METODOLÓGICA	053
3.1. Pré-requisitos	054
3.2. Etapas metodológicas de construção do indicador	055
4. O INDICADOR DE “RECURSOS MATERIAIS” DAS ESCOLAS	063
4.1. Etapas metodológicas de construção do indicador	067
4.2. Construção do indicador de “recursos materiais” das escolas mediante a utilização da Análise Fatorial Múltipla sobre a tabela	

de dados	079
4.3. Validação do indicador de “recursos materiais” das escolas	111
4.4. Uso do indicador e a tipologia associada.	151
5. VALIDAÇÃO DA METODOLOGIA DE ELABORAÇÃO DO INDICADOR	153
5.1. Construção de outros indicadores por formas convencionais.	153
5.2. Comparação dos resultados analisando os elementos de discordância	155
6. CONCLUSÕES E RECOMENDAÇÕES	161
6.1. Conclusões da pesquisa	161
6.2. Recomendações	162
REFERÊNCIAS BIBLIOGRÁFICAS	165
ANEXOS	168

LISTA DE TABELAS

- Tabela 2.1** Distribuição das escolas segundo as condições de funcionamento da biblioteca e do laboratório de informática.
- Tabela 2.2** Perfis da condição de funcionamento dos laboratórios de informática das escolas segundo a condição de funcionamento das bibliotecas.
- Tabela 2.3** Perfis da condição de funcionamento dos bibliotecas das escolas segundo a condição de funcionamento dos laboratórios de informática
- Tabela 2.4** Tabela teórica em condição de independência entre a condição de funcionamento das bibliotecas e dos laboratórios de informática das escolas
- Tabela 4.1** Distribuição de freqüência dos "MISSING VALUES" por escola participante do SAEB97 em todo o País (%)
- Tabela 4.2** Médias do indicador segundo o estado de conservação dos banheiros
- Tabela 4.3** Médias do indicador segundo as condições de funcionamento das quadras de esportes
- Tabela 4.4** Médias do indicador segundo as condições de funcionamento do laboratório de informática
- Tabela 4.5** Distribuição das escolas pelas características do prédio
- Tabela 4.6** Distribuição das escolas pelas características das instalações
- Tabela 4.7** Distribuição das escolas pelas características dos equipamentos
- Tabela 4.8** Distribuição das escolas pelos serviços oferecidos

LISTA DE FIGURAS

Figura 2.1 Esquema da Tabela de Dados Observados

Figura 2.2 Tabela de Contingência

Figura 2.3 Tabela de frequências relativas associada à Tabela de Contingência

Figura 2.4 Decomposição da inércia

Figura 2.5 Coordenadas, contribuições e qualidade da representação dos pontos perfis-coluna

Figura 2.6 Coordenadas, contribuições e qualidade da representação dos pontos perfis-linha

Figura 2.7 Representação gráfica das modalidades das linhas e das colunas no primeiro plano fatorial

Figura 2.8 Tabela de Códigos Condensados TCC (1933 x 6)

Figura 2.9 Tabela TDC (1933x4) correspondente à primeira coluna da tabela TCC (1933x6) estado de conservação dos telhados

Figura 2.10 Tabela Disjuntiva Completa ($n \times K$) correspondente a Tabela de Códigos Condensados ($n \times P$)

Figura 2.11 Variáveis ativas e indivíduos tratados com AFCM

Figura 2.12 Decomposição da inércia

Figura 2.13 Coordenadas, contribuições e cosenos quadrados de algumas escolas (indivíduos).

Figura 2.14 Primeiro plano principal projeção das modalidades.

Figura 2.15 Coordenadas, contribuições e cosenos quadrados das Modalidades

Figura 2.16 Tabela Estatística Múltipla (T^M)

Figura 2.17 Tabela de coordenadas dos indivíduos em R^k

Figura 4.0 Etapas da construção do indicador de recursos materiais

Figura 4.1 Indivíduos incluídos na AFM

Figura 4.2 Variáveis ativas incluídas na AFM

Figura 4.3 Variáveis ilustrativas incluídas na AFM

Figura 4.4 Esquema da tabela de dados

Figura 4.5 Decomposição da inércia na análise global

Figura 4.6 Comparação dos grupos de recursos materiais

Figura 4.7 Coordenadas e ajudas à interpretação dos quatro grupos ativos

Figura 4.8 Projeção das escolas na estrutura conjunta dos grupos de recursos materiais (projeção da nuvem dos centros de gravidade dos indivíduos).

Figura 4.9 Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos

Figura 4.10 Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (situação extrema esquerda)

Figura 4.11 Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (zoom situação extrema esquerda)

Figura 4.12 Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (situação extrema direita)

- Figura 4.13** Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (zoom situação extrema direita)
- Figura 4.14** Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (situação intermédia central)
- Figura 4.15** Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (zoom situação intermédia central)
- Figura 4.16** Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (situação intermédia esquerda)
- Figura 4.17** Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (zoom situação intermédia central)
- Figura 4.18** Projeção das modalidades da variável região (variável ilustrativa) na estrutura conjunta dos grupos de recursos
- Figura 4.19** Projeção das modalidades da variável rede (variável ilustrativa) na estrutura conjunta dos grupos de recursos
- Figura 4.20** Projeção das modalidades da variável localização (variável ilustrativa) na estrutura conjunta dos grupos de recursos
- Figura 4.21** Projeção das modalidades da variável zona (variável ilustrativa) na estrutura conjunta dos grupos de recursos

Figura 4.22 Correlações para os primeiros eixos da AFM, entre as projeções da nuvem de pontos global e as quatro nuvens de pontos associadas a um grupo

Figura 4.23 Representação da situação característica da escola estadual N° 1705 de MT na estrutura conjunta dos recursos

Figura 4.24 Representação da situação característica de MT na estrutura conjunta dos recursos, e das componentes dessa situação segundo os grupos de recursos.

Figura 4.25 Representação da situação característica de SC na estrutura conjunta dos recursos, e das componentes de essa situação segundo os grupos de recursos.

Figura 4.26 Representação dos fatores das análises separadas de cada grupo sobre o primeiro plano da AFM.

Figura 4.27 Índice de semelhança entre as nuvens N^i_j

Figura 4.28 Histograma de frequências do indicador

Figura 4.29 *Boxplot* do indicador para os diferentes estados de conservação dos banheiros

Figura 4.30 *Boxplot* do indicador para as diferentes condições de funcionamento das quadras de esportes

Figura 4.31 *Boxplot* do indicador para as diferentes condições de funcionamento do laboratório de informática

Figura 4.32 Variáveis ativas da ACM do Grupo Prédio

Figura 4.33 Indivíduos estudados na ACM do Grupo Prédio

Figura 4.34 Decomposição da inércia da ACM do Grupo Prédio

Figura 4.35 Projeção das modalidades das variáveis ativas do Grupo Prédio no primeiro plano fatorial

Figura 4.36 Descrição dos eixos (primeiro e segundo) da ACM do Grupo Prédio pelas modalidades ativas

Figura 4.37 Dendrograma das escolas pelas características do prédio

Figura 4.38 Histograma dos índices de nível

Figura 4.39 Projeção das escolas no primeiro plano principal da ACM diferenciando as classes as quais pertencem (prédio).

Figura 4.40 Caracterização das classes pelas modalidades

Figura 4.41 Projeção da tipologia de escolas "condições dos prédios", no espaço das modalidades das variáveis descritivas do estado dos prédios escolares

Figura 4.42 Coordenadas dos centros de classe da partição pelas características do Prédio

Figura 4.43 Distribuição das escolas pela tipologia do prédio.

Figura 4.44 Projeção das escolas no primeiro plano principal da ACM diferenciando as classes as quais pertencem (instalações).

Figura 4.45 Projeção da tipologia de escolas "condições das instalações", no espaço das modalidades das variáveis descritivas das condições das instalações escolares

Figura 4.46 Distribuição das escolas pela tipologia das instalações

Figura 4.47 Projeção das escolas no primeiro plano principal da ACM diferenciando as classes as quais pertencem (equipamentos)

Figura 4.48 Projeção da tipologia de escolas "estado dos equipamentos", no espaço das modalidades das variáveis descritivas do estado dos equipamentos escolares.

Figura 4.49 Distribuição das escolas pela tipologia dos equipamentos.

Figura 4.50 Projeção das escolas no primeiro plano principal da ACM diferenciando as classes as quais pertencem (serviços).

Figura 4.51 Projeção da tipologia de escolas "serviços oferecidos", no espaço das modalidades das variáveis descritivas dos serviços.

Figura 4.52 Distribuição das escolas pela tipologia dos serviços

Figura 4.53 *Boxplot* do indicador segundo as classes da tipologia do estado do Prédio

Figura 4.54 *Boxplot* do indicador segundo as classes da tipologia do estado das Instalações

Figura 4.55 *Boxplot* do indicador segundo as classes da tipologia do estado dos Equipamentos

Figura 4.56 *Boxplot* do indicador segundo as classes da tipologia dos Serviços oferecidos

Figura 4.57 *Boxplot* do indicador segundo a Zona da escola

Figura 4.58 *Boxplot* do indicador segundo a Rede da escola

Figura 4.59 *Boxplot* do indicador segundo a Localização da escola

Figura 4.60 *Boxplot* do indicador segundo a Zona da escola

Figura 5.1 *Boxplot* do indicador alternativo (média simples) segundo as condições de funcionamento das quadras de esportes.

Figura 5.2 *Boxplot* do indicador alternativo (média ponderada) segundo as condições de funcionamento das quadras de esportes.

Figura 5.3 *Boxplot* dos três indicadores para as escolas com boas quadras de esportes.

Figura 5.4 *Boxplot* dos três indicadores para as escolas sem quadras de esportes.

ABREVIações

ACM – Análise de Correspondência Múltiplas

ACP - Análise em Componentes Principais

AFCS - Análise Fatorial de Correspondência Simples

AFM - Análise Fatorial Múltipla

SAEB – Sistema de Avaliação da Educação Básica

TCC – Tabela de Códigos Condensados

TDC – Tabela Disjuntiva Completa

Villamonte, Gretel. PROPOSTA METODOLÓGICA PARA CONSTRUÇÃO DE INDICADORES PARA AVALIAÇÃO DA REDE ESCOLAR BRASILEIRA. Florianópolis, 2001. 198f. Dissertação (mestrado em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção, UFSC, 2001.

RESUMO

Esta pesquisa tem um carácter metodológico ao se propor a seleccionar, ordenar e sistematizar um conjunto de técnicas e procedimentos estatísticos visando construir uma proposta metodológica de instrumentos para a avaliação da rede escolar brasileira usando os dados levantados pelo SAEB.

No primeiro capítulo, discorre sobre as características históricas e operacionais do Sistema de Avaliação do Ensino Básico – SAEB- ressaltando os seus resultados, alcances e limites. Ao relacionar as possíveis utilizações do SAEB como um instrumento de avaliação da rede escolar brasileira apresenta o seguinte problema de pesquisa “Como resumir todas as informações de variáveis não métricas numa variável contínua e única que –dando conta de todas as características iniciais- as ponderem objetivamente...?”. Apresenta dois objetivos básicos o primeiro trata do desenvolvimento de uma metodologia de elaboração de indicadores para avaliação da rede escolar brasileira. Tal metodologia deve garantir um conjunto de propriedades para o indicador: pertinência em relação à finalidade para a qual foi produzido, capacidade descritiva, relação com as diferentes formas de intervenção, comparabilidade no tempo e no espaço, e que seja uma variável única e contínua. O segundo objetivo ao operacionalizar a metodologia proposta visa construir e validar um indicador de recursos materiais das escolas a partir dos dados do SAEB97.

O segundo capítulo apresenta os fundamentos teóricos dos diferentes métodos estatísticos multivariados necessários para construir e validar o indicador. Para o desenvolvimento das técnicas estatísticas selecionou-se uma base de dados observados e disponibilizados pelo SAEB97. Ao realizar o processamento desses dados empíricos ao lado de exemplos teóricos discorreu-se sobre Análise Fatorial de Correspondência Simples - AFCS, Análise de Correspondência Múltipla - ACM e Análise Fatorial Múltipla - AFM destacando os usos e atributos de cada técnica.

O terceiro capítulo apresenta a metodologia a ser adotada na construção do indicador. É neste capítulo que propõe-se o ordenamento e a sistematização dos procedimentos operacionais da metodologia a luz da fundamentação teórica e dos dados observados.

No quarto capítulo sistematiza-se os dados observados a luz da fundamentação teórica e dos procedimentos metodológicos. Do conjunto de análises empreendidas ressalta-se o uso da técnica estatística AFM o que resultou na efetiva elaboração do indicador que define a condição das escolas brasileiras segundo as condições materiais das mesmas.

Finalmente nas conclusões da pesquisa argumenta-se que o indicador de recursos materiais exprime as condições das escolas brasileiras e apresenta as propriedades requisitadas viabilizando uma metodologia exeqüível ao apresentar um indicador que resume as informações não métricas numa variável contínua e única que as pondera objetivamente.

Palavras-Chave Avaliação; Metodologia; Ensino Básico; Gestão da Educação; Indicador; Análise Fatorial Múltipla.

ABSTRACT

This research developed a methodology for elaborating indicators to be used in the evaluation of Brazilian elementary education system. As an application, it builds up and validates an indicator of the material resources of schools using the data collected by the SAEB97. (Basic Education Evaluation System).

This indicator synthesizes a set of observed non-metric characteristics in a single continuous variable that combines all the original characteristics assigning objective weights to them.

The pertinency of the statistical methods employed provides the indicator with properties that assure its intern and extern validity.

Key-words: Evaluation, Methodology, Elementary Education, Education Management, Indicator, Multiple Factorial Analysis.

CAPITULO I

1. INTRODUÇÃO

Sumário

1.1. Antecedentes da pesquisa	001
1.2. O problema da pesquisa	003
1.3. Objetivos da pesquisa	004
1.4. Método e Definições	006
1.4.1 Definições usadas no contexto das análises fatorias	008
1.4.2 Definições das categorias metodológicas	011
1.5. Estrutura do trabalho	014

1.1. Antecedentes da pesquisa

A avaliação da rede escolar é fundamental para fornecer informações que fundamentem a implementação de políticas educacionais. E, para conseguir detectar os efeitos das políticas adotadas, é necessário que essa avaliação seja contínua; portanto um sistema de avaliação deve obter informações periódicas e comparáveis dos diferentes aspectos do sistema educacional. A avaliação na área educacional no Brasil até pouco tempo atrás estava restrita à aprendizagem e desempenho dos alunos. Mas, foi abrangendo outros aspectos relativos a projetos, programas, professores, currículos, instituições e sistemas.

Em 1990 foi implantado o SAEB (Sistema Nacional de Avaliação da Educação Básica), sendo ele responsável pela avaliação do desempenho do sistema educacional. O SAEB foi concebido com o objetivo de fornecer elementos para apoiar a formulação, a reformulação e o monitoramento das políticas voltadas para a melhoria da qualidade da educação no Brasil. É coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais e conta com a participação e apoio das secretarias estaduais e municipais de educação das 27 unidades da federação.

Os levantamentos de dados do SAEB são realizados a cada dois anos em uma amostra probabilística representativa dos 26 estados brasileiros e do distrito federal. A cada aplicação são pesquisados aproximadamente 700 municípios, 3000 escolas públicas e privadas, 25000 professores, 3000 diretores e 220.000 alunos do ensino básico (da 4ª e 8ª séries do ensino fundamental e da 3ª série do ensino médio) nas disciplinas língua portuguesa, matemática e ciências. São aplicados testes de desempenho dos alunos, elaborados com base em matrizes curriculares validadas nacionalmente. Esses testes medem tanto os conteúdos quanto as habilidades (em seus diversos níveis de complexidade) das diversas séries e disciplinas avaliadas. São verificados outros aspectos contextuais e escolares que incidem na qualidade do ensino básico, tais como:

- As características da infra-estrutura e de disponibilidades da unidade escolar (questionário da escola);
- Perfil do diretor e os mecanismos da gestão escolar (questionário do diretor);
- Perfil e a prática docente (questionário do professor);
- As características socioculturais e os hábitos de estudo dos alunos (questionário do aluno).

Os resultados são divulgados tanto em termos gerais (nível nacional) quanto desagregados por região, estado, localização (capital e interior), zona geográfica (urbana e rural), dependência administrativa (estadual, municipal, federal e particular). Os resultados englobam aspectos relacionados ao desempenho dos alunos quanto aspectos referentes às variáveis da escola, do diretor, do professor e dos hábitos de estudo e condições socioculturais dos alunos.

Esses resultados são apresentados de forma descritiva (porcentual, média, etc.) e, também, com algumas associações simples (associação do desempenho dos alunos com a formação do diretor, do professor, participação dos professores em cursos de capacitação, etc.).

As informações produzidas são utilizadas por gestores e administradores de educação, pesquisadores e professores.

As bases de dados do SAEB oferecem os recursos necessários para explorar essa informação com fins de diagnóstico da situação mas também para acompanhar o desempenho da rede. Nesse contexto, o presente trabalho apresenta uma estratégia de tratamento dessas informações com o objetivo de produzir indicadores que facilitem a avaliação do desempenho do sistema de ensino brasileiro.

1.2. O problema da pesquisa

Quando se defronta com a tarefa de construir um indicador a partir de dados do tipo qualitativo tais como os levantados mediante os questionários do SAEB, surge a seguinte pergunta:

Como resumir todas as informações de variáveis não métricas numa variável contínua e única que - dando conta de todas as características iniciais - as ponderem objetivamente?

Por exemplo, se o interesse é construir um indicador de recursos físicos das escolas, o problema é como produzir um indicador que resuma todas as características de infra-estrutura, equipamentos e materiais disponíveis nas escolas (levantados com o questionário da escola apresentado no *Anexo 1*) que geram dados de tipo qualitativo, numa variável sintética que as represente adequadamente?

Para que as informações do SAEB possam subsidiar o processo decisório quanto a ações relacionadas a gestão da rede escolar é necessário produzir sínteses dessa informação.

A estratégia de análise dos dados está destinada à elaboração de conceitos necessários ao diagnóstico a partir do fenômeno observado e, no escopo do presente trabalho, tenta responder a pergunta: como se elaboram os conceitos que servem de fundamento ao diagnóstico objetivo do sistema educativo do ensino básico?

Esta estratégia se ilustrará com a construção do conceito “recursos materiais” do sistema de ensino básico, sintetizando as características dos prédios, instalações, equipamentos e serviços das escolas (37 variáveis observadas) levantadas pelo o SAEB, num indicador que reduz as dimensões do observado sem arbitrariedade.

A utilidade que pretende-se dar a este instrumental de análise é fornecer suporte ao processo decisório em relação a gestão da rede e subsidiar a formulação de políticas de intervenção segundo um determinado ponto de vista gerencial. Pode-se assim contribuir com as discussões referentes a alocação de recursos financeiros para a melhoria da infra-estrutura da rede apontando as prioridades de investimento e verificar o impacto desta intervenção.

1.3. Objetivos da pesquisa

Esta pesquisa foi organizada em torno de dois objetivos centrais:

- 1) Desenvolver uma metodologia de elaboração de indicadores para a avaliação da rede escolar brasileira que assegure-lhe um conjunto de propriedades que permita sua utilização como uma medida de avaliação do ensino básico brasileiro.
- 2) Construir e validar um indicador de recursos físicos das escolas a partir dos dados do SAEB97.

Com respeito ao primeiro objetivo o caminho metodológico a ser percorrido deve estar respaldado em uma fundamentação teórica, em um conjunto de dados observados e no uso de técnicas estatísticas, os quais são ordenados e sistematizados de maneira tal que permitam garantir ao indicador construído as seguintes propriedades: *pertinência em relação à finalidade para o qual foi produzido, capacidade descritiva, que esteja relacionado com as diferentes formas de intervenção, que permita a comparabilidade no tempo e no espaço e que seja uma variável única e contínua.*

O segundo objetivo tem como finalidade a ilustração da metodologia através da construção de um indicador de recursos materiais das escolas, contudo a

seqüência metodológica pode ser aplicada à construção de outros indicadores como por exemplo de recursos humanos das escolas.

A qualidade de um indicador está relacionada aos procedimentos metodológicos utilizados em sua construção, que devem estar teoricamente fundamentados e descritos de tal forma que permita sua replicabilidade e lhe confirmem validade interna e externa. As propriedades aludidas anteriormente podem assim serem conceituadas para o segundo objetivo:

- **Pertinência com respeito à finalidade para a qual foi produzido.** Assim, deve dar conta de todas as características físicas de infra-estrutura e disponibilidades da escola outorgando a cada uma delas ponderadores objetivos de importância.
- **Capacidade descritiva:** deve refletir as diferenças em equipamento das escolas, agrupando aquelas com características semelhantes em valores semelhantes.
- **Que esteja relacionado com as diferentes formas de intervenção que permitem melhorar a realidade representada globalmente pelo indicador, ou seja, deve permitir saber que ações definir para melhorar a infra-estrutura das diferentes escolas.**
- **Que permita a comparabilidade no tempo e no espaço, isto é, deve permitir comparar as condições de infra-estrutura em períodos diferentes (duas aplicações do SAEB), e entre duas áreas geográficas (regiões ou estados). Por exemplo, deve permitir - do ponto de vista do equipamento das escolas - comparar a situação de um estado com respeito a todos os estados do Brasil.**
- **Que seja uma variável única e contínua, associada a uma tipologia das condições observadas nas unidades de estudo.**

1.4. Método e definições

A estratégia metodológica proposta no presente trabalho para o cálculo de indicadores a partir dos dados do SAEB supõe:

- a disposição das bases de dados do SAEB97 para ilustrar a metodologia proposta através da construção de um indicador.
- a utilização de um conjunto de métodos estatísticos multidimensionais.
- o uso de um software que forneça as análises propostas sobre os dados disponíveis.

A base de dados utilizada é o arquivo correspondente às informações levantadas mediante o questionário das escolas (*Anexo 1*) na aplicação de 1997 do SAEB realizada pelo MEC. Ela foi disponibilizada por meio de um CD que contém todas as bases de dados – das escolas, dos diretores, dos professores e dos alunos- assim como também um software de análise chamado ANALISASAEB. A descrição da base das escolas encontra-se no capítulo 4 - O indicador de “recursos materiais” das escolas.

O software utilizado para processar as análises multivariadas foi o SPAD para WINDOWS.

Os métodos estatísticos multidimensionais usados foram os métodos fatoriais cujos fundamentos estão expostos no capítulo 2 - *Fundamentação teórica*. O princípio destes métodos é único, se constroem e representam sobre gráficos as duas nuvens de pontos correspondentes as linhas e as colunas da tabela estudada. As representações das linhas e das colunas se relacionam fortemente entre si (Escofier-Pagès, 1992, p.1). O objetivo dos métodos fatoriais é fornecer representações sintéticas de tabelas de grande tamanho, em geral em forma de visualizações gráficas (Lebart, Morineau, Piron, 1997, p.13-14). Para isso, procura-se reduzir as dimensões da tabela representando as associações entre indivíduos e entre variáveis num espaço de menor dimensão. Atualmente é possível calcular as distâncias entre linhas e colunas de uma tabela retangular de valores numéricos não sendo possível visualizar essas distâncias de forma imediata uma vez que as representações geométricas associadas implicam em

geral representações em espaços multidimensionais; sendo necessário proceder a transformações e aproximações para obter uma representação plana. As análises fatoriais procuram então sub-espacos de menores dimensões (dois ou três, por exemplo) que ajustem melhor à nuvem de pontos-indivíduos e a nuvem de pontos-variáveis, de forma que as proximidades medidas nesses sub-espacos reflitam tanto quanto possível as proximidades reais. Assim se obtém o espaço de representação chamado espaço fatorial.

A geometria da nuvem de pontos e os cálculos de proximidade ou de distâncias diferem de acordo com a natureza das linhas e colunas da tabela analisada. As colunas podem ser variáveis contínuas ou nominais ou categorias no caso de uma tabela de contingência. As linhas podem ser indivíduos ou categorias.

A natureza das informações, sua codificação, as especificidades do domínio de aplicação, vão introduzir variantes nos métodos fatoriais a serem utilizados. As técnicas fundamentais são:

- A Análise em Componentes Principais (ACP) aplica-se a uma tabela de tipo "indivíduos x variáveis" , onde as colunas são variáveis de valores numéricos contínuos e as linhas são indivíduos, observações, objetos. As proximidades de variáveis se interpretam em termos de correlação, as proximidades entre indivíduos em termos de similitudes globais de valores observados.
- A Análise de Correspondências Simples (AFCS) aplica-se a tabelas de contingência, isto é, a tabelas de freqüências obtidas pelo cruzamento de duas variáveis qualitativas nominais. Estas tabelas tem a peculiaridade que as linhas e as colunas desempenham o mesmo papel. A análise fornece representações das associações entre linhas e colunas, baseadas numa distância entre perfis denominada distância de qui-quadrado.
- A Análise de Correspondências Múltiplas (ACM) que representa uma extensão do domínio de aplicação da AFCS com procedimentos de cálculo e regras de interpretação específicas. É adaptada à descrição de grandes tabelas de dados qualitativos nominais. As linhas dessas tabelas são em

geral indivíduos ou observações e as colunas são modalidades de variáveis

- A Análise Fatorial Múltipla (AFM) que se aplica a uma tabela estruturada em temas que compreendem vários grupos de variáveis. As linhas dessas tabelas são indivíduos ou observações e as colunas variáveis quantitativas ou qualitativas agrupadas em temas.

As análises fatoriais são utilizados freqüentemente em conjunto com os métodos de classificação.

1.4.1. Definições usadas no contexto das análises fatoriais

A seguir, apresentam-se - em ordem alfabética- um conjunto de definições usadas no contexto das análises fatoriais (Escofier- Pagès, 1992, p. 279-285):

Ativo: elemento que entra no cálculo dos fatores, em contraposição a elemento suplementar ou ilustrativo.

Baricentro ou centro de gravidade: é o equivalente geométrico ao conceito estatístico de média.

Classificação: o objetivo geral dos métodos de classificação é obter partições de um conjunto de elementos (linhas ou colunas de uma tabela de dados) que agrupem numa mesma classe os elementos que apresentam semelhanças. Os métodos utilizados com maior freqüência se baseiam nas mesmas representações geométricas que os métodos fatoriais (nuvens definidas como em ACP, AFCS, ACM segundo o tipo de tabela). Os algoritmos tendem a obter as partições que maximizam o quociente entre inércia inter-classe e a inércia total (e portanto minimizam o quociente entre a inércia intra-classe e a inércia total). Geralmente o máximo global é inacessível. Frequentemente se trabalha com a tabela dos primeiros fatores, com o qual se considera a projeção da nuvem sobre os primeiros eixos de inércia.

Componente principal: eixo de inércia de uma nuvem de variáveis em ACP.

Contribuição à inércia:

- contribuição de um elemento à inércia de uma nuvem = quociente entre a inércia desse elemento e a inércia da nuvem.
- contribuição de um elemento à inércia de um eixo ou fator = contribuição desse elemento à inércia da nuvem projetada sobre o eixo. Os pontos que mais contribuem à inércia podem determinar o fator.

Coordenada de um elemento sobre um eixo ou coordenada fatorial: norma da projeção ortogonal do elemento sobre o eixo.

Correlação: coeficiente de correlação linear, em ACP as coordenadas das projeções das variáveis sobre os eixos são as correlações com as componentes principais.

Eixo de inércia ou eixo fatorial: nas análises fatoriais as nuvens são projetadas sobre os eixos, esses eixos são os eixos de inércia de uma nuvem: são ortogonais entre si e a inércia da nuvem projetada é máxima sobre o primeiro eixo, depois sobre o segundo e assim sucessivamente; são vetores próprios de uma matriz. Eixo principal de inércia de uma nuvem = eixo de inércia calculado tomando como origem o baricentro.

Fator: conjunto de coordenadas das projeções de uma nuvem de pontos sobre um eixo dessa nuvem.

Inércia: a noção mecânica de inércia de uma nuvem de pontos com respeito a seu baricentro corresponde-se com a noção estatística de variância.

- Inércia de um elemento M de peso p em relação a um ponto O = produto do peso p pelo quadrado da distância entre M e O.

- Inércia de uma nuvem de pontos = soma das inércias dos elementos que a compõem.
- Inércia de um elemento sobre um eixo = inércia da projeção do elemento sobre o eixo.
- Inércia de um eixo ou de um fator = inércia da nuvem projetada sobre o eixo.

Nuvem de pontos = conjunto de pontos munidos de pesos num espaço euclidiano.

Qualidade de representação (cosenos quadrados)

- de um elemento sobre um eixo = quociente entre a inércia do elemento projetado sobre o eixo e a inércia do elemento no espaço (ou inércia total) = quadrado do cosseno do ângulo formado entre os dois vetores que unem a origem com esse ponto e com sua projeção respectivamente.
- Da nuvem sobre o eixo = percentagem de inércia extraída pelo eixo.

Representação simultânea ou representação superposta: das linhas e das colunas de uma tabela em ACP, em AFCS; dos indivíduos caracterizados por diferentes grupos de variáveis em AFM.

Suplementar: um elemento suplementar ou ilustrativo “de peso nulo” se projeta sobre os eixos de inércia da nuvem sem intervir no cálculo dos eixos.

Transição (relações ou fórmulas de transição) = relações entre os fatores obtidos das linhas e os fatores obtidos das colunas.

Valor próprio: em análise fatorial denomina-se valor próprio à inércia de um eixo (ou de um fator); histograma ou diagrama de valores próprios = representação gráfica do decrescimento das inércias dos fatores sucessivos.

1.4.2. Definições das categorias metodológicas

A seguir, apresentam-se as definições das categorias metodológicas.

O conceito de validade

O conceito de validade pode ser analisado segundo dois níveis de profundidade: o primeiro considera o problema de natureza epistemológica e filosófica da qual a mesma noção é originada (problema relacionado à origem do conceito) e o segundo considera o aspecto de carácter operativo conhecido como aplicação empírica do conceito (Marradi, 1990, p. 55-96).

A dupla dimensão do conceito de validade permite distinguir dois níveis de análise: o primeiro refere-se ao exame da validade voltado a verificar se a dimensão do conceito em análise é adequadamente representada pelos indicadores usados; enquanto que o segundo aspecto considera a validade da construção teórica e consiste em comprovar se esta conseguiu propor hipóteses que sejam verificáveis empiricamente.

O primeiro aspecto reporta-se a nível da metodologia da ciência já que tem a ver com a formulação teórica do problema a ser analisado. O segundo aspecto, derivado do primeiro, interessa à metodologia da pesquisa ao envolver técnicas e instrumentos para verificar as hipóteses. O presente trabalho encontra-se delimitado pelo segundo aspecto, pois discute instrumentos de análise a serem utilizados pela metodologia da pesquisa.

A verificação empírica de hipóteses requer a realização de uma série de operações. Uma vez feita a definição nominal do conceito, segue a definição operativa, a qual é o processo de adoção da significação de um termo mediante a especificação das operações necessárias para provar a presença das propriedades as quais o termo se refere. A definição operativa deverá indicar o termo empírico baseado no qual o conceito operativo é observado efetivamente, o que torna possível a revelação do conceito.

Esse aspecto corresponde a um componente do conceito de validade que é a *validade interna*. Ela tende a verificar se cada elemento da definição operativa

está compreendido efetivamente, ou seja se sua definição tem condições de refletir o conceito da forma como ele foi definido nominalmente.

O outro componente da validade é a *validade externa*; com ela se tende a verificar que a definição operativa esteja em condições de definir as relações significativas entre o referente empírico e o conceito em estudo.

Validade interna e externa são dois aspectos intrínsecos da noção de validade. A primeira releva da correção da conceitualização e por isso é necessário examinar que exista uma relação lógica entre a definição operativa e a área de conteúdo da definição nominal. A validação externa consiste em individualizar as técnicas de observação e análise apropriadas ao conceito em estudo e por isso é necessário validar a capacidade da definição operativa de dar resultados empíricos válidos.

Só é possível chegar ao conhecimento de um fenômeno se ele é observável através de manifestações externas, sejam elas indiretas ou diretas. Se assim for, é possível conseguir pôr em evidência suas características e regularidades. Para isso é necessário aprimorar o critério de observação de forma a registrar o fenômeno da forma mais próxima possível da realidade. Quer dizer, que para que um fenômeno seja observado e analisado é necessário um instrumento apropriado, com condições de fornecer uma exata dimensão do fenômeno. A definição usada da validade no sentido proposto, tem a ver com a correspondência entre o instrumento de medida e o fenômeno em análise, e com o fato de que um instrumento é válido quando mede realmente aquilo que se quer medir.

Convalidar um instrumento de medida significa verificar a sua capacidade descritiva e explicativa, no confronto do conceito ou da teoria em análise. BOWERS (1964: p.742) observa que, em diferentes níveis, através da validade pode-se controlar a correção da conceitualização, a aplicabilidade de particulares técnicas de pesquisa e a pertinência do dado em relação ao conceito que se quer mensurar.

A validade da mensuração refere-se concretamente à possibilidade de apresentar a medição de um conceito através da construção de um instrumento que cumpra com essa tarefa.

Como validação explicativa considera-se o resultado da mensuração em função da interpretação que ela sugere do fenômeno em análise.

O conceito de indicador

Em primeiro lugar temos os índices estatísticos que são simplesmente números dos quais foi tirado (corrigido) o efeito da unidade de medida ou o fator de tamanho, através do cálculo de razões entre um fenômeno dado e um segundo fenômeno do qual o primeiro faz parte ou com o qual tem uma relação lógica, ou com o qual adquire um significado particular (Abruzzini-Quinti, 1997, p.21-25).

Por exemplo:

$$\text{índice de natalidade} = \frac{\text{número total de nascidos vivos num período dado}}{\text{população total}}$$

Se considerarmos que um indicador é usualmente um índice estatístico com uma atribuição de significado suplementar dependente do contexto; então um índice sintético corresponde ao esforço de resumir um sistema de relações entre variáveis (índices simples ou indicadores) num "sinal" que possa ser associado a cada unidade de estudo e que seja indicativo da forma com a qual esse sistema de relações manifesta-se em cada unidade. Esse "sinal" pode ser um rótulo, uma ordenação ou uma entidade quantitativa.

Segundo CANNAVÓ(1995, p.10) um indicador é uma construção conceitual, que utiliza fenômenos ou comportamentos observáveis, os quais podem ser postos em evidência, aprendidos indiretamente mediante simbologias, medidas padronizadas singulares ou compostas; que da ausência, presença ou intensidade de tais fenômenos ou comportamentos, deduz a dimensão de um conceito.

1.5. Estrutura do trabalho

O desenvolvimento do presente trabalho foi estruturado em cinco capítulos cujos objetivos são os seguintes:

- Expor uma metodologia de construção de indicadores para a avaliação da rede escolar brasileira.

No Capítulo 2 *Fundamentação teórica* apresentam-se os métodos estatísticos exploratórios multivariados que serão utilizados na construção do indicador e se discute a pertinência deles no contexto da avaliação da rede escolar brasileira.

No Capítulo 3 - *Proposta metodológica*- expõem-se as principais etapas da estratégia de construção de indicadores para a avaliação da rede escolar brasileira.

- Ilustrar a aplicação dessa metodologia construindo um indicador de recursos materiais das escolas.

No Capítulo 4 - *O indicador de recursos materiais*- aplica-se a estratégia metodológica de construção de indicadores apresentada no Capítulo 3.-, à tabela de dados obtidas do levantamento feito pelo SAEB97 mediante o questionário das escolas. para construir um indicador de recursos físicos das escolas.

- Validar o indicador de recursos materiais das escolas.

No Capítulo 4 se valida o indicador obtido interna e externamente e se discutem as limitações do indicador.

- Validar a metodologia utilizada.

No Capítulo 5 se compara o indicador obtido com outros indicadores gerados de forma convencional. .-

- Apresentar as conclusões da pesquisa

No Capítulo 6 se expõem as conclusões da pesquisa e as recomendações.-

CAPITULO II

2. FUNDAMENTAÇÃO TEÓRICA

Sumário

2.1. Fundamentos e funcionamento do método de Análise de Correspondências Simples	017
2.2. Fundamentos e funcionamento do método de Análise Fatorial de Correspondências Múltiplas (AFCM)	031
2.3. Fundamentos e funcionamento do método de Análise Fatorial Múltipla	043
2.3.1. Os dados e a tabela estatística múltipla	044
2.3.2. Ponderação dos grupos de variáveis	045
2.3.3. A AFM em R^K os indivíduos	046
2.3.4. A AFM em R^I as variáveis	049
2.3.5. Fatores comuns aos grupos de variáveis	050
2.3.6. O caso das variáveis qualitativas e das tabelas mistas	051
2.3.7. Ponderação da Tabela Disjuntiva Completa para aplicar a ACP	052

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo se apresentam os fundamentos dos diferentes métodos estatísticos multivariados necessários para construir e validar um indicador; estando esta apresentação restrita a mostrar a adequação dos métodos à resolução do problema de pesquisa. Essa fundamentação desenvolve-se segundo duas lógicas: a primeira apoia-se numa revisão da literatura; a segunda utiliza um conjunto de dados observados -disponibilizados pelo SAEB97- que foram ordenados e sistematizados para gerar exemplos de aplicação da fundamentação

teórica como intuito de dar maior acessibilidade ao conteúdo exposto neste capítulo.

A primeira consideração a ser feita é o que significa construir um indicador quantitativo partindo da observação de p variáveis sobre n unidades de observação. O significado atribuído neste trabalho é ordenar os indivíduos e comparar as suas diferenças segundo a métrica de uma variável única.

Assim, a variabilidade dos indivíduos no indicador é resultante da variabilidade dos mesmos em cada uma das p variáveis observadas.

Nesse contexto estudar a multivariabilidade implica estudar a configuração da nuvem de pontos-indivíduos e a sua inércia num espaço p -dimensional. Isso leva a uma decomposição ortogonal da inércia em eixos de máxima inércia. O primeiro desses eixos - por representar a máxima inércia projetada - pode ser a variável única e quantitativa (o indicador procurado). Essa é a estratégia "clássica" baseada no método de análise em componentes principais.

Se as variáveis observadas não são métricas, não é possível obter uma representação direta da nuvem de pontos num espaço dotado de uma distância euclidiana. Mas, uma transformação adequada da tabela de dados possibilita a representação da dispersão da nuvem de indivíduos num espaço euclidiano, dotado de uma métrica particular: a distância do qui-quadrado, cujas dimensões resultam dos K atributos que apresentam as p variáveis qualitativas consideradas. Essa é a estratégia baseada na análise fatorial de correspondências múltiplas. Estas são as condições segundo as quais se apresentam os dados do SAEB.

Para apresentar a análise fatorial de correspondências múltiplas (ACM) faz-se necessário expor sinteticamente a métrica do qui-quadrado e o método de análise fatorial de correspondências simples (AFCS); porque a ACM não é mais do que uma generalização ao caso p -variado da AFCS.

Considerando que os objetivos deste trabalho não se restringem a criar um indicador que resuma a variabilidade resultante das p variáveis qualitativas, e sim, qualificar um indicador para que as características observadas sobre as escolas reflitam os diferentes aspectos dos recursos materiais das mesmas (características do prédio, das instalações, dos equipamentos e dos serviços), a

ACM não supre esta condição; isto porque ela não leva em consideração o agrupamento em temas das p -variáveis em estudo. Nesse caso, pode não ser conveniente, por razões que serão discutidas posteriormente, tratar esse conjunto de variáveis como se fossem fontes homogêneas de variabilidade.

O método de análise fatorial múltipla constitui uma resposta metodológica ao problema de reduzir a uma dimensão a variabilidade conjunta de p variáveis organizadas em J grupos temáticos.

Veremos que esse problema é tratado por esse método tanto no caso em que os J grupos temáticos estão compostos por variáveis quantitativas, como no caso dos dados do SAEB, no qual esses grupos estão compostos por variáveis qualitativas.

2.1. Fundamentos e funcionamento do método de Análise de Correspondências Simples

A Análise de Correspondências Simples foi concebido para estudar as tabelas de contingência. As tabelas de contingência são elaboradas a partir da observação de duas características qualitativas num conjunto de n unidades de observação.

Sejam duas variáveis qualitativas, por exemplo as condições de funcionamento da biblioteca de uma escola, e as condições de funcionamento do laboratório de informática de uma escola. As condições de funcionamento de qualquer uma das instalações anteriores pode ser boa, regular, inadequada e inexistente, ou seja cada uma das variáveis tem quatro modalidades.

Os dados resultantes da observação dessas características nas 1933 escolas estudadas no SAEB97 produz uma tabela (ver Figura 2.1) do tipo "indivíduos *variáveis qualitativas" com 1933 linhas e 2 colunas.

Figura 2.1. Esquema da tabela de dados observados

Escola	Biblioteca	Laboratório de informática
1	regular	boa
2	inadequada	inexistente
...
...
...
i	x_i	y_i
...
...
...
...
n = 1933	x_n	y_n

Analisar a informação dessa tabela, implica comparar todas as escolas entre si, considerando as semelhanças nas condições de funcionamento da biblioteca e do laboratório de informática.

Sendo que os dados resultantes da observação são códigos; a tabela de dados não tem propriedades numéricas já que a soma em linhas e a soma em colunas não fazem sentido (por tratar-se de códigos). Portanto, a comparação de pares de escolas conduz a contar o número de vezes que aparecem ocorrências similares nos atributos observados, construindo a tabela de contingência.

A tabela de contingência então é uma tabela de frequências obtida cruzando as modalidades de duas variáveis qualitativas definidas sobre a mesma população de indivíduos.

Conhecendo para cada escola a condição de funcionamento da biblioteca e do laboratório de informática a partir da tabela de dados, construí-se a tabela de contingência (ver Tabela 2.1).

Tabela 2.1. Distribuição das escolas segundo as condições de funcionamento da biblioteca e do laboratório de informática.

Variáveis		Laboratório de informática				Total
Modalidades	Bom	Regular	Inadequado	Inexistente		
Biblioteca	Bom	289	23	7	276	595
	Regular	100	31	5	359	495
	Inadequado	15	13	8	247	283
	Inexistente	17	3	0	540	560
	Total	421	70	20	1422	1933

Na Tabela 2.1, perde-se a informação individual das escolas mas se resumem todas as co-ocorrências similares dos dois atributos observados.

Uma tabela de contingência expressa a relação entre duas variáveis qualitativas. No caso estudado tem-se a relação entre as condições de funcionamento da biblioteca e as condições de funcionamento do laboratório de informática das escolas.

Para estudar essa relação, considera-se a tabela de frequências relativas, obtida dividindo cada célula da tabela de contingência entre o total de escolas estudadas.

Figura 2.2. Tabela de contingência (Crivisqui, 1993, p.14)

		Variável J							
		1	2	...	j		J
Variável K	1	n_{11}	n_{12}	...	n_{1j}	n_{1J}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2j}	n_{2J}	$n_{2.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	k	n_{k1}	n_{k2}	...	n_{kj}	n_{kJ}	$n_{k.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	K	n_{K1}	n_{K2}	...	n_{Kj}	n_{KJ}	$n_{K.}$
		$n_{.1}$	$n_{.2}$...	$n_{.j}$	$n_{.J}$	$n_{..}$

Figura 2.3. Tabela de freqüências relativas associada à tabela de contingência

		Variável J							
		1	2	...	j		J
Variável K	1	f_{11}	f_{12}	...	f_{1j}	f_{1J}	$f_{1.}$
	2	f_{21}	f_{22}	...	f_{2j}	f_{2J}	$f_{2.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	k	f_{k1}	f_{k2}	...	f_{kj}	f_{kJ}	$f_{k.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	K	f_{K1}	f_{K2}	...	f_{Kj}	f_{KJ}	$f_{K.}$
		$f_{.1}$	$f_{.2}$...	$f_{.j}$	$f_{.J}$	1

Classicamente diz-se que existe independência quando para todo k e para todo j, cumpre-se a igualdade: $f_{kj} = f_{k.} f_{.j}$

Existe relação entre duas variáveis quando em algumas células da tabela, f_{kj} é diferente ao produto ($f_{k.} \times f_{.j}$). Se f_{kj} é superior a esse produto, diz-se que as modalidades k e j atraem-se. Pelo contrário, se f_{kj} é inferior a esse produto, diz-se que existe repulsão entre essas duas modalidades (Escofier-Pagès, 1992, p.27-29).

A independência expressa-se também considerando a tabela como um conjunto de linhas. A igualdade anterior é equivalente à seguinte: $\frac{f_{kj}}{f_{k.}} = f_{.j}$

A quantidade $f_{.j}$ representa a percentagem da população total que possui a modalidade j , enquanto que ($f_{kj} / f_{k.}$) representa essa mesma percentagem na subpopulação que possui a modalidade k .

Quando existe independência, as subpopulações caracterizadas pelas k modalidades da primeira variável, se repartem segundo as J modalidades da segunda variável nas mesmas percentagens. Todas as linhas são, então, proporcionais. O recíproco também é válido: quando todas as linhas são proporcionais, são proporcionais à marginal $f_{.j}$ e as duas variáveis são independentes. Existe portanto relação sempre que as linhas não sejam todas proporcionais à margem e quando não sejam idênticas desde o ponto de vista de sua associação com o conjunto de colunas. Numa tabela de contingência as linhas e as colunas têm um papel simétrico. A independência expressa-se da mesma forma sobre o conjunto de colunas.

Nesta análise das linhas, a tabela estudada é uma tabela que se calcula dividindo cada célula f_{kj} da linha k pela marginal $f_{k.}$ dessa linha. Essa nova linha é denominada perfil linha. Essa transformação resulta na Tabela 2.2, que é calculada para estudar as relações entre duas variáveis por meio do desvio entre as percentagens em linha.

No caso, permite ver como se distribui a condição de funcionamento das bibliotecas nas escolas que tem um laboratório de informática em estado bom, regular, inadequado ou inexistente.

Tabela 2.2. Perfis da condição de funcionamento dos laboratórios de informática das escolas segundo a condição de funcionamento das bibliotecas.

Variáveis		Laboratório de Informática				Total
Modalidades	Bom	Regular	Inadequado	Inexistente		
Biblioteca	Bom	48,6%	3,9%	1,2%	46,4%	100,0%
	Regular	20,2%	6,3%	1,0%	72,5%	100,0%
	Inadequado	5,3%	4,6%	2,8%	87,3%	100,0%
	Inexistente	3,0%	,5%		96,4%	100,0%
	Total	21,8%	3,6%	1,0%	73,6%	100,0%

Se as condições de funcionamento das bibliotecas fossem independentes das condições de funcionamento dos laboratórios de informática, as linhas desta tabela deveriam ser todas semelhantes entre si e iguais à margem. Como pode-se verificar na Tabela 2.2, a condição de funcionamento dos laboratórios não é semelhante entre as escolas que apresentam boas bibliotecas e aquelas que tem bibliotecas regulares, etc. Dentre as escolas que tem boas bibliotecas, as escolas com laboratórios em boas condições são, proporcionalmente, superiores em número e as que não tem laboratórios são inferiores ao total de escolas. No entanto, dentre as escolas que tem bibliotecas inadequadas, proporcionalmente temos mais escolas sem laboratórios e menos com laboratórios bons quando comparados com os resultados relativos sobre o total de escolas estudadas.

Um raciocínio análogo se faz para as colunas. O estudo das colunas leva a ver como se distribuem as condições de funcionamento das bibliotecas nas escolas que têm laboratórios bons, regulares, inadequados ou inexistentes.

Tabela 2.3. Perfis da condição de funcionamento dos bibliotecas das escolas segundo a condição de funcionamento dos laboratórios de informática

Variáveis		Laboratório de Informática				Total
Modalidades	Bom	Regular	Inadequado	Inexistente		
Biblioteca	Bom	68,6%	32,9%	35,0%	19,4%	30,8%
	Regular	23,8%	44,3%	25,0%	25,2%	25,6%
	Inadequado	3,6%	18,6%	40,0%	17,4%	14,6%
	Inexistente	4,0%	4,3%	0,0%	38,0%	29,0%
	Total	100,0%	100,0%	100,0%	100,0%	100,0%

Dentre as escolas que tem laboratórios bons, temos maior proporção com bibliotecas boas e menor com bibliotecas inadequadas ou inexistentes. Dentre as escolas com laboratórios inadequados temos maior proporção com bibliotecas boas ou inadequadas e menor proporção sem biblioteca; sempre comparando com o total de escolas.

A semelhança entre linhas se expressa de forma análoga à semelhança entre colunas. Duas linhas se consideram próximas entre si ao associarem-se da mesma forma ao conjunto de colunas, ou seja se elas se associam fortemente ou fracamente (muito ou pouco) com as mesmas colunas, considerando em relação a situação de independência. Simetricamente, duas colunas estão próximas entre se associarem-se da mesma forma ao conjunto de linhas.

Esquemáticamente, o estudo do conjunto das linhas consiste em pôr em evidência uma tipologia na qual se busca identificar quais as linhas que se afastam mais da linha que representa o conjunto da população (margem), quais as linhas que se assemelham mais entre si, e quais se opõem. Para relacionar a tipologia das linhas com o conjunto das colunas se caracteriza cada grupo de linhas pelas colunas com as quais esse grupo se associa de maneira forte ou

fraca (muito ou pouco). O estudo das colunas é totalmente análogo. Esta abordagem permite estudar a relação entre as duas variáveis, ou seja o desvio da tabela à hipótese de independência. A análise dessa relação é o objetivo fundamental da AFCS.

No exemplo da Tabela 2.3, que cruza as condições de funcionamento da biblioteca com as condições de funcionamento do laboratório de informática, ambas variáveis estão ordenadas. A relação entre essas duas ordenações (uma boa biblioteca corresponde geralmente a um bom laboratório) explica uma boa parte da relação entre as condições de funcionamento da biblioteca e do laboratório. Mas essa relação não tem porque limitar-se a este único aspecto, poderiam existir outros fenômenos como a associação quase exclusiva de alguns tipos de biblioteca com algum tipo de laboratório. O objetivo da AFCS é decompor a relação entre duas variáveis numa superposição de tendências simples e de medir sua importância relativa, conseguindo ordená-las.

Então, para analisar a informação da tabela de contingência, é preciso (Escofier-Pagès, 1992, p.30)

- avaliar as semelhanças das distribuições dos atributos em coluna (e em linha) nas subpopulações definidas pela partição da característica das linhas (colunas) mediante uma distância entre os elementos em linha e entre os elementos em coluna. Quer dizer que a informação contida na tabela resulta da comparação dos elementos da dupla partição (em linhas e em colunas) das 1933 escolas.
- representar graficamente essas comparações entre os elementos em linha e em coluna.

A distância usada pela AFCS para avaliar sem deformações as semelhanças entre linhas e colunas de uma tabela de contingência permitindo representar graficamente os elementos dessa tabela é a distância do qui-quadrado (χ^2).

$$\text{Distância entre perfis-linha: } d_{\chi^2}(k,l) = \sum_j \frac{1}{f_j} \left(\frac{f_{kj}}{f_k} - \frac{f_{lj}}{f_l} \right)^2$$

$$\text{Distância entre perfis-coluna: } d_{\chi^2}(j, m) = \sum_k \frac{1}{f_{k.}} \left(\frac{f_{kj}}{f_{.j}} - \frac{f_{km}}{f_{.m}} \right)^2$$

Nestas relações, a distância entre duas linhas depende essencialmente das diferenças termo a termo entre os dois perfis. Essa distância é a soma ponderada dos quadrados das citadas diferenças. A ponderação equilibra a influência das colunas sobre a distância entre linhas aumentando os termos *a priori* mais fracos.

Na tabela estudada, ambos os espaços tem quatro dimensões; o que torna possível calcular as distâncias, mas é impossível representar graficamente as semelhanças entre os pontos perfis-linha no espaço das colunas e os pontos perfis-coluna no espaço das linhas.

A AFCS permite fazer isso para quaisquer que sejam as dimensões da tabela, achando um plano chamado plano fatorial sobre o qual se projeta cada nuvem de pontos perfis de forma que a configuração de pontos obtida seja o mais próxima possível da configuração da nuvem de pontos perfis no espaço original. Para conseguir isso é preciso considerar cada perfil-linha como um conjunto de J valores numéricos que pode ser representado por um ponto no espaço \mathfrak{R}^J no qual cada uma das J dimensões está associada a uma modalidade da segunda variável.

A distância qui-quadrado $(\chi)^2$ que mede a semelhança entre as linhas confere a \mathfrak{R}^J a estrutura de espaço euclidiano. Esta distância conduz a uma atribuição à j -ésima dimensão de \mathfrak{R}^J do peso $1/f_{.j}$. A soma das coordenadas de cada perfil-linha vale 1. Os pesos de cada ponto da nuvem vem determinados: o ponto k tem um peso igual à frequência marginal $f_{k.}$.

O baricentro dos pontos de \mathfrak{R}^K dotados de peso denota-se por G_k e sua coordenada é igual à frequência marginal $f_{.j}$. Interpreta-se como o perfil médio. No exemplo estudado o baricentro é o perfil das condições de funcionamento das bibliotecas no conjunto de escolas estudadas, considerando todas as condições dos laboratórios acumulados.

Assim, estudar em que medida e de que forma uma classe de escolas k se diferencia do conjunto de escolas conduz a estudar o desvio entre esse perfil e o perfil do conjunto da população (médio). Estudar a dispersão da nuvem entorno a seu baricentro leva a estudar o desvio entre os perfis-linha e a marginal, portanto a relação entre as duas variáveis.

Define-se a inércia de uma nuvem de pontos como a dispersão dessa nuvem em relação ao seu baricentro, ou seja:

$$\text{Inércia}(N^K) = \sum_k \text{Inércia}(k) = \sum_k f_k \sum_j \frac{1}{f_j} \left(\frac{f_{kj}}{f_k} - f_j \right)^2$$

Achar esse plano de projeção, tal que, globalmente, os conjuntos de distâncias para cada grupo de pontos perfis se aproximem o máximo possível do conjunto de distâncias reais medidas em ambos espaços originais, leva a descompor ortogonalmente a inércia de uma nuvem de pontos. A AFCS consiste em buscar um conjunto de eixos ortogonais sobre os quais se projeta a nuvem N^K . Cada eixo possui a propriedade de fazer máxima a inércia projetada da nuvem com a restrição de ortogonalidade entre os eixos já encontrados. Já que a classe definida pela modalidade k está representada por um ponto cuja coordenada sobre o j -ésimo eixo vale $f_{kj}/f_k - f_j$, a posição deste ponto expressa a diferença entre a distribuição da população total sobre o conjunto de modalidades da segunda variável.

Assim, buscar as direções de máxima inércia da nuvem centrada conduz a pôr em evidência as classes que mais se diferenciam do conjunto da população. Cada perfil está dotado de um peso, o que faz com que cada modalidade possua uma importância proporcional à população que representa. Desta maneira, a iguais disparidades de perfis, os eixos põem em maior evidência os fenômenos relativos a uma proporção importante da população. As modalidades com menor peso, para as quais os perfis correm o risco de serem menos confiáveis, intervêm menos na construção dos eixos.

O ajuste da nuvem de perfis-colunas apresenta-se nos mesmos termos e tem as mesmas propriedades que o ajuste da nuvem de perfis-linhas.

Denomina-se fator ao conjunto de coordenadas das projeções dos pontos de uma nuvem sobre um dos eixos fatoriais: os fatores sobre as linhas são, portanto, as projeções de \mathfrak{X}^K e os fatores sobre as colunas são as projeções de \mathfrak{X}^J . A ordem de um fator é a ordem do eixo fatorial correspondente. Demonstra-se que as inércias associadas aos eixos da mesma ordem de cada uma das nuvens são iguais.

Tradicionalmente, ao estudar uma tabela de contingência, ou seja, um conjunto de n indivíduos através de duas variáveis qualitativas, mede-se a relação entre essas variáveis pela estatística qui-quadrado (χ^2). Ela mede o desvio entre as frequências observadas e as frequências esperadas em situação de independência. No exemplo, essa tabela teórica com as frequências que seriam observadas se a condição de funcionamento da biblioteca das escolas fosse independente da condição de funcionamento de laboratório de informática seria a seguinte:

Tabela 2.4. Tabela teórica em condição de independência entre a condição de funcionamento das bibliotecas e dos laboratórios de informática das escolas

Variáveis		Laboratório de Informática				Total
Biblioteca	Modalidades	Bom	Regular	Inadequado	Inexistente	
	Bom	129,6	21,5	6,2	437,7	595
	Regular	107,8	17,9	5,1	364,1	495
	Inadequado	61,6	10,2	2,9	208,2	283
	Inexistente	122,0	20,3	5,8	412,0	560
	Total	421	70	20	1422	1933

A estatística qui-quadrado (χ^2) calcula-se da seguinte forma:

$$\chi^2 = \sum_{kj} \frac{(\text{frequências observadas} - \text{frequências teóricas})^2}{\text{frequências teóricas}} = n \times \sum_{kj} \frac{(f_{kj} - f_k \cdot f_j)^2}{f_k \cdot f_j}$$

Ela é igual, a exceção do coeficiente n, à inércia total em relação ao baricentro de uma e outra das nuvens \mathcal{N}^K e \mathcal{N}^J .

$$\chi^2 = n \times [\text{Inercia}(\mathcal{N}^K)] = n \times [\text{Inercia}(\mathcal{N}^J)]$$

Desta dupla igualdade deduz-se que a inércia total de cada uma das nuvens representa de duas formas diferentes o mesmo fenômeno: a relação entre as duas variáveis (Escofier-Pagès, 1992, p.39). A AFCS decompõe ortogonalmente a inércia das nuvens, portanto a soma das inércias de uma nuvem sobre cada um dos eixos é igual a inércia total da nuvem. Sendo a inércia total o índice de desvio à situação de independência da tabela, a AFCS proporciona uma decomposição desta estatística e cada fator representa uma parte da relação entre as variáveis.

A inércia do primeiro fator tem um significado que em termos absolutos mede a importância da parte da relação que ela representa. O primeiro fator aporta a aproximação mais importante ao desvio da independência da tabela de contingência.

O valor da estatística qui-quadrado $(\chi)^2$ para este exemplo é 468,925. Dividindo-o por 1933 obtemos o valor da inércia. A decomposição da inércia está representada na Figura 2.4.

Figura 2.4. Decomposição da inércia¹

```

EIGENVALUES
PROPORTION OF THE VARIANCE : TRACE AVANT DIAGONALISATION .. 0.2426
SOMME DES VALEURS PROPRES .... 0.2426
PROGRAMME DES 3 PREMIERES VALEURS PROPRES
-----+-----
NUMERO | VALEUR | POURCENT. | POURCENT. |
        | PROPRE |           | CUMULE    |
-----+-----
1      | 0.2237 | 92.23    | 92.23    | *****
2      | 0.0154 | 6.33     | 98.57    | *****
3      | 0.0035 | 1.43     | 100.00   | **
-----+-----
    
```

Na Figura 2.4, a inércia total vale 0.2426 e a inércia projetada no primeiro eixo 0.2237. O valor próprio associado ao primeiro eixo –que se denota como λ_1 - é a inércia projetada sobre o primeiro eixo. A inércia total resulta da soma dos valores próprios ($\sum \lambda$). Essa variância representa a maior parte da associação entre as condições de funcionamento das bibliotecas e as condições de funcionamento dos laboratórios de informática das escolas estudadas.

Na tabela de decomposição da inércia, a coluna POURCENT indica a percentagem da inércia total que é captada por cada eixo e resulta do cálculo do quociente $\lambda_\alpha / \sum \lambda$. E a coluna POURCENT CUMULE acumula - para cada eixo α - as percentagens de inércia absorvidas pelos eixos anteriores a ele (incluindo-o).

¹ O histograma dos valores próprios é utilizado para estudar o decrescimento destes valores. O princípio de leitura deste diagrama é o seguinte: se dois fatores estão associados a valores próprios quase iguais, representam a mesma parte da variabilidade e não existem motivos -a priori- para escolher um deles para interpretar. Recíprocamente, um forte decrescimento entre dois valores próprios sucessivos leva a interpretar só os fatores precedentes a esse decrescimento. (Escofier-Pagès, 1992, p.235).

Na Figura 2.5, apresentam-se os elementos de ajuda à interpretação dos gráficos para as colunas da tabela: peso, distância total ao centro de gravidade da nuvem, coordenadas, contribuições à inércia e os cossenos quadrados que dão uma medida da qualidade de representação (ver capítulo 1).

Figura 2.5. Coordenadas, contribuições e qualidade da representação dos pontos perfis-coluna

COORDONNEES, CONTRIBUTIONS DES FREQUENCES SUR LES AXES 1 A 3
FREQUENCES ACTIVES

FREQUENCES			COORDONNEES			CONTRIBUTIONS			COSINUS CARRES		
IDEN - LIBELLE COURT	P.REL	DISTO	1	2	3	1	2	3	1	2	3
Labo. Bom	21.78	0.77	-0.87	0.05	-0.01	74.3	3.8	0.2	1.00	0.00	0.00
Labo. Regular	3.62	0.36	-0.25	-0.51	0.18	1.0	61.6	33.7	0.18	0.73	0.09
Labo. Inadequado	1.03	0.73	-0.16	-0.70	-0.47	0.1	32.7	66.1	0.04	0.66	0.30
Labo. Inexistente	73.56	0.08	0.27	0.02	0.00	24.6	1.8	0.0	0.99	0.01	0.00

A tabela da Figura 2.6 apresenta os mesmos elementos para as linhas.

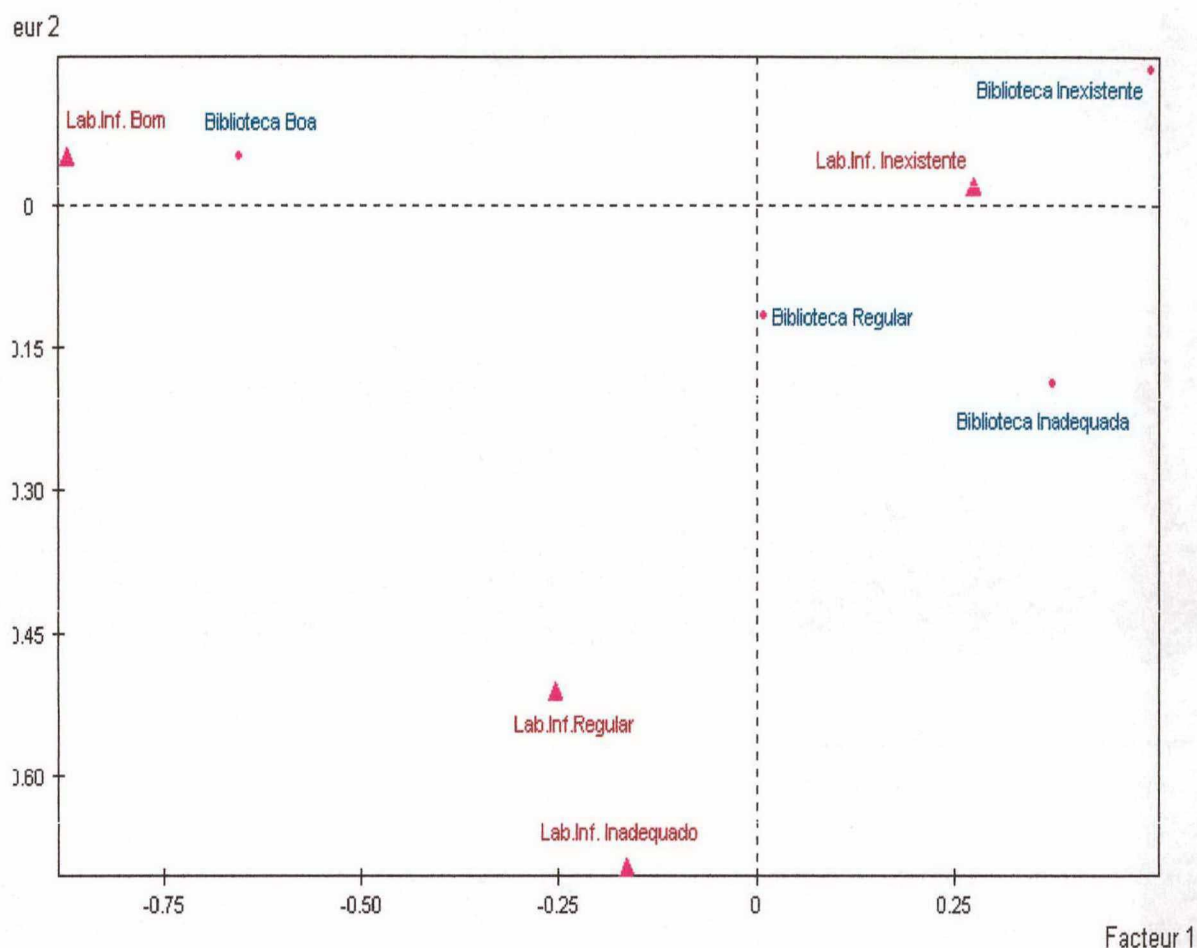
Figura 2.6. Coordenadas, contribuições e qualidade da representação dos pontos perfis-linha

COORDONNEES, CONTRIBUTIONS ET COSINUS CARRES DES INDIVIDUS
AXES 1 A 3

INDIVIDUS			COORDONNEES			CONTRIBUTIONS			COSINUS CARRES		
IDENTIFICATEUR	P.REL	DISTO	1	2	3	1	2	3	1	2	3
Biblioteca boa	30.78	0.43	-0.65	0.05	-0.02	58.8	5.2	5.3	0.99	0.01	0.00
Biblioteca regular	25.61	0.02	0.01	-0.12	0.08	0.0	22.4	52.0	0.00	0.65	0.34
Biblioteca inadequada	14.64	0.18	0.37	-0.19	-0.10	9.1	33.7	42.6	0.75	0.19	0.06
Biblioteca inexistente	28.97	0.27	0.50	0.14	0.00	32.2	38.8	0.1	0.92	0.08	0.00

A Figura 2.6 aporta a representação gráfica das modalidades das linhas e das colunas no primeiro plano fatorial.

Figura 2.7. Representação gráfica das modalidades das linhas e das colunas no primeiro plano fatorial



2.2. Fundamentos e funcionamento do método de Análise Fatorial de Correspondências Múltiplas (ACM)

No caso do estudo das escolas realizado pelo SAEB, se o objetivo é comparar as escolas segundo um conjunto de recursos físicos superior a duas características, a AFCS não resolve esse problema. Sejam, por exemplo, o estado de conservação dos telhados, das cozinhas, dos banheiros, das instalações elétricas e das condições de funcionamento das bibliotecas e dos laboratórios de informática; pretendendo avaliar as semelhanças das escolas no conjunto dessas

seis características, sendo todas elas qualitativas e cada uma delas com 4 modalidades: bom, regular, inadequado e inexistente. O método estatístico que permite estudar uma população de n indivíduos descritos por p variáveis qualitativas é a análise fatorial de correspondências múltiplas.

Os objetivos da análise de uma tabela deste tipo seriam (Crivisqui, 1999, p.107):

- Construir tipologias das unidades de observação mediante a comparação de todas as unidades através de todas as modalidades das características observadas.
- Estudar a relação existente entre todas as modalidades das características observadas.
- Resumir o conjunto de características observadas num pequeno número de variáveis quantitativas relacionadas com o conjunto de variáveis qualitativas estudado.
- Estudar a relação existente entre todas as características observadas.

No exemplo utilizado, procura-se realizar uma tipologia das escolas, isto é, definir, nas 1933 escolas, uma partição em classes de escolas que se assemelhem no conjunto de recursos físicos observados. Procura-se, também, resumir essa informação numa variável quantitativa relacionada com as seis variáveis qualitativas observadas. Essa variável quantitativa seria um indicador dos recursos físicos das escolas. Também interessa estudar a relação existente entre todos os recursos físicos observados, isto é, realizar um balanço das relações entre o estado de conservação dos telhados, banheiros, etc. Com maior detalhe, o balanço é feito em nível das modalidades e não das variáveis.

A tabela de dados observados é uma tabela de dimensões 1933×6 ; representando cada linha uma escola e cada coluna uma característica observada. Cada linha dessa tabela contém todos os códigos correspondentes às modalidades atribuídas a uma escola em cada uma das características observadas.

Figura 2.8. Tabela de códigos condensados: TCC (1933 x 6)

Escola	Telhado	Banheiro	Cozinha	Inst. elétr.	Biblioteca	Lab. Inf.
1	1	2	1	1	2	2
2	1	1	1	2	2	1
...
...
i	k_{i1}	k_{i2}	k_{i3}	k_{i4}	k_{i5}	k_{i6}
...
...
$n = 1933$	k_{n1}	k_{n2}	k_{n3}	k_{n4}	k_{n5}	k_{n6}

Esta tabela, chamada tabela de códigos condensados, não possui propriedades numéricas, portanto não é possível de ser analisada diretamente. É necessário transformá-la em outra tabela que possua propriedades numéricas mas que resuma o fenômeno observado nos mesmos termos. Por meio de uma recodificação, passa-se à tabela lógica, chamada também tabela disjuntiva completa (TDC). Nesta tabela, as linhas representam os indivíduos, no caso - as escolas; e as colunas representam as modalidades das variáveis. Na interseção da linha i com a coluna j encontra-se x_{ij} que vale 1 ou 0 indicando que o indivíduo i possui ou não a modalidade.

No exemplo, consideramos a variável estado de conservação dos telhados. Ela apresenta quatro modalidades (bom, regular, inadequado ou inexistente). Na TCC, a informação dessa variável está armazenada numa coluna que possui para cada escola o código correspondente à modalidade observada na mesma. Na TDC, a informação do estado de conservação dos telhados está armazenada em quatro colunas. Cada escola vai apresentar o valor 1 na modalidade observada correspondente e 0 nas restantes (nas modalidades não observadas).

Figura 2.9. Tabela TDC (1933x4) correspondente à primeira coluna da tabela TCC (1933x6): estado de conservação dos telhados

Escola	Bom	Regular	Inadequado	Inexistente	Soma em Linha
1	1	0	0	0	1
2	0	1	0	0	1
i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	1
n = 1933	1	0	0	0	1
Soma em coluna	n.1	n.2	n.3	n.4	n

Recodificando os dados dessa forma, chega-se à TDC, tabela na qual as colunas são funções numéricas, definidas sobre o conjunto de indivíduos, denominadas indicatrizes da modalidade. A origem do nome da tabela é a seguinte: o conjunto de valores x_{ij} de um indivíduo para as modalidades da mesma variável está formado pelo valor 1 uma vez (completa) e só uma vez (disjuntiva).

Figura 2.10. Tabela disjuntiva completa($n \times K$) correspondente a tabela de códigos condensados($n \times P$)

	Var.1 modalidades $1_1 \ 1_2 \ \dots \ 1_r$...	Var.J modalidades $J_1 \ J_2 \ \dots \ J_j \ \dots \ J_s$...	Var.P modalidades $P_1 \ P_2 \ \dots \ P_s \ \dots \ P_t$	
	1 2	(...) j	(...) K	
1	1 0 ... 0	(...)	0 1 ... 0 ... 0	(...)	1 0 ... 0 ... 0	p
2	1 0 ... 0	(...)	0 1 ... 0 ... 0	(...)	1 0 ... 0 ... 0	p
3	1 0 ... 0	(...)	0 1 ... 0 ... 0	(...)	1 0 ... 0 ... 0	p
...	(...)	(...)	p
i	1 0 ... 0	(...)	0 1 ... x_{ij} ... 0	(...)	1 0 ... 0 ... 0	p
...	(...)	(...)	p
...	(...)	(...)	p
n	1 0 ... 0	(...)	0 1 ... x_{nj} ... 0	(...)	1 0 ... 0 ... 0	p
	$n_1 \ n_2 \ \dots \ \dots$	(...) n_j	(...) n_K	np

Sendo:

K : total de modalidades das P características observadas.

Esta tabela tem propriedades numéricas: a margem coluna vale p (número de variáveis) e a margem em linhas é a distribuição de frequências das p características observadas. Pode ser considerada como uma tabela de contingência de tipo particular. Os cálculos da análise fatorial de correspondências podem ser aplicados à tabela disjuntiva completa, mas devem ser interpretados adequadamente, atendendo a natureza particular desta tabela e também às propriedades numéricas particulares. Esses cálculos com essa nova interpretação constituem um método chamado análise de correspondências múltiplas (ACM).

Considerando a tabela disjuntiva completa como uma tabela de contingência, calcula-se a distancia de χ^2 entre as linhas, para avaliar a semelhança entre os

indivíduos:
$$d^2(i,r) = \frac{1}{p} \sum_{j=1}^K \frac{n}{n_j} (x_{ij} - x_{rj})^2 .$$

A expressão entre parênteses $(x_{ij} - x_{i'j})^2$ vale 1 só se apenas um dos indivíduos i e i' apresentar a modalidade j . Vale 0 quando os indivíduos i e i' apresentam ou não conjuntamente a modalidade j .

A distância entre dois indivíduos cresce na medida que aumenta o número de modalidades diferentes que apresentam esses indivíduos. Cada modalidade intervém no cálculo da distância com o fator n / n_j , inversa de sua freqüência. Desse modo, a distância entre dois indivíduos aumenta também se apresentam divergências em modalidades pouco freqüentes. A proximidade de dois indivíduos sobre um plano fatorial vai traduzir graficamente o fato observado que ambos os indivíduos apresentam um número suficientemente grande de atributos semelhantes nas p variáveis estudadas.

Aplicando a distancia de χ^2 entre as colunas da tabela disjuntiva completa,

calcula-se a distância entre as modalidades:
$$d_{(j,k)}^2 = \sqrt{\sum_{i=1}^n \left(\frac{x_{ij}}{n_j} - \frac{x_{ik}}{n_k} \right)^2}.$$

O quadrado da distância entre duas modalidades é igual a soma das seguintes proporções: indivíduos que possuem a modalidade j mas não a k' e indivíduos que possuem a modalidade k' mas não a j . Isso faz com que duas modalidades da mesma variável estejam obrigatoriamente afastadas uma da outra, duas modalidades presentes nos mesmos indivíduos se superpõem, as modalidades raras estão afastadas de todas as outras.

A proximidade entre indicatrizes mede sua associação mútua. Essa proximidade entre duas modalidades sobre um plano fatorial traduz graficamente o fato que os atributos j e k' foram apresentados conjuntamente por um grande número de indivíduos observados. Na prática, interpreta-se a proximidade entre modalidades de variáveis diferentes como associação de modalidades, e a proximidade entre modalidades de uma mesma variável como semelhança entre classes de indivíduos (Lebart-Morineau-Piron, 1997, p.120).

Um elemento linha ou coluna influi na construção dos eixos por meio de sua inércia com respeito ao centro de gravidade. Um cálculo simples leva a obter a

inércia da modalidade j em relação ao centro de gravidade G .

$$Contr._j al_G = \frac{1}{p} \left(1 - \frac{n_j}{n} \right)$$

Por isso é freqüente em ACM ver os primeiros fatores determinados quase exclusivamente por modalidades muito raras compartilhadas pelos mesmos indivíduos. Como é mais interessante pôr em evidência os fenômenos mais gerais do que os fenômenos pontuais, procura-se evitar as modalidades raras.

As variáveis qualitativas não são introduzidas explicitamente na análise fatorial de correspondências de uma tabela disjuntiva completa. Elas aparecem através do conjunto de suas modalidades. Como um dos objetivos da ACM é obter um pequeno número de variáveis sintéticas, relacionadas o máximo possível com o conjunto de variáveis iniciais, pode-se mostrar que os fatores do ACM são essas variáveis sintéticas usando a razão de correlação. Ela mede a relação entre uma variável numérica (o fator) e uma variável qualitativa. Uma variável qualitativa define uma partição sobre o conjunto de indivíduos em tantas classes quanto o número de modalidades. A inércia total (ou variância) de uma variável numérica pode ser descomposta na soma da inércia inter (inércia dos centros de gravidade das classes) e das inércias intra (inércia dos indivíduos ao respeito do centro de gravidade da classe à qual pertencem). A razão de correlação é o quociente entre a inércia inter e a inércia total. Quando ela é próxima de 1, os indivíduos de uma mesma classe estão muito reagrupados e as classes estão muito separadas entre si, o que indica uma relação muito forte entre a variável qualitativa e a variável numérica. Quando está próxima de 0, as médias das classes estão muito próximas à média geral e os indivíduos de uma classe estão muito dispersos, isto é a variável qualitativa e a variável numérica não estão relacionadas. A razão de correlação entre um fator α e uma variável qualitativa J é:

$$\eta^2(F_\alpha, J) = \frac{\sum_j \text{inércia das modalidades } j \text{ projetadas sobre o eixo de ordem } \alpha}{\text{inércia total}}$$

A quantidade maximizada pelos eixos fatoriais no espaço R^1 é a inércia projetada da nuvem do conjunto de modalidades. Reagrupando as modalidades de uma mesma variável, este critério é a média das razões de correlação entre o fator e cada uma das variáveis. Portanto, os fatores da ACM são as variáveis numéricas mais relacionadas com o conjunto de variáveis qualitativas estudadas, constituindo uma síntese dessas informações. A contribuição de uma variável à inércia de um fator é a soma das contribuições de todas as suas modalidades.

Enquanto a inércia associada a um eixo fatorial representa para a nuvem de modalidades - sendo os eixos fatoriais combinações lineares de todas as modalidades observadas- a parte assumida por esse eixo da associação observada entre as modalidades das variáveis da TDC; para a nuvem de indivíduos representa a parte assumida por esse eixo do resultado das comparações de todos os indivíduos considerando todas as variáveis da TDC.

Para analisar a tabela de dados que descreve as escolas pelo estado de conservação dos telhados, das cozinhas, dos banheiros, das instalações elétricas e das condições de funcionamento das bibliotecas e dos laboratórios de informática - resumindo essas características numa variável quantitativa que represente todas as variáveis observadas - a primeira etapa é realizar uma ACM. A Figura 2.11 informa quais os indivíduos e as variáveis consideradas na ACM.

Figura 2.11. Variáveis ativas e indivíduos tratados com ACM

```

SELECTION DES INDIVIDUS ET DES VARIABLES UTILES
VARIABLES NOMINALES ACTIVES
  6 VARIABLES          24 MODALITES ASSOCIEES
-----
  6 . Telhado          ( 4 MODALITES )
 10 . Banheiros       ( 4 MODALITES )
 11 . Cozinha         ( 4 MODALITES )
 13 . Instalações elétricas ( 4 MODALITES )
 19 . Biblioteca      ( 4 MODALITES )
 21 . Laboratorio de Informatica ( 4 MODALITES )
-----
INDIVIDUS
-----
POIDS DES INDIVIDUS: Poids des individus, uniforme egal a 1.
RETENUS ..... NITOT = 1933      PITOT = 1933.000
ACTIFS ..... NIACT = 1933      PIACT = 1933.000
SUPPLEMENTAIRES .... NISUP = 0      PISUP = 0.000
    
```

Na decomposição da inércia mostrada na Figura 2.12 pode-se perceber a importância do primeiro eixo, em relação aos outros e, particularmente, em relação ao segundo. Esse primeiro eixo representa a parte mais importante da relação entre as diversas condições de funcionamento das bibliotecas e laboratórios e do estado de conservação dos telhados, banheiros, cozinhas e instalações elétricas (no espaço das colunas da TDC).

No espaço dos indivíduos da TDC esse eixo representa o resultado das comparações de todas as escolas através de todas as características consideradas na ACM.

Figura 2.12. Decomposição da inércia

HISTOGRAMME DES 14 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.5059	21.68	21.68	*****
2	0.2505	10.73	32.42	*****
3	0.1907	8.17	40.59	*****
4	0.1750	7.50	48.09	*****
5	0.1637	7.02	55.10	*****
6	0.1605	6.88	61.98	*****
7	0.1504	6.45	68.43	*****
8	0.1412	6.05	74.48	*****
9	0.1335	5.72	80.20	*****
10	0.1237	5.30	85.51	*****
11	0.1071	4.59	90.10	*****
12	0.0866	3.71	93.81	*****
13	0.0800	3.43	97.24	*****
14	0.0645	2.76	100.00	*****

As projeções das escolas nesse primeiro eixo podem ser representadas graficamente, mas esse gráfico não se apresenta porque o grande número de escolas dificulta a sua leitura. Na Figura 2.13 temos as coordenadas de algumas das escolas no primeiro eixo. Esse valor representa a parte mais importante do resultado das comparações de todas as escolas nas seis características.

Se duas escolas têm valores próximos, suas projeções situam-se próximas no primeiro eixo e esse fato indica que trata-se de escolas que apresentam um número importante de modalidades semelhantes, quer dizer que tem condições de conservação e de funcionamento similares nas características incluídas como critério de comparação.

As relações de transição entre os dois espaços de uma TDC mostram que a projeção de um indivíduo sobre um eixo fatorial α corresponde ao baricentro das modalidades que nele foram observadas.

Assim, por exemplo, as escolas com rótulos 2_3_RO_24 e 2_3_RO_26 tem as seguintes coordenadas no primeiro eixo: -0.61 e -0.65. Elas são próximas no primeiro eixo, mas não são idênticas porque a distância total ao centro de gravidade não é igual para ambas (1.82 e 2.19), mas devem ter algumas características semelhantes.

Os resultados para algumas das escolas são apresentados na figura 2.13.

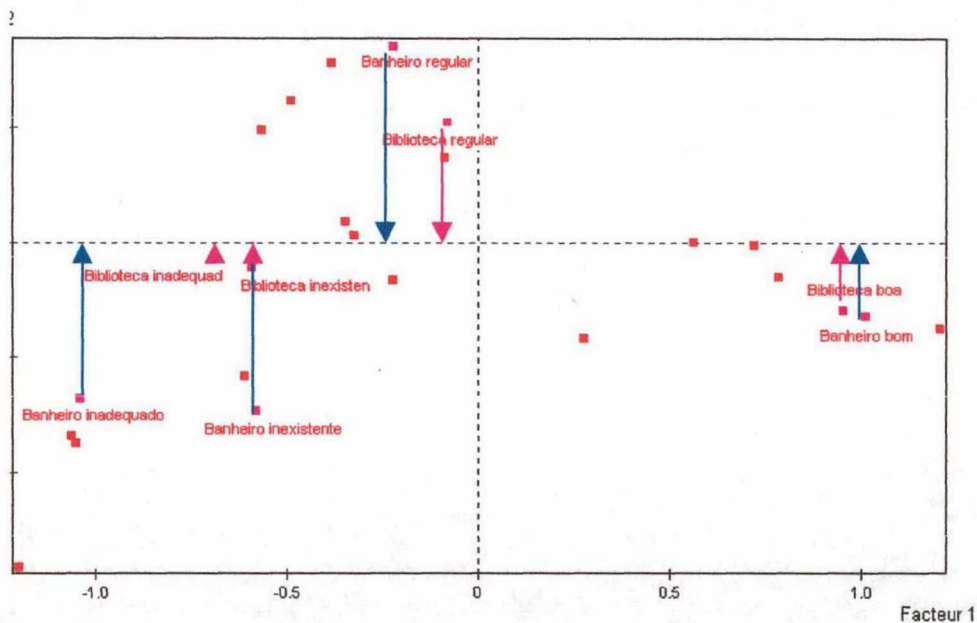
Figura 2.13. Coordenadas, contribuições e cosenos quadrados de algumas escolas (indivíduos).

COORDONNEES, CONTRIBUTIONS ET COSINUS CARRES DES INDIVIDUS - AXES 1 A 5																	
INDIVIDUS			COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENT.	P.REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
3_3_RO_1	0.05	2.03	-0.10	0.44	-0.06	0.28	0.31	0.0	0.0	0.0	0.0	0.0	0.00	0.10	0.00	0.04	0.05
3_3_RO_2	0.05	2.39	0.55	0.07	0.06	0.56	0.50	0.0	0.0	0.0	0.1	0.1	0.12	0.00	0.00	0.13	0.10
1_3_RO_3	0.05	1.58	0.35	0.20	0.12	0.30	-0.43	0.0	0.0	0.0	0.0	0.1	0.08	0.03	0.01	0.06	0.12
1_3_RO_4	0.05	1.76	-0.22	0.80	0.17	-0.05	-0.45	0.0	0.1	0.0	0.0	0.1	0.03	0.37	0.02	0.00	0.12
3_3_RO_5	0.05	1.79	1.22	-0.39	-0.01	0.18	0.04	0.2	0.0	0.0	0.0	0.0	0.83	0.09	0.00	0.02	0.00
1_3_RO_6	0.05	2.84	-0.98	-0.15	0.26	0.63	0.34	0.1	0.0	0.0	0.1	0.0	0.34	0.01	0.02	0.14	0.04
1_3_RO_7	0.05	1.26	0.57	0.15	-0.27	0.01	0.04	0.0	0.0	0.0	0.0	0.0	0.26	0.02	0.06	0.00	0.00
1_3_RO_8	0.05	2.02	-0.33	-0.38	0.05	-0.04	-0.22	0.0	0.0	0.0	0.0	0.0	0.05	0.07	0.00	0.00	0.02
1_3_RO_9	0.05	1.61	0.03	0.63	-0.13	0.21	0.45	0.0	0.1	0.0	0.0	0.1	0.00	0.24	0.01	0.03	0.13
2_3_RO_10	0.05	1.37	0.61	0.04	0.03	0.14	-0.62	0.0	0.0	0.0	0.0	0.1	0.28	0.00	0.00	0.01	0.28
1_3_RO_11	0.05	1.79	-0.25	0.84	-0.04	-0.02	0.35	0.0	0.1	0.0	0.0	0.0	0.03	0.39	0.00	0.00	0.07
3_3_RO_12	0.05	1.26	0.86	-0.24	-0.22	0.14	-0.05	0.1	0.0	0.0	0.0	0.0	0.58	0.05	0.04	0.01	0.00
2_3_RO_13	0.05	1.59	0.06	0.59	0.08	0.18	-0.35	0.0	0.1	0.0	0.0	0.0	0.00	0.22	0.00	0.02	0.08
2_3_RO_14	0.05	2.05	-0.68	0.61	0.30	0.12	-0.29	0.0	0.1	0.0	0.0	0.0	0.23	0.18	0.04	0.01	0.04
2_3_RO_15	0.05	2.08	-0.08	0.39	-0.06	0.68	0.59	0.0	0.0	0.0	0.1	0.1	0.00	0.07	0.00	0.22	0.17
3_3_RO_16	0.05	7.94	0.13	0.18	1.71	-0.91	1.10	0.0	0.0	0.8	0.2	0.4	0.00	0.00	0.37	0.10	0.15
1_3_RO_17	0.05	1.76	0.07	0.41	0.21	0.08	-0.54	0.0	0.0	0.0	0.0	0.1	0.00	0.10	0.03	0.00	0.16
1_3_RO_18	0.05	1.77	-0.16	0.12	-0.04	0.34	0.38	0.0	0.0	0.0	0.0	0.0	0.01	0.01	0.00	0.07	0.08
2_3_RO_19	0.05	1.85	0.47	-0.16	-0.10	0.63	0.33	0.0	0.0	0.0	0.1	0.0	0.12	0.01	0.01	0.22	0.06
1_3_RO_20	0.05	1.99	0.20	0.15	-0.05	0.67	0.56	0.0	0.0	0.0	0.1	0.1	0.02	0.01	0.00	0.23	0.16
2_3_RO_21	0.05	1.97	-0.48	0.25	-0.61	-0.20	0.14	0.0	0.0	0.1	0.0	0.0	0.12	0.03	0.19	0.02	0.01
1_3_RO_22	0.05	1.50	0.06	0.74	0.04	0.05	-0.30	0.0	0.1	0.0	0.0	0.0	0.00	0.36	0.00	0.00	0.00
2_3_RO_23	0.05	2.48	-0.90	-0.14	-0.42	-0.02	0.01	0.1	0.0	0.0	0.0	0.0	0.33	0.01	0.07	0.00	0.06
2_3_RO_24	0.05	1.82	-0.61	0.90	-0.43	-0.27	0.13	0.0	0.2	0.0	0.0	0.0	0.21	0.44	0.10	0.04	0.01
2_3_RO_25	0.05	1.29	0.49	-0.18	-0.62	-0.12	-0.26	0.0	0.0	0.1	0.0	0.0	0.19	0.02	0.29	0.01	0.05
2_3_RO_26	0.05	2.19	-0.65	0.56	0.32	0.00	-0.54	0.0	0.1	0.0	0.0	0.1	0.19	0.14	0.05	0.00	0.13
1_3_RO_27	0.05	2.13	-0.57	-0.11	0.31	-0.03	-0.79	0.0	0.0	0.0	0.0	0.2	0.15	0.01	0.04	0.00	0.29
1_3_RO_28	0.05	1.26	0.86	-0.24	-0.22	0.14	-0.05	0.1	0.0	0.0	0.0	0.0	0.58	0.05	0.04	0.01	0.00
3_3_RO_29	0.05	1.26	0.86	-0.24	-0.22	0.14	-0.05	0.1	0.0	0.0	0.0	0.0	0.58	0.05	0.04	0.01	0.00
1_3_RO_30	0.05	1.85	0.47	-0.16	-0.10	0.63	0.33	0.0	0.0	0.0	0.1	0.0	0.12	0.01	0.01	0.22	0.06
3_3_RO_31	0.05	1.90	0.98	-0.12	0.24	0.18	-0.53	0.1	0.0	0.0	0.0	0.1	0.50	0.01	0.03	0.02	0.15
1_3_RO_32	0.05	2.68	-0.79	0.36	0.18	0.50	0.40	0.1	0.0	0.0	0.1	0.1	0.23	0.05	0.01	0.09	0.06
3_3_RO_33	0.05	1.26	0.86	-0.24	-0.22	0.14	-0.05	0.1	0.0	0.0	0.0	0.0	0.58	0.05	0.04	0.01	0.00
2_3_RO_34	0.05	1.42	0.22	0.13	-0.56	-0.09	-0.03	0.0	0.0	0.1	0.0	0.0	0.03	0.01	0.22	0.01	0.00
2_3_RO_35	0.05	2.64	-1.09	-0.64	-0.34	0.11	-0.06	0.1	0.1	0.0	0.0	0.0	0.45	0.16	0.04	0.00	0.00
2_3_RO_36	0.05	1.82	-0.61	0.90	-0.43	-0.27	0.13	0.0	0.2	0.0	0.0	0.0	0.21	0.44	0.10	0.04	0.01
3_3_RO_37	0.05	5.87	0.52	-0.08	1.38	0.43	0.23	0.0	0.0	0.5	0.1	0.0	0.05	0.00	0.32	0.03	0.01
3_3_RO_38	0.05	2.53	-0.50	0.26	-0.10	0.55	0.74	0.0	0.0	0.0	0.1	0.2	0.10	0.03	0.00	0.12	0.21
3_3_RO_39	0.05	1.79	1.22	-0.39	-0.01	0.18	0.04	0.2	0.0	0.0	0.0	0.0	0.83	0.09	0.00	0.02	0.00
3_3_RO_40	0.05	6.18	-0.03	0.45	1.52	0.24	0.36	0.0	0.0	0.6	0.0	0.0	0.00	0.03	0.37	0.01	0.02
3_3_RO_41	0.05	1.79	1.22	-0.39	-0.01	0.18	0.04	0.2	0.0	0.0	0.0	0.0	0.83	0.09	0.00	0.02	0.00
3_3_RO_42	0.05	1.26	0.86	-0.24	-0.22	0.14	-0.05	0.1	0.0	0.0	0.0	0.0	0.58	0.05	0.04	0.01	0.00
2_3_RO_43	0.05	1.39	0.58	0.07	-0.17	0.17	0.19	0.0	0.0	0.0	0.0	0.0	0.25	0.00	0.02	0.02	0.03
1_3_RO_44	0.05	1.76	-0.54	0.23	-0.44	-0.31	-0.12	0.0	0.0	0.1	0.0	0.0	0.16	0.03	0.11	0.05	0.01
1_3_RO_45	0.05	1.72	-0.21	0.90	0.13	0.21	-0.12	0.0	0.2	0.0	0.0	0.0	0.03	0.47	0.01	0.03	0.01
2_3_RO_46	0.05	1.39	0.58	0.07	-0.17	0.17	0.19	0.0	0.0	0.0	0.0	0.0	0.25	0.00	0.02	0.02	0.03

2_3_RO_47	0.05	1.54	0.04	0.64	0.08	-0.22	-0.63	0.0	0.1	0.0	0.0	0.1	0.00	0.27	0.00	0.03	0.26
2_3_RO_48	0.05	1.39	0.58	0.07	-0.17	0.17	0.19	0.0	0.0	0.0	0.0	0.0	0.25	0.00	0.02	0.02	0.03
1_3_RO_49	0.05	2.35	-0.84	0.06	0.40	0.13	-0.60	0.1	0.0	0.0	0.0	0.1	0.30	0.00	0.07	0.01	0.16
1_3_RO_50	0.05	2.35	-0.84	0.06	0.40	0.13	-0.60	0.1	0.0	0.0	0.0	0.1	0.30	0.00	0.07	0.01	0.16
1_3_RO_51	0.05	2.98	-0.51	-0.28	0.36	-0.11	-0.68	0.0	0.0	0.0	0.0	0.1	0.09	0.03	0.04	0.00	0.16
3_3_RO_52	0.05	1.79	1.22	-0.39	-0.01	0.18	0.04	0.2	0.0	0.0	0.0	0.0	0.83	0.09	0.00	0.02	0.00
2_3_RO_53	0.05	1.26	0.86	-0.24	-0.22	0.14	-0.05	0.1	0.0	0.0	0.0	0.0	0.58	0.05	0.04	0.01	0.00

As relações de transição entre os dois espaços também mostram que a projeção sobre um eixo α de uma modalidade determinada está representada pela coordenada média dos indivíduos observados que apresentaram essa modalidade. Para compreender em que se assemelham as escolas cujas projeções são próximas, deve-se consultar os resultados da ACM para as colunas da TDC e a sua correspondente representação gráfica. No primeiro plano fatorial seguinte, estão projetadas todas as modalidades. Como exemplo, indicam-se as projeções dos estados de conservação dos banheiros e das condições de funcionamento das bibliotecas.

Figura 2.14. Primeiro plano principal: projeção das modalidades.



O primeiro eixo opõe as escolas com banheiro inadequado ou inexistente (a esquerda) das escolas com banheiros regulares ou bons (a direita). Em relação a biblioteca, a oposição é semelhante. Não podemos dizer que o primeiro eixo

ordena as escolas pela condição da biblioteca porque temos uma inversão: as escolas com bibliotecas inadequadas se situam mais a esquerda (coordenada -0.69) que aquelas com bibliotecas inexistentes (coordenada -0.59). Por outra parte, a proximidade sobre um eixo fatorial α , de duas modalidades, por exemplo banheiro bom e biblioteca boa, traduz uma forte associação entre essas variáveis indicatrizes.

Conforme exposto anteriormente, o primeiro eixo contém informação de todas as modalidades estudadas, mas elas contribuem diferentemente na geração da inércia projetada por ele.

Na Figura 2.15- que apresenta os resultados para as colunas da TDC - podem-se consultar essas contribuições.

Figura 2.15. Coordenadas, contribuições e cosenos quadrados das Modalidades

MODALITES			COORDONNEES					CONTRIBUTIONS					COSINUS CARRES					
IDEN	LIB	P.REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
6 . Telhado																		
AF_1	- P1_B	9.40	0.77	0.56	0.00	-0.14	-0.09	-0.12	5.9	0.0	1.0	0.4	0.9	0.41	0.00	0.03	0.01	0.02
AF_2	- P1_R	5.40	2.09	-0.57	0.49	0.10	0.33	0.32	3.4	5.1	0.3	3.4	3.3	0.15	0.11	0.00	0.05	0.05
AF_3	- P1_Ina	1.87	7.91	-1.19	-1.41	0.44	-0.51	-0.29	5.3	14.9	1.9	2.8	0.9	0.18	0.25	0.02	0.03	0.01
									CONTRIBUTION CUMULEE = 14.6 20.1 3.1 6.6 5.1									
10 . Banheiros																		
AJ_1	- P5_B	6.10	1.73	0.99	-0.34	0.02	0.11	-0.08	11.5	2.8	0.0	0.4	0.3	0.57	0.07	0.00	0.01	0.00
AJ_2	- P5_R	6.06	1.75	-0.23	0.84	-0.10	-0.20	0.12	0.6	17.2	0.3	1.4	0.5	0.03	0.41	0.01	0.02	0.01
AJ_3	- P5_Ina	4.50	2.70	-1.04	-0.68	0.11	0.13	-0.05	9.6	8.3	0.3	0.4	0.1	0.40	0.17	0.00	0.01	0.00
									CONTRIBUTION CUMULEE = 22.1 28.3 0.6 2.3 0.8									
11 . Cozinha																		
AK_1	- P6_B	7.44	1.24	0.72	-0.02	-0.14	0.38	-0.10	7.6	0.0	0.8	6.1	0.4	0.42	0.00	0.02	0.12	0.01
AK_2	- P6_R	5.06	2.29	-0.49	0.62	0.10	-0.19	-0.34	2.4	7.8	0.3	1.1	3.6	0.10	0.17	0.00	0.02	0.05
AK_3	- P6_Ina	3.03	4.51	-1.06	-0.84	-0.14	0.40	0.12	6.8	8.5	0.3	2.7	0.3	0.25	0.16	0.00	0.04	0.00
AK_4	- P6_Ine	1.14	13.64	0.28	-0.42	0.85	-2.67	1.82	0.2	0.8	4.4	46.4	23.0	0.01	0.01	0.05	0.52	0.24
									CONTRIBUTION CUMULEE = 17.0 17.1 5.7 56.4 27.3									
13 . Instalações elétricas																		
AM_1	- P8_B	7.59	1.20	0.77	-0.16	-0.13	-0.06	-0.15	8.9	0.8	0.7	0.1	1.0	0.49	0.02	0.01	0.00	0.02
AM_2	- P8_R	5.57	1.99	-0.39	0.77	0.01	0.03	0.42	1.6	13.2	0.0	0.0	6.0	0.08	0.30	0.00	0.00	0.09
AM_3	- P8_Ina	3.51	3.75	-1.05	-0.88	0.26	0.07	-0.35	7.6	10.9	1.3	0.1	2.7	0.29	0.21	0.02	0.00	0.03
									CONTRIBUTION CUMULEE = 18.1 24.8 1.9 0.3 9.7									
19 . Biblioteca																		
AS_1	- I14_B	5.13	2.25	0.95	-0.30	0.08	0.00	0.37	9.2	1.8	0.2	0.0	4.4	0.40	0.04	0.00	0.00	0.06
AS_2	- I14_R	4.27	2.91	-0.08	0.53	0.75	0.00	-1.01	0.1	4.8	12.6	0.0	26.4	0.00	0.10	0.19	0.00	0.35
AS_3	- I14_Ina	2.44	5.83	-0.69	-0.07	0.39	1.26	1.28	2.3	0.1	2.0	22.0	24.5	0.08	0.00	0.03	0.27	0.28
AS_4	- I14_Ine	4.83	2.45	-0.59	-0.11	-0.95	-0.64	-0.16	3.3	0.2	22.7	11.3	0.7	0.14	0.01	0.37	0.17	0.01
									CONTRIBUTION CUMULEE = 14.9 6.9 37.4 33.3 56.0									
21 . Laboratório de Informática																		
AU_1	- I16_B	3.65	3.57	1.20	-0.38	0.28	0.11	0.17	10.4	2.1	1.5	0.3	0.7	0.40	0.04	0.02	0.00	0.01
AU_2	- I16_R	0.66	24.43	-0.12	0.36	3.61	-0.51	-0.28	0.0	0.3	44.8	1.0	0.3	0.00	0.01	0.53	0.01	0.00
AU_4	- I16_Ine	12.36	0.35	-0.35	0.09	-0.27	-0.01	-0.04	3.0	0.4	4.9	0.0	0.1	0.35	0.02	0.22	0.00	0.00
									CONTRIBUTION CUMULEE = 13.4 2.8 51.2 1.2 1.1									

Na Figura 2.15 pode-se aferir qual foi a contribuição de cada uma das variáveis na geração do primeiro eixo. A ordem de contribuição de cada uma das características é a seguinte: estado de conservação dos banheiros (22.1), das

instalações elétricas (18.1), das cozinhas (17.0) , condição de funcionamento das bibliotecas (14.9), estado dos telhados (14.5) e finalmente condição dos laboratórios (13.4). Essas contribuições indicam a intensidade da relação entre as variáveis qualitativas consideradas e o primeiro eixo fatorial, variável quantitativa que representa a parte mais importante da associação entre as variáveis indicatrizes.

2.3. Fundamentos e funcionamento do método de Análise Fatorial Múltipla

A Análise Fatorial Múltipla (AFM) é um método fatorial adaptado ao tratamento de tabelas nas quais um conjunto de indivíduos é descrito por vários grupos de variáveis. O primeiro problema suscitado pelo tratamento simultâneo de vários grupos de variáveis é a ponderação desses grupos. A análise fatorial múltipla pode ser considerada como uma ACP ou uma ACM onde a influência dos grupos de variáveis está equilibrada. Aliás, ao considerar a estrutura em grupos de um conjunto de variáveis, o objetivo do estudo fica enriquecido, da mesma forma quando se procura comparar os indivíduos (em termos de semelhança), ou as variáveis (em termos de relações). Pode-se comparar globalmente os grupos de variáveis e a existência de grupos de variáveis conduz a querer descrever os indivíduos através de cada um dos grupos (Crivisqui-Villamonte, 1999, p.2-3).

No exemplo anterior, que foi tratado com ACM, a tabela de estudo tinha nas linhas as 1933 escolas e nas colunas as 6 características qualitativas (estado de conservação dos telhados, das cozinhas, dos banheiros, das instalações elétricas e das condições de funcionamento das bibliotecas e dos laboratórios de informática). Se considerarmos que as primeiras quatro variáveis correspondem as características do prédio da escola e as outras duas, às características das suas instalações, teríamos definido na tabela dois grupos de variáveis; de tal modo que , a análise de correspondências múltiplas não é adequada para estudar essa tabela.

2.3.1. Os dados e a tabela estatística múltipla

Os dados estão constituídos por um conjunto de indivíduos descritos por vários grupos de variáveis. Em primeiro lugar, apresenta-se o caso de variáveis quantitativas, estudando as variáveis qualitativas em separado. A cada grupo de variáveis corresponde uma tabela. Todos os grupos de variáveis estão definidas sobre o mesmo conjunto de indivíduos, de tal forma que todas as tabelas possam ser justapostas formando uma única tabela. O conjunto inicial de várias tabelas aparece como uma única tabela estruturada em subtabelas.

Figura 2.16. Tabela Estatística Múltipla (T^M)

	X_1		
	1	j	K_1
1			
i		$v_j(i)$	
n			

...

	X_j		
	1	k	K_j
1			
i		$v_k(i)$	
n			

...

	X_K		
	1	m	K_K
1			
i		$v_m(i)$	
n			

O i -ésimo indivíduo está descrito por K valores. O valor $v_k(i)$ corresponde ao i -ésimo valor da k -ésima variável do grupo k_j associado à tabela X_j .

- X : tabela completa
- K : conjunto de variáveis de todas as tabelas
- I : conjunto de indivíduos
- J : conjunto de tabelas constitutivas de X (subtabelas)

- K_j : conjunto de variáveis da tabela X_j .
- X_j : tabela associada ao conjunto K_j .

A existência de grupos de variáveis leva a estudar a semelhança dos indivíduos através dos grupos de variáveis. Para que sejam comparáveis é preciso resolver o problema técnico que será tratado na seção seguinte.

2.3.2. Ponderação dos grupos de variáveis

Na AFM modificam-se os pesos iniciais das variáveis. A AFM baseia-se na análise de componentes principais da tabela X transformada. A transformação visa a resolver dois problemas relativos a importância que um grupo de variáveis tem na análise global:

- número de variáveis que compõem o grupo: quanto maior é o número de variáveis, maior é a influência do grupo.
- a estrutura interna do grupo: quanto mais forte é a estrutura do grupo, ou seja, quanto mais relacionadas estão as variáveis, a sua influência será mais determinante na construção dos principais fatores de variabilidade.

Assim, a transformação realizada sobre a tabela X tem por objetivo equilibrar a importância relativa de cada grupo X_k na análise global. A AFM atribui o mesmo peso a todas as variáveis de um mesmo grupo para conservar a estrutura interna do grupo. Esse peso, dado a cada uma das variáveis de um grupo, é igual ao inverso da inércia da primeira componente principal efetuado sobre as variáveis desse grupo. A ponderação leva em conta o número de variáveis e suas correlações. Uma ponderação que não levasse em conta as relações entre as variáveis diminuiria a inércia de um grupo composto por muitas variáveis independentes e aumentaria, relativamente, a inércia em uma direção de um grupo composto por poucas variáveis muito correlacionadas.

Como consequência dessa ponderação, resultante de multiplicar as variáveis das tabelas X_j pelo inverso do primeiro valor próprio correspondente ($1 / \lambda_1^j$):

- a inércia da nuvem associada fica multiplicada por esse peso em cada direção do espaço.

- a inércia da primeira componente principal de cada grupo de variáveis (inércia máxima de cada tabela X_j ponderada) é igual a 1 e a soma das inércias das variáveis de um mesmo grupo sobre um eixo qualquer do espaço é inferior a 1.
- a estrutura interna do grupo se conserva já que o coeficiente é idêntico para todas as variáveis do mesmo grupo.

Dessa forma o papel de cada grupo se equilibra no sentido que nenhum deles pode determinar por si só a primeira componente principal do conjunto.

Esta análise proporciona uma representação dos indivíduos e das variáveis que se interpreta segundo as regras usuais da análise em componentes principais (ACP).

2.3.3. A AFM em R^K : os indivíduos

O espaço R^K contém as representações dos indivíduos. Cada uma das dimensões se associa a uma variável. Partindo desse espaço procura-se:

- Uma representação gráfica da nuvem de indivíduos caracterizados pelo conjunto de variáveis. Ela é obtida mediante uma ACP da tabela X , com as variáveis ponderadas.
 - Uma representação superposta das J nuvens de indivíduos caracterizadas cada uma por um grupo de variáveis. Esse gráfico é obtido com ajuda das projeções das linhas suplementares na ACP anterior.
- 1) A comparação dos indivíduos na análise global se faz mediante a distância calculada em R^K . Nessa nuvem N_j , o quadrado dessa distância entre os pontos i e i' , é a soma dos quadrados da sua distância nas N_j^j .

$$d^2(i, i') = \sum_{k \in K} m_k (v_k(i) - v_k(i'))^2 = \sum_{j \in J} \sum_{k \in K_j} m_k (v_k(i) - v_k(i'))^2 = \sum_{j \in J} d^2(i^j, i'^j)$$

onde i^j é o ponto que representa ao indivíduo i na nuvem N_j^j e m_k ao peso da variável k (Escofier- Pagès, 1992, p.144).

A influência de cada grupo de variáveis nessa somatória estaria equilibrada só se as distâncias em cada nuvem parcial de indivíduos são da mesma ordem. Multiplicar os pesos iniciais das variáveis do grupo j por um coeficiente é a forma de equilibrar a influência dos grupos calculando-se a distância desta forma:

$$d^2(i, i') = \sum_{j \in J} \frac{1}{\lambda_j} d^2(i, i')$$

Com essa ponderação nenhum grupo pode ser preponderante na primeira direção de inércia. Mas, o número de direções de N_i sobre as quais o grupo j influencia, cresce com a dimensão de N_i^j .

A representação da nuvem de pontos N_i (indivíduos) num subespaço mais reduzido se faz classicamente com a ACP da tabela X ponderada.

- 2) As coordenadas em R^k_j dos n indivíduos da nuvem de pontos N_i^j estão contidas na tabela X_j . As coordenadas desses pontos em R^k estão contidas na tabela obtida anulando todas as coordenadas que correspondem aos outros grupos de variáveis (completando com zeros):

Figura 2.17. Tabela de coordenadas dos indivíduos em R^k

	X_1		X_k			X_j	
	1	K_1	1	j	K_k	1	K
$\tilde{X}_j :$	0			$V_j(i)$		0	

Sejam: N_i^* , nuvem dos centros de gravidade, dos n pontos i^j que representam o mesmo indivíduo i nas N_i^j nuvem de pontos.

Essa nuvem é a nuvem média para as N_i^j . Sua representação inscreve-se na problemática clássica da análise fatorial: procura-se projetar a nuvem sobre um subespaço de dimensão menor de forma que a projeção obtida seja o mais semelhante possível à nuvem original. Para isso realiza-se uma ACP da tabela X . A única peculiaridade desta ACP é que as variáveis são ponderadas para equilibrar a influência dos grupos.

A representação superposta das nuvens N_i^j se obtém por projeção em elementos suplementares das mesmas sobre os eixos fatoriais de N_i .

Como todas as nuvens N_i^j estão situadas em R^k , é possível obter uma representação simultânea por projeção sobre um mesmo subespaço. A escolha do subespaço tenta satisfazer duas condições essenciais para que essa representação permita comparar a posição de um indivíduo nas diferentes nuvens:

- Cada nuvem N_i^j deve estar bem representada. Para isso se escolhe uma representação de máxima inércia total para cada nuvem N_i^j .
- As representações das nuvens N_i^j devem assemelhar-se entre elas. Como não é possível comparar as posições de um mesmo ponto nas diferentes

nuvens, estas representações são em conjunto muito diferentes; é necessário que pontos homólogos - que representam o mesmo indivíduo - sejam o mais próximos possível um do outro. Para que os pontos associados a um mesmo indivíduo estejam próximos entre si, procura-se minimizar a inércia-intra projetada de cada N_i^j .

Para satisfazer simultaneamente esses dois critérios, o espaço de projeção das J nuvens N_i^j deve maximizar a inércia total e minimizar a inércia intra de cada N_i^j . Estas duas propriedades são geralmente incompatíveis: a qualidade de representação das nuvens e a semelhança entre estas representações não podem-se otimizar simultaneamente. Dando uma importância equivalente as duas propriedades, o teorema de Huygens (inércia total = inércia inter + inércia intra) sugere maximizar a inércia-inter projetada.

A qualidade da representação de cada nuvem mede-se classicamente mediante a razão entre a inércia projetada e a inércia total da nuvem.

Essa qualidade de representação sempre é fraca, motivo pelo qual esse indicador dá uma medida pessimista da qualidade de representação, no sentido que a forma da nuvem pode ser respeitada mesmo sendo esta razão pequena.

Para avaliar a utilidade da representação conjunta calcula-se um índice de semelhança entre as projeções das nuvens, para cada eixo, a razão entre inércia inter e inércia total (Escofier-Pagès, 1992, p.148).

Se esta razão é próxima de 1, todas as nuvens tem suficientes caracteres comuns que justificam um estudo detalhado das diferenças. Se é próxima de 0, as diferenças de forma são tão importantes que um estudo, mesmo aprofundado, não resulta de utilidade.

2.3.4. A AFM em R^l : as variáveis

O espaço R^l permite obter as representações das variáveis iniciais, podendo situar-se também as componentes principais de cada um dos grupos em separado.

- A representação das variáveis é obtida diretamente da ACP da tabela completa X . Ela constitui uma ajuda à interpretação da nuvem dos indivíduos. E é uma representação ótima das correlações entre as variáveis.
- As componentes principais fazem máxima a inércia projetada de todas as variáveis. A inércia projetada num eixo das variáveis de um grupo pode ser interpretada como a contribuição desse grupo à inércia nessa direção.
- Um estudo sistemático das correlações entre as primeiras componentes de cada grupo aporta elementos interessantes para a comparação dos grupos. Essa representação é obtida por projeção das componentes principais de cada grupo nos eixos fatoriais da AFM. Assim, essas primeiras componentes principais são consideradas variáveis suplementares da tabela X .

2.3.5. Fatores comuns aos grupos de variáveis

Na ACP procura-se obter variáveis que resumam o conjunto de variáveis estudadas. Introduzindo em ACP as noções de grupos de variáveis, e de equilíbrio entre os grupos; trata-se também de resumir esses grupos. Esse é o objetivo dos fatores comuns.

Julgar o caráter verdadeiramente comum de um fator implica medir a semelhança entre as projeções das nuvens. Para isso, calcula-se o coeficiente de correlação entre cada um dos fatores F_s^j e o fator global F_s .

Esse coeficiente pode tomar um valor elevado para o conjunto dos grupos, ou para alguns deles. Assim, a AFM permite pôr em evidência os fatores comuns a um conjunto de grupos, os fatores comuns a alguns grupos e os fatores específicos de um grupo. Esse coeficiente de correlação indica em que medida o fator comum F_s está efetivamente presente no grupo K_j .

Quando existe uma direção de dispersão comum, é interessante medir e comparar a sua importância nos diferentes grupos. Isso se mede pela inércia acumulada das variáveis do grupo sobre esse fator, denominando-se também peso do fator no grupo.

A representação dos grupos pode ser considerada como uma ajuda na interpretação dos outros gráficos e como uma imagem ótima da nuvem dos grupos entre si.

Os cálculos necessários se deduzem diretamente dos resultados da ACP de X. A coordenada de um grupo sobre um eixo é a inércia acumulada das variáveis do grupo sobre o eixo correspondente da AFM.

Como consequência da ponderação dos grupos, suas coordenadas estão entre 0 e 1. Podem ser consideradas como uma medida da associação entre o grupo de variáveis e o eixo. Uma coordenada próxima de 1 implica que essa direção é uma direção de inércia importante para o grupo.

O gráfico dos pontos-grupo interpreta-se como a projeção ortogonal de uma nuvem de pontos no qual cada ponto representa um grupo.

Dois pontos-grupo próximos indicam que a estrutura definida por essas variáveis sobre o conjunto dos indivíduos são semelhantes.

2.3.6. O caso das variáveis qualitativas e das tabelas mistas

A problemática analisada no caso das variáveis quantitativas estende-se sem modificações importantes as variáveis qualitativas. A essência da aplicação da AFM nas tabelas disjuntivas completas, nas quais as variáveis estão estruturadas em grupos, reside na equivalência entre a análise de correspondências múltiplas e a análise em componentes principais aplicadas às variáveis indicatrizes ponderadas adequadamente.

Considerando a Tabela Disjuntiva Completa em perfis; os pesos dos indivíduos como $1/n$; e os pesos das modalidades como n_k/np ; as propriedades da ACM são:

- as modalidades de uma mesma variável são ortogonais entre si; a transformação em perfis não muda a direção
- cada modalidade possui a mesma inércia em relação a origem.

$$Inércia_o(k) = \frac{n_k}{np} \sum_i n \left(\frac{x_{ik}}{n_k} \right)^2 = \frac{1}{p}$$

2.3.7. Ponderação da Tabela Disjuntiva Completa para aplicar a ACP

Seja a nuvem de variáveis indicatrizes em R^n , sem centrar, e divididas pelo desvio padrão. Se atribui a cada variável indicatriz o mesmo peso: $\left(\frac{n - n_k}{n}\right)$

A direção das indicatrizes não é modificada pela divisão pelo desvio padrão, e também possuem a mesma inércia em relação a origem.

$$\text{Inércia}_o(k) = \frac{n - n_k}{np} \sum_i \frac{1}{n} \times \frac{x_{ik}^2}{\frac{n_k(n - n_k)}{n^2}} = 1.$$

Aplicado a esta nuvem de modalidades a operação centrada da ACM interpreta-se como em ACP, portanto uma ACP formada das indicatrizes ponderadas conduz aos mesmos fatores que verificados em uma ACM.

Devido a essa equivalência, é possível aplicar a AFM a variáveis qualitativas, com a condição de introduzir estas por meio de suas indicatrizes ponderadas adequadamente. Aporta uma solução técnica para a abordagem da problemática associada ao estudo simultâneo de vários grupos de variáveis quando estas são qualitativas.

Em conseqüência, na AFM pode-se trabalhar simultaneamente com grupos de variáveis dos dois tipos: grupos de variáveis qualitativas e grupos de variáveis quantitativas; nesse caso, a tabela de dados chama-se tabela mista. Desde que cada grupo compreende variáveis do mesmo tipo, a ponderação da AFM que equilibra a inércia e portanto a influência dos grupos, elimina os desequilíbrios entre grupos provenientes tanto por diferenças de estrutura quanto por diferenças de tipos de variáveis.

CAPITULO III**3. A PROPOSTA METODOLÓGICA****Sumário**

3.1. Pré-requisitos	054
3.1.1. As bases de dados do SAEB necessárias para o cálculo do indicador.	054
3.1.2. O problema dos valores ausentes ("missing values").	055
3.2. Etapas metodológicas de construção do indicador	056
3.2.1. Definir as finalidades do indicador	056
3.2.2. Reunir os pré-requisitos do estudo	057
3.2.3. Construção da tabela de dados observados	057
3.2.4. Construção do indicador	059
3.2.5. Validação do indicador	060
3.2.6. Uso do indicador e a tipologia associada	062
3.2.7. Validação da metodologia de elaboração do indicador	062

A PROPOSTA METODOLÓGICA

Neste capítulo, descreve-se a estratégia a ser adotada na construção da metodologia para sistematização e análise dos dados observados, sua disponibilização em uma base de dados e a ordenação e operacionalização de um conjunto de técnicas estatísticas, visando criar as condições necessárias para responder a pergunta de pesquisa e atender aos objetivos do presente estudo.

3.1. Pré-requisitos

3.1.1. Dispor das bases de dados do SAEB necessárias para o cálculo do indicador.

As bases de dados do SAEB contém as informações dos questionários (escola, diretor, professor e aluno) e testes aplicados, sendo possível relacionar as informações das diferentes bases. Os dados levantados com os questionários são ingressados e armazenados nas bases num formato alfabético. Para serem processados estatisticamente devem ser transformados em códigos numéricos. Essas bases de dados contém as observações brutas, sem nenhuma correção. Os erros presentes nas bases do SAEB são dos seguintes tipos:

- Valores ausentes
- Valores inconsistentes
- Valores errôneos (erros de entrada dos dados).

Globalmente, esses “erros sistemáticos” podem representar um erro de observação que é superior ao “erro aleatório” de amostragem, daí a importância de realizar uma validação prévia da informação. Não se coloca como um pré-requisito a correção da informação das bases porque alguns dos erros devem ser identificados e corrigidos em diferentes momentos do processo de coleta dos dados ou na recepção dos questionários, mas sempre previamente a difusão das bases. Dentre desses erros formais, essas bases têm itens mal definidos que geram um número elevado de “*missing values*”, ou respostas idênticas (constantes); escalas (de opinião ou de preferência) insuficientemente definidas, o que leva as respostas não serem comparáveis; e, finalmente, a carência de codificação que distinga a não resposta, a rejeição a responder e a resposta que não corresponde.

3.1.2. Ter estudado e resolvido nas bases de dados a serem utilizadas o problema dos valores ausentes ("*missing values*").

Das etapas de controle da qualidade de uma base de dados, a última é a imputação de "*missing values*". Para mostrar o impacto da presença de "*missing values*" numa base de dados é preciso considerar que inevitavelmente, a observação padrão de p características (qualitativas e/ou quantitativas) sobre n unidades de observação produz:

- $N_b < n \times p$ itens corretamente observados.
- N_r itens que se apresentam com frequência muito baixa.
- N_e itens com observações errôneas (que não correspondem com o que se queria observar).
- N_m itens que não foram observados, chamados valores ausentes ou "*missing values*").

A taxa de erros de observação é definida assim:

$$\tau_{\text{errosobs}} = \frac{N_m + N_e + N_r}{n \times p} \times 100$$

A etapa de ingresso dos questionários, transforma essas observações nos seguintes tipos de "dados disponíveis" :

- D_c (em geral $D_c < N_b + N_r$); dados corretos.
- D_m (em geral $D_m > N_m$); valores ausentes (*missing cases*).
- D_i (em geral $D_i > N_e$); dados inconsistentes.
- D_e ; dados errôneos (erros de digitação ou de ingresso de dados)...

Na última etapa de preparação dos dados (tratamento de valores ausentes), temos:

$$D_{\text{base}} = n \times p = D'_{\text{corr}} + D_{\text{missing}}$$

Sendo:

- D'_{corr} = dados corretos + dados corrigidos
- $D_{missing}$ = dados ausentes

A metodologia de amostragem utilizada permite “controlar” o erro das estimações que depende do fato de se ter observado uma amostra e não toda a população (erro aleatório). Mas, já que os erros são sempre cumulativos, as estimações feitas vão ter um erro global que vale:

$$\text{erro global} = \text{erro aleatório} + \text{erro de observação}$$

Se o erro de observação é muito maior que o erro aleatório, as estimações carecem de valor; portanto não pode-se subestimar a verificação e correção da informação das bases de dados. Contudo, a imputação de valores ausentes pode ser feita pelo pesquisador que tem por objetivo a construção do indicador. Mas, para selecionar um procedimento adequado de imputação dos dados ausentes é necessário levar em consideração:

- A taxa de valores ausentes $\tau_{missing} = \frac{D_{missing}}{n \cdot p} \times 100$
- A distribuição dos valores ausentes na base de dados.

É por isso que não pode-se recomendar *a priori* um procedimento de imputação adequado a qualquer situação.

3.2. Etapas metodológicas de construção do indicador

3.2.1. Definir as finalidades do indicador

O indicador é uma síntese de um conjunto de informações que descrevem um fenômeno. Portanto, o indicador deve diferenciar as diferentes situações das unidades de estudo em relação ao fenômeno por ele representado. Esse objetivo

deve ser claramente estabelecido, especificando quais as situações que se quer diferenciar.

Por outra parte, as informações do SAEB devem servir ao diagnóstico objetivo da situação de ensino básico e devem fornecer elementos para planejar políticas de intervenção em diferentes níveis de decisão.

Em conseqüência, o indicador deve atingir objetivos operacionais que tem que ser definidos, especificando quais as questões avaliativas que o indicador deve responder, de que forma e por quem podem ser usadas as informações sintetizadas através do indicador.

3.2.2. Reunir os pré-requisitos do estudo

Evidentemente, uma etapa importante se caracteriza por reunir os pré-requisitos detalhados no item 3.1.

3.2.3. Construção da tabela de dados observados

A construção da tabela e a definição do seu desenho requer as seguintes ações:

- Escolher um conjunto homogêneo de unidades de observação que serão as linhas da tabela de dados X ($n \times p$). As linhas das bases de dados do SAEB correspondem a unidades de observação diferentes (série, turma, aluno). Portanto, uma vez definida a unidade de análise da pesquisa, devem-se produzir as agregações necessárias para obter a tabela que será objeto de estudo.
- Escolher as características observadas (variáveis) necessárias para descrever o objeto do estudo, que serão as colunas ativas da tabela de dados X ($n \times p$). Esta escolha vai depender dos objetivos definidos para o indicador, ou seja de qual o aspecto da realidade observada através de um conjunto de características que o indicador deve sintetizar. Essas características serão as variáveis ativas.

- Escolher as variáveis necessárias para validar o indicador, que serão as colunas ilustrativas da tabela de dados X ($n \times p$). A escolha das variáveis ilustrativas depende do objetivo do indicador e das informações disponíveis na base de dados. Nas bases de dados do SAEB estão contidas as variáveis de estratificação das escolas: zona (urbana ou rural), localização (capital ou interior), rede (estadual, municipal e particular), região e unidade da federação. Essas variáveis que identificam a região, estado, zona e localização não devem participar na comparação dos indivíduos (sejam eles escolas, diretores, professores ou alunos), intervindo somente de forma ilustrativa ou explicativa. A forma de validar externamente os resultados do estudo - com as informações disponíveis- é efetuar o cruzamento do indicador obtido com as variáveis de estratificação; verificando portanto se o que ele está descrevendo é coerente com o que se conhece das unidades por região, estado, rede, etc.
- Determinar os "valores" que vão figurar nas células da tabela, ou seja nos cruzes das linhas e colunas da X ($n \times p$). Para isso, tem que se produzir as agregações de dados necessários. Por exemplo: se a unidade de estudo escolhida fosse a escola, e houvesse interesse em comparar as escolas pela proficiência dos alunos delas: tem que se definir a proficiência da escola, já que o que foi avaliado no levantamento do SAEB foram as proficiências dos alunos.
- Definir - segundo o ponto de vista do pesquisador - a estrutura temática da tabela. As p variáveis devem ser organizadas em temas que respondem aos objetivos do estudo. Analisando os questionários é possível identificar que as variáveis que caracterizam os diferentes aspectos do fenômeno observado estão estruturadas em grupos. Por exemplo; temos um conjunto de características observadas relacionadas com a formação do professor, com a situação socio-econômica dos alunos, com os recursos materiais de uma escola, etc.

3.2.4. Construção do indicador

Em primeiro lugar requer a realização de uma análise fatorial múltipla da tabela de dados em concordância com os temas já definidos.

Em geral, os distintos aspectos que fazem parte do fenômeno observado determinam uma estrutura temática da tabela de dados na qual os temas ou aspectos definidos formam grupos com diferentes número de variáveis.

Também é comum que aconteça que as análises parciais desses grupos de variáveis revelem notórias diferenças nas estruturas dos mesmos.

Em consequência, esses grupos de variáveis terão uma influência diferente na composição dos eixos explicativos da variabilidade entre as diferentes características que formam a tabela $X(n \times p)$.

A técnica apropriada para o tratamento de uma tabela estruturada em grupos de variáveis é a técnica de análise fatorial múltipla (AFM), que elimina o risco de que algum grupo influencie de forma preponderante, neutralizando a influência dos demais grupos e permitindo que a influência dos grupos de variáveis seja equilibrada entre si.

O AFM da tabela $X(n \times p)$ permite, ainda, enriquecer o estudo ao comparar as unidades globalmente pelos grupos de variáveis, viabilizando a descrição de suas semelhanças, segundo os grupos que aglutinam o conjunto das variáveis ativas.

Em segundo lugar se faz necessário verificar o grau de presença do primeiro fator - resultante da análise anterior - em cada grupo de variáveis

Isso determinará se é possível resumir a informação num indicador contínuo único ou não.

A AFM mostra se existe um fator comum aos grupos e qual é esse fator. Se assim for, isto é, se essa caracterização global é adequada, é possível considerar o primeiro fator como um resumo do conjunto de características observadas nas unidades de estudo. E assim, torna-se possível observar quais são as variáveis de cada grupo que estão mais relacionadas com ele, ou seja, quais são, dentro do conjunto de características estudadas (variáveis ativas), as que mais diferenciam as unidades.

Em terceiro lugar descreve-se o indicador mostrando sua distribuição no conjunto de unidades. Trata-se de mostrar a sua distribuição em subconjuntos de unidades gerados pelas partições definidas pelas variáveis mais contributivas ao eixo para conhecer que valores do indicador correspondem a unidades que pertencem a mesma modalidade dessas variáveis.

3.2.5. Validação do indicador

Inicialmente procurar-se-á validar internamente o indicador. Trata-se de verificar se as propriedades desejadas estão sendo contempladas pelo indicador usando a informação com a qual foi gerado o próprio indicador.

Assim, para cada tema escolhido, deve-se construir uma tipologia das unidades de observação, com cada grupo de variáveis. A construção de uma tipologia das unidades de observação de uma tabela de dados do tipo "indivíduos x variáveis qualitativas" supõe a combinação da ACM com as técnicas de classificação e partição.

A análise fatorial permite determinar o número de dimensões pertinentes (fatores contínuos) que devem ser consideradas para descrever o essencial da estrutura da informação da tabela. Nesse sentido opera como um filtro das variações aleatórias ou não-sistemáticas dos dados. E é indispensável para verificar a classificabilidade das unidades já que permite observar a organização espacial das classes de indivíduos. Não é possível postular a existência de classes num conjunto de observações porque os elementos de uma tabela qualquer não necessariamente são classificáveis.

As técnicas de classificação permitem determinar o número de classes que correspondem a uma boa partição das n unidades observadas no seguinte sentido: unidades que pertencem a uma classe se assemelham e unidades de classes diferentes se diferenciam. As distâncias iniciais entre as n unidades se calculam no espaço dos primeiros eixos fatoriais já selecionados mediante a ACM.

Mediante os métodos de partição, se classificam as n unidades em um número de classes definido anteriormente, construindo dessa forma uma partição.

Imediatamente, se descrevem as classes pelas modalidades que mais as caracterizam. O processo compreende, então, as seguintes etapas:

- Realizar uma Análise de Correspondências Múltiplas, determinando o número de eixos que produzem um resumo suficiente da informação.
- Realizar uma Classificação com os eixos obtidos e determinar o número de classes razoáveis nas quais agrupar as unidades.
- Realizar uma Partição com o número de classes desejado.
- Caracterizar as classes obtidas, e determinar quais as unidades que pertencem a cada classe.

A Análise de Correspondências Múltiplas unida à Classificação e Partição de cada grupo de variáveis, resulta em tantas tipologias quanto o número de temas nos quais se agruparam as variáveis ativas da tabela. O estudo dessas partições conduz a conhecer quais as características que predominam nessas classes de unidades, e a detectar quais as características que as diferenciam.

Para comprovar se o indicador atribui valores semelhantes a unidades que pertencem à mesma classe cruzam-se os valores do indicador produzido com cada uma das tipologias associadas.

Finalmente procurar-se-á validar externamente o indicador. Trata-se de mostrar se o que o indicador está medindo corresponde com as informações conhecidas das unidades. Para isso, deve-se cruzar o indicador com variáveis que não foram usadas na sua construção, como as variáveis de estratificação (rede, localização, zona, região, estado).

Com as informações disponíveis, analisa-se a relação entre essa estrutura da informação sintetizada pelo indicador e as variáveis de estratificação; verificando, portanto, se o resultado obtido é coerente com o que se conhece das unidades por região, estado, rede, etc. As variáveis de estratificação não participaram na construção dos fatores de variabilidade, intervindo somente de forma ilustrativa ou explicativa.

3.2.6 Uso do indicador e a tipologia associada

A tipologia e o indicador podem ser empregados como um modelo de base sobre o observado para determinar o impacto previsível de diferentes ações de correção da situação global das unidades. As possibilidades de uso são variadas mas dependem obviamente do fenômeno que o indicador pretende refletir.

3.2.7. Validação da metodologia de elaboração do indicador

Trata-se de comparar o comportamento do indicador obtido pela metodologia proposta com indicadores obtidos de outras formas; sempre referindo-se ao cumprimento dos objetivos para os quais foi calculado o indicador. Para tanto se faz necessário:

- Construir outros indicadores por formas convencionais tais como a combinação linear de códigos - média simples e ou média ponderada por tema.
- Comparar os resultados analisando os elementos de discordância. Comparar como representam os diferentes indicadores as unidades em situações extremas (melhores e piores) e em situações intermédias. Observar como ficam segundo os diferentes indicadores as unidades de diferentes redes, regiões, estados, etc.

CAPITULO IV**4. O INDICADOR DE "RECURSOS MATERIAIS" DAS ESCOLAS****Sumário**

4.1. Etapas metodológicas de construção do indicador	067
4.1.1 Definir as finalidades do indicador	072
4.1.2. Verificação das condições prévias indispensáveis em os dados de observação disponíveis	072
4.1.3. Construção da tabela de dados para a construção do indicador	074
4.1.3.1. Seleção do conjunto homogêneo de escolas	074
4.1.3.2. Seleção das características observadas necessárias	075
4.1.3.3. Seleção das variáveis necessárias para validar o indicador	076
4.1.3.4. Determinar os "valores" que vão figurar nas celas da tabela de dados	077
4.1.3.5. Definir a estrutura temática da tabela	077
4.2. Construção do indicador de "recursos materiais" das escolas mediante a utilização da Análise Fatorial Múltipla sobre a tabela de dados	079
4.2.1. Análise Fatorial Múltipla da tabela em concordância com os temas já definidos	080
4.2.2. Influência dos recursos na diferenciação das escolas	081
4.2.3. Comparação das escolas segundo os recursos	083
4.2.4. Estudo das variáveis de estratificação na estrutura conjunta	094
4.2.5. A análise "local" da situação dos recursos materiais em uma escola na estrutura conjunta	098
4.2.6. Exemplo de análise local da situação dos recursos materiais de um conjunto de escolas (escolas de um estado)	101
4.2.7. Estudo das relações entre a análises da estrutura conjunta	

e as análises da estrutura de cada grupo de recursos	104
4.2.7.1. Grau de presença do primeiro fator dessa estrutura conjunta em cada grupo de variáveis	105
4.2.7.2. Descrição do indicador de "recursos materiais" das escolas	105
4.3. Validação do indicador de "recursos materiais" das escolas	111
4.3.1. Validação interna do indicador de "recursos materiais" das escolas	111
4.3.1.1. Tipologia das escolas caracterizadas pelos seus prédios	113
4.3.1.1.1. Primeira etapa da comparação das escolas pelo conjunto de descritores do estado do prédio Análise de Correspondências Múltiplas	113
4.3.1.1.2. Segunda etapa Classificação hierárquica mista das escolas baseada nos primeiros eixos da ACM do Grupo Prédio	116
4.3.1.1.3. Terceira etapa Partição do conjunto de escolas em classes com o método de nuvens dinâmicas	119
4.3.1.1.4. Caracterização das classes da tipologia de escolas pelas "condições dos prédios"	120
4.3.1.2. Tipologia das escolas descritas pelas instalações	127
4.3.1.2.1. Construção da tipologia de "condições das instalações" das escolas	127
4.3.1.2.2. Caracterização das classes pela tipologia de "condições dessas instalações" das escolas	128
4.3.1.3. Tipologia das escolas descritas pelos equipamentos	131
4.3.1.3.1. Construção da tipologia de "condições dos equipamentos" das escolas	131
4.3.1.3.2. Caracterização das classes pela tipologia de "condições dos equipamentos" das escolas	132
4.3.1.4. Tipologia das escolas descritas pelos serviços oferecidos	135
4.3.1.4.1. Construção da tipologia de "condições dos serviços" das escolas	135

4.3.1.4.2. Caracterização das classes pela tipologia de “condições dos serviços” das escolas	136
4.3.1.5 Cruzamento do indicador de “recursos materiais” com as tipologias das escolas por tipo de recursos	139
4.3.2. Validação externa do indicador de “recursos materiais das escolas	143
4.3.3 Limitações do indicador	147
4.3.3.1 Aspectos que limitam a validade do indicador de recursos materiais das escolas	148
4.3.3.2 Validação do indicador em termos da estabilidade da estrutura estudada	150
4.4. Uso do indicador e a tipologia associada.	151

O INDICADOR DE “RECURSOS MATERIAIS” DAS ESCOLAS

Para que o conjunto de informações a serem extraídas do banco de dados do SAEB pudessem subsidiar o processo de gestão da rede escolar (em nível federal, estadual ou municipal) fez-se necessário definir objetivos.

Na definição desses objetivos foram colocadas algumas “questões-tipo” que hipoteticamente poderiam ser formuladas pelo gestor do sistema educativo, para serem “resolvidas” a partir das informações disponíveis nas bases de dados do SAEB. O primeiro objetivo operacional proposto foi « construir » o conceito de recursos materiais das escolas - descritas pelo conjunto das variáveis observadas que caracterizam as instalações físicas e os equipamentos das escolas - e sintetizar essas informações num indicador.

Evidentemente, o conjunto de escolas do Brasil não dispõem dos mesmos recursos materiais. Se assim fosse, não seria necessário fazer uma avaliação dos recursos dessas escolas. Por isso, o gestor do sistema educativo poderia se interessar em saber como se assemelham e se diferenciam as escolas brasileiras segundo a situação de seus recursos materiais. A avaliação da situação dos recursos materiais do conjunto das escolas brasileiras interessa ao “gestor”

desses recursos em nível federal sempre que através de uma comparação objetiva das semelhanças e diferenças entre as escolas pudesse produzir um diagnóstico da situação de tais recursos o que responde a três preocupações maiores:

- Identificação das escolas que apresentam situações semelhantes em seus recursos materiais.
- Identificação dos recursos materiais que caracterizam os tipos de situações semelhantes que foram observados.
- Definição de prioridades entre as múltiplas ações corretivas que possam ser definidas a partir desse diagnóstico.

Essa identificação pode ser facilitada se o conjunto de recursos materiais fossem sintetizados num indicador que consiga diferenciar as situações mencionadas, facilitando a possibilidade de “dimensionar” e “focalizar” as políticas de correção de desigualdades na distribuição de recursos materiais.

Essas finalidades levaram a uma opção metodológica de tratamento estatístico da informação disponível, que conduziu a uma análise estrutural das informações do SAEB. Com efeito, reconhecendo os limites de uma modelização *a priori* da informação, com fins preditivos, procurou-se definir uma maneira de analisar a informação tendo por princípio o respeito as exigências dos níveis de comparabilidade que interessam ao decisor, privilegiando a modelização a partir da compreensão do fenômeno observado.

Assim, do reconhecimento de grupos de variáveis, definiram-se unidades temáticas que possibilitaram a construção dos conceitos operacionais (no caso recursos materiais).

O estudo da estrutura da Tabela Múltipla de “recursos materiais” do SAEB97, no espaço das escolas da amostra observada permite:

- Comparação das escolas da amostra segundo a situação dos prédios, das instalações, dos equipamentos e dos serviços.
- Comparação das escolas da amostra segundo o conjunto destes critérios, equilibrando a importância acordada aos mesmos.

- Construção de uma variável numérica que sintetize as comparações das escolas, ordenando-as segundo a sua situação de recursos materiais.
- Análise da distribuição dessa variável segundo as variáveis de estratificação.

O estudo da estrutura da Tabela Múltipla de “recursos materiais” do SAEB97, nos espaços: das Tabelas que compõem a Tabela Múltipla; das variáveis dessa tabela; das modalidades de presença-ausência de cada componente dos recursos materiais observados permite:

- Estudo da influência de cada tipo de recursos (prédios, instalações, equipamentos e serviços) na diferenciação observada nas escolas.
- Análise da contribuição da presença ou ausência de cada item observado de recursos, à diferenciação observada das escolas.
- Comparação, na estrutura da Tabela Múltipla, da situação característica dos recursos materiais das escolas, segundo as regiões, estados e redes de ensino.

A seqüência dos conteúdos das principais etapas da construção de um indicador de “recursos materiais” das escolas pode ser apresentada em forma esquemática.

4.1. Etapas metodológicas de construção do indicador

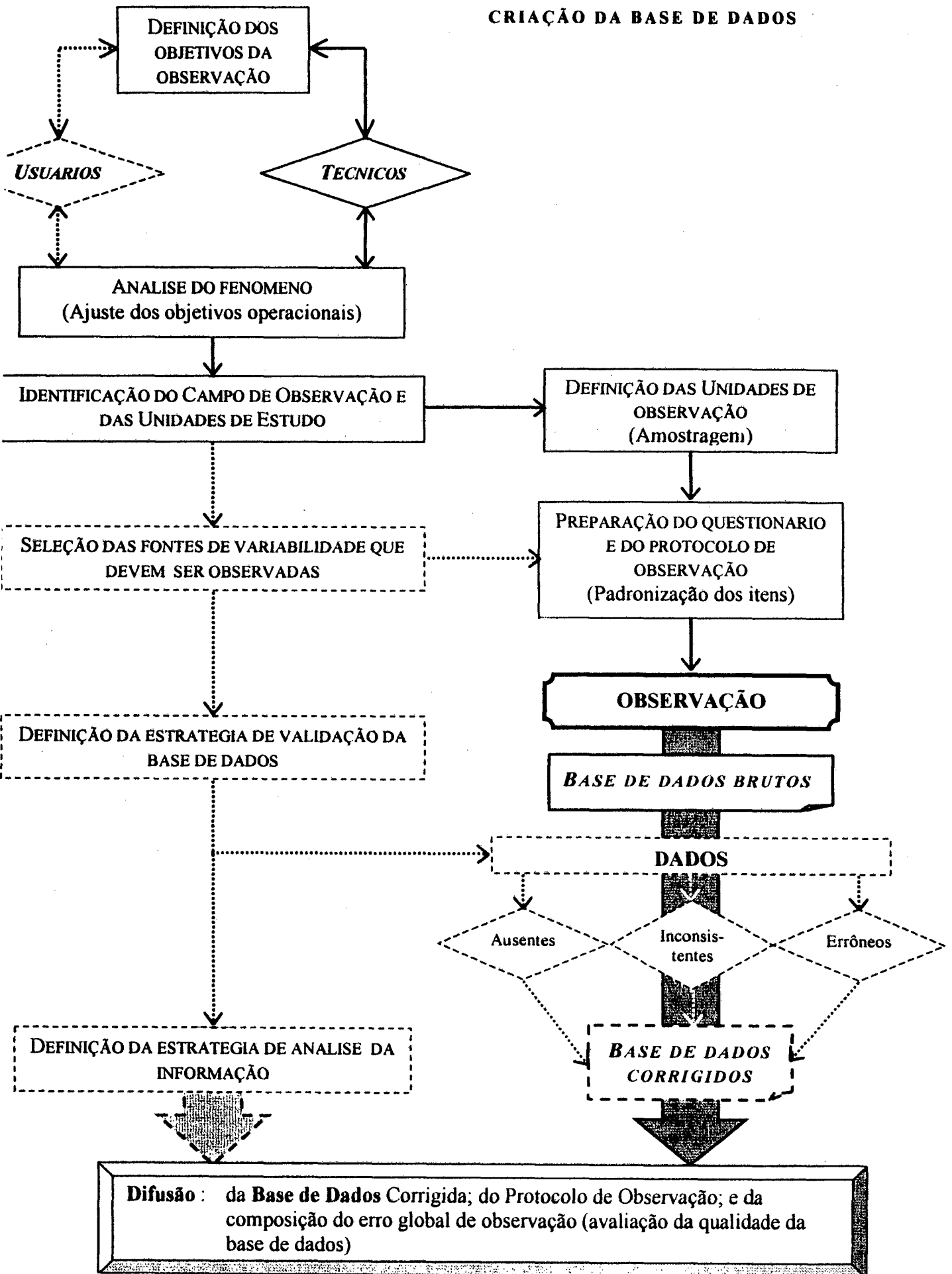
Neste item será desenvolvida a metodologia proposta no Capítulo 3 para construção do indicador. Inicialmente dispôs-se das bases de dados do SAEB97 observando os seus limites e realizando o tratamento de valores faltantes na base de dados escolhida.

A estratégia de construção do indicador propriamente dita inicia-se, primeiramente, com a definição de suas finalidades e com a construção da tabela de dados observados. Tal construção requer que seja escolhido um conjunto homogêneo de unidades de observação, requer a escolha das variáveis necessárias para descrever o objeto de estudo (variáveis ativas), a escolha das variáveis para validação (variáveis de estratificação), a determinação dos valores

das células da tabela e a definição de sua estrutura temática. Em segundo lugar dá-se a construção do indicador através da utilização da técnica de AFM para definição do primeiro fator e de sua importância para o conjunto das informações, verificando o grau de preservação desse primeiro fator e descrevendo-o. Em terceiro lugar, com a descrição do primeiro fator – tido então como o indicador pretendido – procura-se efetuar sua validação. Tal validação se dá segundo uma lógica interna e outra externa. A primeira compreende a combinação das técnicas estatísticas de ACM, Classificação e Partição, para identificar os agrupamentos de escolas segundo cada tipo de recursos materiais e um estudo da coerência entre os valores atribuídos pelo indicador e esses agrupamentos. A segunda, a validação externa, é verificada através do estudo da correspondência entre as informações sintetizadas pelo indicador e as variáveis ilustrativas de estratificação das escolas. Finalizando, descreve-se o comportamento do indicador segundo a relação entre os possíveis usos do mesmo.

Algumas das etapas preliminares são parcialmente cumpridas pelo SAEB. Esquemáticamente, a seqüência dos conteúdos das principais etapas da construção de um indicador de "recursos materiais" das escolas é a seguinte.

Figura 4.0. Etapas da construção do indicador de recursos materiais



BASE DE DADOS CORRIGIDA

FENÔMENO OBSERVADO

ETAPAS

PRIMEIRA ETAPA: PREPARAÇÃO DAS TABELAS DE DADOS

SEGUNDA ETAPA: CONSTRUÇÃO DA VARIÁVEL INDICADORA

TERCEIRA ETAPA: ANÁLISE ESTRUTURAL DOS RECURSOS MATERIAIS

DEFINIÇÃO DOS OBJETIVOS DA OBSERVAÇÃO

USUÁRIOS

TÉCNICOS

FINALIDADES DO INDICADOR DE « RECURSOS MATERIAIS » DAS ESCOLAS
(Identificação dos usos previstos do indicador)

OPERACÕES

- Seleção de um conjunto homogêneo de escolas;
- Seleção dos itens observados necessários;
- Seleção das variáveis de validação do indicador ;
- Definição da estrutura temática da Tabela $T(T_1, \dots, T_K)$ de Recursos Materiais.

- Análise Fatorial Múltipla da Tabela $T(T_1, \dots, T_K)$.

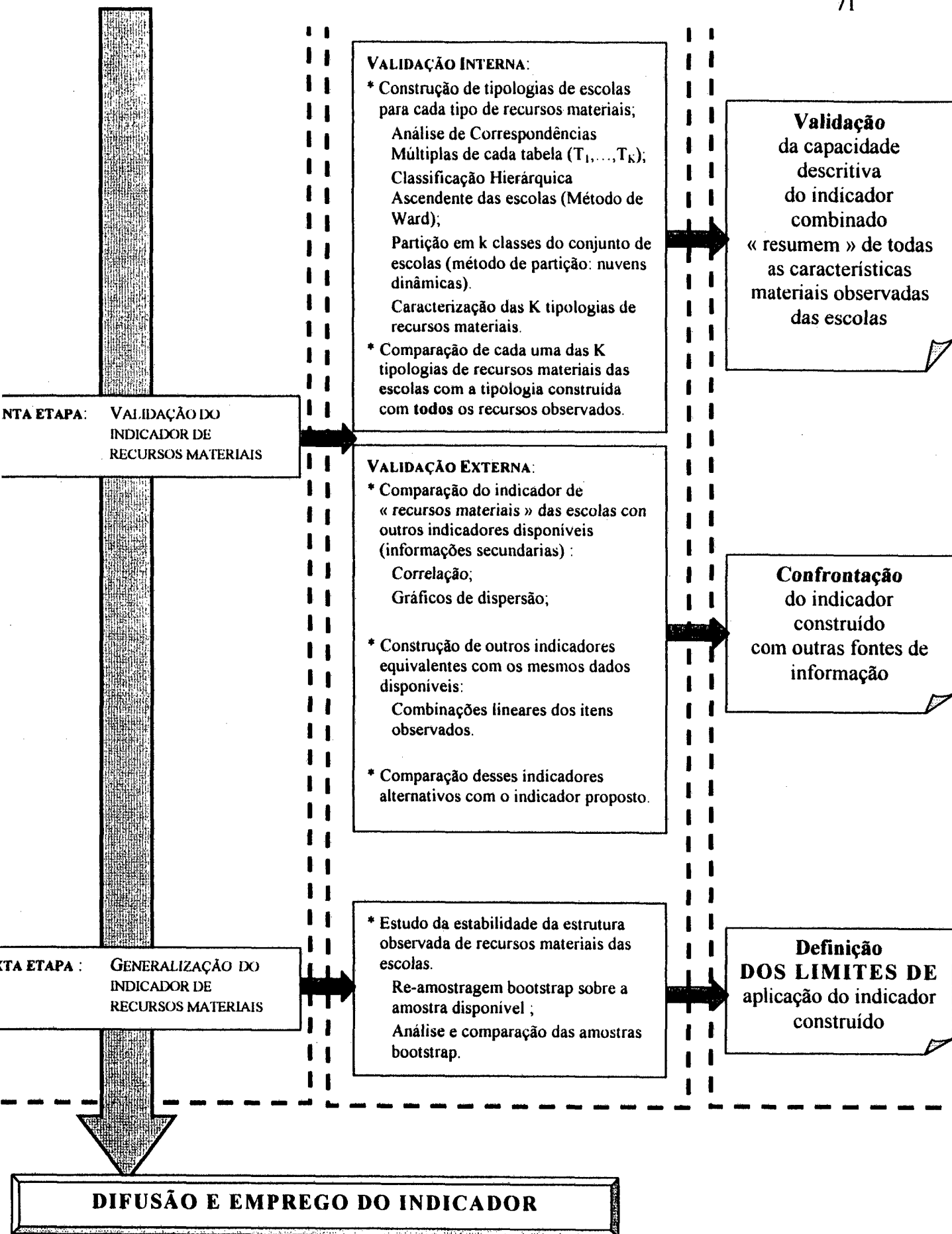
- Comparações das distribuições de recursos materiais entre grupos específicos de escolas ;
- Análise local ;
- etc... ;
- Interpretação do conjunto dos resultados da AFM.

PRODUTO

Tabela $T(T_1, \dots, T_K)$
Tabela de Dados Estruturada segundo los grupos temáticos

Variável « recursos materiais » das escolas
Primeiro eixo fatorial da AFM da Tabela $T(T_1, \dots, T_K)$

Resolução dos objetivos específicos de diferentes usuários da informação



4.1.1 Definir as finalidades do indicador

Um indicador de recursos materiais das escolas observadas no SAEB97 deve cumprir com as seguintes propriedades; já descritas no capítulo 1:

- " pertinência com respeito à finalidade para a qual foi produzido. Assim, deve dar conta de todas as características físicas de infra-estrutura e disponibilidades da escola outorgando a cada uma delas ponderadores objetivos de importância.
- capacidade descritiva: deve refletir as diferenças em equipamento das escolas, agrupando aquelas com características físicas semelhantes em valores semelhantes.
- que esteja relacionado com as diferentes formas de intervenção que permitem melhorar a realidade representada globalmente pelo indicador; ou seja que deve permitir saber que ações definir para melhorar a infra-estrutura das diferentes escolas.
- que permita a comparabilidade no tempo e no espaço: deve permitir comparar as condições de infra-estrutura em períodos diferentes (duas aplicações do SAEB), e entre duas áreas geográficas (regiões, estados). Por exemplo deve permitir - do ponto de vista do equipamento das escolas- comparar a situação de um estado com respeito a todos os estados do Brasil.
- que seja uma variável única e contínua, associada a uma tipologia das condições de infra-estrutura das escolas."

4.1.2. Verificação das condições prévias indispensáveis nos dados de observação disponíveis

Um problema metodológico a tratar é a presença de valores faltantes ou "*missing values*", depois de unificar as observações repetidas. O uso de uma metodologia de tratamento dos valores faltantes - imputação multivariada usando o software *SOLAS for Missing Data Analysis* - permitiu utilizar como indivíduos

ativos da análise as 1875 escolas que apresentaram até seis "missing values", o que corresponde a 97% das escolas que participaram do exame do SAEB97. A exposição dos métodos de imputação para o tratamento dos valores faltantes foge aos limites desse trabalho. Para maiores esclarecimentos sobre o tema sugerimos consultar a seguinte bibliografia: Crivisqui, Eduardo (1994) " Les processus d'observation systematique des phénomènes sociaux- Place et fonction des méthodes statistiques. Thèse, Institute de Statistique et de Recherche Opérationelle, Université Libre de Bruxelles, février 1994, 261 p., Belgique.

A Tabela 4.1 apresenta a distribuição do número de "missing values" por escola, as linhas ressaltadas mostram o corte efetuado pelo pesquisador quando da escolha dos indivíduos que farão parte da sua base de dados.

Tabela 4.1. Distribuição de freqüências dos "MISSING VALUES" nas escolas participantes do SAEB97 em todo o País (%)

MISSING VALUES	FREQUÊNCIA	PORCENTAGEM	PORCENTAGEM ACUMULADA
0	1395	72,2	72,2
1	361	18,7	90,8
2	57	2,9	93,8
3	26	1,3	95,1
4	13	,7	95,8
5	9	,5	96,3
6	14	,7	97,0
7	4	,2	97,2
8	5	,3	97,5
9	3	,2	97,6
10	2	,1	97,7
11	5	,3	98,0
12	4	,2	98,2
13	2	,1	98,3
14	2	,1	98,4
16	1	,1	98,4
18	1	,1	98,5
19	2	,1	98,6
21	1	,1	98,7
23	1	,1	98,7
34	1	,1	98,8
36	1	,1	98,8
37	23	1,2	100,0
Total	1933	100,0	

4.1.3. Construção da tabela de dados para a construção do indicador

Para atender aos objetivos propostos, deve-se preparar uma tabela estatística adequada: $X(n \times p)$, tendo nas linhas as unidades de estudo (indivíduos) e nas colunas as características observadas (variáveis).

4.1.3.1 Seleção do conjunto homogêneo de escolas

Em primeiro lugar, se deve escolher os elementos que comporão as linhas da tabela de dados $X (n \times p)$, isto é, o conjunto homogêneo das unidades de observação.

A escolha da unidade de observação recaiu sobre as escolas tendo em vista que:

- A utilização do indicador a ser construído está voltada para o processo de gestão dos recursos disponibilizados para os estabelecimentos de ensino.
- O exemplo utilizado para construção do indicador baseou-se nos recursos materiais dos estabelecimentos de ensino compreendendo um conjunto de informações que possibilita a elaboração de um diagnóstico das condições da distribuição e conservação desses recursos junto a rede escolar do país.
- a escola é uma unidade administrativa no processo de gestão dos recursos disponibilizados.

Conforme o exposto acima e os objetivos propostos definiu-se como unidade de análise a escola. Sendo que as unidades de observação que participaram da amostra observada pelo SAEB97 foram 1933.

Mas, como elementos ativos da análise, (em conformidade com o estudo dos "missing values") serão utilizadas 1875 escolas que são aquelas que tem até 6 "missing values" do total das 37 características observadas.

Figura 4.1. Indivíduos incluídos na AFM

INDIVIDUS	NOMBRE	POIDS
POIDS DES INDIVIDUS: Poids des individus, uniforme egal a 1. UNIF		
RETENUS NITOT =	1933	PITOT = 1933.000
SELECTION APRES FILTRAGE		
ACTIFS NIACT =	1875	PIACT = 1875.000
SUPPLEMENTAIRES NISUP =	58	PISUP = 58.000

4.1.3.2. Seleção das características observadas necessárias

Deve-se selecionar as variáveis observadas que são necessárias para descrever o objetivo do estudo. Essas variáveis serão as colunas ativas da tabela de dados X (n x p).

Para construir o conceito de recursos físicos das escolas se fez necessário reunir as características físicas das escolas na tabela de dados X(n x p).As características físicas das escolas estão armazenadas na base ESC97.DBF. A base contém 37 itens relativos ao estado de conservação do prédio, das instalações, dos equipamentos e dos serviços oferecidos. Essas informações foram levantadas através da aplicação do questionário da escola por ocasião da realização das provas de proficiência. Essas 37 variáveis serão usadas para comparar as escolas.

Figura 4.2. Variáveis ativas incluídas na AFM

VARIABLES NOMINALES ACTIVES		
37 VARIABLES	133 MODALITES ASSOCIEES	
6 . Telhado	(4 MODALITES)	
7 . Paredes	(4 MODALITES)	
8 . Piso	(4 MODALITES)	
9 . Portas e janelas	(4 MODALITES)	
10 . Banheiros	(4 MODALITES)	
11 . Cozinha	(4 MODALITES)	
12 . Instalações hidráulicas	(4 MODALITES)	
13 . Instalações elétricas	(4 MODALITES)	
14 . Ventilação	(4 MODALITES)	
15 . Iluminação	(2 MODALITES)	
16 . Ruidos	(2 MODALITES)	
17 . Area externa(patio, jardins)	(2 MODALITES)	
18 . Espaço para recreação	(2 MODALITES)	
19 . Biblioteca	(4 MODALITES)	
20 . Laboratorio de Ciências	(4 MODALITES)	
21 . Laboratorio de Informatica	(4 MODALITES)	
22 . Auditorio	(4 MODALITES)	
23 . Quadra de esportes ou Ginasio	(4 MODALITES)	
24 . Vestiarios	(4 MODALITES)	
25 . Sala de Professores	(4 MODALITES)	
26 . Sala de Aula	(4 MODALITES)	
27 . Qualidade de livros	(3 MODALITES)	
28 . Quadro negro	(4 MODALITES)	
29 . Carteiras	(4 MODALITES)	
30 . Mesas	(4 MODALITES)	
31 . Televisão	(4 MODALITES)	
32 . Video-cassete	(4 MODALITES)	
33 . Mimeografo	(4 MODALITES)	
34 . Maquina fotocopadora	(4 MODALITES)	
35 . Projetor de slides	(4 MODALITES)	
36 . Retroprojetor	(4 MODALITES)	
37 . Maquinas de datilografia	(4 MODALITES)	
38 . Computador	(4 MODALITES)	
39 . Serviços médicos	(3 MODALITES)	
40 . Serviços odontologicos	(3 MODALITES)	
41 . Transporte	(3 MODALITES)	
42 . Alimentação	(3 MODALITES)	

4.1.3.3. Seleção das variáveis necessárias para validar o indicador

Essas variáveis compõem as colunas ilustrativas da tabela de dados

X (n x p).

Na base ESC97.DBF também estão contidas as variáveis de estratificação das escolas: zona (urbana ou rural), localização (capital ou interior), rede (estadual, municipal e particular), região e unidade da federação.

Essas variáveis que identificam a região, estado, zona e localização não devem participar da comparação das escolas, intervindo somente de forma ilustrativa ou explicativa.

Figura 4.3. Variáveis ilustrativas incluídas na AFM

VARIABLES NOMINALES ILLUSTRATIVES	
7 VARIABLES	49 MODALITES ASSOCIEES
1 . Zona	(2 MODALITES)
2 . Rede	(3 MODALITES)
3 . Localização	(2 MODALITES)
4 . Região	(5 MODALITES)
5 . Unidades da Federação	(27 MODALITES)

4.1.3.4. Determinar os "valores" que vão figurar nas células da tabela de dados

Deve-se definir os valores que figurarão nos cruzes das linhas e colunas da tabela X (n x p).

Como a tabela em estudo tem nas linhas as escolas e nas colunas as 37 características de recursos observadas por meio do questionário das escolas, e as variáveis de estratificação; os valores que vão figurar nas células são os códigos numéricos correspondentes a modalidade observada em cada uma das variáveis.

4.1.3.5. Definir a estrutura temática da tabela.

Segundo o ponto de vista do pesquisador, as p variáveis devem ser organizadas em temas que respondem aos objetivos do estudo.

Do ponto de vista do estudo proposto aqui, teríamos um conjunto de características de recursos materiais pelas quais vamos comparar as escolas (variáveis ativas), e outro das variáveis de estratificação (variáveis ilustrativas).

Alem disso, do estudo do questionário é possível identificar que as variáveis ativas que caracterizam os recursos materiais estão estruturadas nos seguintes grupos :

- Grupo 1: *Prédio*
13 itens relativos ao estado de conservação de telhados, paredes, pisos, etc. do prédio e de aspectos relacionados com a ventilação, iluminação, etc.(itens 1 à 13 do questionário).

- **Grupo 2: Instalações**
9 itens referentes às condições de funcionamento da biblioteca, laboratórios de ciências e informática, etc. (itens 14 à 22).
- **Grupo 3: Equipamentos**
11 itens relativos ao estado de conservação dos equipamentos da escola (quadro, mesas, mimeógrafo, retroprojctor, etc.). Correspondem aos itens 23 à 33 do questionário.
- **Grupo 4: Serviços**
4 itens referentes aos serviços oferecidos pela escola: médicos, serviços odontológicos, transporte e alimentação. (itens 34 à 37)

Assim, a estrutura temática da tabela pode vir a indicar diferentes formas possíveis de intervenção a partir da relação que o decisor venha a fazer entre os aspectos que compõem os diferentes grupos de recursos, e o conjunto de aspectos passíveis de alterações por parte do decisor e os recursos disponíveis.

A Figura 4.4 apresenta um esquema sobre a tabela de dados estruturada segundo os grupos temáticos definidos.

Figura 4. 4. Esquema da tabela de dados

INDIVÍDUOS	GRUPO PRÉDIO					GRUPO INSTALAÇÕES				GRUPO EQUIPAMENTOS					GRUPO SERVIÇOS																									
	TELHADO	PAREDES	PISO	PORTAS E JANELAS	BANHEIROS	COZINHA	INSTALAÇÕES HIDRAULICAS	INSTALAÇÕES ELETRICAS	VENTILAÇÃO	ILLUMINAÇÃO	RUIDOS	ASPECTOS DA AREA EXTERNA (PATIO, JARDINS)	ESPAÇA PARA RECREAÇÃO INFANTIL	BIBLIOTECA	LABORATORIO DE CIÊNCIAS	LABORATORIO DE INFORMATICA	AUDITORIO	QUADRA DE ESPORTES OU GINÁSIO	VESTIARIOS	SALA DE PROFESSORES	SALAS DE AULA	QUALIDADE DOS LIVROS	QUADRO NEGRO	CARTEIRAS	MESAS	TELEVISÃO	VIDEO-CASSETTE	MIMEOGRAFO	MAQUINA FOTOCOPIADORA	PROJETOR DE SLIDES	RETROPROJETOR	MAQUINAS DE DATILOGRAFIA	MÉDICOS	ODONTOLÓGICOS	TRANSPORTE	ALIMENTAÇÃO				
	1V_1	...	1V_7	...	$^1V_{11}$	2V_1	...	2V_3	...	2V_9	3V_1	...	3V_3	...	$^3V_{11}$	4V_1	...	4V_4																						
1	3	...	2	...	4	1	...	3	...	2	2	...	4	...	1	1	...	3																						
...																																								
i	$^1X_{i7}$	$^2X_{i3}$	$^3X_{i5}$	$^4X_{i4}$																						
...																																								
n	$^1X_{n7}$	$^2X_{n3}$	$^3X_{n5}$	$^4X_{n4}$																						

4.2. Construção do indicador de "recursos materiais" das escolas mediante a utilização da Análise Fatorial Múltipla

O indicador procurado é resultado da análise da estrutura da tabela múltipla previamente definida. O primeiro eixo resultante da Análise Fatorial Múltipla é uma variável numérica que resume o conjunto de características físicas observadas e estruturadas nos grupos definidos. Trata-se de mostrar que esse primeiro eixo tem as propriedades desejadas para o indicador.

4.2.1. Análise Fatorial Múltipla da tabela em concordância com os temas já definidos

Em primeiro lugar vemos - na Tabela 4.5 - que o primeiro eixo é uma direção de inércia importante; portanto ele capta grande parte dos resultados das comparações das escolas através de todas as características dos 37 recursos físicos observadas e estruturadas em 4 grupos.

Figura 4.5. Decomposição da inércia (na análise global)

HISTOGRAMME DES 5 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	2.7560	12.63	12.63	*****//
2	0.9167	4.20	16.83	*****
3	0.8510	3.90	20.73	*****
4	0.8023	3.68	24.41	*****
5	0.6981	3.20	27.61	*****

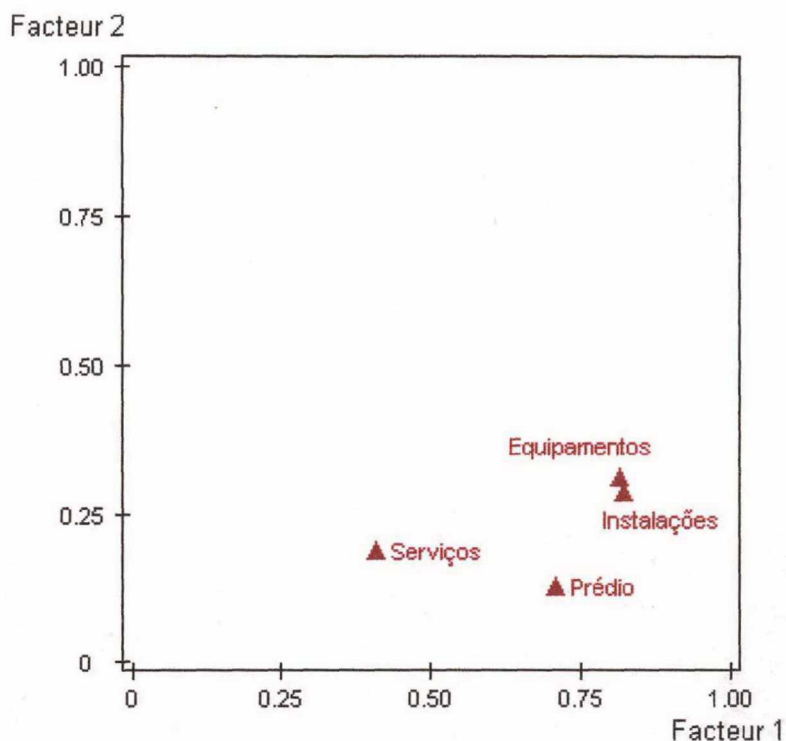
Um dos objetivos da pesquisa era conseguir um indicador que fosse uma variável contínua e única que refletisse a estrutura de recursos físicos das escolas. O uso da AFM assegura que a variabilidade de todos os recursos físicos dos quatro grupos está sendo considerada pelo primeiro eixo; é uma variável contínua por construção; e é de máxima variabilidade, também por construção. Dada a importância do primeiro eixo – vista na decomposição da inércia- pode tornar-se o indicador procurado, no sentido de ser único.

Como se mostrará na seção 4.2.3., o primeiro eixo ordena as escolas de esquerda a direita em forma crescente de valores da escala. As escolas com valores menores do indicador proposto - aquelas com coordenadas baixas no primeiro eixo- são as que apresentam as piores situações de recursos materiais e as escolas com valores maiores do indicador proposto –aquelas com coordenadas altas no primeiro eixo- são as que estão em melhores condições de recursos materiais. Através do estudo realizado na seção 4.2.3 se mostraram as diferentes situações intermédias de recursos materiais das escolas representadas na escala do fator.

4.2.2. Influência dos recursos na diferenciação das escolas

Para estudar a *influência de cada tipo de recursos* (prédios, instalações, equipamentos e serviços) na diferenciação observada das escolas, comparam-se os grupos de recursos materiais na estrutura global da Tabela Múltipla. Nesta análise, dois grupos de recursos se assemelham se duas escolas próximas em um grupo, também são próximas em outro. Por exemplo: o grupo « prédio » será próximo do grupo « instalações » se as escolas que se assemelham pelas condições dos prédios também se assemelham pelas condições das instalações. Assim, dois grupos são mais próximos quanto mais se assemelhem as estruturas definidas por eles no conjunto das escolas.

Para realizar a leitura do gráfico seguinte (Figura 4.6) temos que considerar que a coordenada de um grupo de variáveis sobre um eixo pode ser considerada como uma medida da relação entre o grupo e o fator correspondente. Os valores dessa coordenada estão compreendidos entre 0 e 1. Se a coordenada de um grupo é próxima de 1 quer dizer que esse fator está ligado a um grande número de variáveis do grupo porque esta coordenada representa a variabilidade acumulada das variáveis do grupo sobre o eixo correspondente.

Figura 4.6. Comparação dos grupos de recursos materiais

O primeiro eixo é – por construção- a direção de máxima variabilidade. Para os quatro grupos ativos o primeiro eixo é uma direção importante. Os grupos mais importantes na determinação do primeiro eixo são os grupos referentes as instalações e aos equipamentos seguidos pelo grupo de características do prédio. O grupo menos importante na formação da variabilidade do primeiro eixo é serviços que apresenta a menor relação com ele e, portanto, pesa menos na diferenciação das escolas. Por outra parte, o grupo de equipamentos está muito próximo do grupo das instalações, o que mostra que as escolas que se assemelham nos equipamentos também se assemelham pelas instalações.

Essa situação é diferente para o segundo eixo pois as coordenadas dos quatro grupos são pequenas: os grupos são praticamente unidimensionais, e do ponto de vista prático, as diferenças das escolas serão detectadas pelo primeiro eixo. No segundo eixo não vão aparecer as diferenças nas características referentes aos grupos dos serviços e dos prédios. Isto é, praticamente toda a variabilidade entre as variáveis que compõem esses grupos são detectadas no primeiro eixo. Para os grupos referentes as instalações e aos equipamentos o segundo eixo representa

uma direção de variabilidade complementar porque apresentam variabilidade nos dois eixos sendo considerada alta no primeiro e baixa no segundo. Os valores das coordenadas dos grupos sobre os dois primeiros eixos figuram na Figura 4.7.

Figura 4.7. Coodenadas e ajudas à interpretação dos quatro grupos ativos

AXES		1					A					5										
GRP.	P.REL	DISTO	COORDONNEES					CONTRIBUTIONS					COSINUS CARRES									
			1	2	3	4	5	1	2	3	4	5	1	2	3	4	5					
GR 1	0.25	1.79	0.71	0.13	0.16	0.02	0.19	25.8	14.0	19.4	2.3	27.0	0.28	0.01	0.02	0.00	0.02					
GR 2	0.25	1.99	0.82	0.29	0.10	0.01	0.08	29.8	31.3	12.2	1.6	11.3	0.34	0.04	0.01	0.00	0.00					
GR 3	0.25	2.56	0.81	0.31	0.10	0.04	0.12	29.6	34.0	12.3	4.5	17.0	0.26	0.04	0.00	0.00	0.01					
GR 4	0.25	2.91	0.41	0.19	0.48	0.73	0.31	14.8	20.7	56.1	91.5	44.8	0.06	0.01	0.08	0.19	0.03					
			ENSEMBLE					=100.0					100. 100. 100. 100. 100.					+ 0.22 0.03 0.03 0.06 0.02 +				

Para o primeiro eixo (inércia=2,75), as inércias das variáveis dos três primeiros grupos são: 0,71, 0,82, 0,81; esse primeiro eixo está muito relacionado com os grupos: prédio, instalações e equipamentos, no sentido que constitui uma direção de inércia importante para esses grupos. Não acontece isso com o quarto grupo: os serviços, com coordenada 0,41. Enquanto as contribuições à inércia do segundo eixo são baixas para todos os grupos, mas a inércia recolhida por esse segundo eixo (0, 91) já é baixa em relação ao primeiro, como já visto na Tabela 4.5. de decomposição da inércia.

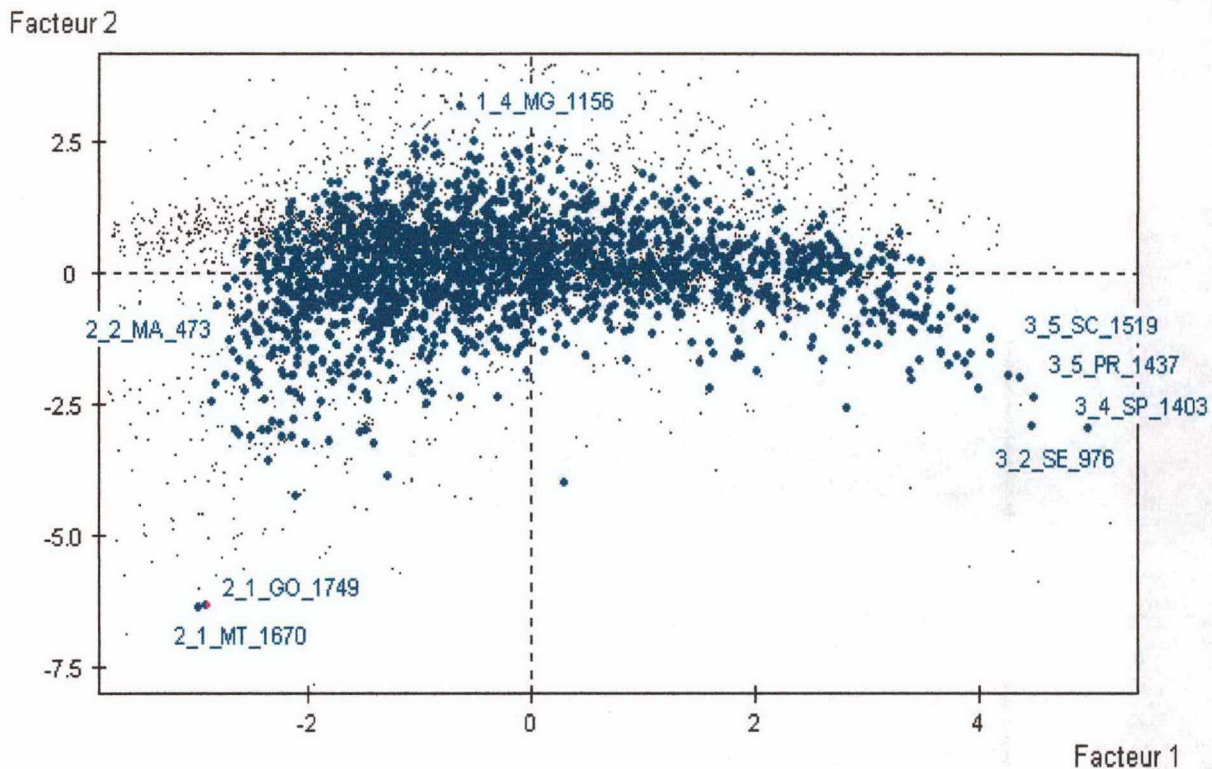
Uma das propriedades requisitadas do indicador é que leve em consideração todas as características de infra-estrutura e disponibilidades da escola, outorgando a elas ponderadores objetivos de importância. Esses pesos para os grupos de variáveis são objetivos no sentido que a relação deles com o eixo depende da capacidade de diferenciar as escolas que as variáveis do grupo tem.

4.2.3. Comparação das escolas segundo os recursos

O estudo da estrutura da Tabela Múltipla de "recursos materiais" do SAEB97, no espaço das escolas da amostra observada permitirá *comparar as escolas segundo a situação dos prédios, das instalações, dos equipamentos e dos serviços*. Na AFM essa comparação equilibra a importância acordada aos grupos de variáveis usadas como critério de comparação. O gráfico seguinte mostra a

posição das escolas no primeiro plano fatorial da análise global. A proximidade de duas escolas indica a semelhança delas na estrutura de recursos físicos.

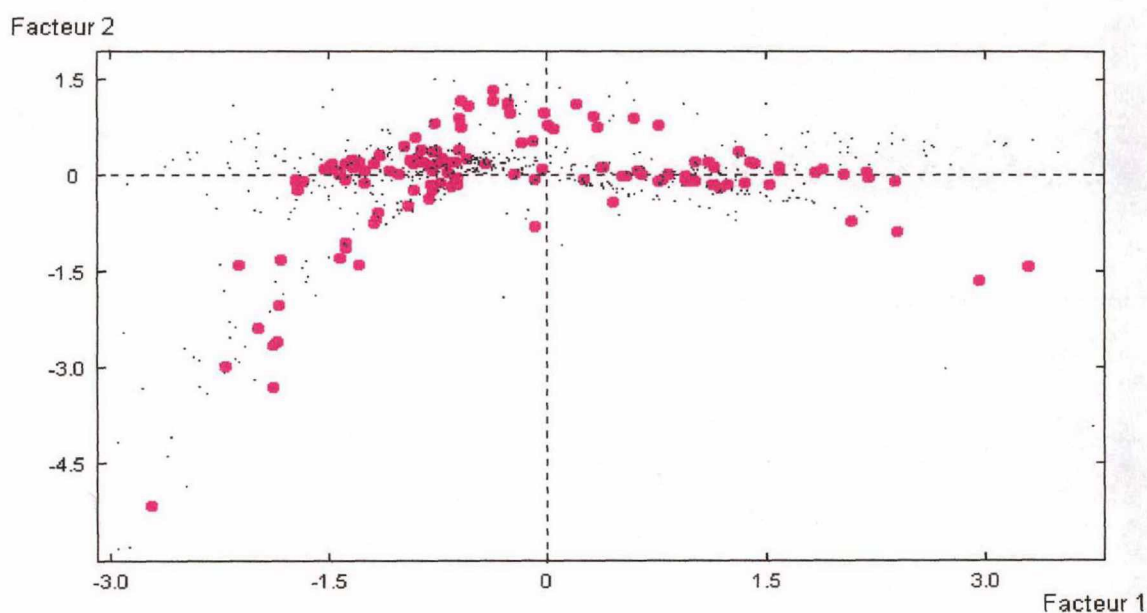
Figura 4.8. Projeção das escolas na estrutura conjunta dos grupos de recursos materiais (projeção da nuvem dos centros de gravidade dos indivíduos).



Foram identificadas algumas das escolas em situações extremas. O identificador de cada escola compõe-se de: o código da rede, o código da região, a sigla da unidade da federação e o número outorgado à escola na base de dados. Assim o símbolo 3_5_SC_1519 identifica a escola 1519 da rede particular, região sul, estado de Santa Catarina. Essa codificação está fornecida na documentação do SAEB97 citada na bibliografia. Temos a direita quatro escolas que pertencem à rede particular dos estados PR, SC, SP e SE; na extrema esquerda inferior duas escolas da rede municipal de MT e GO. Este gráfico resulta muito intrincado devido a quantidade de escolas. A proximidade dessas escolas obedece a sua semelhança na estrutura de seus recursos materiais. Para poder interpretar a situação de recursos materiais das escolas temos que identificar as

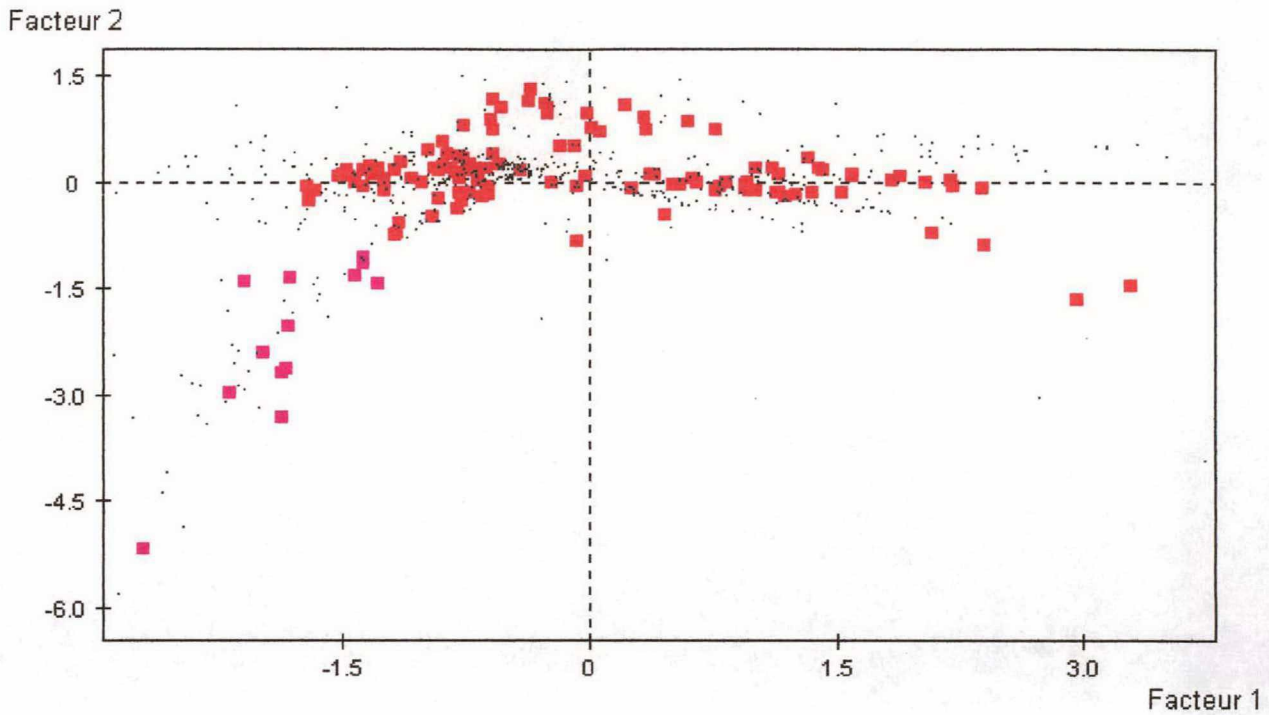
características pelas quais se assemelham as escolas que estavam próximas no gráfico anterior (figura 4.8). E para determinar em que se assemelham as escolas que aparecem próximas nesta representação, é preciso analisar qual a *contribuição da presença ou ausência de cada item observado de recursos, a essa diferenciação observada nas escolas*. Isso leva também ao estudo das relações entre as modalidades das variáveis de recursos na estrutura conjunta dos grupos de recursos.

Figura 4.9. Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos



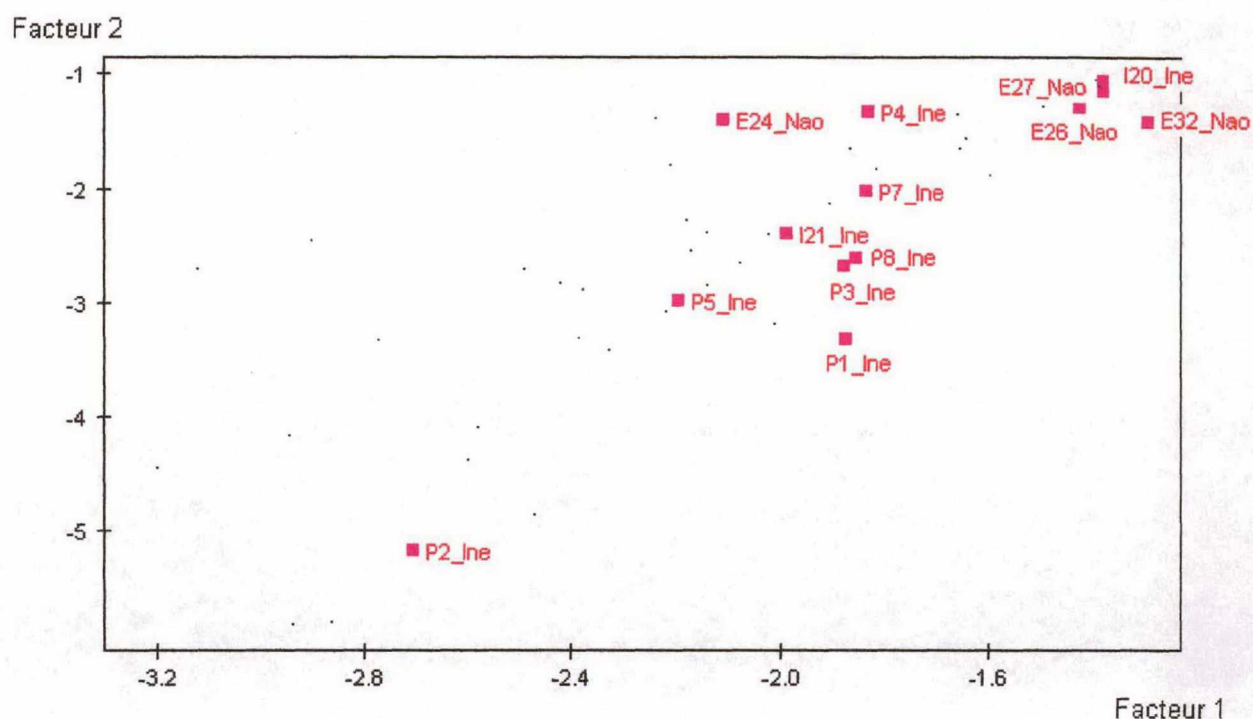
Para interpretar este gráfico faremos um "zoom" das diferentes zonas que o compõe. Como o primeiro eixo revela as principais características que opõem as escolas, vamos decompor esse gráfico pela oposição fundamental, que vai mostrar como são as escolas que se encontram no extremo direito e no extremo esquerdo do primeiro eixo no gráfico anterior, ou seja, começaremos pelas situações extremas de estrutura de recursos materiais para identificar quais os tipos de recursos que caracterizam essas situações extremas.

Figura 4.10. Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (situação extrema esquerda)



Os rótulos das modalidades identificam pela primeira letra o grupo ao qual pertencem (prédio, instalações, equipamentos, serviços); o número corresponde ao item do questionário e as letras seguintes identificam a modalidade. Por exemplo, P2_Ine identifica o item 2 do questionário (paredes) como inexistente e esse item faz parte do grupo Prédio.

Figura 4.11. Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (“zoom” situação extrema esquerda)



A Figura 4.11 é uma ampliação da Figura 4.10 contendo apenas as modalidades situadas no extremo esquerdo inferior do plano. Nessa zona do plano o que caracteriza as escolas é o fato de não ter paredes, banheiros, piso, telhado, instalações elétricas e hidráulicas, portas e janelas (referente ao Prédio); não ter salas de professores nem de aula (referente as Instalações); não ter carteiras, televisão, videocassete e nem máquinas de datilografia (referente aos Equipamentos). Note-se que nenhum dos serviços caracterizam as escolas desta zona do plano, e que as características das instalações e equipamentos que diferenciam estas escolas são as que poderíamos chamar de básicas (salas, carteiras). É a pior situação de recursos materiais das escolas observadas.

Para esse conjunto de escolas pode-se recomendar – conforme o diagnóstico acima - uma ação que priorize a conservação e construção de prédios.

Figura 4.12. Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (situação extrema direita)

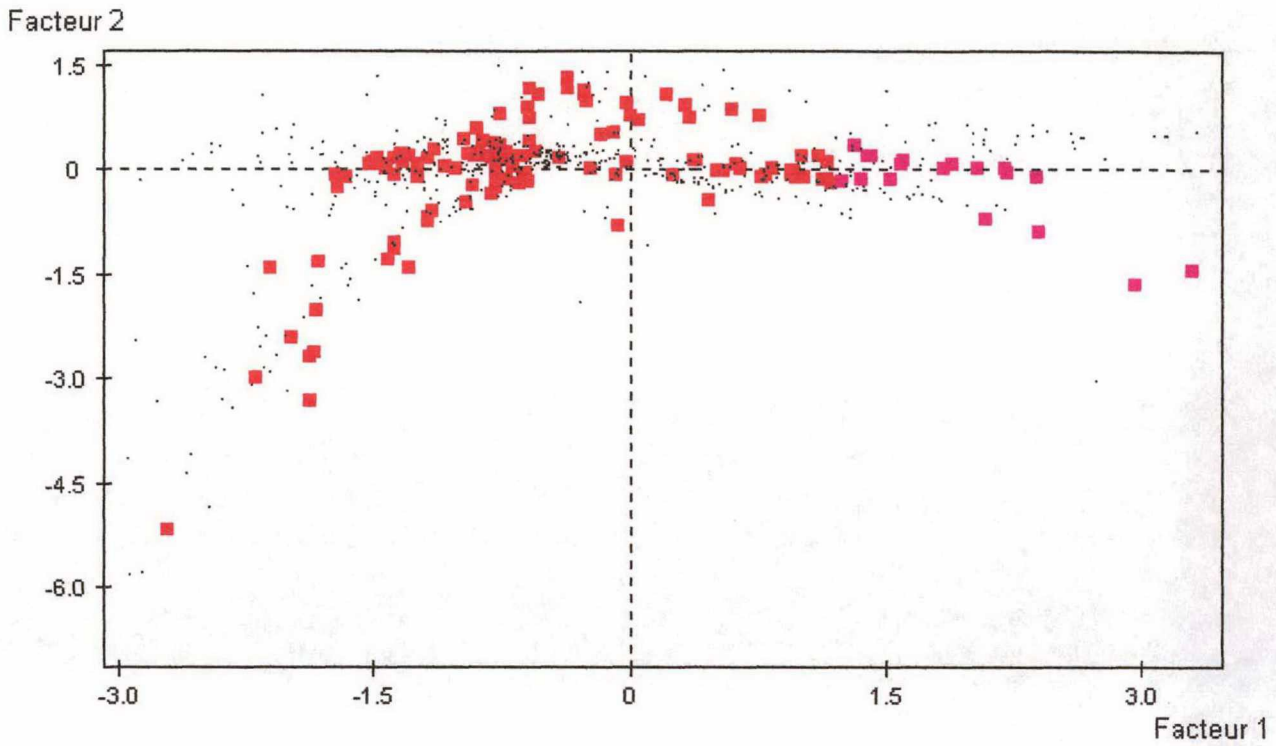
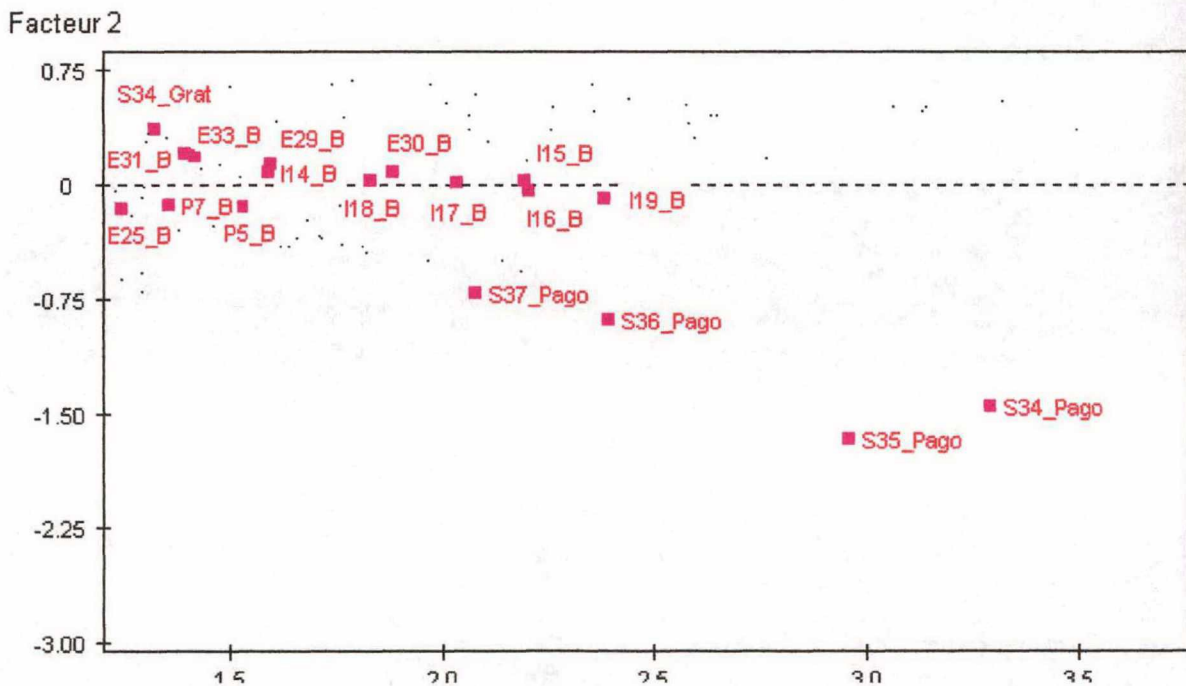


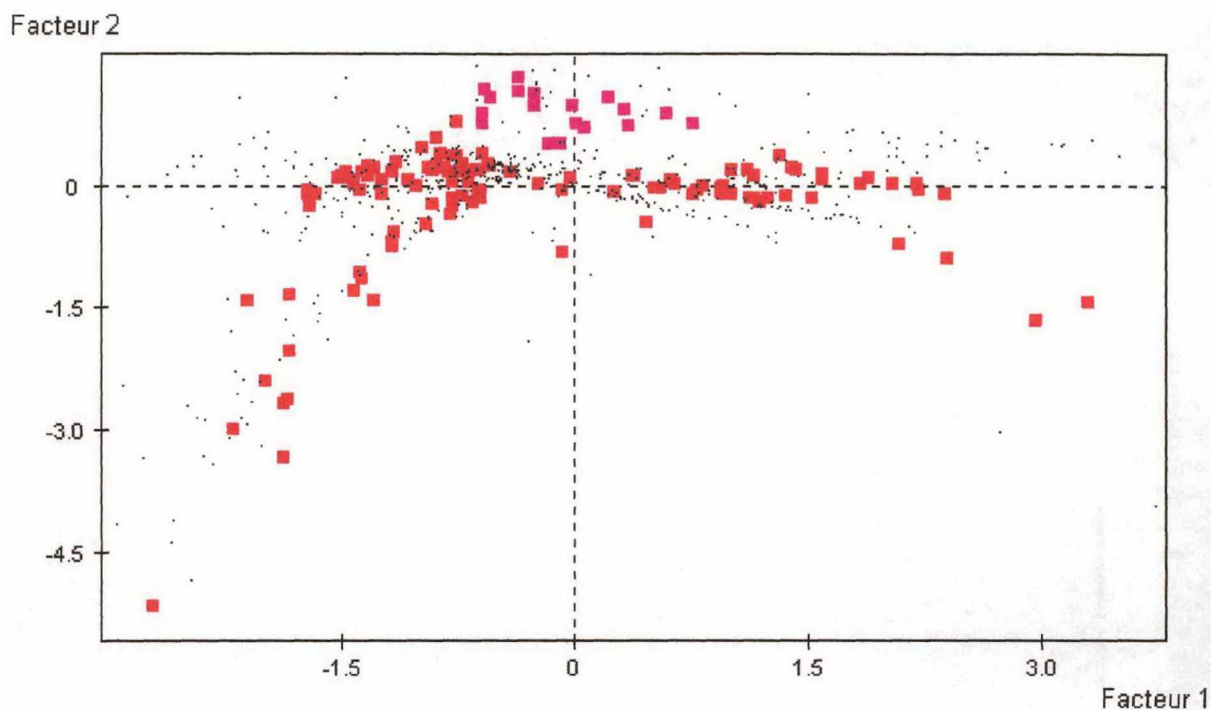
Figura 4.13. Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos ("zoom" situação extrema direita)



As características mais marcantes são: ter todos os serviços pagos; vestiários, laboratório de ciências e informática, auditório, quadra de esportes e biblioteca bons (referente as Instalações); banheiro bom (referente ao Prédio); projetor de slides, máquina fotocopadora, computador, retroprojetor, e mesas em bom estado (referente aos Equipamentos). Também é uma característica das escolas desta zona do plano o serviço médico gratuito. Note-se que as únicas características do prédio que diferenciam estas escolas das restantes são os banheiros e as instalações hidráulicas, as demais variáveis referentes ao prédio, tais como paredes, telhado, janelas, etc. não são as variáveis que as diferenciam das demais escolas. Da mesma forma podemos observar que com relação aos grupos de variáveis referentes aos equipamentos e instalações o que diferenciam essas escolas são as características que demandam uma tecnologia mais avançada, isto é, elas não levam em conta as variáveis referentes as salas de aulas, salas de professores, quadros, etc., considerados recursos de baixa tecnologia.

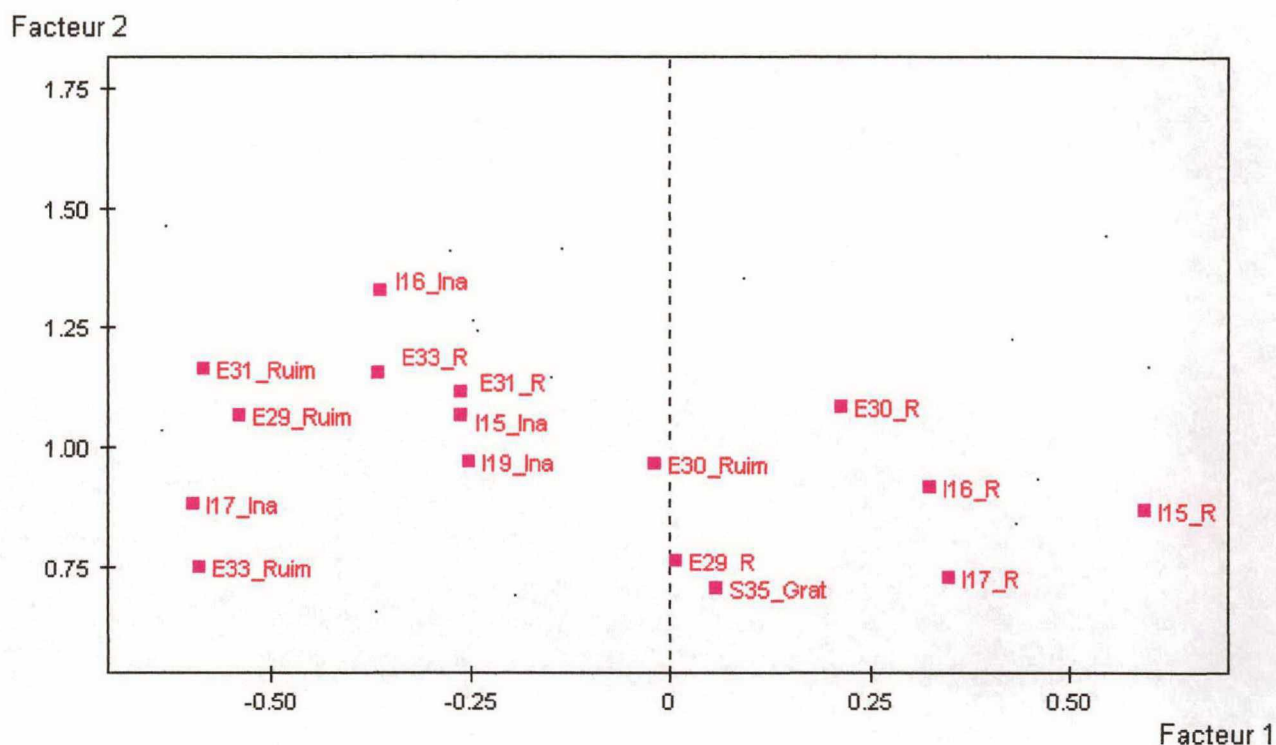
Uma vez identificadas as características que discriminam as situações extremas; passamos a analisar as diferentes situações intermediárias na estrutura de recursos materiais.

Figura 4.14. Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (situação intermédia central)



Trata-se de um tipo de situação intermediária no conjunto de recursos; com o "zoom" do gráfico seguinte identificaremos quais as características mais relevantes presentes nesta situação.

Figura 4.15. Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos ("zoom" situação intermédia central)



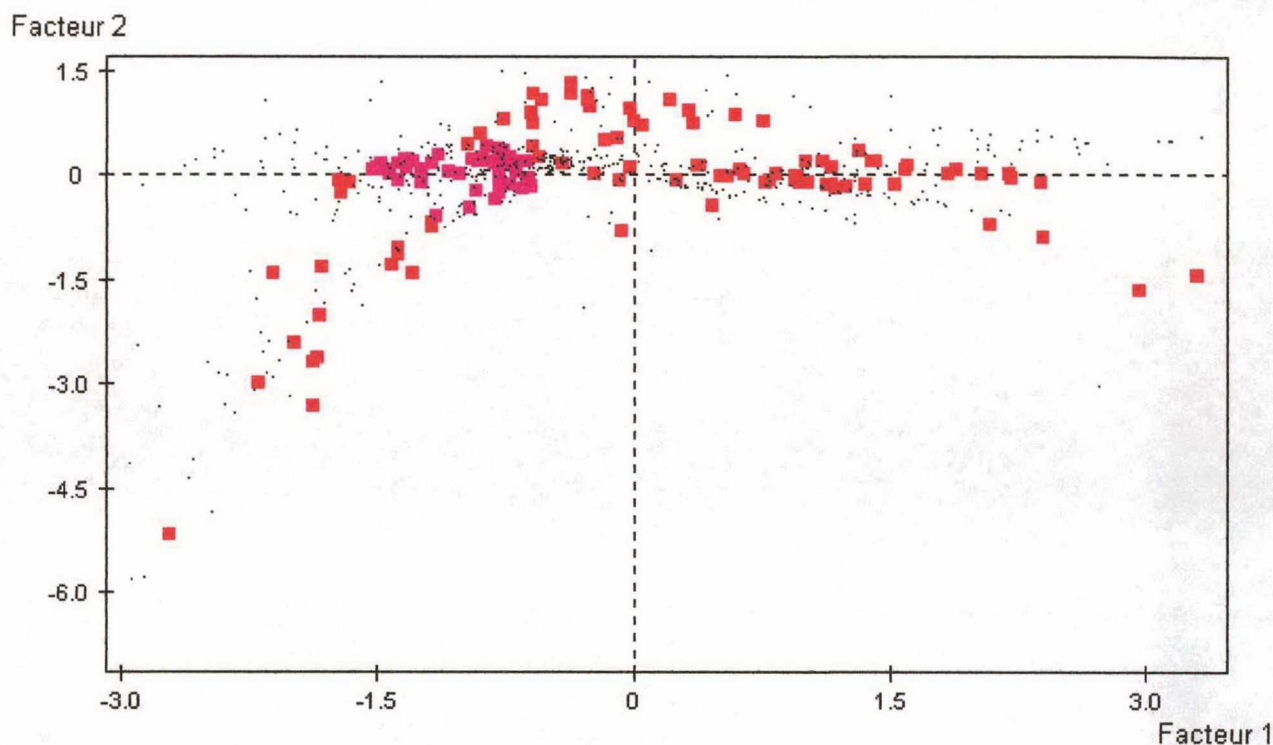
Este gráfico é uma ampliação do gráfico anterior contendo apenas as modalidades situadas na parte intermediária superior do plano.

A direita do segundo eixo, encontramos laboratório de ciências e informática, e auditório regulares (referente ao grupo Instalações); projetor de slides regular ou ruim, máquina fotocopadora regular (referente ao grupo Equipamentos) e serviços odontológicos gratuitos. Note-se que não temos aqui nenhuma das características do grupo Prédio e nenhuma das características básicas ou mais simples do grupo de Instalações ou Equipamentos (baixa tecnologia).

A esquerda do segundo eixo a situação muda; a diferença com a zona anterior tem a ver com o estado predominante nos equipamentos e instalações que são inadequados ou ruins; assim temos escolas com laboratório de ciências e informática, auditório e vestiários inadequados (referente ao grupo instalações); computador e retroprojetor regular ou ruim e máquina fotocopadora ruim (referente ao grupo equipamentos).

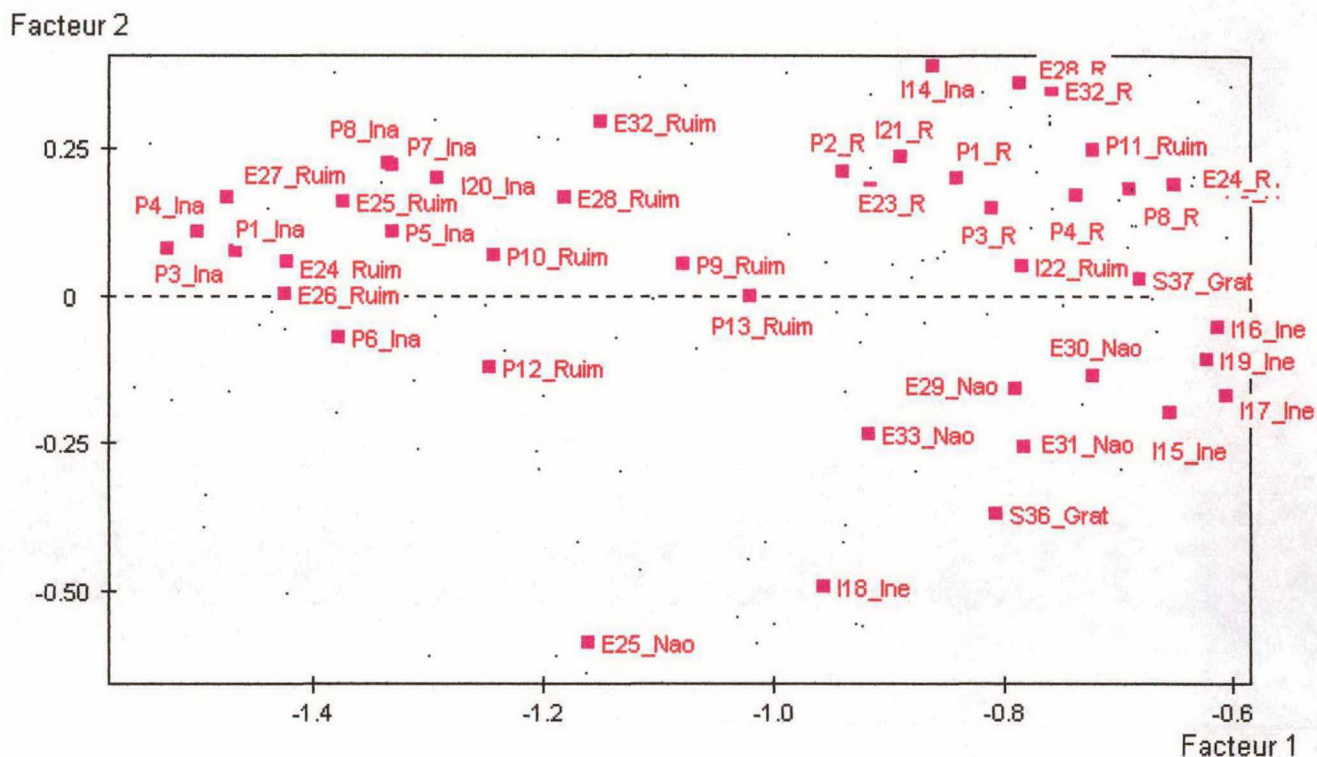
Para esse conjunto de escolas, além de outras ações concernentes, o maior impacto diante a rede como um todo, seria priorizar investimentos em instalações e equipamentos.

Figura 4.16. Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos (situação intermédia esquerda)



Trata-se de uma segunda situação intermediária no conjunto de recursos; com o "zoom" do gráfico seguinte identificaremos quais as características mais relevantes desta situação.

Figura 4.17. Projeção das modalidades das variáveis de recursos no plano principal da estrutura conjunta dos grupos de recursos ("zoom" situação intermédia central)



Nesta situação intermediária de recursos, as características do prédio estão presentes, sendo que o estado predominante entre essas escolas é o inadequado e mais próximo da origem o estado é considerado regular. Os equipamentos existentes nas escolas mais a esquerda estão em estado ruim e os mais a direita em estado regular (acima do primeiro eixo) ou inexistentes (abaixo do primeiro eixo); no entanto as instalações se caracterizam como inexistentes nesse grupo de escolas.

Em síntese, para interpretar as posições relativas das escolas no primeiro plano da estrutura conjunta de recursos materiais (Figura 4.8); no qual as escolas se agrupam segundo as suas semelhanças; houve uma análise da representação no primeiro plano das diversas modalidades de recursos (Figura 4.9). Pode-se concluir que o primeiro eixo ordena as escolas da esquerda à direita, isto é

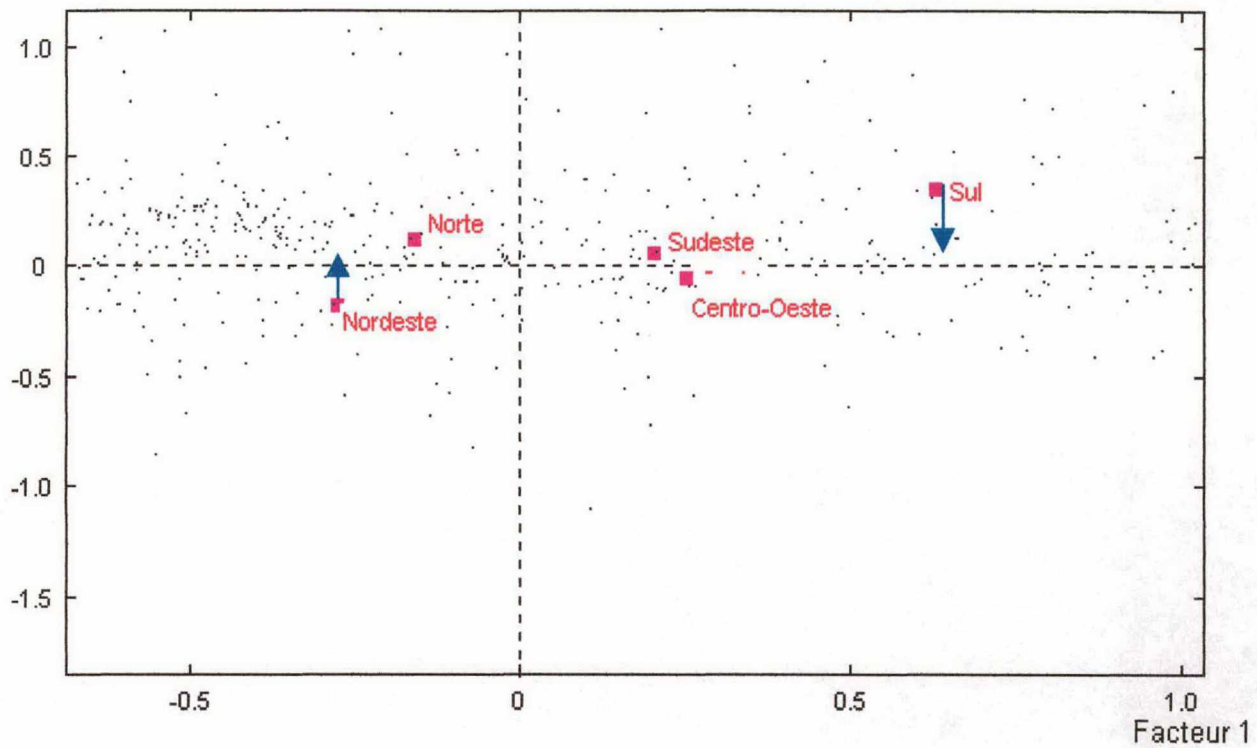
daquelas que estão nas piores situações à aquelas que estão nas melhores situações de recursos materiais e que o segundo eixo diferencia as escolas em situações intermediárias entre as piores e as melhores, sendo caracterizadas essas diversas situações pelas modalidades presentes nelas, conforme detalhado.

Ao longo desta seção foram identificados grupos de escolas com estruturas de recursos similares, e também quais os recursos que caracterizam esses grupos. O indicador ordena as escolas segundo as similitudes detectadas atribuindo valores semelhantes a escolas em situações similares. Desta forma atende as propriedades de capacidade descritiva e pertinência em relação a finalidade para a qual foi construído. Possibilita-se o diagnóstico das situações deficitárias de cada grupo de escolas. Permite recomendar ações, para cada classe de escolas que visem a corrigir tais situações subsidiando o processo de planejamento e acompanhamento de políticas voltadas para a melhoria das condições dos recursos físicos da rede escolar. Esta propriedade atende um dos requisitos para construção do indicador demonstrando portanto a sua utilidade enquanto instrumento de diagnóstico e sua relação com formas de intervenção.

4.2.4. Estudo das variáveis de estratificação na estrutura conjunta

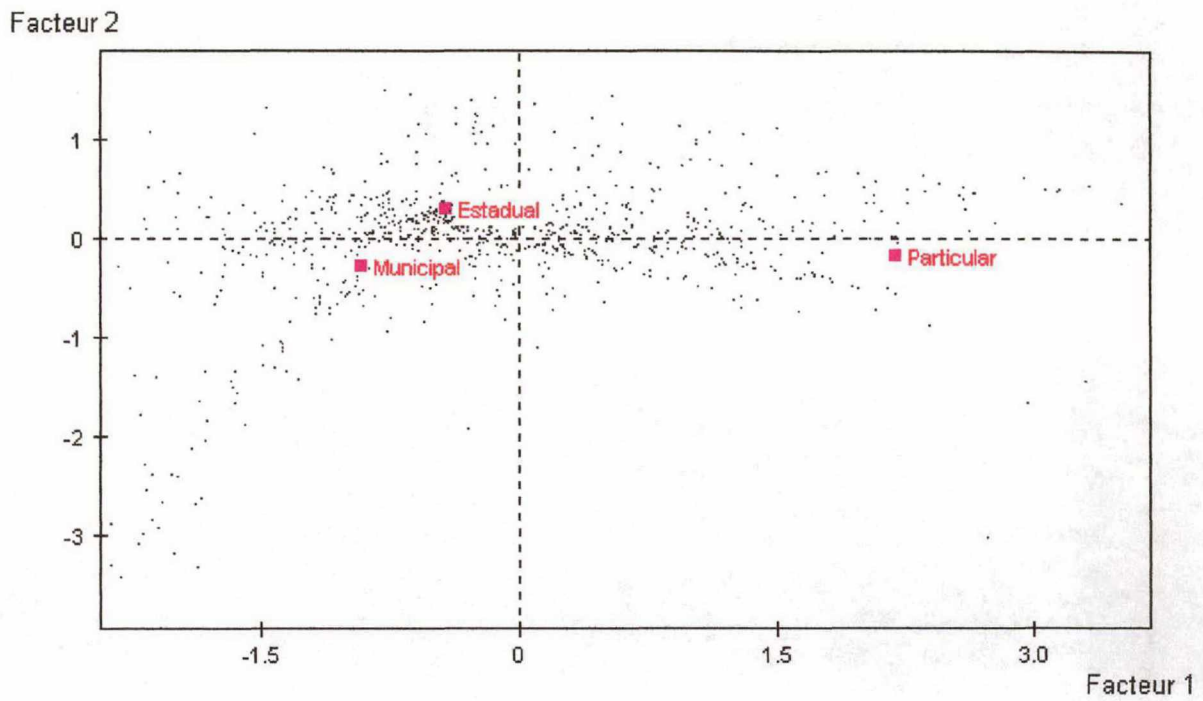
A *comparação*, na estrutura da Tabela Múltipla, da “*situação característica*” dos recursos materiais das escolas, segundo as regiões, estados e redes de ensino; se consegue mediante a projeção das modalidades das variáveis de estratificação (variáveis ilustrativas) na estrutura conjunta dos grupos de recursos. A projeção de uma modalidade ilustrativa, por ex. uma região, permite representar a “situação característica” das escolas dessa região já que o ponto de projeção representa a situação baricêntrica das escolas dessa região e pode ser considerado como a escola baricentro (ou escola “média”) da região.

Figura 4.18. Projeção das modalidades da variável região (variável ilustrativa) na estrutura conjunta dos grupos de recursos



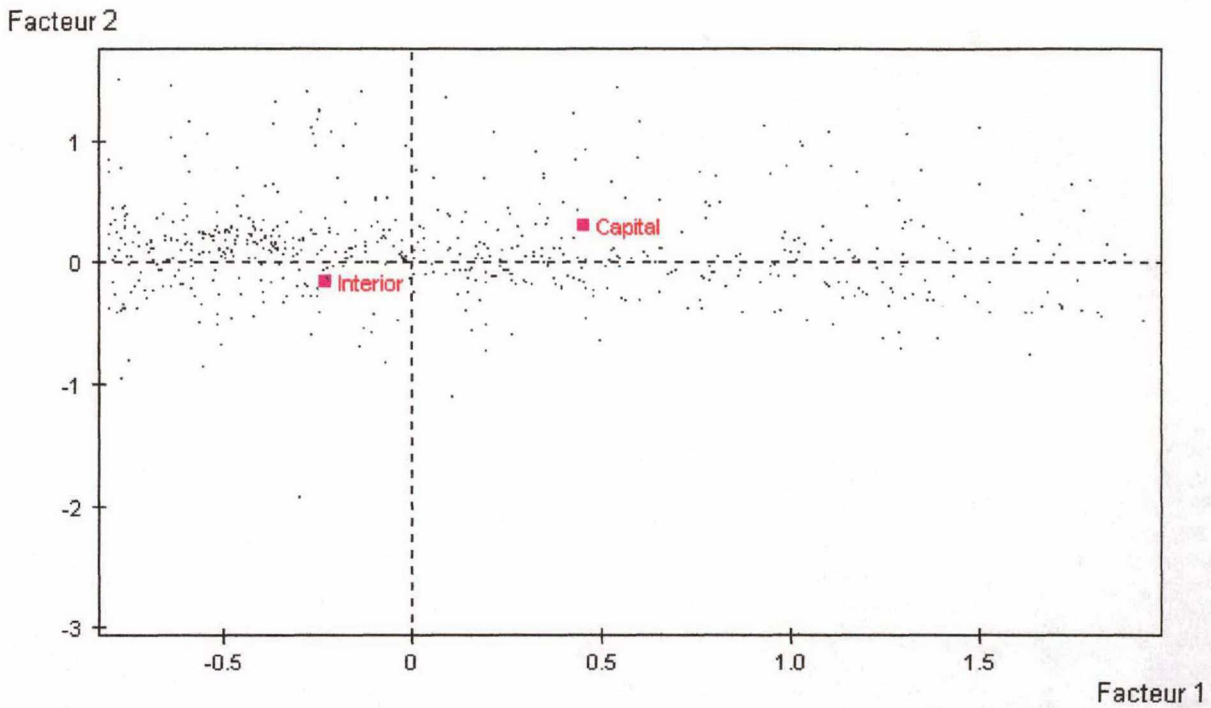
A melhor situação de recursos materiais referente aos aspectos mencionados, comparativamente, situa-se a direita do gráfico. Neste caso, a região que apresenta a pior situação característica de recursos materiais comparativamente as outras regiões do Brasil, é região nordeste, e, em ordem crescente apresentam-se as regiões norte, sudeste, centro-oeste e sul.

Figura 4.19 Projeção das modalidades da variável rede (variável ilustrativa) na estrutura conjunta dos grupos de recursos



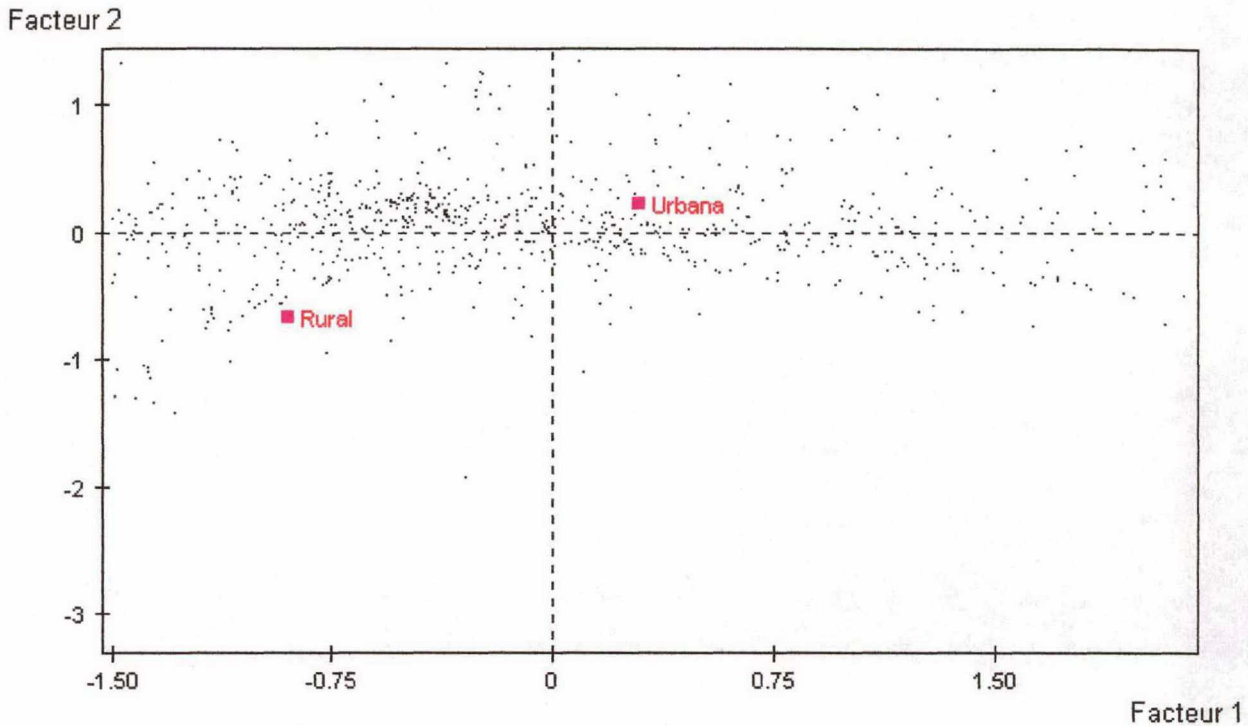
A melhor situação característica de recursos é à da rede particular oposta às situações da rede estadual e municipal.

Figura 4.20. Projeção das modalidades da variável localização (variável ilustrativa) na estrutura conjunta dos grupos de recursos



As escolas das capitais apresentam melhor situação de recursos que as escolas do interior.

Figura 4.21. Projeção das modalidades da variável zona (variável ilustrativa) na estrutura conjunta dos grupos de recursos



A situação de recursos que caracteriza as escolas da zona urbana é melhor que à da área rural.

Observando a projeção das modalidades das variáveis de estratificação ao longo do primeiro eixo, pode-se verificar a coerência do diagnóstico realizado por meio da análise da estrutura. As informações provenientes dessa análise coincidem com a percepção dos especialistas em gestão de educação.

4.2.5. A análise "local" da situação dos recursos materiais em uma escola na estrutura conjunta

Um resultado da AFM que não tem equivalência nos métodos clássicos de análise fatorial (ACP, AFCS, ACM) é a *representação superposta dos indivíduos*. Ela permite ver a situação geral de uma escola no conjunto de recursos em relação às componentes de essa situação (sendo as componentes as condições dos prédios, instalações, equipamentos e serviços). Esta representação se obtém

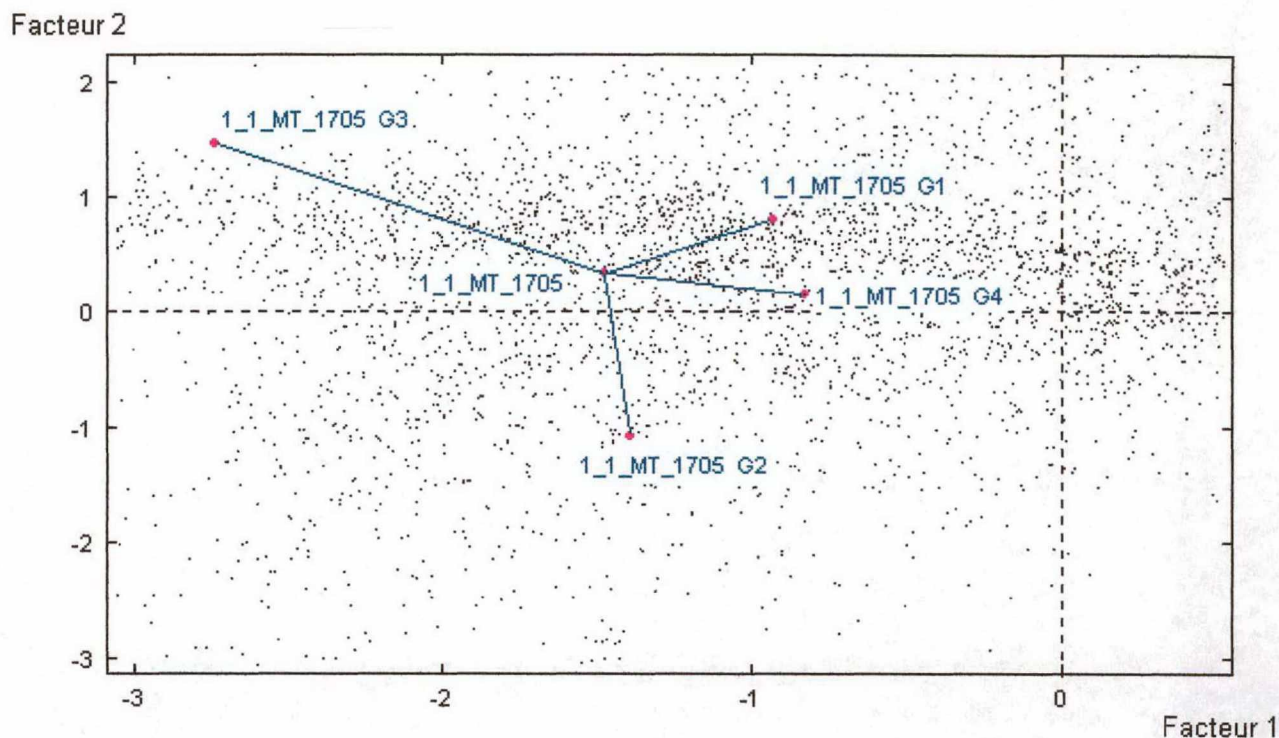
por projeção, em elementos suplementares, das quatro nuvens sobre os eixos da AFM. Para medir a semelhança entre as projeções das quatro nuvens sobre o mesmo eixo, calcula-se o coeficiente de correlação entre cada uma dessas projeções e a da nuvem global.

Figura 4.22. Correlações para os primeiros eixos da AFM, entre as projeções da nuvem de pontos global e as quatro nuvens de pontos associadas a um grupo

AXES 1 A 5					
CORRELATIONS					
FAC.	1	2	3	4	5
GR 1	0.85	0.52	0.43	0.19	0.56
GR 2	0.91	0.73	0.51	0.16	0.46
GR 3	0.90	0.76	0.51	0.26	0.55
GR 4	0.67	0.48	0.73	0.94	0.67

O primeiro eixo é um fator comum aos quatro grupos de variáveis: os três primeiros grupos apresentam uma correlação muito elevada com ele, e a correlação com o quarto grupo é razoável. A representação superposta só é interessante se existir uma direção comum, como acontece neste caso. A maior quantidade das escolas vão estar representadas por quatro pontos próximos ao primeiro eixo, como se pode deduzir da razão inércia inter/ inércia total que para o primeiro eixo é 0.7 (comentada na próxima etapa). Se os pontos que representam uma escola vista por cada um dos grupos de recursos não estão muito próximos, então a situação da mesma é diferente no estado dos prédios, das instalações, dos equipamentos e dos serviços.

Figura 4.23. Representação da situação característica da escola estadual N° 1705 de MT na estrutura conjunta dos recursos



O grupo de variáveis referente a equipamentos apresenta-se como a principal ponderação para situar a escola estadual N° 1705 de MT na região esquerda do primeiro eixo. Em segundo lugar contribui o grupo referente as instalações. Sendo que os grupos referentes aos serviços e, posteriormente, aos prédios tendem a situá-la na extremidade à direita do primeiro eixo de observação da variabilidade.

Depreende-se, portanto, que para a escola estadual N° 1705 de MT, as condições que contribuem para que esteja situada na extremidade esquerda do primeiro eixo, comparativamente as demais escolas brasileiras situadas a sua direita ao longo do primeiro eixo, são os aspectos referentes aos equipamentos e as instalações respectivamente, sendo que, comparativamente as demais escolas da rede, uma possível alteração na situação dessa escola requer uma intervenção em relação as variáveis referentes ao grupo de equipamentos, dada a importância que esse grupo tem na situação da escola, que possui uma distribuição bem

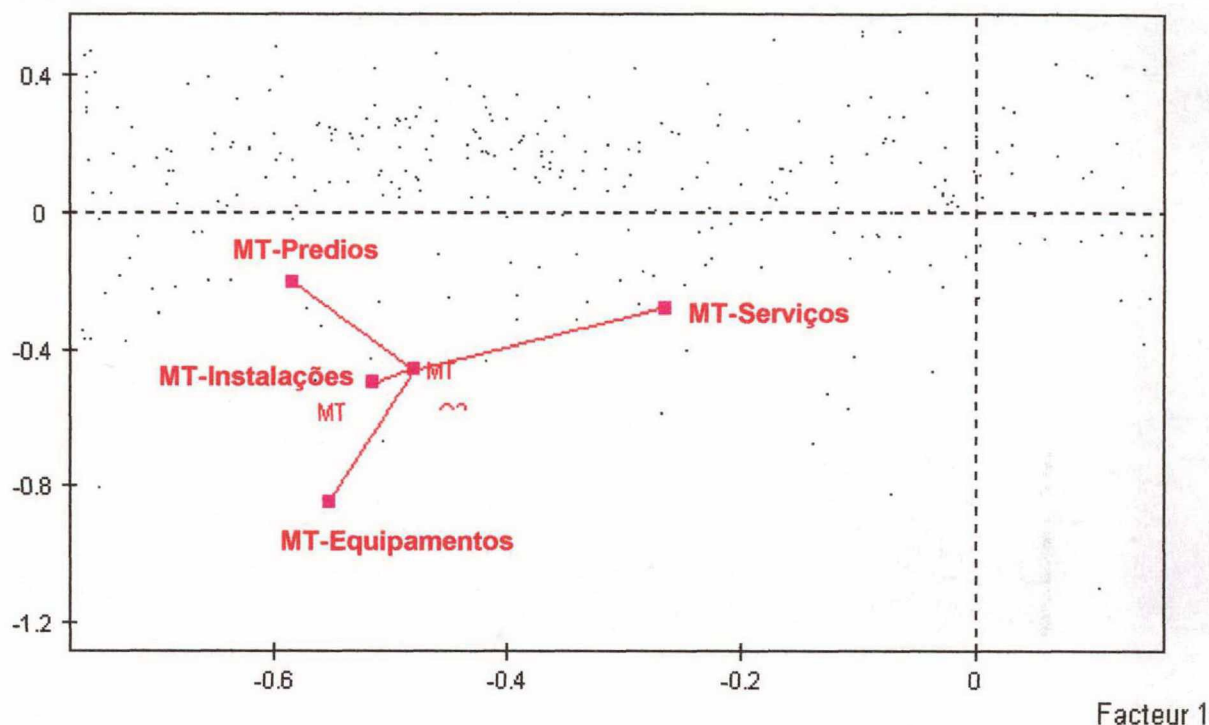
heterogênea, como pode-se observar pelo posicionamento do grupo que "arrasta" essa escola para o conjunto das escolas regulares .

A situação de uma escola em relação a cada um dos tipos de recursos pode interessar a um diretor. Neste caso, o gestor da escola poderá tomar a decisão, a partir desse diagnóstico, de prorizar a aplicação de recursos na aquisição de equipamentos e de melhorias nas instalações. Tendo em vista que a margem de ação de um diretor da rede pública, é limitada pela pouca disponibilidade de recursos que gere diretamente deverá o mesmo gestionar junto a instâncias superiores para inclusão no planejamento de recursos que priorize tal situação.

4.2.6. Exemplo de análise local da situação dos recursos materiais de um conjunto de escolas (escolas de um estado)

Do ponto de vista de um decisor a nível estadual interessa conhecer, por exemplo a situação representativa das escolas do seu estado segundo os tipos de recursos.

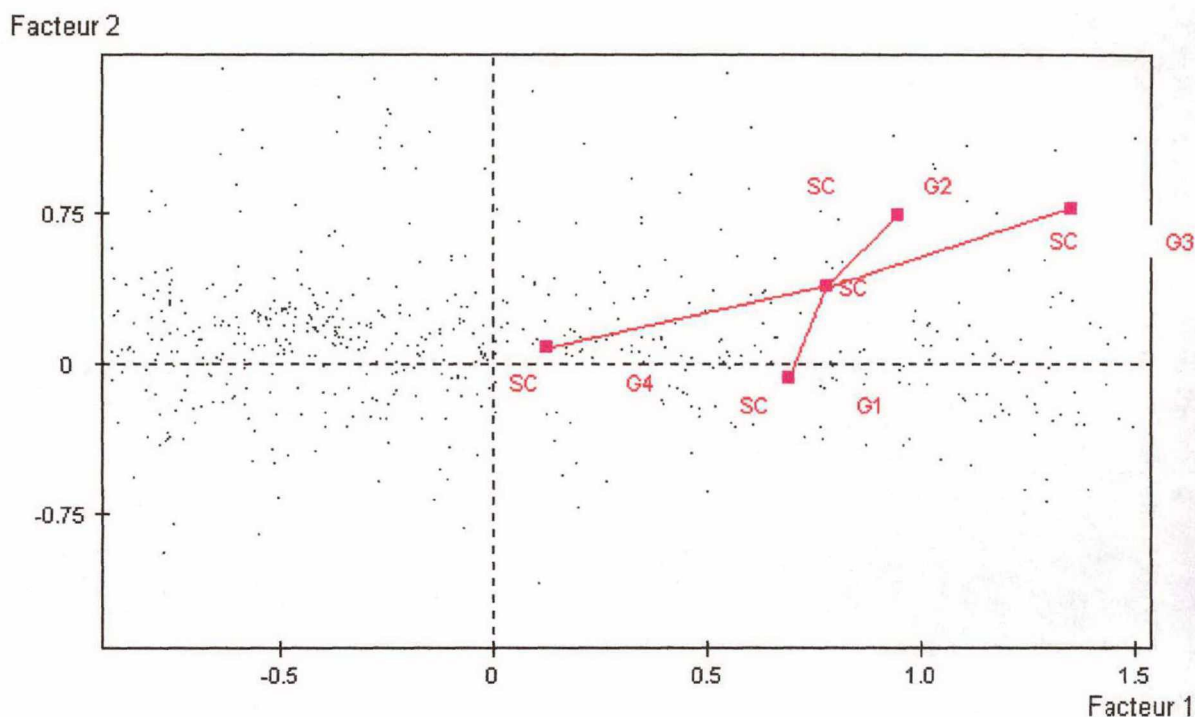
Figura 4.24. Representação da situação característica de MT na estrutura conjunta dos recursos, e das componentes dessa situação segundo os grupos de recursos.



Dentre os diferentes tipos de recursos, aqueles relativos aos equipamentos são os que levam a situação representativa de MT se aproximar das piores condições. As características do prédio também influenciam.

Diante desse diagnóstico pode-se recomendar ao gestor estadual a formulação de um programa que viabilize a aquisição e manutenção de equipamentos para o conjunto das escolas do Estado. Uma ação específica e de menor intensidade deve ser efetuada para lograr o melhoramento dos prédios junto ao conjunto de escolas que apresentam deficiências nesse quesito. Tal recomendação busca maximizar o impacto em relação ao posicionamento das escolas do Estado em relação a outras unidades da Federação esperando, conseqüentemente, uma melhoria na qualidade de ensino.

Figura 4.25. Representação da situação característica de SC na estrutura conjunta dos recursos, e das componentes de essa situação segundo os grupos de recursos.



A situação característica das escolas de SC é das melhores, exercendo uma grande influência nessa posição as condições dos seus equipamentos (grupo 3).

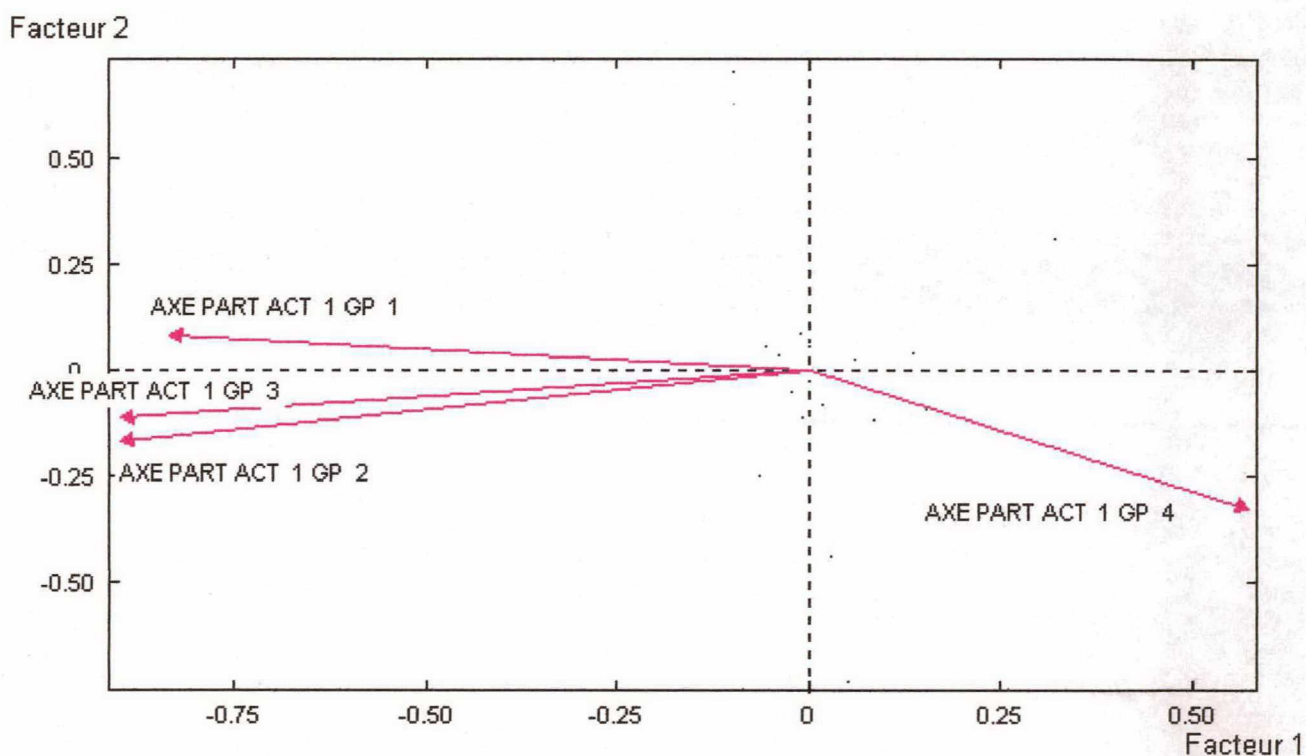
Diante da situação das escolas pode-se sugerir ao gestor estadual as seguintes ações para melhorar o desempenho das rede escolar do Estado em relação a rede nacional e outras unidades da Federação:

- Manutenção dos níveis de equipamentos e instalações sendo que o investimento deve-se pautar pelo incremento médio de investimentos em outras regiões;
- Redirecionar e aumentar a inversão de recursos para o programa de prestação de serviços considerando que neste quesito o impacto em relação a melhoria no posicionamento da rede escolar é maximizado, tendo em vista que é esta condição que apresenta o menor desempenho da rede escolar catarinense.

4.2.7. Estudo das relações entre a análise da estrutura conjunta e as análises da estrutura de cada grupo de recursos

As projeções dos fatores normados das ACM de cada um dos quatro grupos sobre o plano dos dois primeiros fatores da AFM permite observar as relações entre os fatores destas análises.

Figura 4.26. Representação dos fatores das análises separadas de cada grupo sobre o primeiro plano da AFM.



O primeiro eixo dos análises separadas dos grupos prédio, instalações e equipamentos tem uma alta correlação com o primeiro eixo da AFM.

Esse primeiro eixo, está para esses três grupos, muito próximo a sua principal direção de dispersão. A correlação do primeiro eixo do grupo de serviços não é tão elevada (0.6) quanto as outras.

Este resultado precisa a importância relativa do primeiro eixo da AFM nos grupos de variáveis.

4.2.7.1. Grau de presença do primeiro fator dessa estrutura conjunta em cada grupo de variáveis

O grau de presença do primeiro fator determinará se é possível resumir a informação num indicador contínuo único ou não.

Para isso calcula-se um índice de semelhança entre as nuvens N_j^1 apresentado na seção da fundamentação teórica (ver §2.3.3.) que é uma medida da utilidade da representação conjunta.

Pode-se apreciar que o valor desse índice para o primeiro eixo - que é o indicador procurado - é alto (0.70), portanto os grupos de variáveis tem suficientes características comuns.

Figura 4.27. Índice de semelhança entre as nuvens N_j^1

RAPPORT : INERTIE INTER/INERTIE TOTALE					
AXES 1 A 5					
RAPPORTS					
FAC.	1	2	3	4	5
	0.70	0.39	0.27	0.25	0.30

O fato que os grupos de variáveis tenham suficientes características comuns, indica que os grupos de recursos definidos podem ser analisados conjuntamente e portanto a dimensão do conceito de recursos materiais das escolas é adequadamente representado pelo indicador obtido: o primeiro eixo da AFM .

4.2.7.2. Descrição do indicador de "recursos materiais" das escolas

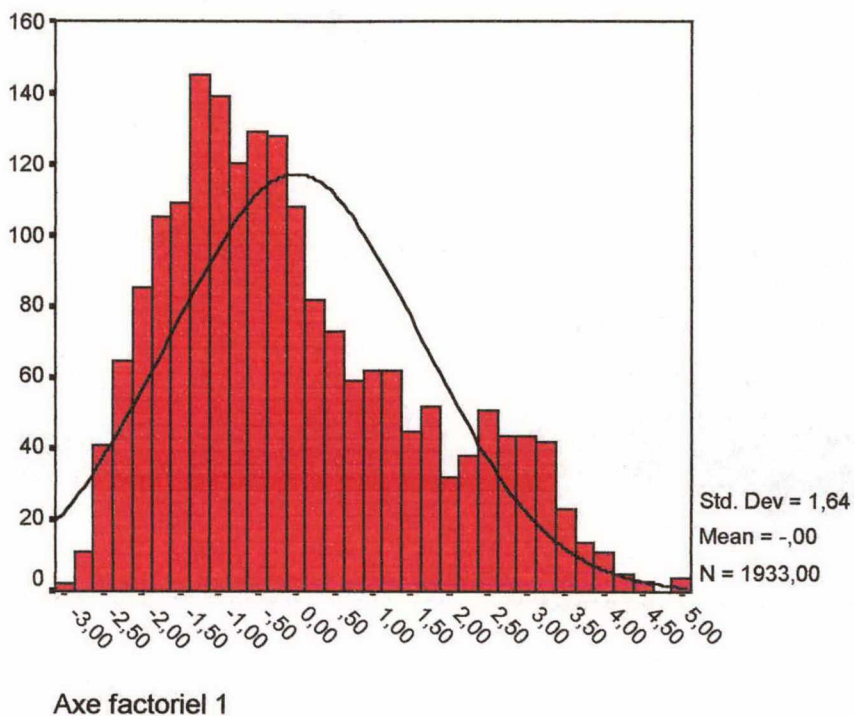
Nesta seção mostra-se a distribuição do indicador em subconjuntos de escolas gerados pelas partições definidas pelas variáveis mais contributivas ao eixo.

Mediante a AFM foi gerado um indicador de recursos materiais das escolas e mostrou-se que esse indicador cumpre com as propriedades desejadas. Essas

propriedades podem ser postas em evidência mediante o uso de ferramentas estatísticas simples.

Um histograma de freqüências do indicador (coordenadas das escolas no primeiro eixo da AFM) permite visualizar a distribuição do indicador no conjunto das escolas observadas.

Figura 4.28. Histograma de freqüências do indicador



Assim, pode-se ver a sua capacidade de diferenciar as escolas: a distribuição não tem descontinuidades. Como visto anteriormente, os valores baixos do indicador correspondem a escolas que tem as piores situações de recursos físicos. Sendo a distribuição de indicador assimétrica (como pode-se verificar no histograma) temos maior quantidade de escolas agrupadas nos valores baixos do indicador, fato que indica a existência de um maior número de escolas em situação deficitária do que escolas com boa situação de recursos materiais.

A análise feita na subseção 4.2.3. permite identificar quais características correspondem a diferentes valores do eixo.

Ao mostrar a distribuição do indicador em subconjuntos de escolas que pertencem as mesmas modalidades das variáveis mais contributivas ao primeiro eixo; é uma forma de conhecer quais os valores que ele atribui a escolas semelhantes nessas características.

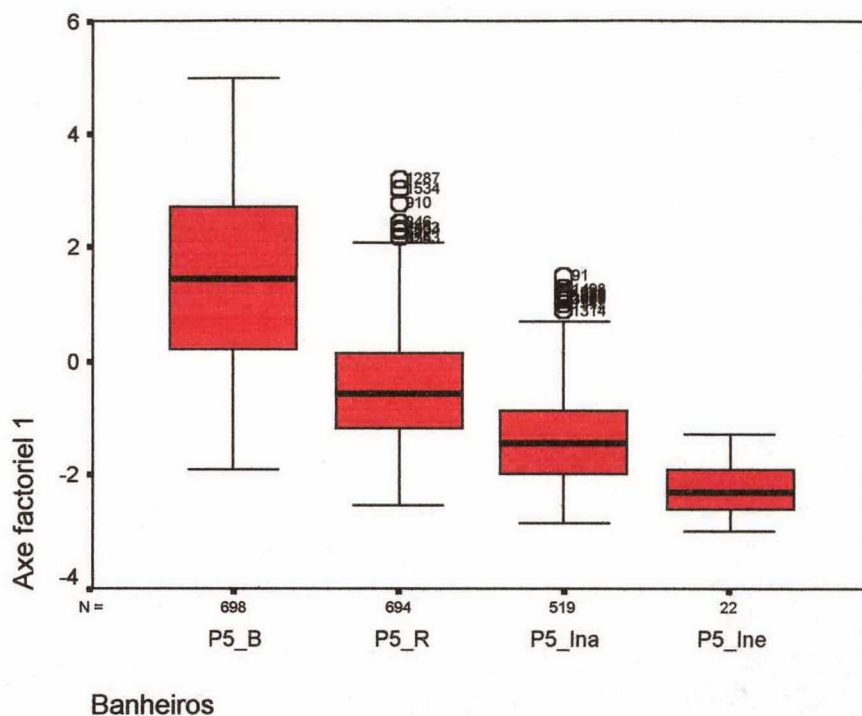
As características que contribuem com maior força à variabilidade refletida pelo primeiro eixo são: estados de conservação dos banheiros, condições de funcionamento das quadras de esportes, biblioteca e laboratório de informática e o estado de conservação do projetor de slides e dos computadores.

Apresenta-se a seguir as tabelas com as médias e os *boxplots* do indicador para cada um dos diferentes estados de conservação dos banheiros, cada uma das diversas condições de funcionamento das quadras de esportes, e do laboratório de informática; já que a análise apresenta os mesmos tipos de resultados para as outras características.

Tabela 4.2. Médias do indicador segundo o estado de conservação dos banheiros

Banheiros	Mean	N	Std. Deviation
P5_B	1,4536	698	1,4657
P5_R	-,4079	694	1,0442
P5_Ina	-1,3339	519	,8162
P5_Ine	-2,1953	22	,5038
Total	-0,0046	1933	1,6442

Figura 4.29. *Boxplot* do indicador para os diferentes estados de conservação dos banheiros

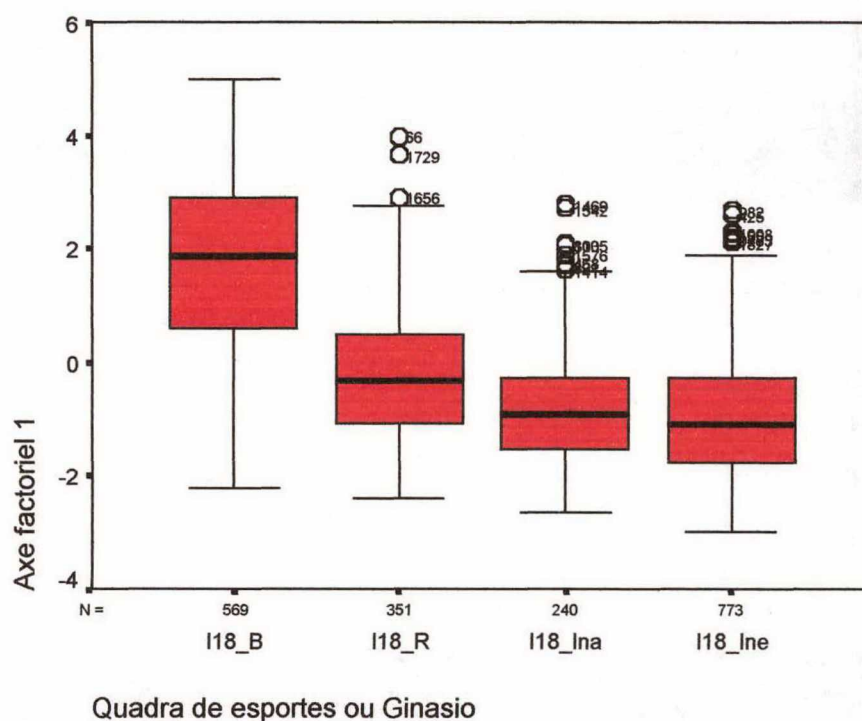


No *boxplot* da Figura 4.29 pode-se visualizar que as escolas com diferentes estados de conservação dos banheiros se diferenciam claramente nos valores do indicador. A maior variabilidade nos valores do indicador se apresenta nas escolas com banheiro bom. As escolas sem banheiro assumem os valores mais baixos do indicador, com pouca variabilidade, o que quer dizer que essas escolas tem as piores situações gerais de recursos físicos e que o banheiro é mais um recurso ruim dentre eles. As escolas com banheiro regular ou inadequado apresentam valores intermédios do indicador. Nesses grupos algumas escolas apresentam valores do indicador maiores que o grupo ao qual pertencem. Como o indicador representa o estado geral de recursos físicos, essas escolas que se diferenciam do seu grupo, apresentam uma situação geral melhor que as outras do grupo, e nelas o banheiro representa um item deficitário em relação aos outros itens.

Tabela 4.3. Médias do indicador segundo as condições de funcionamento das quadras de esportes

Quadra de esportes ou Ginasio	Mean	N	Std. Deviation
I18_B	1,7168	569	1,4101
I18_R	-,1685	351	1,1404
I18_Ina	-,7675	240	1,0508
I18_Ine	-,9606	773	1,0387
Total	-0,0046	1933	1,6442

Figura 4.30. *Boxplot* do indicador para as diferentes condições de funcionamento das quadras de esportes



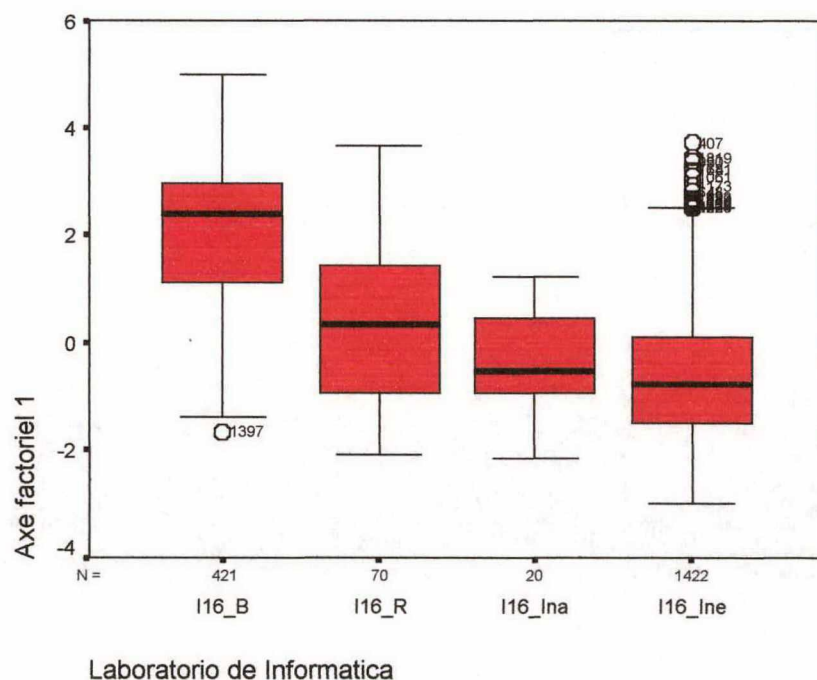
Os valores do indicador se diferenciam para as escolas com distintas condições de funcionamento das quadras de esportes, mas essa diferenciação não é semelhante a diferenciação pelo estado dos banheiros. Os valores mais altos do indicador correspondem a escolas com quadras boas. Mesmo apresentando variabilidade nos valores do indicador, essas escolas tem situações

gerais de recursos boas e bem melhores que aquelas com quadras regulares, inadequadas ou inexistentes. Além disso, dentre as escolas sem quadras de esportes, temos algumas que apresentam valores do indicador semelhantes aos das escolas com quadras boas. Essas escolas apresentam uma situação geral de recursos boa mas com déficit nas quadras.

Tabela 4.4. Médias do indicador segundo as condições de funcionamento do laboratório de informática

Laboratorio de Informatica	Mean	N	Std. Deviation
I16_B	2,0201	421	1,3859
I16_R	,3259	70	1,3099
I16_Ina	-,3620	20	1,0458
I16_Ine	-,6154	1422	1,1896
Total	-0,00467	1933	1,6442

Figura 4.31. *Boxplot* do indicador para as diferentes condições de funcionamento do laboratório de informática



O indicador diferencia os quatro grupos de escolas segundo as condições de funcionamento do laboratório de informática. Mas os quatro grupos de escolas não

são tão homogêneos na situação geral de recursos físicos. Dentre as escolas que não tem laboratório de informática, algumas tem valores do indicador semelhantes as que tem um bom laboratório. Essas escolas apresentam uma situação geral de recursos boa não tendo laboratório de informática.

A variabilidade na situação geral é maior – comparativamente - nas escolas que não tem laboratório de informática que entre as que não tem banheiros ou ginásio de esportes.

4.3. Validação do indicador de “recursos materiais” das escolas

O processo de validação do indicador compreende a validação interna e a validação externa.

4.3.1. Validação interna do indicador de “recursos materiais” das escolas

Para verificar se o indicador possui as propriedades - usando a informação com a qual foi gerado o próprio indicador - aplica-se a seqüência metodológica sugerida na proposta apresentada na seção 3.

Esse processo compreende as seguintes etapas:

- Realizar uma Análise de Correspondências Múltiplas, determinando o número de eixos que produzem um resumo suficiente da informação.
- Realizar uma Classificação com os eixos obtidos e determinar o número de classes razoáveis nas quais agrupar as escolas.
- Realizar uma Partição com o número de classes desejado.
- Caracterizar as classes obtidas, descrevendo-lhas pelas modalidades e determinar quais as escolas que pertencem a cada classe.

Esta seqüência aplicada a cada um dos grupos de variáveis, dá como resultado tantas tipologias quanto número de temas foram definidos. O estudo dessas

partições conduz a conhecer quais as características que predominam nessas classes de escolas, e a detectar quais as características que as diferenciam.

Realizam-se então quatro análises separadas, uma para cada grupo de variáveis: Prédio, Instalações, Equipamentos e Serviços. A seqüência de técnicas estatísticas é a mesma nas quatro análises parciais: Análise de Correspondências Múltiplas seguida de Classificação e Partição, portanto só se apresenta o funcionamento dessa seqüência para o primeiro grupo de variáveis (Prédio), mostrando-se para os outros apenas os resultados.

Como resultados de cada um desses análises; conseguiremos responder (para cada grupo de variáveis) as perguntas do tipo:

- Quais as escolas que se assemelham nas características de seus prédios (instalações, equipamentos, serviços)?
- Existe uma tipologia das escolas, ou seja, classes homogêneas de escolas?
- Quais são as características dos prédios (instalações, equipamentos, serviços) predominantes nessas classes de escolas?

Finalmente, se cruzam os valores do indicador com cada uma das tipologias geradas para comprovar se os valores atribuídos pelo indicador a escolas que pertencem à mesma classe, são semelhantes.

4.3.1.1. Tipologia das escolas caracterizadas pelos seus prédios

Esta tipologia é resultante do estudo das semelhanças das escolas considerando os seus prédios, isto é, os itens 1 até o 13 do questionário, que são:

- Telhado
- Paredes
- Piso
- Portas e janelas
- Banheiros
- Cozinha
- Instalações hidráulicas
- Instalações elétricas

- Ventilação
- Iluminação
- Ruídos
- Aspectos da área externa (pátio, jardins)
- Espaço para recreação infantil

Os itens 1 à 8 foram classificados como: Bom, Regular, Inadequado ou Inexistente; os itens 9 à 13 como: Bom ou Ruim de acordo com o questionário apresentado no *Anexo 1*.

4.3.1.1.1. Primeira etapa da comparação das escolas pelo conjunto de descritores do estado do prédio: Análise de Correspondências Múltiplas

A ACM efetua a comparação das escolas pelo conjunto de descritores do estado do prédio. Com ele consegue-se: selecionar o número de dimensões necessárias para captar a essência da estrutura da informação; e observar a configuração espacial dos indivíduos aos efeitos de explorar a classificabilidade das escolas verificando se é pertinente organizar as escolas em grupos.

Figura 4.32. Variáveis ativas da ACM do Grupo Prédio

VARIABLES NOMINALES ACTIVES	
13 VARIABLES	42 MODALITES ASSOCIEES
6 . Telhado	(4 MODALITES)
7 . Paredes	(4 MODALITES)
8 . Piso	(4 MODALITES)
9 . Portas e janelas	(4 MODALITES)
10 . Banheiros	(4 MODALITES)
11 . Cozinha	(4 MODALITES)
12 . Instalações hidráulicas	(4 MODALITES)
13 . Instalações elétricas	(4 MODALITES)
14 . Ventilação	(2 MODALITES)
15 . Iluminação	(2 MODALITES)
16 . Ruídos	(2 MODALITES)
17 . Area externa(patio, jardins)	(2 MODALITES)
18 . Espaço para recreação	(2 MODALITES)

Figura 4.33. Indivíduos estudados na ACM do Grupo Prédio

```

INDIVIDUS
----- NOMBRE ----- POIDS ---
POIDS DES INDIVIDUS: Poids des individus, uniforme egal a 1. UNIF
RETENUS ..... NITOT = 1933 PITOT = 1933.000

SELECTION APRES FILTRAGE
ACTIFS ..... NIACT = 1875 PIACT = 1875.000
SUPPLEMENTAIRES .... NISUP = 58 PISUP = 58.000
    
```

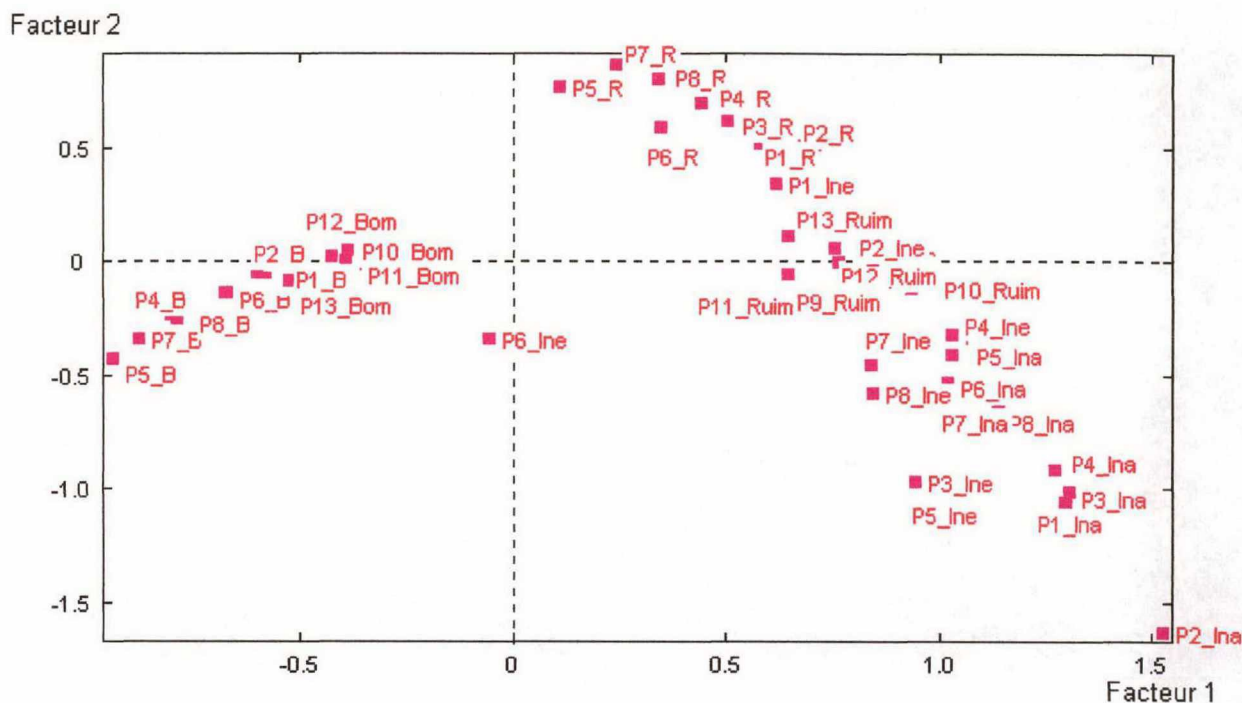
Figura 4.34. Decomposição da inércia da ACM do Grupo Prédio

ISTOGRAMME DES 23 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.4596	25.98	25.98	*****//
2	0.1806	10.21	36.18	*****
3	0.0954	5.39	41.58	*****
4	0.0855	4.83	46.41	*****
5	0.0846	4.78	51.19	*****
6	0.0768	4.34	55.53	*****
7	0.0702	3.97	59.50	*****
8	0.0659	3.72	63.22	*****
9	0.0654	3.69	66.92	*****
10	0.0609	3.44	70.36	*****
11	0.0593	3.35	73.71	*****
12	0.0573	3.24	76.95	*****
13	0.0541	3.06	80.01	*****
14	0.0532	3.01	83.02	*****
15	0.0472	2.67	85.68	*****
16	0.0419	2.37	88.05	*****
17	0.0394	2.23	90.28	*****
18	0.0353	1.99	92.27	*****
19	0.0317	1.79	94.07	*****
20	0.0301	1.70	95.77	*****
21	0.0283	1.60	97.36	*****
22	0.0270	1.53	98.89	*****
23	0.0196	1.11	100.00	*****

A forma da decomposição da inércia – com um decrescimento regular a partir do sexto eixo - permite escolher o número de dimensões necessárias para representar a estrutura da tabela e – consequentemente - o número de eixos a serem usados para classificar as escolas. No caso seis eixos consideram-se suficientes para resumir a informação da tabela.

Figura 4.35. Projeção das modalidades das variáveis ativas do Grupo Prédio no primeiro plano fatorial



A Figura 4.35 permite reconhecer as modalidades que se apresentam conjuntamente nas mesmas escolas: identificando aquelas cujas projeções estão próximas no primeiro plano fatorial.

A Figura 4.36 uma ajuda para a leitura deste plano já que - por tratar-se de um número grande de modalidades - facilita a identificação das oposições fundamentais no primeiro e no segundo eixo, listando as modalidades mais contributivas à inércia de cada eixo ordenadas por seu “valor-test”. (a explicação do “valor-test” encontra-se nas notas de curso de Eduardo Crivisqui sobre os *Métodos de Classificação* no site do Programa PRESTA citado nas referências bibliográficas).

Figura 4.36. Descrição dos eixos (primeiro e segundo) da ACM do Grupo Prédio pelas modalidades ativas

DESCRIPTION DES AXES FACTORIELS

DESCRIPTION DU FACTEUR 1
PAR LES MODALITES ACTIVES

ID.	V.TEST	LIBELLE MODALITE	LIBELLE DE LA VARIABLE	POIDS	NUMERO
AI_1	-32.39	P4_B	Portas e janelas	871.00	1
AL_1	-31.49	P7_B	Instalações hidráulicas	764.00	2
AM_1	-31.29	P8_B	Instalações elétricas	857.00	3
AG_1	-30.56	P2_B	Paredes	1089.00	4
Z O N E C E N T R A L E					
AL_3	25.49	P7_Ina	Instalações hidráulicas	425.00	39
AQ_2	25.88	P12_Ruim	Area externa (patio, jardins)	627.00	40
AO_2	26.04	P10_Ruim	Iluminação	549.00	41
AJ_3	27.37	P5_Ina	Banheiros	512.00	42

DESCRIPTION DU FACTEUR 2
PAR LES MODALITES ACTIVES

ID.	V.TEST	LIBELLE MODALITE	LIBELLE DE LA VARIABLE	POIDS	NUMERO
AG_3	-21.14	P2_Ina	Paredes	153.00	1
AI_3	-17.37	P4_Ina	Portas e janelas	300.00	2
AH_3	-17.05	P3_Ina	Piso	246.00	3
AF_3	-16.44	P1_Ina	Telhado	213.00	4
Z O N E C E N T R A L E					
AI_2	23.10	P4_R	Portas e janelas	692.00	39
AM_2	23.75	P8_R	Instalações elétricas	598.00	40
AJ_2	24.94	P5_R	Banheiros	678.00	41
AL_2	26.26	P7_R	Instalações hidráulicas	617.00	42

Em linhas gerais o primeiro eixo opõe condições boas do prédio às ruins e o segundo eixo diferencia as condições regulares das inadequadas. Nesta tabela só figuram as modalidades mais contributivas à variabilidade captada pelos eixos.

Esta representação, mesmo sendo das modalidades, também permite apreciar a forma da nuvem, verificando a classificabilidade das escolas.

4.3.1.1.2. Segunda etapa: Classificação hierárquica mista das escolas baseada nos primeiros eixos da ACM do Grupo Prédio

De acordo com a etapa anterior, as distâncias iniciais entre as escolas são calculadas no espaço dos 6 primeiros eixos da ACM. O objetivo desta etapa é

determinar o número de classes que dão como resultado uma "boa" partição do conjunto das escolas., no sentido que as classes devem estar formadas por escolas que se assemelhem entre elas e que se diferenciem das escolas das outras classes.

Duas ferramentas possibilitam determinar esse número de classes: o dendrograma e o histograma dos índices de nível, apresentados a seguir (Crivisqui, 1999, p.52).

Figura 4.37. Dendrograma das escolas pelas características do prédio

Classificação das escolas pelas características dos Prédios

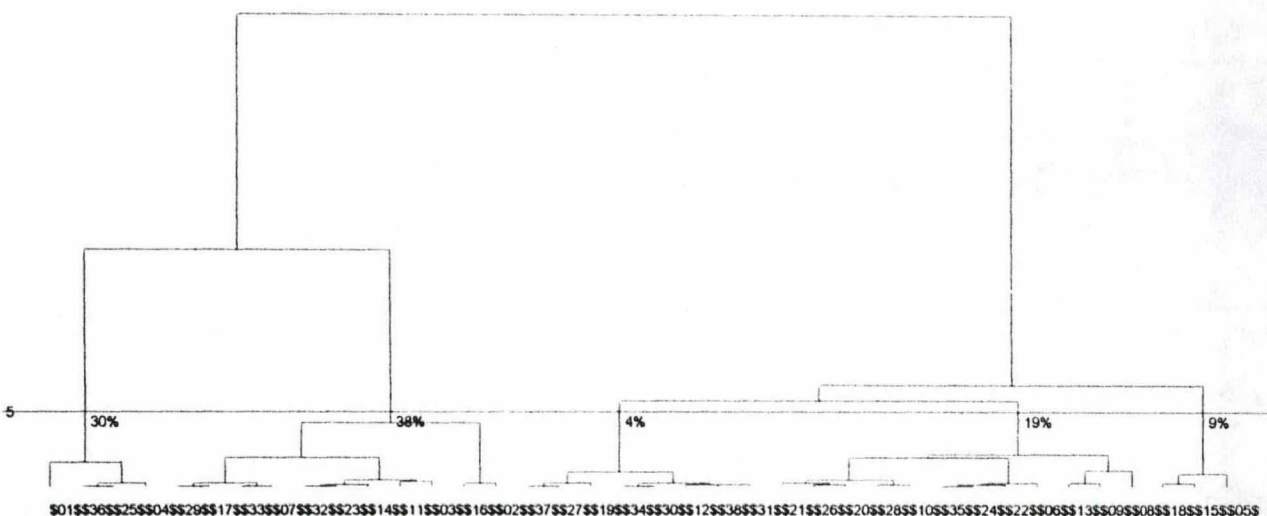


Tabela 4.38. Histograma dos índices de nível

DESCRIPTION DES NOEUDS				HISTOGRAMME DES INDICES DE NIVEAU		
NUM.	AINE	BENJ	EFF.	POIDS	INDICE	
39	25	36	2	7.00	0.00010	*
40	24	35	2	9.00	0.00018	*
41	27	37	2	7.00	0.00028	*
42	30	34	2	5.00	0.00029	*
43	17	29	2	26.00	0.00043	*
44	7	33	2	88.00	0.00043	*
45	23	32	2	12.00	0.00048	*
46	31	38	2	5.00	0.00054	*
47	10	28	2	48.00	0.00078	*
48	20	26	2	20.00	0.00092	*
49	46	12	3	40.00	0.00112	*
50	22	40	3	19.00	0.00122	*
51	14	45	3	44.00	0.00129	*
52	6	50	4	118.00	0.00175	*
53	15	18	2	46.00	0.00177	*
54	49	42	5	45.00	0.00195	*
55	4	39	3	137.00	0.00215	*
56	44	43	4	114.00	0.00223	*
57	19	41	3	22.00	0.00224	*
58	48	21	3	33.00	0.00235	*
59	9	13	2	84.00	0.00246	*
60	2	16	2	276.00	0.00248	*
61	3	11	2	272.00	0.00299	*
62	61	51	5	316.00	0.00431	**
63	47	58	5	81.00	0.00444	**
64	5	53	3	172.00	0.00802	***
65	54	57	8	67.00	0.00912	***
66	8	59	3	163.00	0.01033	***
67	55	1	4	568.00	0.01489	****
68	62	56	9	430.00	0.01774	*****
69	52	63	9	199.00	0.01811	*****
70	66	69	12	362.00	0.01992	*****
71	60	68	11	706.00	0.04022	*****
72	70	65	20	429.00	0.05443	*****
73	64	72	23	601.00	0.06404	*****
74	71	67	15	1274.00	0.14960	*****
75	73	74	38	1875.00	0.30231	*****
SOMME DES INDICES DE NIVEAU =				0.74791		****

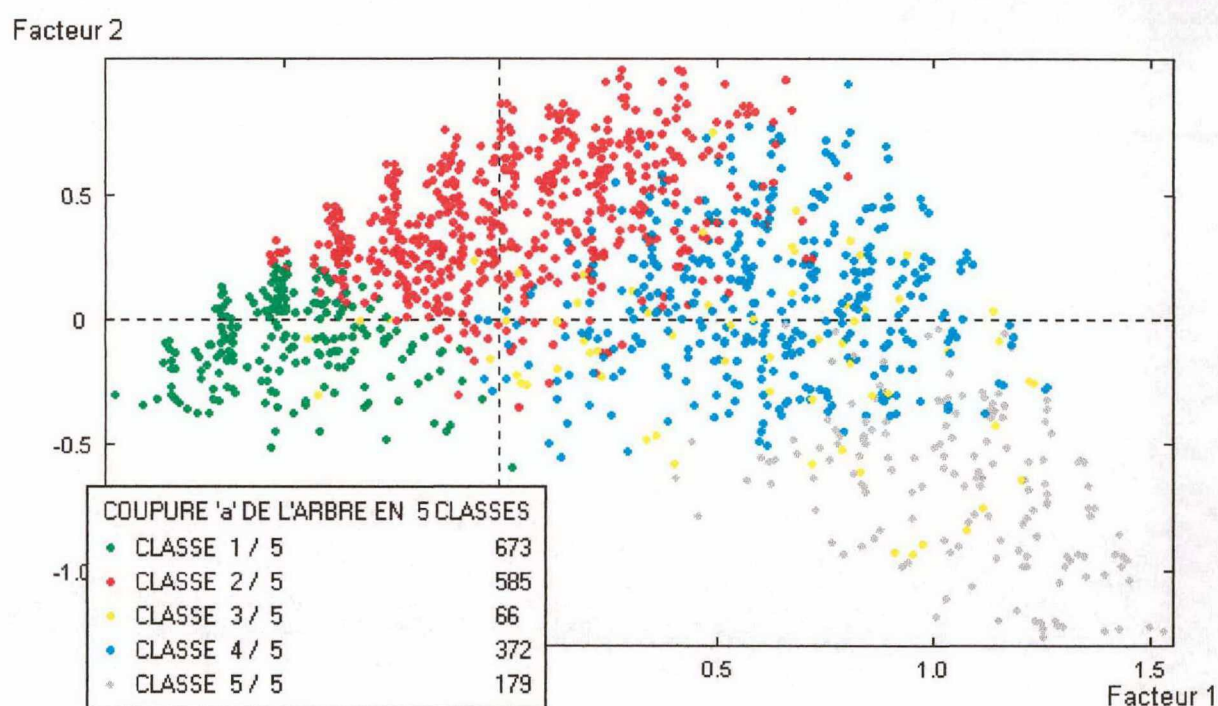
Para que a partição seja "boa" deve-se escolher um nível de agregação que transforme fracamente as distancias iniciais entre as escolas; ou seja um nível no qual o valor do índice não seja muito elevado. Uma partição em 3 classes resultaria insuficiente porque só distingue as escolas com situações boas de prédio daquelas com situações ruins, mas não consegue diferenciar as situações intermédias. Por tanto se considera que uma partição em 5 classes seria adequada.

4.3.1.1.3. Terceira etapa: Partição do conjunto de escolas em classes com o método de nuvens dinâmicas

O resultado da partição das escolas em 5 classes classifica as escolas da seguinte forma:

- Classe 1: 673 escolas
- Classe 2: 585 escolas
- Classe 3: 66 escolas
- Classe 4: 372 escolas
- Classe 5: 179 escolas

Figura 4.39. Projecção das escolas no primeiro plano principal da ACM diferenciando as classes as quais pertencem (prédio).



A Figura 4.39 distingue pelas cores as escolas que pertencem a distintas classes. Dessa forma os pontos da mesma cor representam escolas que se

assemelham pelo conjunto de características do prédio, podendo ser identificado cada um deles de forma de saber qual a escola que ele representa.

Mesmo tendo em consideração que a partição está baseada nos 6 primeiros eixos e neste gráfico as escolas estão sendo projetadas nos dois primeiros; pode-se verificar a dispersão das classes e a densidade das mesmas. Assim, para julgar a homogeneidade das classes e da diferenciação delas deve-se considerar que esta projeção está sendo realizada num espaço de dimensão menor. A classe com maior dispersão é a classe 3, representada com amarelo. Dentro dessa classe existem escolas cujas projeções se misturam com as da classe 5 – que apresenta as piores condições do prédio - e as da classe 4, que apresentam condições regulares mas com banheiros e instalações elétricas inadequadas.

4.3.1.1.4. Caracterização das classes da tipologia de escolas pelas "condições dos prédios"

O processo de classificação criou uma variável tipológica que afeta cada escola a uma Classe. O objetivo desta etapa é descrever cada classe em função das modalidades das características observadas.

A tabela seguinte contém – para cada classe - um conjunto de informações que permitem atingir o objetivo de conhecer quais são as características predominantes em cada uma das classes de escolas. Para cada modalidade característica da classe, a Figura 4.40 informa sobre as seguintes proporções:

CLA/MOD: é uma proporção (expressada como percentagem) que indica do total de escolas que apresentam uma determinada modalidade qual a quantidade que pertencem à classe. Assim por exemplo para a classe 1, do total de escolas com instalações elétricas boas 72.93% pertencem a esta classe.

MOD/CLA: é uma proporção (expressa em forma de percentagem) que informa do total de escolas que pertencem a uma classe e qual a quantidade que apresenta essa modalidade. A título de exemplo, do total de escolas da classe 1, 92.87% apresentam instalações elétricas boas.

GLOBAL: representa o peso da modalidade no conjunto de escolas. As escolas com instalações elétricas boas são 45.71% do total.

Essas percentagens, junto com o "valor-test", são os elementos usados para caracterizar as classes. A Figura 4.40 informa então as modalidades com maior poder de caracterização das classes, tanto para as variáveis ativas (características do prédio) quanto para as variáveis ilustrativas (variáveis de estratificação no caso estudado).

Figura 4.40. Caracterização das classes pelas modalidades

CHARACTERISATION PAR LES MODALITES DES CLASSES OU MODALITES
COUPURE 'a' DE L'ARBRE EN 5 CLASSES

ASSE 1 / 5

TEST	PROBA	CLA/	MOD/	POURCENTAGES	GLOBAL	MODALITES		IDEN	POIDS
		MOD	CLA	CLA/	MOD/	CHARACTERISTIQUES	DES VARIABLES		
		35.89			CLASSE 1 / 5				
4.89	0.000	80.24	91.08	40.75	P7_B	Instalações hidráulicas		aa1a	673
2.50	0.000	72.93	92.87	45.71	P8_B	Instalações elétricas		AL_1	764
2.29	0.000	82.96	81.72	35.36	P5_B	Banheiros		AM_1	857
0.76	0.000	70.61	91.38	46.45	P4_B	Portas e janelas		AJ_1	663
7.16	0.000	63.34	91.38	51.79	P3_B	Piso		AI_1	871
5.32	0.000	58.03	93.91	58.08	P2_B	Paredes		AH_1	971
3.92	0.000	57.58	91.38	56.96	P1_B	Telhado		AG_1	1089
3.34	0.000	64.71	78.75	43.68	P6_B	Cozinha		AF_1	1068
1.90	0.000	56.81	87.37	55.20	P13_Bom	Espaço para recreação		AK_1	819
1.40	0.000	80.29	50.22	22.45	Particular	Rede		AR_1	1035
0.84	0.000	50.96	94.50	66.56	P12_Bom	Area externa(patio, jardins)		AB_3	421
9.66	0.000	48.64	95.84	70.72	P10_Bom	Iluminação		AQ_1	1248
7.85	0.000	49.19	90.79	66.24	P9_Bom	Ventilação		AO_1	1326
0.90	0.000	42.84	87.96	73.71	P11_Bom	Ruidos		AN_1	1242
7.16	0.000	40.32	84.84	75.52	Urbana	Zona		AP_1	1382
4.47	0.000	42.99	40.12	33.49	Capital	Localização		AA_2	1416
2.63	0.004	45.40	11.74	9.28	Sul	Região		AC_1	628
								AD_5	174

ASSE 2 / 5

TEST	PROBA	CLA/	MOD/	POURCENTAGES	GLOBAL	MODALITES		IDEN	POIDS
		MOD	CLA	CLA/	MOD/	CHARACTERISTIQUES	DES VARIABLES		
		31.20			CLASSE 2 / 5				
5.74	0.000	70.66	74.53	32.91	P7_R	Instalações hidráulicas		aa2a	585
2.63	0.000	63.42	73.50	36.16	P5_R	Banheiros		AL_2	617
1.47	0.000	65.22	66.67	31.89	P8_R	Instalações elétricas		AJ_2	678
2.66	0.000	49.13	58.12	36.91	P4_R	Portas e janelas		AM_2	598
9.09	0.000	44.82	49.57	34.51	P3_R	Piso		AI_2	692
9.09	0.000	46.03	45.64	30.93	P6_R	Cozinha		AH_2	647
5.60	0.000	40.24	40.51	31.41	P1_R	Telhado		AG_2	589
4.50	0.000	38.10	41.03	33.60	P2_R	Paredes		AF_2	589
4.36	0.000	36.51	51.79	44.27	Estadual	Rede		AK_2	580
4.05	0.000	36.07	51.79	44.80	P13_Ruim	Espaço para recreação		AB_1	830
3.97	0.000	33.94	76.92	70.72	P10_Bom	Iluminação		AR_2	840
								AO_1	1326

ASSE 3 / 5

TEST	PROBA	POURCENTAGES			MODALITES	IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL			
		3.52	CLASSE 3 / 5			aa3a	66
1.08	0.000	88.41	92.42	3.68	P7_Ine Instalações hidraulicas	AL_4	69
9.36	0.000	11.33	78.79	24.48	Rural Zona	AA_1	459
7.31	0.000	8.13	77.27	33.44	P12_Ruim Area externa(patio, jardins)	AQ_2	627
7.07	0.000	8.01	75.76	33.28	Municipal Rede	AB_2	624
6.69	0.000	17.56	34.85	6.99	P6_Ine Cozinha	AK_4	131
4.91	0.000	15.62	22.73	5.12	CE Unidades da Federação	AE_6	96
4.90	0.000	4.89	92.42	66.51	Interior Localização	AC_2	1247
4.55	0.000	5.71	72.73	44.80	P13_Ruim Espaço para recreação	AR_2	840
3.93	0.000	10.46	24.24	8.16	P2_Ina Paredes	AG_3	153
3.41	0.000	5.69	54.55	33.76	P9_Ruim Ventilação	AN_2	633
3.23	0.001	5.83	48.48	29.28	P10_Ruim Iluminação	AO_2	549
3.12	0.001	6.67	34.85	18.40	P6_Ina Cozinha	AK_3	345
2.84	0.002	6.67	30.30	16.00	P4_Ina Portas e janelas	AI_3	300
2.74	0.003	11.11	12.12	3.84	MT Unidades da Federação	AE13	72
2.63	0.004	10.53	12.12	4.05	GO Unidades da Federação	AE_9	76
2.51	0.006	5.10	50.00	34.51	P3_R Piso	AH_2	647
2.40	0.008	5.06	48.48	33.71	Nordeste Região	AD_2	632
2.35	0.009	6.16	27.27	15.57	Centro-Oeste Região	AD_1	292

ASSE 4 / 5

TEST	PROBA	POURCENTAGES			MODALITES	IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL			
		19.84	CLASSE 4 / 5			aa4a	372
9.98	0.000	56.47	64.52	22.67	P7_Ina Instalações hidraulicas	AL_3	425
8.30	0.000	48.83	67.20	27.31	P5_Ina Banheiros	AJ_3	512
6.78	0.000	52.28	55.38	21.01	P8_Ina Instalações elétricas	AM_3	394
6.70	0.000	42.06	71.24	33.60	P2_R Paredes	AG_2	630
3.77	0.000	40.44	59.68	29.28	P10_Ruim Iluminação	AO_2	549
3.00	0.000	38.20	60.48	31.41	P1_R Telhado	AF_2	589
2.52	0.000	36.17	62.90	34.51	P3_R Piso	AH_2	647
2.07	0.000	34.68	64.52	36.91	P4_R Portas e janelas	AI_2	692
1.34	0.000	31.43	70.97	44.80	P13_Ruim Espaço para recreação	AR_2	840
1.22	0.000	34.76	59.14	33.76	P9_Ruim Ventilação	AN_2	633
1.02	0.000	34.61	58.33	33.44	P12_Ruim Area externa(patio, jardins)	AQ_2	627
8.82	0.000	32.41	50.54	30.93	P6_R Cozinha	AK_2	580
8.57	0.000	33.67	44.62	26.29	P11_Ruim Ruidos	AP_2	493
6.42	0.000	33.04	30.65	18.40	P6_Ina Cozinha	AK_3	345
5.88	0.000	27.72	46.51	33.28	Municipal Rede	AB_2	624
5.14	0.000	31.33	25.27	16.00	P4_Ina Portas e janelas	AI_3	300
2.66	0.004	22.65	50.54	44.27	Estadual Rede	AB_1	830
2.51	0.006	33.33	5.91	3.52	RJ Unidades da Federação	AE19	66

ASSE 5 / 5

TEST	PROBA	POURCENTAGES			MODALITES		IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES	DES VARIABLES		
		9.55	CLASSE	5 / 5			aa5a	179
1.41	0.000	75.16	64.25	8.16	P2_Ina	Paredes	AG_3	153
1.02	0.000	55.28	75.98	13.12	P3_Ina	Piso	AH_3	246
0.65	0.000	48.00	80.45	16.00	P4_Ina	Portas e janelas	AI_3	300
8.85	0.000	54.93	65.36	11.36	P1_Ina	Telhado	AF_3	213
6.93	0.000	33.18	78.77	22.67	P7_Ina	Instalações hidráulicas	AL_3	425
6.27	0.000	33.76	74.30	21.01	P8_Ina	Instalações elétricas	AM_3	394
5.99	0.000	28.52	81.56	27.31	P5_Ina	Banheiros	AJ_3	512
4.77	0.000	33.91	65.36	18.40	P6_Ina	Cozinha	AK_3	345
2.93	0.000	24.04	73.74	29.28	P10_Ruim	Iluminação	AO_2	549
1.56	0.000	21.05	73.74	33.44	P12_Ruim	Area externa (patio, jardins)	AQ_2	627
0.80	0.000	20.22	71.51	33.76	P9_Ruim	Ventilação	AN_2	633
9.51	0.000	16.67	78.21	44.80	P13_Ruim	Espaço para recreação	AR_2	840
7.97	0.000	19.27	53.07	26.29	P11_Ruim	Ruidos	AP_2	493
3.98	0.000	12.65	58.66	44.27	Estadual	Rede	AB_1	830
2.47	0.007	20.00	7.26	3.47	AM	Unidades da Federação	AE_3	65

Em síntese, as características das classes de escolas segundo o estado do prédio são as seguintes:

- Classe 1

Está composta por escolas que se caracterizam por ter um bom estado de conservação do prédio todo. A metade delas pertencem à Rede Particular, 40 % delas estão localizadas na Capital, 85% delas são urbanas, 80% das escolas particulares estão nesta classe e das escolas da Região Sul 46% pertencem a esta classe.

- Classe 2

Uma proporção importante das escolas apresentam em condições regulares as: instalações hidráulicas, elétricas e os banheiros; enquanto o estado regular das portas e janelas, cozinha, telhado e paredes; caracteriza a quase a metade das escolas da classe. Um pouco mais da metade delas pertencem à Rede Estadual.

- Classe 3

Está formada por escolas nas quais não existem instalações hidráulicas, suas áreas externas são ruins assim como os espaços para recreação. Uma proporção importante delas apresentam ventilação e iluminação ruins; paredes, portas e janelas inadequadas, não tendo cozinha. Quase 80% delas estão na Zona Rural, 75% delas são da Rede Municipal, 92% delas estão localizadas no interior, quase

a metade delas pertencem ao Nordeste, enquanto que 27% pertencem à Região Centro-Oeste. Os estados que mais caracterizam esta classe são: CE (23%), MT (12%) e GO (12%).

- Classe 4

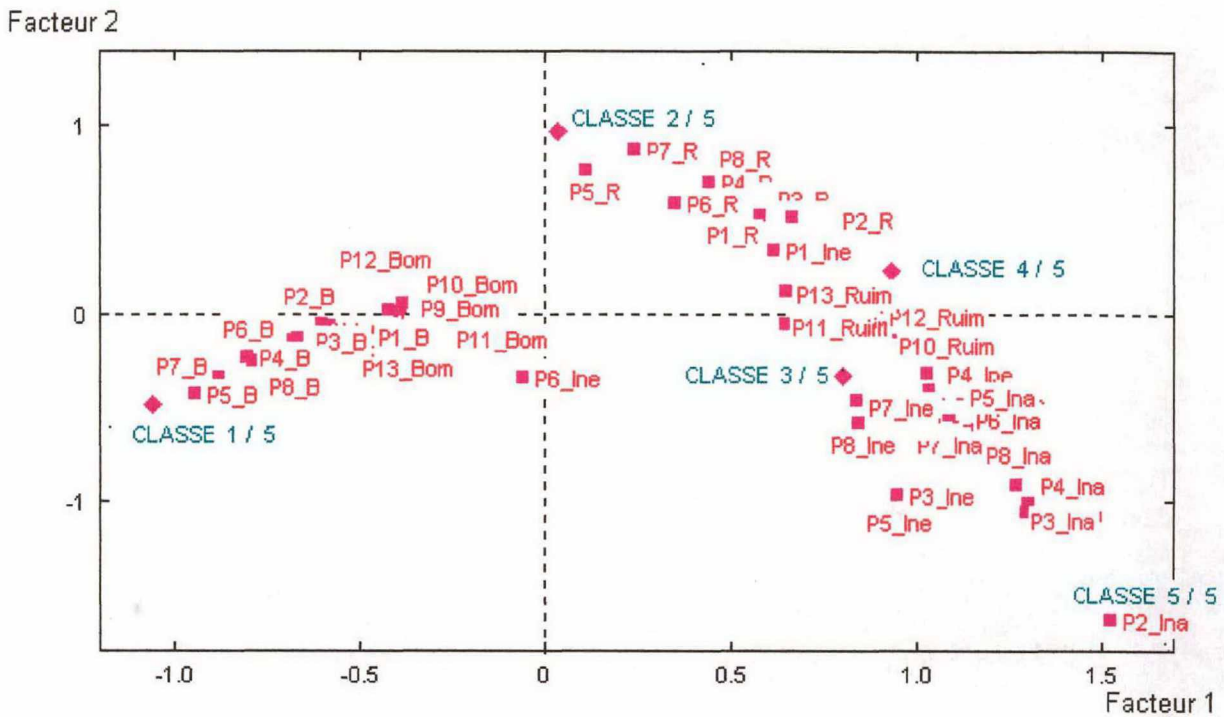
Trata-se de escolas que apresentam-se regulares em alguns aspectos: paredes, telhado, piso, portas e janelas, cozinha; enquanto outros são inadequados: banheiros, instalações elétricas e hidráulicas. A metade delas são da Rede Estadual e 47% da Rede Municipal. Pertencem a esta classe a terceira parte das escolas de RJ.

- Classe 5

São escolas nas quais o estado de conservação do prédio é inadequado, 59% delas pertencem à Rede Estadual, sendo que pertencem a esta classe 20% das escolas de AM.

Finalmente, a cada classe corresponde um centro de gravidade ou escola-tipo média de cada classe. Os centros de gravidade de cada classe podem ser projetados no primeiro plano fatorial na qualidade de indivíduo ilustrativo. Essa projeção ajuda a compreender quais são as características mais marcantes de cada grupo de escolas as quais esse centro de gravidade representa.

Figura 4.41. Projeção da tipologia de escolas "condições dos prédios", no espaço das modalidades das variáveis descritivas do estado dos prédios escolares



Se projetarmos os centros de gravidade de cada classe sobre o primeiro eixo, eles ficariam ordenados da classe 1 à classe 5 (de esquerda a direita); e sabendo que esse eixo opõe as melhores condições do prédio às piores, teríamos uma interpretação simples das características das escolas de cada classe.

Os valores dessas projeções figuram na Figura 4.42.

Figura 4.42. Coordenadas dos centros de classe da partição pelas características do Prédio

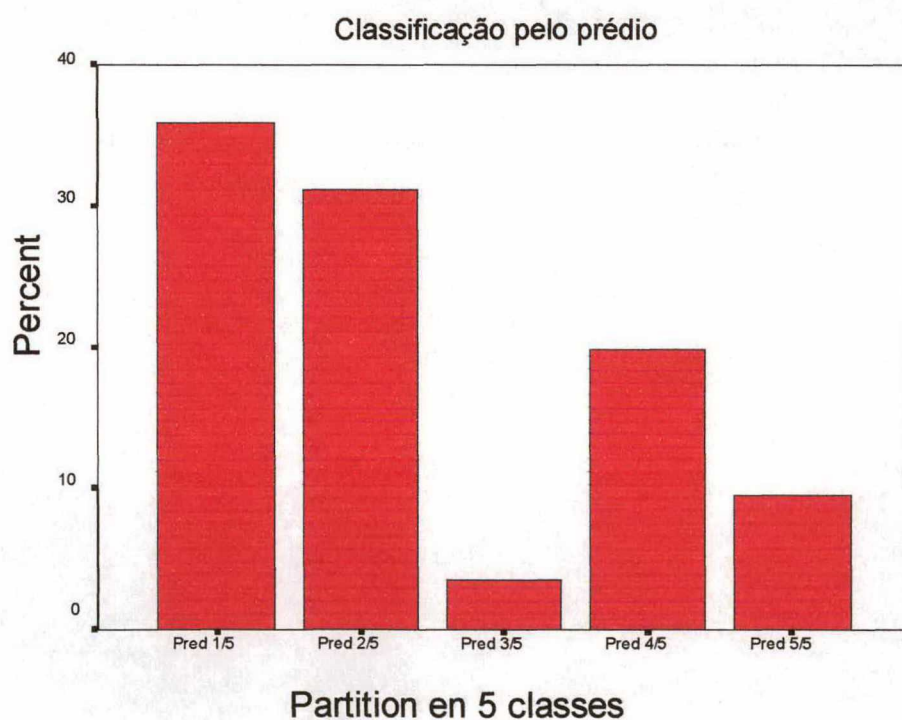
COORDONNEES ET VALEURS-TEST APRES CONSOLIDATION													
AXES 1 A 5													
CLASSES			VALEURS-TEST					COORDONNEES					
IDEN-LIBELLE	EFF.P.ABS		1	2	3	4	5	1	2	3	4	5	DISTO.
COUPURE 'a' DE L'ARBRE EN 5 CLASSES													
aa1a-CLASSE 1/5	673	673.0	-34.3	-15.7	3.2	3.5	1.9	-0.72	-0.21	0.03	0.03	0.02	0.56
aa2a-CLASSE 2/5	585	585.0	1.0	28.2	-8.5	-6.1	-11.3	0.02	0.41	-0.09	-0.06	-0.11	0.19
aa3a-CLASSE 3/5	66	66.0	6.6	-2.8	-23.5	3.7	21.3	0.54	-0.14	-0.88	0.13	0.75	1.69
aa4a-CLASSE 4/5	372	372.0	20.1	4.8	18.4	6.5	8.8	0.63	0.10	0.26	0.09	0.12	0.50
aa5a-CLASSE 5/5	179	179.0	23.1	-23.7	-2.1	-7.4	-10.4	1.11	-0.72	-0.05	-0.15	-0.22	1.82

Analisadas as características predominantes nas escolas de cada classe, interessa avaliar qual o peso que tem cada uma delas no conjunto das escolas brasileiras observadas no SAEB97. A distribuição das escolas pela classe de prédio apresenta-se na tabela e gráfico seguintes:

Tabela 4.5. Distribuição das escolas pelas características do prédio

	Frequency	Percent
Pred 1/5	673	35,9
Pred 2/5	585	31,2
Pred 3/5	66	3,5
Pred 4/5	372	19,8
Pred 5/5	179	9,5
Total	1875	100,0

Figura 4.43. Distribuição das escolas pela tipologia do prédio.



Esta distribuição das escolas observadas segundo as características dos seus prédios permite dimensionar a situação dos prédios no total.

4.3.1.2. Tipologia das escolas descritas pelas instalações

4.3.1.2.1. Construção da tipologia de "condições das instalações" das escolas

Essa tipologia é o produto do estudo das semelhanças das escolas pelas condições de funcionamento das instalações, incluídas nos itens 14 a 22 do questionário, que são:

- Biblioteca e qualidade de livros
- Laboratório de Ciências
- Laboratório de Informática
- Auditório
- Quadra de esportes o Ginásio
- Vestiários
- Sala de professores
- Sala de aula
- Qualidade de livros

Cada uma das instalações foi classificada no questionário como: Bom, Regular, Inadequado ou Inexistente, com exceção da qualidade dos livros que se classificou em três: boa, má ou não tem.

Processando uma análise de correspondências múltiplas seguida de uma classificação e partição obtém-se uma tipologia constituída por 5 classes de escolas:

Classe 1: 463 escolas

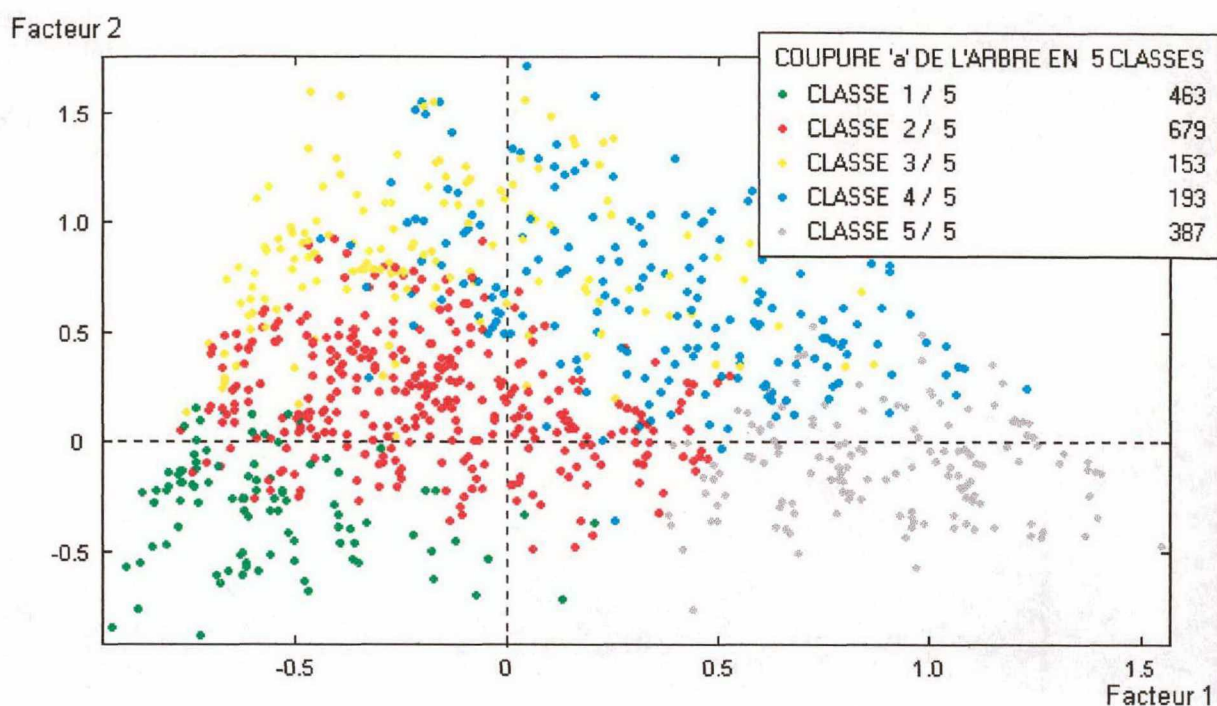
Classe 2: 679 escolas

Classe 3: 153 escolas

Classe 4: 193 escolas

Classe 5: 387 escolas

Figura 4.44. Projeção das escolas no primeiro plano principal da ACM diferenciando as classes as quais pertencem (instalações).



4.3.1.2.2. Caracterização das classes pela tipologia de "condições dessas instalações" das escolas

- Classe 1

Está formada por escolas que não tem biblioteca, quadra de esportes, laboratório de ciências e de informática, auditório, vestiários. Uma grande proporção delas não tem sala de professores e 57% pertencem à Rede Municipal, 83% estão localizadas no interior e 46% são da Região Nordeste. Pertencem a essa classe 44% das escolas do CE, 44% das escolas do MA, 43% das escolas de MG e 42% das escolas do MT.

- Classe 2

Está composta por escolas que não tem laboratórios de ciências e informática, quadra de esportes, vestiários, auditório; as bibliotecas são inadequadas ou

regulares; as salas de aula e de professores também são regulares e 34% são da Região Norte, 52% pertencem à Rede Estadual e 38% à Rede Municipal. Pertencem a esta classe 53% das escolas de TO, 55% das escolas de AP e 51% das escolas de RO.

- Classe 3

Está formada por escolas que apresentam - em importante proporção - laboratórios de ciências, auditórios, quadra de esportes, vestiários inadequados e - em menor proporção - bibliotecas e salas de aula inadequadas, salas de professores regulares, e livros sem qualidade didática, 87% delas não possuem laboratório de informática, 65% são da Rede Estadual, 14% das escolas da Região Sul pertencem a esta classe; sendo que 18% das escolas de RS são desta classe.

- Classe 4

Trata-se de escolas que apresentam condições regulares para os laboratórios de ciências, de informática, auditório, quadra de esportes e vestiários. A biblioteca ou é regular ou é boa e são boas as salas de professores, 59% delas pertencem à Rede Estadual, 47% estão localizadas na Capital. Pertencem a esta classe 20% das escolas do DF e 16% das escolas da Região Sul.

- Classe 5

Estão nesta classe as escolas que apresentam boas todas em suas instalações, 70% delas pertencem à Rede Particular, 95% são urbanas e 47% estão localizadas na Capital. Pertencem a esta classe 27% das escolas da Região Sudeste, 27% da Região Centro-Oeste; 33% das escolas de SP e 33% das escolas do DF.

Figura 4.45. Projeção da tipologia de escolas "condições das instalações", no espaço das modalidades das variáveis descritivas das condições das instalações escolares

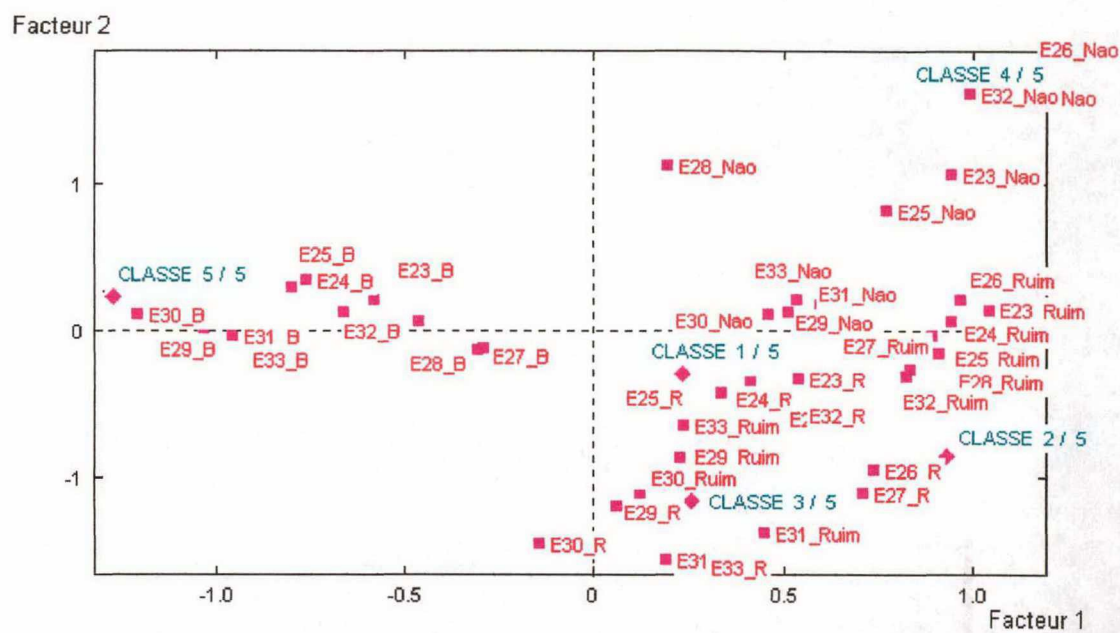
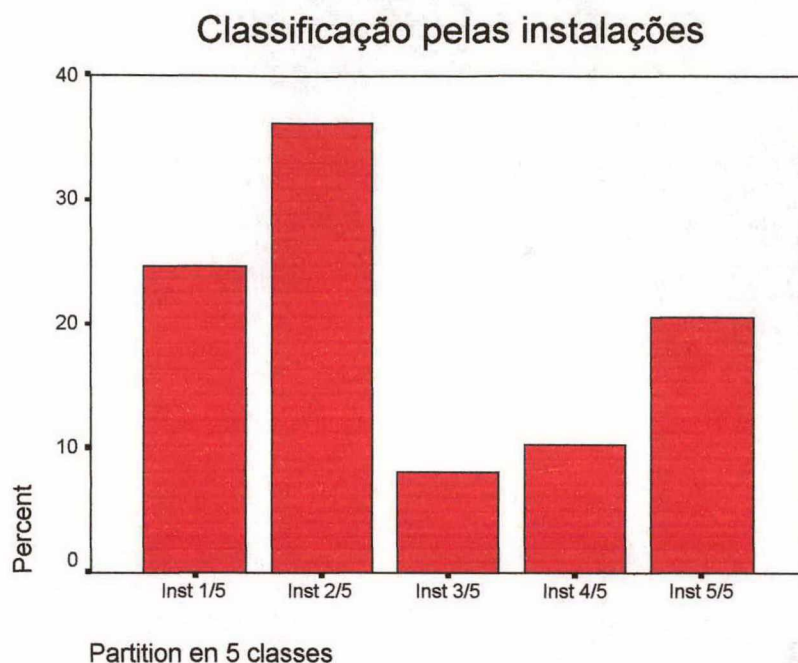


Tabela 4.6. Distribuição das escolas pelas características das instalações

	Frequency	Percent
Inst 1/5	463	24,7
Inst 2/5	679	36,2
Inst 3/5	153	8,2
Inst 4/5	193	10,3
Inst 5/5	387	20,6
Total	1875	100,0

Figura 4.46. Distribuição das escolas pela tipologia das instalações

4.3.1.3. Tipologia das escolas descritas pelos equipamentos

4.3.1.3.1. Construção da tipologia de "condições dos equipamentos" das escolas

Esta tipologia é resultante do estudo das semelhanças das escolas considerando os seus equipamentos considerados nos itens 23 até o 33 do questionário, que são:

- Quadro negro
- Carteiras
- Mesas
- Televisão
- Videocassete
- Mimeógrafo
- Máquina fotocopadora
- Projetor de slides
- Retroprojetor
- Máquinas de datilografia

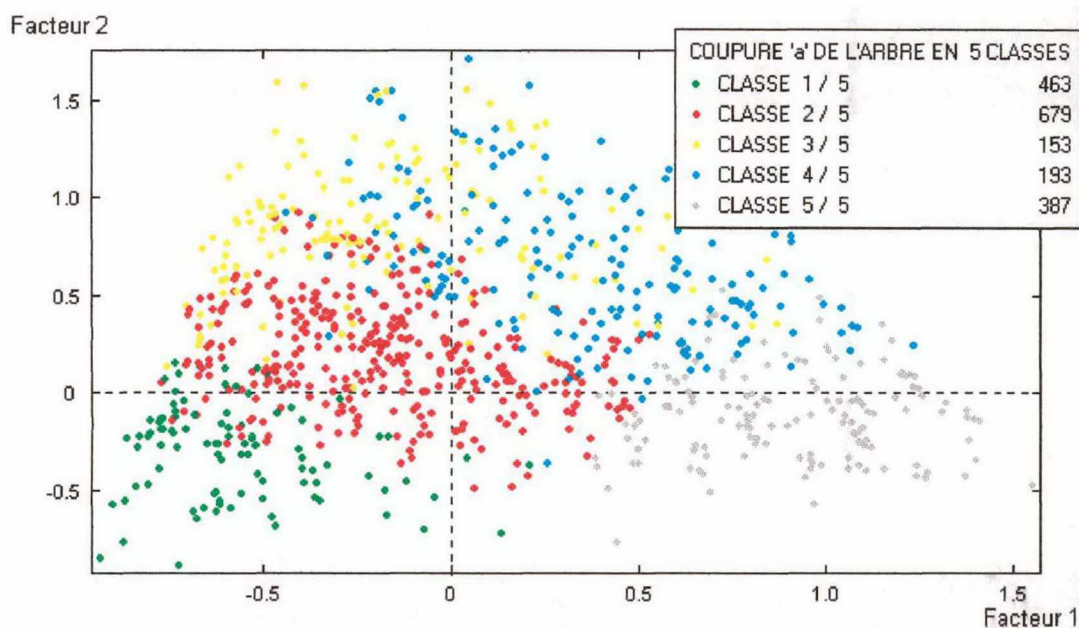
- Computador

Todos os itens foram classificados como: Bom, Regular, Ruim e Não existe.

O resultado do processamento estatístico da análise de correspondências múltiplas seguido de uma classificação e partição é uma tipologia constituída por 5 classes de escolas:

- Classe 1: 766 escolas
- Classe 2: 253 escolas
- Classe 3: 74 escolas
- Classe 4: 231 escolas
- Classe 5: 551 escolas

Figura 4.47. Projeção das escolas no primeiro plano principal da ACM diferenciando as classes as quais pertencem (equipamentos)



4.3.1.3.2. Caracterização das classes pela tipologia de "condições dos equipamentos" das escolas

- Classe 1

Está formada por escolas que tem carteiras, mesas, quadro negro, mimeógrafo e máquina de datilografia em estado regular; e que não tem projetor de slides,

máquina fotocopadora, retroprojektor, computador; mas tem televisão e videocassete em bom estado., 57 % delas são da Rede Estadual e 38% da Rede Municipal, 28% são rurais,31% estão localizadas na Região Norte e 79% das escolas de RR estão nesta classe.

- Classe 2

Uma parte importante das escolas desta classe tem carteiras, mesas, quadro negro, máquinas de datilografia e mimeógrafo em estado ruim; televisão e videocassete em estado regular e não tem retroprojektor, projetor de slides e nem computador, 57% delas são da Rede Estadual e 28% das escolas de SE pertencem a esta classe.

- Classe 3

Está composta por escolas que tem retroprojektor e projetor de slides em estado ruim, 72% delas pertencem a Rede Estadual. A metade delas estão localizadas na Capital, 88% são urbanas e 12% pertencem ao DF.

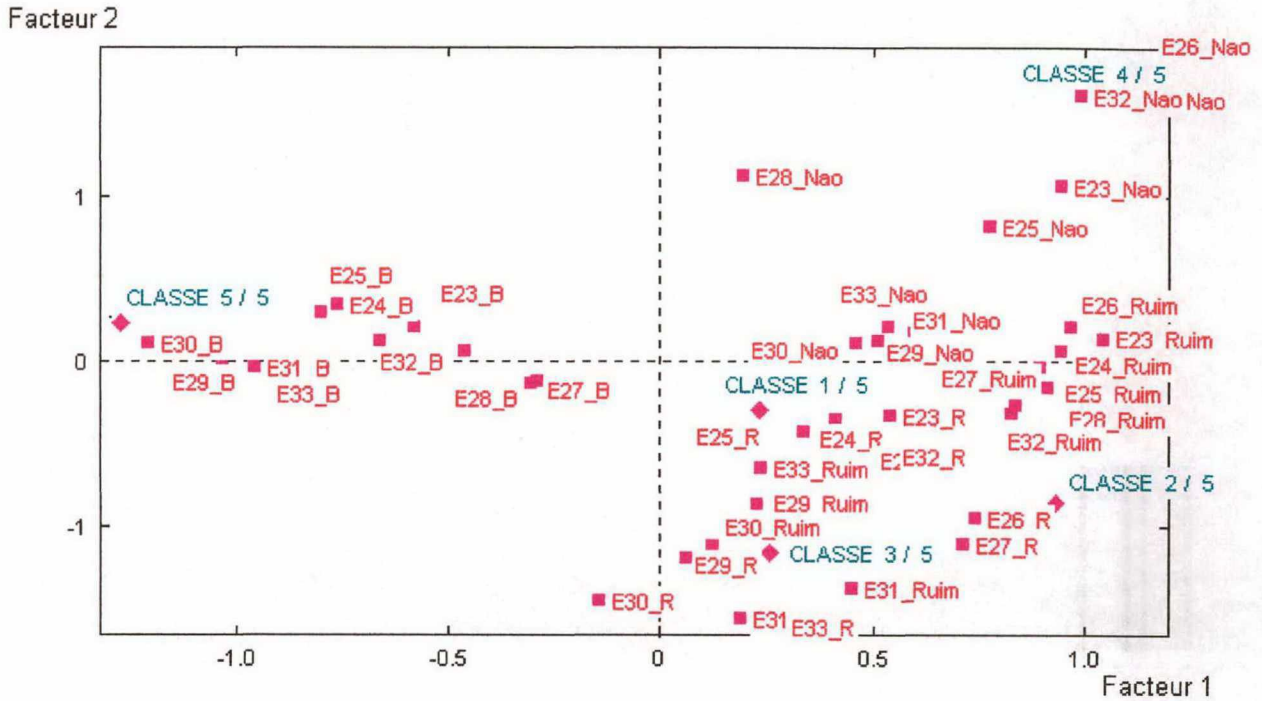
- Classe 4

As escolas desta classe caracterizam-se por não ter praticamente nenhum equipamento. A metade delas estão na Zona Rural, 71% delas são da Rede Municipal, 91% delas estão localizadas no interior, 54% delas pertencem à Região Nordeste sendo 28% das escolas da PB, 25% das escolas do CE, 26% das escolas da BA e 24% das escolas do MT.

- Classe 5

São escolas que tem todo seu equipamento em bom estado sendo que 62% delas são particulares, 91% são urbanas e 44% estão localizadas na capital. Pertencem a esta classe 52% das escolas da Região Sul, 39% das escolas da Região Centro-Oeste e 37% das escolas da Região Sudeste. Estão nesta classe 58% das escolas de SC, 55% das escolas de RS, 51% das escolas do DF, 47% das escolas de SP e 43% das escolas de MS.

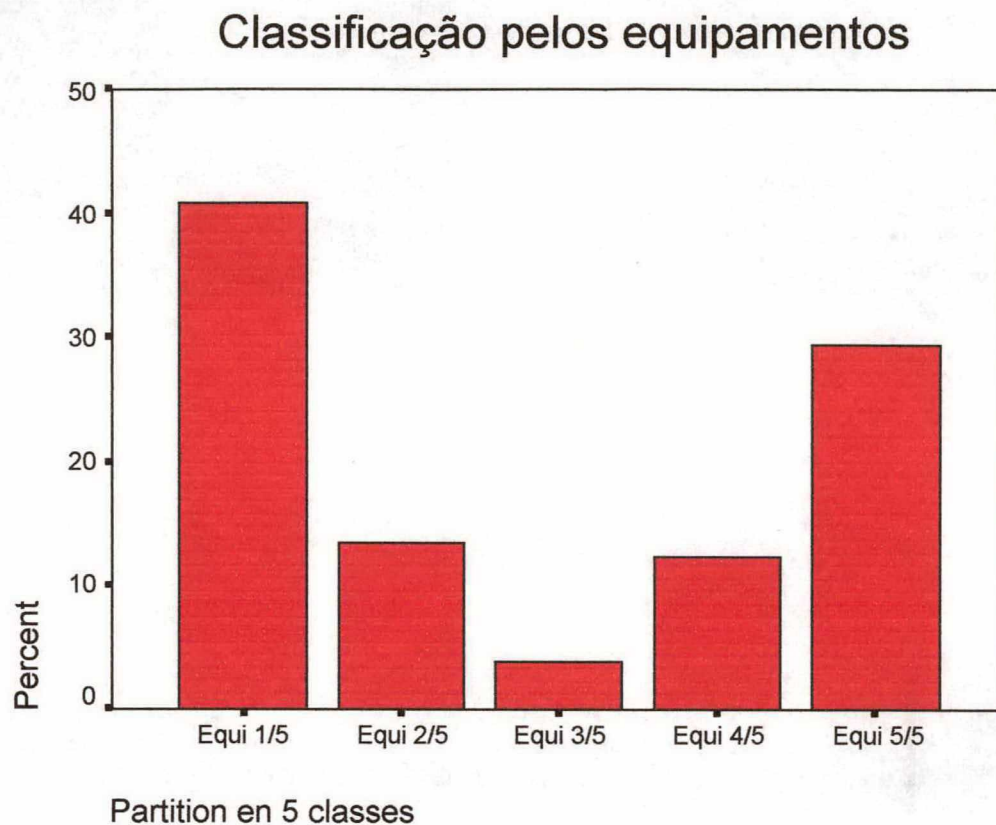
Figura 4.48. Projeção da tipologia de escolas "estado dos equipamentos", no espaço das modalidades das variáveis descritivas do estado dos equipamentos escolares.



A distribuição das escolas pela classe de equipamentos é a seguinte:

Tabela 4.7. Distribuição das escolas pelas características dos equipamentos

	Frequency	Percent
Equi 1/5	766	40,9
Equi 2/5	253	13,5
Equi 3/5	74	3,9
Equi 4/5	231	12,3
Equi 5/5	551	29,4
Total	1875	100,0

Figura 4.49. Distribuição das escolas pela tipologia dos equipamentos.

4.3.1.4. Tipologia das escolas descritas pelos serviços oferecidos

4.3.1.4.1. Construção da tipologia de "condições dos serviços" das escolas

Esta tipologia é resultante do estudo das semelhanças das escolas considerando os serviços oferecidos, incluídos nos itens 35 até o 37 do questionário. Eles são:

- Médicos
- Odontológicos
- Transporte
- Alimentação

Todos os itens foram classificados em 3 modalidades: Pago, Gratuito e Não oferece.

O resultado da aplicação da análise de correspondências múltiplas seguido de uma classificação e partição mostra uma tipologia constituída por 5 classes de escolas:

Classe 1: 1006 escolas

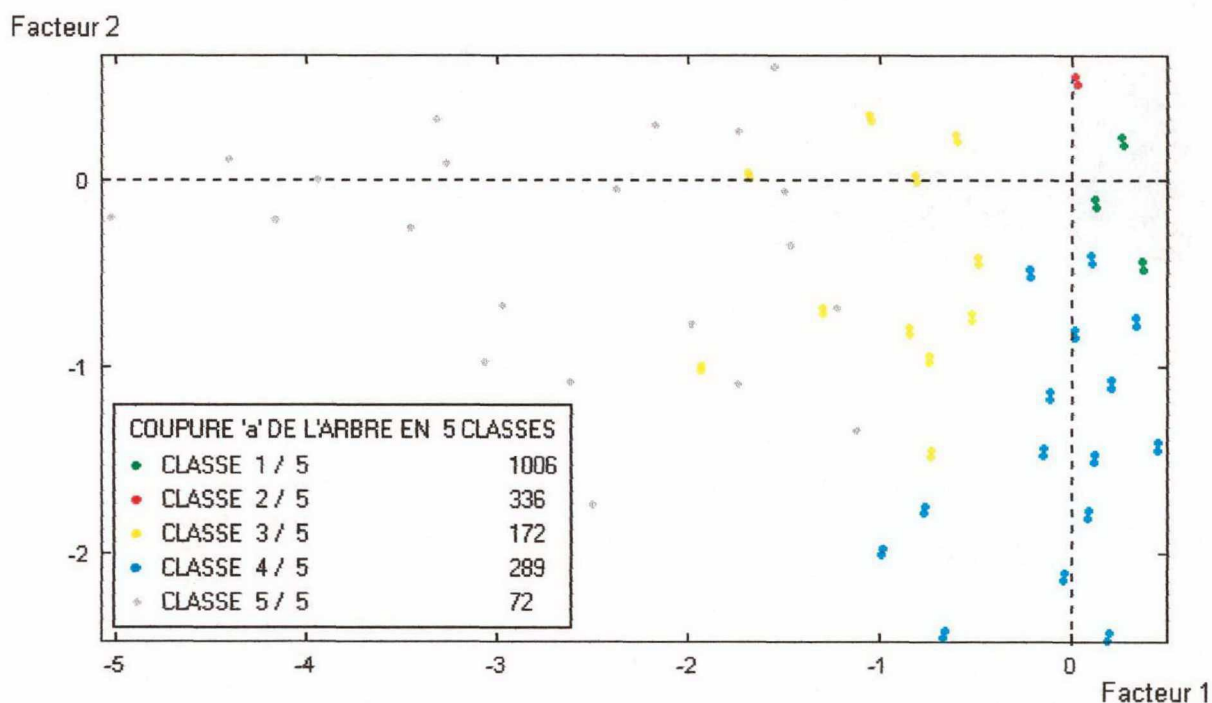
Classe 2: 336 escolas

Classe 3: 172 escolas

Classe 4: 289 escolas

Classe 5: 72 escolas

Figura 4.50. Projeção das escolas no primeiro plano principal da ACM diferenciando as classes as quais pertencem (serviços).



4.3.1.4.2. Caracterização das classes pela tipologia de "condições dos serviços" das escolas

- Classe 1

Está formada por escolas que não oferecem serviços odontológicos nem médicos e o serviço de alimentação é gratuito. Sendo que 44% delas pertencem à

Rede Municipal, 53% à Rede Estadual, 34% são rurais e 72% estão localizadas no interior. Das escolas da MT 70% pertencem a esta classe.

- Classe 2

São escolas que não oferecem nenhum serviço. A metade delas pertencem à Rede Particular e quase todas são urbanas, sendo que 42% estão na capital. Pertencem a esta classe 36% das escolas do MA e 32% das escolas do DF.

- Classe 3

Trata-se de escolas que tem serviço de alimentação e transporte pagos e que não oferecem serviço odontológico. Sendo que 58% delas são da Rede Particular. Os estados que apresentam escolas deste tipo são: SP (18%) e RS (19%).

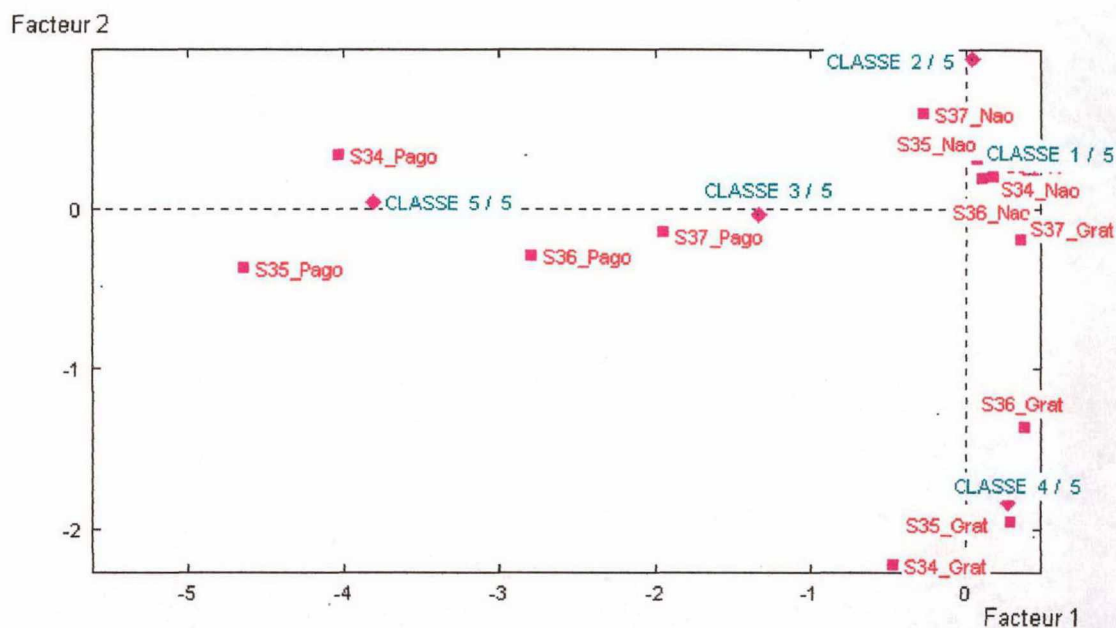
- Classe 4

Uma grande quantidade de escolas desta classe oferecem serviços odontológicos e de alimentação gratuitos. Quase a metade oferecem serviços médicos gratuitos em algumas o serviço de transporte é gratuito. Pertencem a esta classe 41% das escolas de MS, 30% das escolas de RR e 84% delas são urbanas.

- Classe 5

Todas as escolas que oferecem serviços odontológicos e médicos pagos pertencem a esta classe. Uma proporção importante das escolas dessa classe oferecem serviço médico pago e uma proporção menor serviço odontológicos e de transporte pagos. São escolas que ou oferecem serviço pago de alimentação ou não o oferecem. Sendo que 82% delas pertencem à Rede Particular e 92% são urbanas.

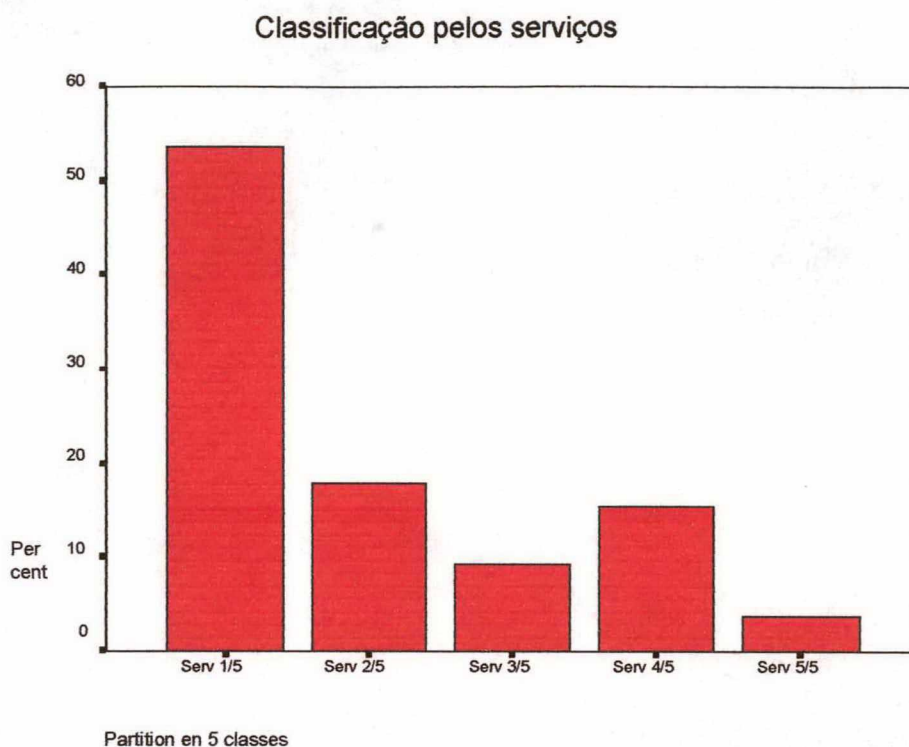
Figura 4.51. Projeção da tipologia de escolas "serviços oferecidos", no espaço das modalidades das variáveis descritivas dos serviços.



A distribuição das escolas pela classe de serviços é a seguinte:

Tabela 4.8. Distribuição das escolas pelos serviços oferecidos

	Frequency	Percent
Serv 1/5	1006	53,7
Serv 2/5	336	17,9
Serv 3/5	172	9,2
Serv 4/5	289	15,4
Serv 5/5	72	3,8
Total	1875	100,0

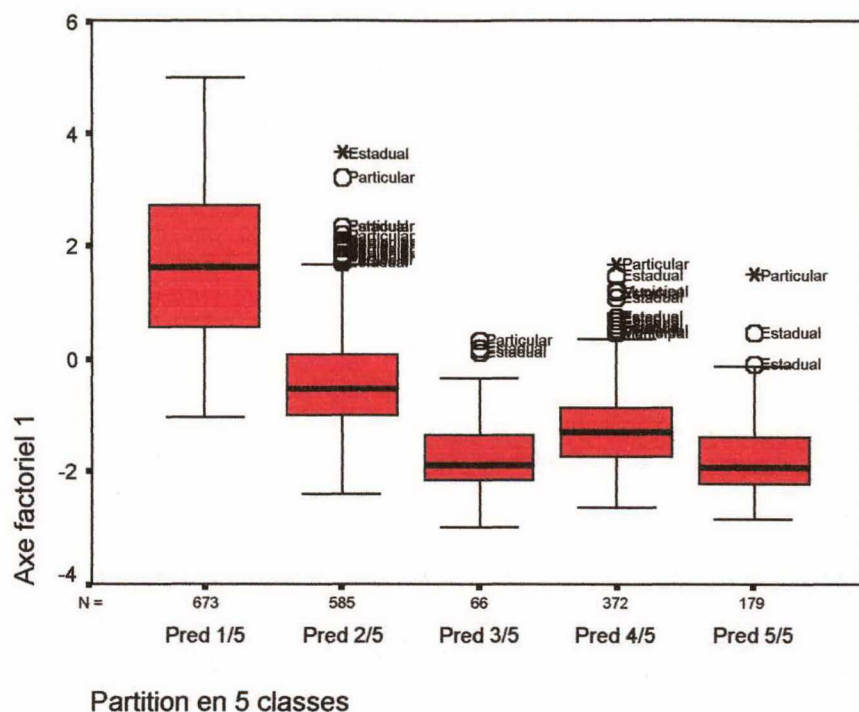
Figura 4.52. Distribuição das escolas pela tipologia dos serviços

4.3.1.5 Cruzamento do indicador de "recursos materiais" com as tipologias das escolas por tipo de recursos

Uma vez construídas as quatro tipologias, importa ver quais são os valores atribuídos pelo indicador as escolas que pertencem a mesma classe (em cada uma das tipologias).

O objetivo é comprovar se as escolas que se assemelham pelas características de seus prédios – aquelas que pertencem a mesma classe da tipologia do estado do prédio - assumem valores semelhantes no indicador. E assim para as restantes das tipologias: instalações, equipamentos e serviços. Para isso podem ser usadas as ferramentas básicas da estatística exploratória que permitem visualizar a distribuição do indicador segundo as modalidades da tipologia.

Figura 4.53. *Boxplot* do indicador segundo as classes da tipologia do estado do Prédio



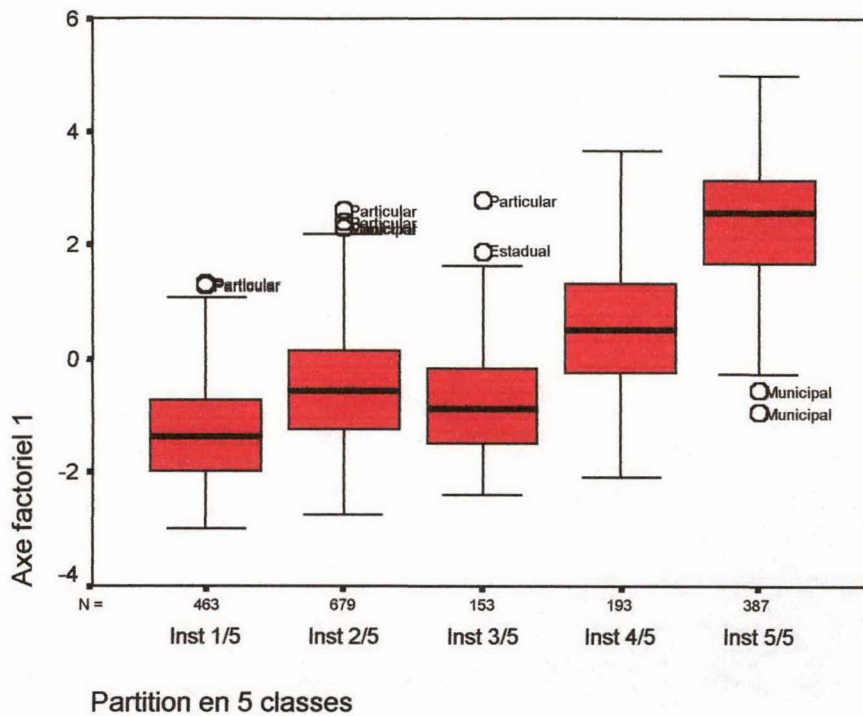
Em linhas gerais, o indicador consegue diferenciar as classes de escolas que se diferenciam pelo estado do prédio e atribui valores semelhantes as escolas que pertencem à mesma classe.

A classe que tem mais variabilidade (nos valores do indicador) é a número 1 que é aquela com um bom estado geral do prédio; essa classe assume os valores mais altos do indicador. Segue a ela – nos valores do indicador - a Classe 2 que tem banheiros e instalações elétricas regulares e a Classe 4 que tem prédios regulares com banheiros e instalações elétricas inadequadas.

A Classe 3 – cujas características essenciais são as áreas externas ruins e não ter cozinha, não se diferencia muito – nos valores do indicador - da Classe 5 na qual o estado geral do prédio é ruim. Mas – pelo visto anteriormente na Figura 4.25 - a Classe 3 era a mais heterogênea da partição.

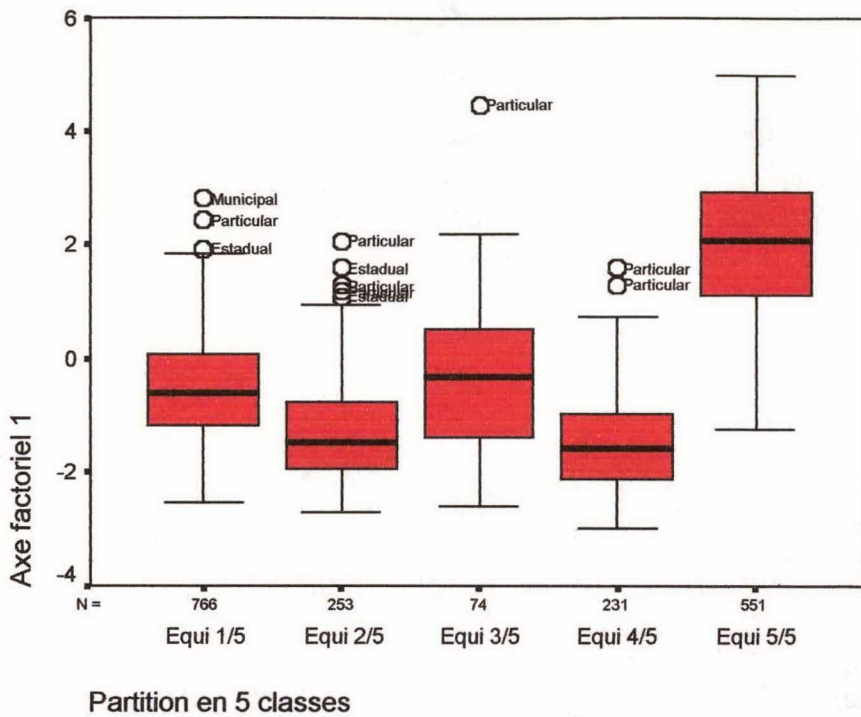
Note-se que as escolas às quais o indicador atribui valores maiores que as de sua classe são ou particulares ou estaduais.

Figura 4.54. *Boxplot* do indicador segundo as classes da tipologia do estado das Instalações



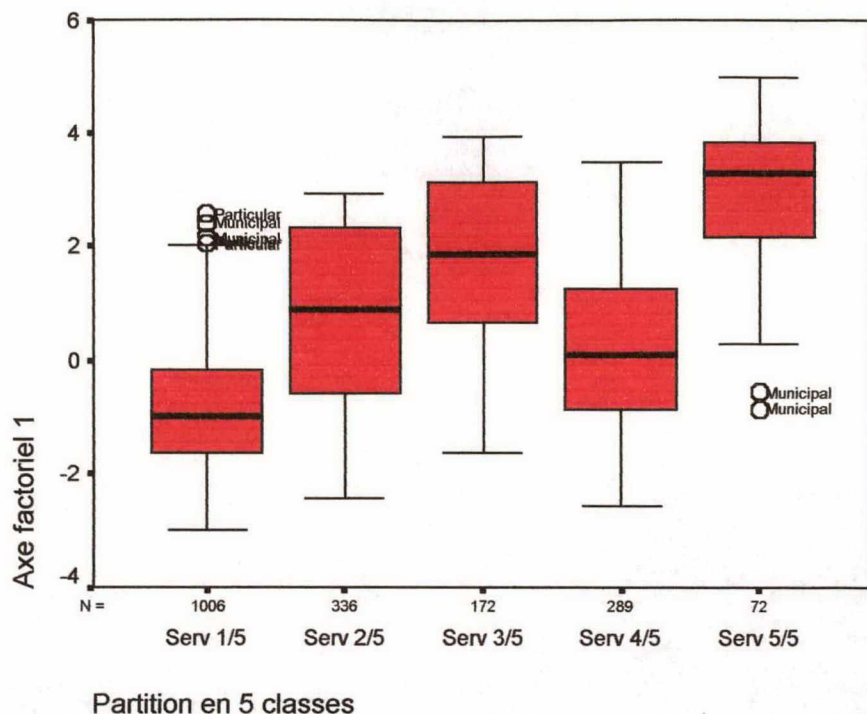
O indicador diferencia as escolas que pertencem as diferentes classes da tipologia das instalações. Os valores mais altos do indicador são atribuídos as escolas que tem todas as instalações boas. Dentre elas existem duas escolas municipais que assumem valores do indicador menores que as do seu grupo. As escolas do Grupo 1 - que são aquelas sem instalações - são as que assumem os menores valores do indicador, seguidas das do Grupo 3 que tem laboratórios e bibliotecas inadequados. As escolas do Grupo 2, que não tem laboratórios nem biblioteca, assumem valores maiores que as do Grupo 3. Existem poucas escolas nas quais o indicador assume valores muito diferentes em relação a classe correspondente.

Figura 4.55. *Boxplot* do indicador segundo as classes da tipologia do estado dos Equipamentos



O indicador diferencia claramente as escolas que tem um bom equipamento (Classe 5) das classes restantes. Elas tem os maiores valores, seguidas das da Classe 3 que tem computadores regulares. Com valores menores aos anteriores aparece a Classe 1, que tem equipamentos regulares mas sem computador e retroprojektor. As Classes 2 e 4 assumem os menores valores do indicador e bastante semelhantes; ambas as classes apresentam equipamentos ruins com pequenas diferenças entre as mesmas.

Figura 4.56. *Boxplot* do indicador segundo as classes da tipologia dos Serviços oferecidos



O indicador diferencia as escolas que pertencem as diferentes classes, mas as classes tem variabilidade nos valores do indicador. As classes ficam ordenadas em forma crescente da seguinte maneira: Classe 1, que oferece alimentação gratuita, Classe 4, que oferece alimentação e serviço odontológico grátis; Classe 2 que não oferece nada; Classe 3, que oferece alimentação e transporte pago e serviços médicos e odontológicos gratuitos; Classe 5 que oferece serviços pagos.

4.3.2. Validação externa do indicador de “recursos materiais” das escolas

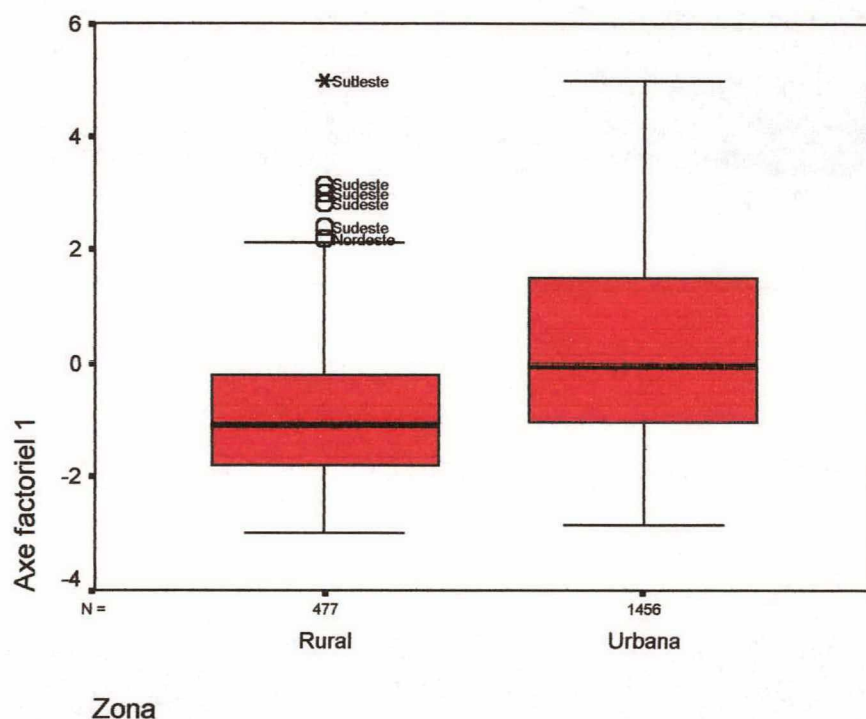
O objetivo desta seção é mostrar se o que o indicador está medindo corresponde com o que se conhece dos recursos materiais das escolas; usando informação externa; quer dizer, informação que não foi considerada na construção do indicador. A única informação disponível que não foi usada para gerar o indicador é a que refere-se a rede escolar quanto a localização da escola, da zona

a qual pertence e da região. Essas características correspondem às variáveis de estratificação das escolas segundo o SAEB97.

A construção do indicador foi feita por meio de uma AFM da tabela de características materiais das escolas estruturadas em quatro grupos. Nessa análise as variáveis de estratificação foram usadas como ilustrativas; de modo que na seção §4.2.4. já aparece uma primeira validação do indicador projetando essas variáveis ilustrativas na estrutura conjunta dos recursos materiais das escolas.

Apresenta-se a seguir, o cruzamento do indicador com as variáveis de estratificação para analisar as distribuições dele nas diversas modalidades das variáveis de estratificação.

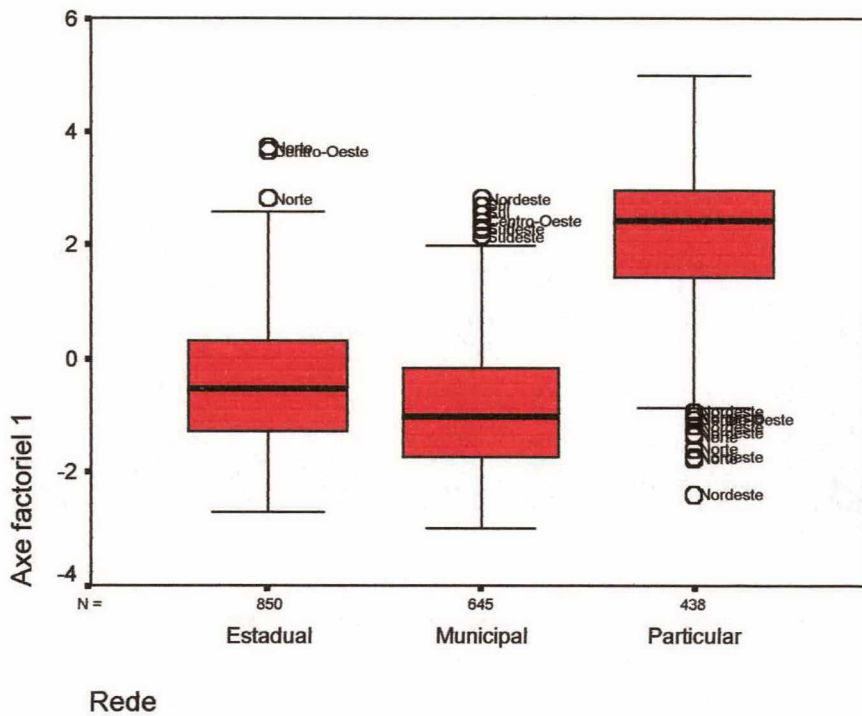
Figura 4.57. *Boxplot* do indicador segundo a Zona da escola



Os valores que o indicador de recursos materiais atribui as escolas da área rural são inferiores aos atribuídos as da área urbana. A distribuição do indicador é mais homogênea nas escolas da área rural. Isso indica que as escolas da área rural tem piores condições físicas que as da área urbana; e que a diversidade na situação de recursos materiais é maior nas escolas da área urbana. Dentre as

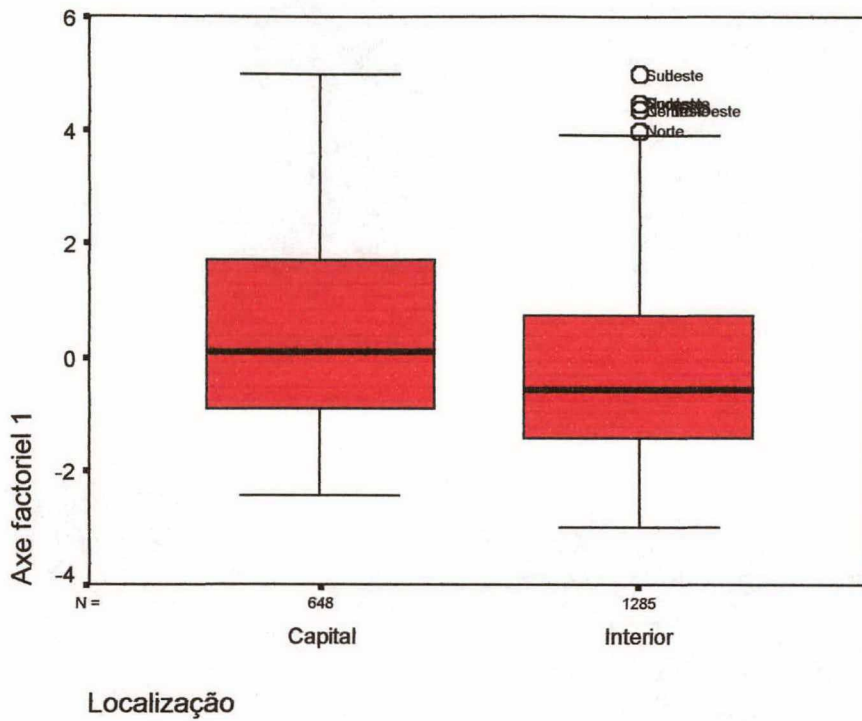
escolas da área rural existem algumas poucas que tem as melhores condições físicas; correspondendo com valores do indicador do mesmo nível que as melhores da área urbana. Elas situam-se na região Sudeste.

Figura 4.58. *Boxplot* do indicador segundo a Rede da escola

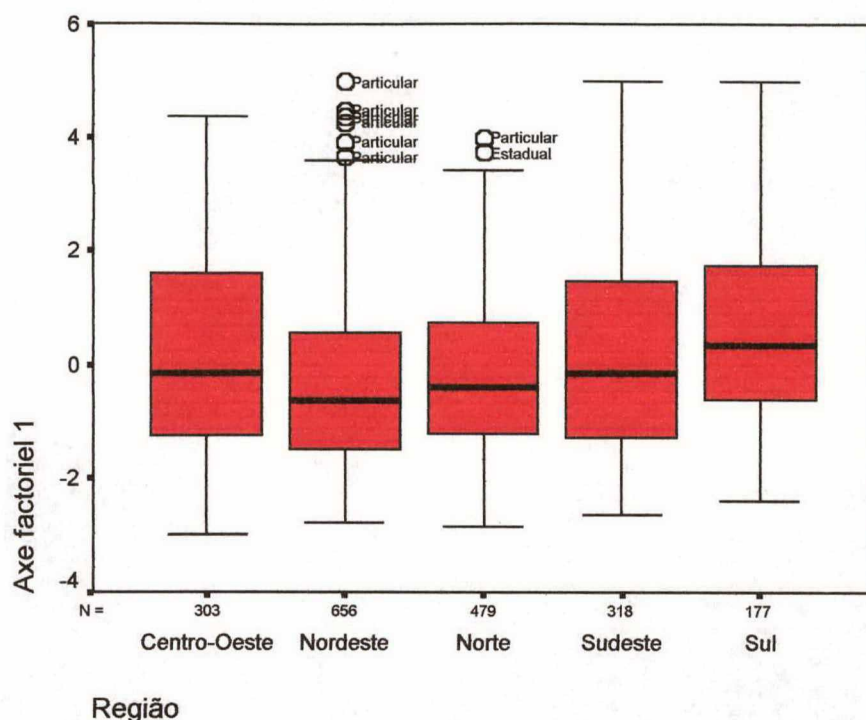


O indicador diferencia claramente as escolas da rede particular que apresentam valores sensivelmente maiores que as da rede estadual e municipal. Portanto, a situação de recursos das escolas da rede particular é bem melhor que a da rede pública. A diferença entre as escolas da rede estadual e municipal é menor estando a rede estadual em melhor situação que a municipal.

Dentre as escolas da rede particular, as que tem condições físicas notoriamente inferiores são as da região Nordeste.

Figura 4.59. *Boxplot* do indicador segundo a Localização da escola

Segundo os valores do indicador, a estrutura de recursos materiais das escolas das capitais é melhor que a das escolas do interior. Contudo, a variabilidade dentro de cada grupo é grande.

Figura 4.60. *Boxplot* do indicador segundo a Zona da escola

As diferenças na situação geral dos recursos - no total de escolas - não é tão grande entre as escolas das diferentes regiões porque todas as regiões apresentam situações boas e ruins de recursos. Contudo, os conjuntos de escolas que poderiam se chamadas de típicas de cada região, que representam a metade central; podem ser ordenadas pela situação de recursos físicos em forma ascendente (no sentido de pior a melhor estrutura de recursos): Nordeste, Norte, Sudeste, Centro-Oeste e Sul.

4.3.3. Limitações do indicador

A estratégia de tratamento dos dados, permitiu atingir os objetivos da pesquisa. No entanto; a validade do indicador de recursos materiais construído, está limitada por questões relativas à forma em que foi efetuado o processo de observação do SAEB; tanto do ponto de vista da amostra estudada quanto à

forma em que foram levantados e armazenados e divulgados os dados. Discutem-se a seguir esses aspectos que limitam a validade do indicador de recursos materiais construído porque impedem inferir ao universo de referência as conclusões extraídas da análise da estrutura da amostra de escolas. No entanto discute-se também a possibilidade de aplicar a metodologia que permite generalizar o indicador calculado sobre a amostra à população de escolas brasileiras.

4.3.3.1 Aspectos que limitam a validade do indicador de recursos materiais das escolas.

Discutem-se a seguir os principais aspectos metodológicos da pesquisa do SAEB97 que impedem ou perturbam a definição de operadores de inferência do indicador de "recursos materiais" das escolas construído a partir da amostra observada.

- **Qualidade da base de dados**

As bases de dados disponibilizadas pelo SAEB não estão acompanhadas da respectiva medida da qualidade. Como exposto anteriormente, essas bases contém erros sistemáticos de diferentes tipos (valores ausentes, valores inconsistentes, valores errôneos) que podem representar um erro de observação superior ao erro de amostragem.

No caso da base de dados usada no presente trabalho para construir o conceito de recursos materiais das escolas sintetizando esse conceito por meio do indicador construído; a base tem:

- escalas de opinião insuficientemente definidas o que traz como consequência à falta de comparabilidade das respostas. Por exemplo: a pergunta 22 "*Caso a escola possua biblioteca, há livros de boa qualidade para utilização didática e pedagógica?*", tem como respostas possíveis *Sim* ou *Não*. O questionário não padroniza as respostas, portanto não se pode garantir que as 1933 respostas sejam comparáveis.

- Carência de codificação que distinga a não resposta, a rejeição a responder e a resposta que não corresponde: todas essas situações aparecem na base como valores ausentes.
- Valores ausentes. O tratamento deles foi feito no marco desta pesquisa; mas para imputar esses valores foram usados critérios que podem ser discutíveis. Mesmo não sendo apresentada a forma de imputação dos "*missing values*" - porque foge aos limites deste estudo- esse processo de imputação deve ser realizado antes de difundir as bases de dados.
- **Escolha da estrutura temática da tabela objeto de análise.**

Para construir o conceito de recursos materiais das escolas, utilizou-se a informação levantada mediante o questionário das escolas, constituída por 37 aspectos observados. Essas 37 características foram estruturadas em 4 temas que agrupam as características observadas em aspectos relativos ao estado de conservação do prédio, as condições de funcionamento das instalações, ao estado dos equipamentos e aos serviços oferecidos. Esse agrupamento corresponde a um ponto de vista em relação ao objeto de estudo. O ponto de vista adotado é o gerenciamento dos recursos da rede o que o difere do ponto de vista pedagógico. Se para o primeiro, o agrupamento das variáveis em dois grupos distintos equipamentos e instalações faz sentido, já para o segundo, esse agrupamento poderia reunir todas as características em apenas um grupo, pois do ponto de vista pedagógico não faz sentido a existência de laboratórios sem os respectivos equipamentos. Planejar políticas de intervenção em relação a aquisição de mesas, computadores, comparativamente a construção de instalações para laboratórios demandam informações e processo de gestão diferenciados na administração pública, o que justifica a desagregação quando o ponto de vista predominante é o da gestão, o que gera, portanto uma limitação na análise dos dados. Outra limitação está circunscrita aos níveis estratégicos de decisão (Diretor, Secretarias Estaduais e Municipais de Educação, Ministério de Educação, etc.), neste caso, cada nível estratégico deve selecionar as variáveis disponíveis em dois grandes grupos:

- variáveis não controláveis (correspondem a realidades que não podem ser modificadas através da intervenção pelo nível de decisão correspondente).
- variáveis controláveis (correspondem a realidades que podem ser modificadas através da intervenção pelo nível de decisão correspondente).

Nessa perspectiva, os grupos temáticos dependem das possibilidades de intervenção do decisor. Programas tais como da merenda escolar, melhoramento da infra-estrutura, podem estar relacionados a variáveis tais como existência ou não de cozinhas nas escolas, existência ou não de laboratórios, existência ou não de equipamentos didáticos (vídeo, televisão), existência ou não de banheiros adequados, etc. permitindo um controle do gestor da rede ou do respectivo programa disponibilizado à rede.

- **Limitações da amostra estudada**

A amostra do SAEB é uma amostra probabilística de alunos tendo como variável mestre a proficiência e está estratificada segundo a Rede a qual pertença (municipal, estadual, privada), localização (capital e interior) e zona (urbana e rural). Permite a desagregação em nível da federação e da região pelas variáveis de estratificação.

Tomando como unidade de observação a escola, os resultados obtidos a partir dessa nova base de dados não nos permite fazer inferências estatísticas com precisão conhecida, e a mudança da unidade coloca em discussão a representatividade da amostra.

4.3.3.2. Validação do indicador em termos da estabilidade da estrutura estudada

Os objetivos destinados ao análise não se referem a "valores" dos quais se necessite estimar a importância no conjunto de escolas do país ou de um estado. Pelo tamanho da amostra é possível (e necessário) provar a estabilidade das estruturas observadas na informação disponível sobre as escolas, validando em consequência os resultados obtidos.

As estruturas estudadas pelas análises fatoriais não exigem uma representatividade tão estrita quanto as estimações de percentagens ou momentos de ordem 1. Obviamente uma amostra na qual alguns aspectos da população não estejam presentes, não fornecerá resultados extrapoláveis; mesmo sendo as configurações obtidas estáveis. No caso do SAEB97, dada a forma como foi realizada a distribuição da amostra nas regiões: por rede, por localização e por zona, fez com que a mesma possuía internamente a diversidade compreendida pelo conjunto de escolas do País. Portanto a não realização de um estudo de estabilidade representa uma limitação de caráter metodológico que questiona a possibilidade de generalização do indicador ao conjunto das escolas brasileiras.

4.4. Uso do indicador e a tipologia associada

A tipologia identifica classes de escolas com situações similares de recursos materiais e o indicador obtido reflete essas diferenças observadas. Também consegue-se identificar quais os aspectos de recursos que são deficitários em cada grupo de escolas. Desta forma, o indicador pode ser usado com fins de diagnóstico, oportunizando que se proponha uma política de intervenção diferenciada para cada tipo de escolas segundo as suas principais características e conseqüentemente suas carências. Alguns exemplos foram apresentados na seção 4.2.3..

Assim, o indicador aporta uma ampla avaliação dos recursos para o conjunto da rede escolar brasileira, constituindo um subsídio apropriado na orientação das atividades e planos de ação do sistema educativo.

Também poder-se-ia mostrar o emprego do indicador e a tipologia associada como um modelo de base sobre o observado para determinar o impacto previsível de diferentes ações de correção da situação global das escolas. Simulando políticas de intervenção em diferentes níveis de decisão poder-se-ia exemplificar o uso dessas informações com fins de predição de impactos de políticas. Mas, essa simulação necessita de uma definição mais clara das possibilidades de intervenção dos diferentes níveis de decisão do sistema educativo. As formas de

simular os impactos de possíveis políticas de intervenção seriam basicamente duas:

- produzindo teoricamente mudanças nas escolas de um estado em algumas das características observadas, é possível: comparar a trajetória das escolas melhoradas em relação as escolas observadas, estudar as mudanças de casse das escolas melhoradas na tipologia das escolas.
- definindo objetivos globais teóricos de mudanças em algumas das características observadas das escolas do Brasil; é possível mostrar o efeito dessas mudanças em relação as escolas de outras unidades ou redes.

CAPÍTULO V

5. VALIDAÇÃO DA METODOLOGIA DE ELABORAÇÃO DO INDICADOR

5.1. Construção de outros indicadores por formas convencionais.	153
5.2. Comparação dos resultados analisando os elementos de discordância	155

Nesta seção se compara o indicador obtido pela metodologia proposta neste trabalho com indicadores obtidos de outras formas. Essa comparação é abordada desde o ponto de vista dos objetivos que deve atingir o indicador enquanto uma medida que resume um fenômeno observado.

5.1. Construção de outros indicadores por formas convencionais.

Uma forma tradicional de sintetizar um conjunto de informações é através do cálculo da média simples. Assim, calcula-se um indicador por uma combinação linear dos índices observados que outorga a cada uma dessas características o mesmo peso. No caso do conceito que se quer medir – os recursos materiais das escolas brasileiras - os dados observados pelo SAEB apresentam as seguintes peculiaridades:

- São 37 características de recursos físicos portanto cada uma delas teria um peso de $1/37$.
- Os dados são códigos, que não possuem propriedades numéricas.

Duas observações surgem antes de comparar o comportamento da “média por indivíduo das características observadas” com o indicador obtido pela metodologia proposta nesta pesquisa. A primeira é em relação à natureza dos dados: o cálculo da média se realiza por uma combinação linear de códigos, portanto essa forma de resumo não é adequada para com a informação de tipo qualitativo que se pretende sintetizar. O único requisito desses códigos é que respeitem a relação de

ordem das modalidades das variáveis, portanto qualquer uma das duas codificações seguintes seria adequada.

	Primeira opção	Segunda opção
Bom	1	1
Regular	2	8
Inadequado	3	27
Inexistente	4	64

A média leva em conta os valores numéricos, assim os resultados seriam diferentes segundo se utilize a primeira opção de codificação ou a segunda.

A fórmula da média por indivíduo das características observadas seria:

$$\text{média_simples} = \frac{1}{37} \sum_{j=1}^{37} k_j$$

sendo k_j cada um dos códigos das 37 características de recursos materiais.

A segunda refere-se aos pesos atribuídos a cada um dos itens: com o cálculo da média outorga-se a mesma importância a cada um dos recursos, podendo acontecer que uma característica que não tenha variabilidade entre as escolas, pese tanto quanto outra característica que diferencia as escolas.

Outra forma – também convencional- de sintetizar informações seria o cálculo dessa média por indivíduo, ponderada pela importância de cada tema. No caso apresentado o segundo indicador alternativo foi obtido da forma seguinte:

- Calcular uma média simples para cada tema, obtendo uma média para cada um dos quatro grupos de variáveis.
- Calcular a média simples das quatro médias obtidas no passo anterior.

A fórmula da média por indivíduo ponderada pela importância dos temas:

$$\text{média_por_temas} = \frac{1}{4} \sum_{i=1}^4 \left(\frac{1}{n_i} \sum_{j=1}^{n_i} k_j \right)$$

sendo k_j cada um dos códigos das 37 características de recursos materiais, n_i o número de características de cada um dos quatro temas.

Corresponde a este cálculo a mesma observação que à primeira proposta alternativa (média simples): resulta inadequado por tratar-se de uma combinação linear de códigos sem propriedades numéricas. Em relação aos ponderadores; a diferença desta proposta é que no interior de cada tema as características tem a mesma importância.

As duas formas de cálculo alternativas trazem como resultado que esses indicadores não cumpram com o objetivo de *pertinência em relação à finalidade* para a qual foi concebido. Procura-se que o indicador sirva como um elemento de diagnóstico refletindo a situação de recursos observadas nas escolas. O fato que os ponderadores dê importância para cada uma das características dos recursos a priori, faz com que esses pesos independam da observação do fenômeno.

5.2. Comparação dos resultados analisando os elementos de discordância

Uma vez calculados os indicadores alternativos corresponde investigar como eles representam as situações extremas e as situações intermédias; e também observar como cada indicador representa as escolas de diferentes redes, regiões, etc. É com o uso das técnicas estatísticas simples que se consegue comparar o comportamento dos indicadores alternativos.

Como os procedimentos são similares, apresenta-se a título de exemplo o cruzamento dos indicadores com uma característica usada para a sua construção: condição de funcionamento da quadra de esportes. Esse cruzamento efetua-se de duas formas diferentes. Em primeiro lugar observa-se o comportamento de cada

um dos indicadores no conjunto de todas as escolas. Em segundo lugar compare-se a capacidade discriminante dos três indicadores sobre sub-grupos de escolas definidos pelas diferentes condições de funcionamento da quadra de esportes.

Figura 5.1. *Boxplot* do indicador alternativo (média simples) segundo as condições de funcionamento das quadras de esportes.

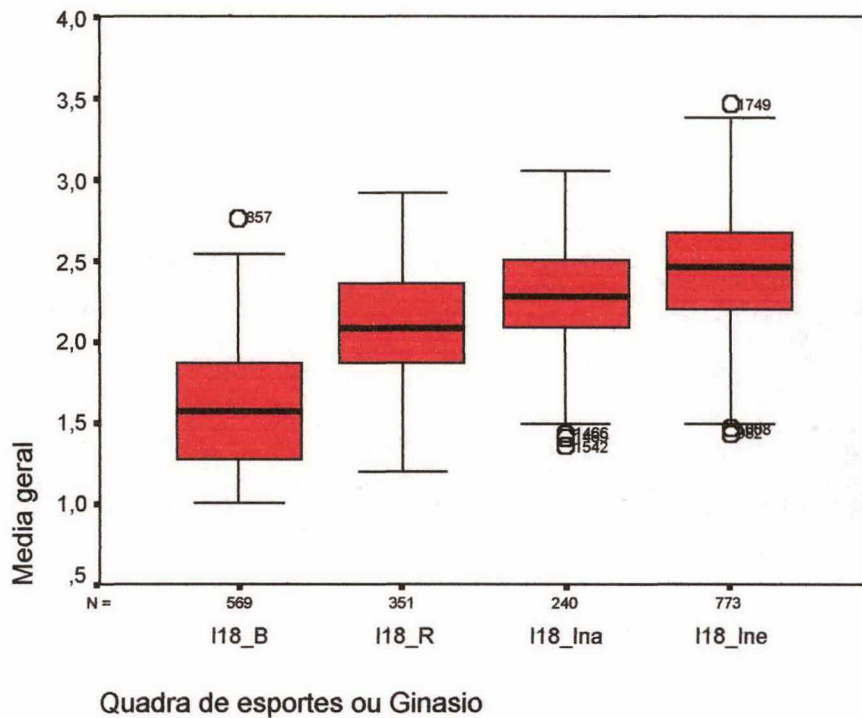
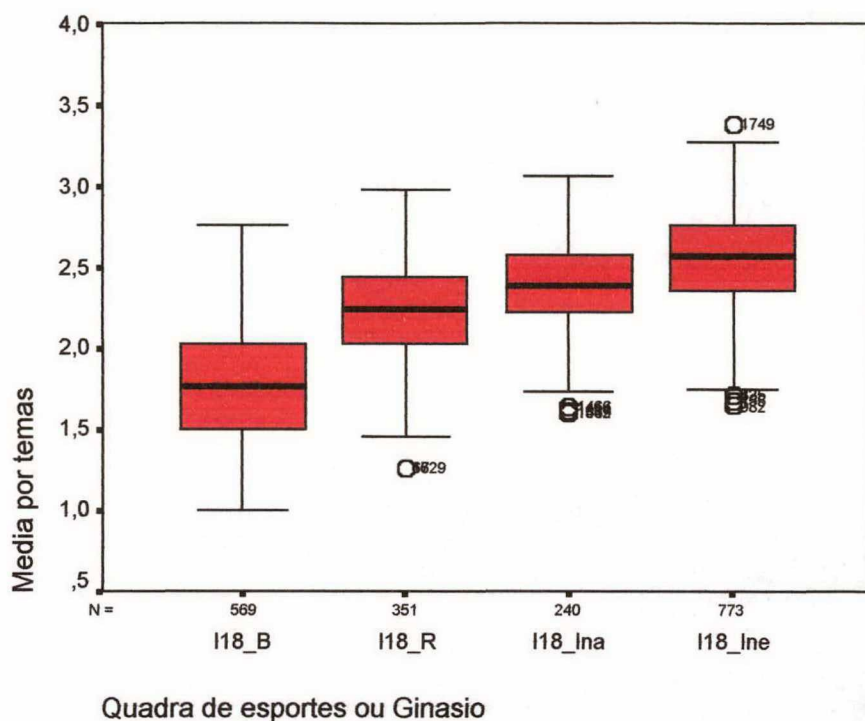


Figura 5.2. *Boxplot* do indicador alternativo (média ponderada) segundo as condições de funcionamento das quadras de esportes.



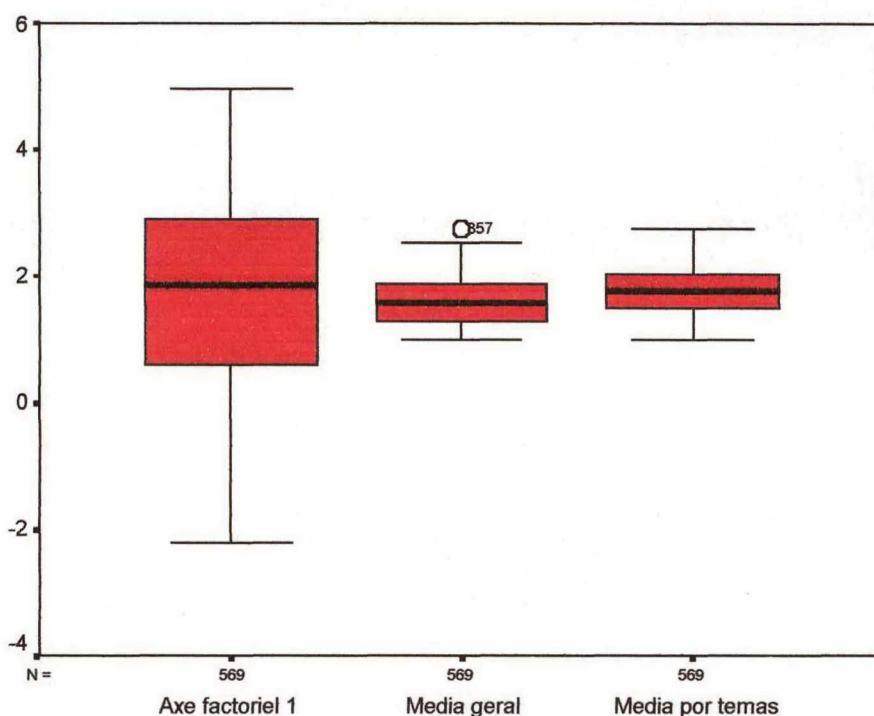
Para realizar a comparação entre os valores atribuídos pelos diferentes indicadores às escolas que tem diferentes condições de funcionamento de suas quadras de esportes, temos a esclarecer que:

- Os valores do indicador proposto (primeiro eixo da AFM) que correspondem a situações boas de recursos são os valores altos, enquanto que nos indicadores alternativos acontece o inverso. Dada que a codificação adotada é de 1 até 4 representando as modalidades Bom até Inexistente; os valores baixos do indicador correspondem as melhores situações.
- A comparação através dos *boxplot* deve considerar os comentários realizados na Figura 4.23 que corresponde ao *boxplot* do indicador proposto assim como a amplitude dos três indicadores.

Da comparação dos três indicadores desprende-se que os três diferenciam as escolas com diferentes condições das quadras de esportes embora de formas diferentes.

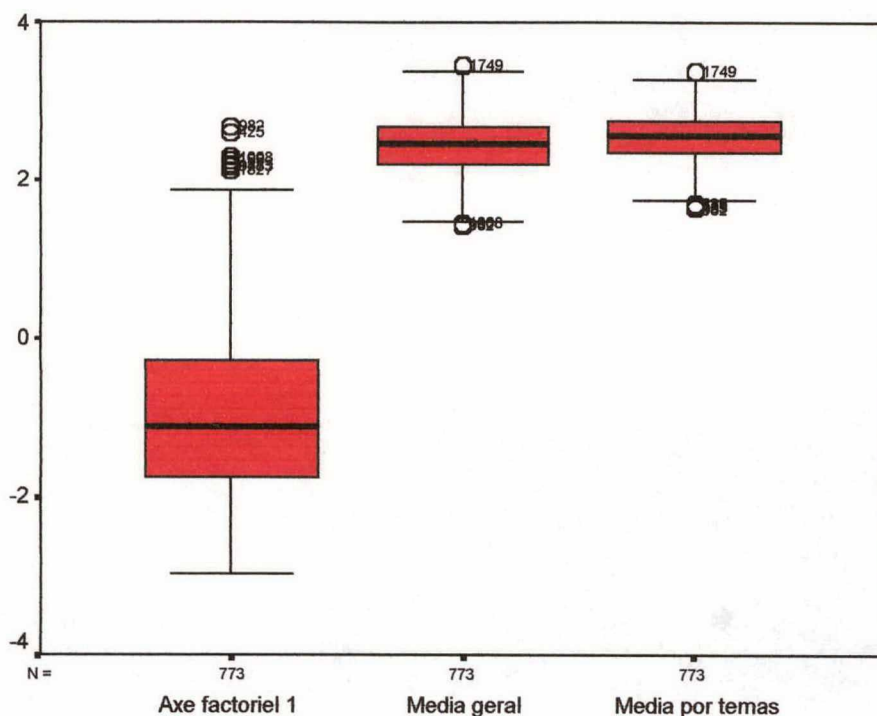
Nos indicadores alternativos, o indicador ordena de forma quase linear as escolas com diferentes situações das quadras: quanto melhores as quadras maiores são os valores atribuídos ao indicador, portanto a situação geral das escolas também são melhores. Enquanto o indicador proposto diferencia claramente aquelas que tem boas quadras das restantes. Isto é, as escolas com boas quadras apresentam uma situação geral de recursos bem melhor que as outras escolas com quadras regulares ou mesmo inexistentes.

Figura 5.3. *Boxplot* dos três indicadores para as escolas com boas quadras de esportes.



Outra diferença é que a variabilidade na situação geral representada pelos indicadores (mesmo tendo em conta a diversidade das amplitudes) no caso das escolas que tem quadras boas; é maior para o indicador proposto que para os outros.

Figura 5.4. *Boxplot* dos três indicadores para as escolas sem quadras de esportes.



Por outra parte, segundo o indicador proposto, existem muitas escolas que não tendo quadras de esportes apresentam uma situação geral de recursos boa. Com os indicadores alternativos a situação geral de recursos é mais homogênea nas escolas sem quadras de esportes.

As diferenças entre o comportamento dos três indicadores não são notórias porque o espaço conjunto da AFM (cujo primeiro eixo é o indicador proposto) apresenta uma forte correlação com o primeiro eixo dos espaços de cada grupo de variáveis. Assim, todos esses espaços estão orientados segundo o primeiro eixo que sempre é um eixo de escala. Portanto o indicador proposto não pode-se diferenciar demasiado de qualquer outra combinação linear de atributos das escolas. Mas através deste exemplo detectamos grupos de escolas nos quais os indicadores divergem, e essa divergência ocorre porque os indicadores alternativos expressam somente as relações lineares entre as variáveis, enquanto que o indicador obtido pela AFM consegue captar também as relações não

lineares. O indicador proposto distingue melhor as configurações particulares de recursos materiais adequando-se aos objetivos da análise.

Poder-se-ia realizar uma análise semelhante para cada uma das 37 características de recursos e para cada uma das variáveis de estratificação. Mas, mais importante que o comportamento dos indicadores é a verificação que eles conseguem atender ou não as propriedades requisitadas. O fato da possibilidade de associar aos diferentes valores do indicador proposto, uma tipologia das escolas segundo as condições de seus diversos recursos materiais; lhe confere uma capacidade descritiva que os outros não tem. Os indicadores alternativos não conseguem realizar a correspondência entre os valores do indicador e as classes de escolas que se assemelham por seus recursos materiais. Assim, também não podem se relacionar com diferentes formas de intervenção, dada sua incapacidade de detectar as características deficitárias nos diferentes tipos de escolas.

CAPÍTULO VI

6. CONCLUSÕES E RECOMENDAÇÕES

6.1. Conclusões da pesquisa	161
6.2. Recomendações	162
6.2.1. Recomendações metodológicas	162
6.2.2. Recomendações sobre a aplicabilidade do indicador	163

6.1. Conclusões da pesquisa

O presente estudo tem um caráter eminentemente metodológico. Tais estudos tem por pressuposto elucidar um problema de pesquisa que exige do pesquisador o desenvolvimento de procedimentos sistemáticos com um enfoque analítico e comparativo entre o fenômeno observado e a rigurosidade do método científico.

Ao defrontar-se com o seguinte problema de pesquisa:

" Como resumir todas as informações de variáveis não métricas numa variável continua e única que - dando conta de todas as características iniciais- as ponderem objetivamente?" pressupôs-se que o desencadeamento de uma investigação que permitisse responder este questionamento obrigatoriamente teria dois caminhos a seguir, que complementarmente, deram origem aos dois objetivos:

- "Desenvolver uma metodologia de elaboração de indicadores para a avaliação da rede escolar brasileira que assegure-lhe um conjunto de propriedades que permita sua utilização como uma medida de avaliação do ensino básico brasileiro.
- Construir e validar um indicador de recursos físicos das escolas a partir dos dados do SAEB97."

Para desenvolver uma metodologia seria necessário o desenvolvimento de um indicador que a exemplificasse; para desenvolver o indicador necessário se faz a construção de uma metodologia. Desta relação que surge a simbiose entre teoria e prática. Diante ao conjunto de dados observados e disponibilizados pelo SAEB e

as possibilidades teóricas dos métodos estatísticos multivariados permitiram delimitar um conjunto de propriedades ao indicador. Tanto a conceituação destas propriedades quanto a construção do indicador, passa então a ter por pressuposto a análise criteriosa dos dados observados, que tratados com as técnicas apropriadas revelaram estruturas que permitiram a sua construção, a partir de uma relação teórica-empírica. Isto induziu à não postulação de modelos matemáticos a "*priori*".

A metodologia apresentada privilegia a forma de tratar a informação propondo uma estratégia de análise dos dados que permite explorar as relações entre aspectos factuais da realidade e construir os conceitos como uma dedução do observado. Além do mais, essas relações são exploradas a partir da comparação do conjunto das unidades de observação sem se limitar ao estudo das relações entre as variáveis.

Conclui-se argumentando que diante do exposto nos Capítulos IV e no V o indicador de recursos materiais exprime as condições das escolas brasileiras e apresenta as propriedades requisitadas. Tal fato demonstra a pertinência da proposta metodológica o que, indubitavelmente, atende aos objetivos do presente estudo e responde a pergunta de pesquisa. Isto é, esta pesquisa apresenta uma metodologia exeqüível e validada ao possibilitar a construção de um indicador que resume as informações não métricas numa variável contínua e única e que as pondera objetivamente.

6.2. Recomendações

As recomendações atentam especialmente para o desenvolvimento de novos estudos de ordem metodológica ou de aplicabilidade como suporte para a formulação de políticas públicas.

6.2.1. Recomendações metodológicas

Como o indicador obtido resulta do primeiro eixo de uma AFM, recomenda-se o estudo da estabilidade desse eixo por meio de um procedimento de re-amostragem baseado em técnicas de simulação de amostras partindo de uma só

amostra. Esse estudo de estabilidade basicamente consiste em repetir as análises sobre as diferentes amostras simuladas e estudar as flutuações dos resultados obtidos. Vários métodos de validação permitem gerar de forma diferentes as amostras artificiais. A técnica de Bootstrap se aplica a numerosos problemas onde não é possível estimar analiticamente a variabilidade de um parâmetro. A técnica consiste em simular m (em geral superior a 30) amostras do mesmo tamanho que a amostra inicial. Elas são obtidas por seleção aleatória com reposição dos n indivíduos observados, cada um deles com a mesma probabilidade $1/n$ de ser selecionado. Alguns indivíduos terão assim um peso elevado e outros peso nulo. Sobre cada amostra simulada efetua-se a análise fatorial, e depois se estabelece a distribuição de freqüências para os eixos. As zonas de confiança obtidas são úteis para provar a qualidade dos resultados e portanto a possibilidade de generalizar o indicador calculado sobre a amostra de escolas à população de referência (Lebart-Morineau-Piron, 1997, p. 389).

Recomenda-se também a realização de estudos sobre a replicabilidade da construção do indicador de recursos materiais, junto a outras amostras do SAEB observadas em outros períodos. Cita-se como exemplo o fato de que os dispositivos de informações apresentados para 1999 incluídos no questionários das escolas relativo a suas características diferem do utilizado em 1997 o que pode alterar substancialmente a validade longitudinal do indicador calculado. No entanto a estratégia metodológica proposta pode permanecer validada.

Finalmente, recomenda-se que se realize estudos similares com outras dimensões das bases de dados do SAEB procurando construir indicadores sobre os aspectos sócio-econômicos dos alunos, formação dos docentes e recursos humanos disponibilizados a rede escolar.

6.2.2. Recomendações sobre a aplicabilidade do indicador

Recomenda-se a realização de estudos utilizando o indicador como um modelo de predição de impactos das políticas públicas setoriais ou globais. Isto é, planejar uma dada intervenção – aquisição de equipamentos para as escolas da região nordeste – e, com base na estrutura dos dados observados – predizer qual

o impacto dessa intervenção na melhoria do posicionamento dessas escolas diante a rede escolar brasileira.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABBRUZZINI, S. y QUINTI, G. (1997) **"Estrategias metodológicas y métodos estadísticos para la construcción de indicadores e índices"**, Notas del curso del seminario temático PRESTA realizado em Concepción, Chile, del 3 a 7 de marzo de 1997. Lab. de Méthodologie du Traitement des Données, ULB, Bruxelles.
- CANNAVÒ, L. (1984) **"Definizioni operative, concettualizzazione e logica della ricerca"**, in Sociología e ricerca sociale, V, 13, pp.99-140.
- CANNAVÒ, L. (1995) **"Il primato della pragmatica. Il senso degli indicatori nella ricerca sociale"**, in Sociología e ricerca sociale, XVI, 47-48, pp.7-25.
- CRIVISQUI, E. **"Les processus d'observation systématique des phénomènes sociaux – Place et fonction des méthodes statistiques"**. Thèse, Institut de Statistique et de Recherche Opérationnelle, Université Libre de Bruxelles, février 1994, 261p. Bruxelles, Belgique.
- CRIVISQUI, E. **Análisis en Componentes Principales**. Notas de curso de los "Seminarios Locales de Métodos Exploratorios de Estadística Multivariada", site do Programa PRESTA (www.ulb.ac.be/assoc/presta), 1999.
- CRIVISQUI, E. **Análisis Factorial de Correspondencias Simples y Múltiples**. Notas de curso de los "Seminarios Locales de Métodos Exploratorios de Estadística Multivariada", site do Programa PRESTA (www.ulb.ac.be/assoc/presta), 1999.
- CRIVISQUI, E. **Análisis Factorial de Correspondencias**. Edición del Laboratorio de Informática Social, Universidad Católica de Asunción, 1993.
- CRIVISQUI, E. **Métodos de clasificación**. Notas de curso de los "Seminarios Locales de Métodos Exploratorios de Estadística Multivariada", site do Programa PRESTA (www.ulb.ac.be/assoc/presta), 1999.
- CRIVISQUI, E. y VILLAMONTE, G. **"Introducción al análisis exploratorio de tablas múltiples: la estrategia exploratoria del AFM"**, notas de curso das Primeiras Jornadas Latino-americanas de Estatística Aplicada, UFScarm, 16 a 19 de novembro de 1999.

- DUPIORIER, E. et PARODI, J-L. (1997) "Les indicateurs socio-politiques aujourd'hui", in Actes du Colloque de l'Association Française de Science Politique et de l'Observatoire Interrégional du Politique, janvier 1996. Ed; L'Harmattan, Paris.
- EFRON, B. **The Jackknife, the Bootstrap et other Resampling Plans**. SIAM, Philadelphia, 1982.
- EFRON, B., TIBSHIRANI, R.J. **An introduction to the Bootstrap**. Chapman and Hall, New York, 1993.
- ESCOFIER, B., PAGÉS, J. **Análisis Factoriales Simples y Múltiples**. Objetivos, métodos e interpretación., Servicio Editorial de la Universidad del País Vasco, Bilbao, 1992.
- GREENACRE, M. **Theory and Applications of Correspondance Analysis**. Academic press, London, 1984.
- GRIFFITH, J. **The relation of organizational process orientation to effectiveness and efficiency in elementary public schools in Social Psychology of Education**, Netherlands, 1998, .297-321
- LEBART, L., MORINEAU, A., FÉNELON, J-P. **Tratamiento Estadístico de Datos**. Marcombo, Barcelona, 1995.
- LEBART, L., MORINEAU, A., PIRON, M. **Statistique exploratoire multidimensionnelle**. Dunod, Paris, 1997, 2^a edition.
- MAMBRINI, J., CESAR, C.C., SOARES, J.F. **Fatores Determinantes do Desempenho dos Alunos Mineiros no SAEB de 1995**. Instituto de Ciências Exatas, Departamento de Estatística, UFMG. Jfsoares@est.ufmg.br, agosto de 1999.
- MARRADI, A. (1990), "Fedeltà di un dato, affidabilità di una definizione operativa", in *Rassegna Italiana di Sociologia*, XXXI, 1, pp.55-96.
- McCARTY, T.A , YAIASAWARNG, S. **Technical Efficiency in New Jersey School Districts**. in *The Measurement of Productive Efficiency. Techniques and Applications*. 1993, New York; Oxford, 271-287.
- MEC/INEP. **Análise do Resultado da Avaliação SAEB97**. Documentação do software ANALISASAEB

MEC/INEP. Documentos disponibilizados no site do SAEB do INEP

(www.inep.gov.br/saeb/):

- Histórico
- Descrição do SAEB
- SAEB95
- SAEB97.

MEC/INEP. **Questionários utilizados no SAEB97.** Documentação do software
ANALISASAEB

MEC/INEP. **Sistema Nacional de Avaliação da Educação Básica.** Visão geral
do SAEB. . Documentação do software ANALISASAEB.

MEMOLI, R. Y SAPORITI, A. (1995) **Disegno de la ricerca e analisi dei dati**, La
Goliardica Ed. Univ. Di Roma, Roma.

PAWSON, A. (1989), **A measure for measures**, Routledge, London.

SARTORI, G. (1984), **Social Science Concepts. A Systematic Analysis**, Sage
Publications Inc., Beverly Hills-London.

ANEXO 1

QUESTIONÁRIO DE ESCOLA

Indique o estado de conservação dos seguintes itens referentes ao prédio da escola.

Itens	Bom	Regular	Inadequado	Inexistente
1. Telhado.....	(A)	(B)	(C)	(D)
2. Paredes.....	(A)	(B)	(C)	(D)
3. Piso.....	(A)	(B)	(C)	(D)
4. Portas e janelas.....	(A)	(B)	(C)	(D)
5. Banheiros.....	(A)	(B)	(C)	(D)
6. Cozinha.....	(A)	(B)	(C)	(D)
7. Instalações hidráulicas.....	(A)	(B)	(C)	(D)
8. Instalações elétricas.....	(A)	(B)	(C)	(D)

Bom = adequado. **Regular** = necessita de pequena reforma. **Inadequado** = necessita de grande reforma.

Analise as instalações quanto aos seguintes aspectos:

Aspectos	Bom	Ruim
9. Ventilação.....	(A)	(B)
10. Iluminação.....	(A)	(B)
11. Ruídos.....	(A)	(B)
12. Aspectos da área externa (pátio, jardins).....	(A)	(B)
13. Espaço para recreação infantil.....	(A)	(B)

Bom = adequado. **Ruim** = Inadequado.

Indique as condições de funcionamento das seguintes instalações:

Instalações	Bom	Regular	Inadequado	Inexistente
14. Biblioteca.....	(A)	(B)	(C)	(D)
15. Laboratório de Ciências.....	(A)	(B)	(C)	(D)
16. Laboratório de Informática.....	(A)	(B)	(C)	(D)
17. Auditório.....	(A)	(B)	(C)	(D)
18. Quadra de esportes ou Ginásio.....	(A)	(B)	(C)	(D)
19. Vestiários.....	(A)	(B)	(C)	(D)
20. Sala de professores.....	(A)	(B)	(C)	(D)
21. Sala de aula.....	(A)	(B)	(C)	(D)

Bom = adequado. **Regular** = necessita de pequena reforma. **Inadequado** = necessita de grande reforma.

22. Caso a escola possua biblioteca, há livros de boa qualidade para utilização didática e pedagógica?	Sim (A)	Não (B)
--	------------	------------

Indique o estado de conservação dos seguintes equipamentos da escola.

Instalações	Bom	Regular	Ruim	Não existe
23. Quadro negro.....	(A)	(B)	(C)	(D)
24. Carteiras.....	(A)	(B)	(C)	(D)
25. Mesas.....	(A)	(B)	(C)	(D)
26. Televisão.....	(A)	(B)	(C)	(D)
27. Vídeo-cassete.....	(A)	(B)	(C)	(D)
28. Mimeógrafo.....	(A)	(B)	(C)	(D)
29. Máquina fotocopadora.....	(A)	(B)	(C)	(D)
30. Projetor de slides.....	(A)	(B)	(C)	(D)
31. Retroprojektor.....	(A)	(B)	(C)	(D)
32. Máquinas de datilografia.....	(A)	(B)	(C)	(D)
33. Computador.....	(A)	(B)	(C)	(D)

Regular = necessita de manutenção. **Ruim** = Irrecuperável.

Indique se a escola oferece os seguintes serviços:

Serviços	Sim, pago	Sim, gratuito	Não oferece
34. Médicos	(A)	(B)	(C)
35. Odontológicos	(A)	(B)	(C)
36. Transporte	(A)	(B)	(C)
37. Alimentação	(A)	(B)	(C)