Estudo de n Variáveis do Sistema Educacional: um Estudo da Análise Multivariada de Dados para a Validação das Estruturas Físicas e Pedagógicas das Escolas Públicas do Paraná Pertencentes a Região Sul

Anselmo Chaves Neto [1]
Antônio Carlos Machnicki [2]
Liliana Madalena Gramani [3]

RESUMO

O sistema educacional brasileiro abrange as esferas federal, estadual e municipal e é composto de instituições públicas e privadas. O trabalho investiga quais as variáveis estruturais, pedagógicas e auxiliares (sociais) que melhor se correlacionam com o ensino e como cada estrutura escolar responde aos interesses da aprendizagem. Avaliou-se doze variáveis estruturais, doze variáveis pedagógicas e trinta e quatro variáveis auxiliares. As pesquisas foram feitas, com amostra estratificada, em vinte e duas escolas públicas da região sul do Paraná. As ferramentas de análise são a estatística multivariada de dados com análise fatorial e análise de regressão. Os resultados apontam os conjuntos de variáveis que melhor se correlacionam, citadas nos fatores, e indicam quais instituições de ensino apresentam os melhores modelos de estrutura física, pedagógica e auxiliar (social) que serve de parâmetro tanto para a alocação de recursos financeiros como para o gerenciamento escolar.
Palavras-chave. Educacional. Escolar. Análise multivariada.

ABSTRACT

The brazilian educational system includes the federal, state and municipal levels and is composed by public and private institutions. This work investigates the structural, pedagogical and auxiliary (social) variables that best correlate with education and how each school structure responds to the interests of learning. We evaluated twelve structural variables, twelve pedagogical variables and thirty-four auxiliary variables. The research was conducted with a stratified sample on twenty-two publics schools in southern Parana. The tools of analysis are multivariate statistical data with factor analysis and regression analysis. The results indicate the sets of variables that best correlate, listed under factor, and indicate which educational institutions are the best models of physical, pedagogical and auxiliary (social) infrastructure, which serves as a parameter for both financial resource allocation and for the school management.
Keywords. Educational. School. Multivariate analysis.

1 Universidade Federal do Paraná (UFPR), Curitiba, PR. Professor titular. Dr. e-mail anselmo@ufpr.br
2 Universidade Federal do Paraná (UFPR), Curitiba, PR. Estudante de mestrado. e-mail machnicki@uol.com.br
3 Universidade Federal do Paraná (UFPR), Curitiba, PR. Professora titular. Drª. e-mail l.gramani@gmail.com.br

1 Introdução

No mundo globalizado o acesso rápido a informações e o refino de tais informações são de fundamental importância. É comum, rotineiro e necessário os meios de comunicações informarem a previsão do tempo e o fluxo de carros nas rodovias. Jornais estão recheados de indicadores econômicos e de previsões de toda natureza. Empresas investem em publicidade como forma de divulgar seus produtos e qualificá-las no mercado. No campo social ou político as informações ganham fundamental importância a medida que a constituição brasileira garante, a todo cidadão, que os serviços públicos prestados sejam de forma eficiente e com qualidade. A educação é foco de estudos e pesquisas pela comparação entre instituições de diferentes estados, municípios e regiões, medindo sua estrutura de funcionamento, de pessoal e de produção intelectual. Qualificar os profissionais das áreas exatas é um desafio que exige estudo, empenho e conhecimento das ferramentas matemáticas e de programas computacionais que estão a nossa disposição.

O objetivo do trabalho é encontrar uma forma de avaliar e verificar de que maneira o estudante é influenciado pela tecnologia, pelos recursos alocados as escolas, pelas condições de trabalho e formação do professor e pelas condições físicas do estabelecimento que ele frequenta. Tende a apontar, segundo pesquisa direcionada a colégios de ensino, quais são as variáveis mais importantes no rendimento estudantil e que poderão influenciar positivamente na melhoria dos índices regionais e brasileiros em educação.

A partir da questão inicial outro objetivo que surge é o de como direcionar os recursos oriundos dos fundos educacionais de forma a minimizar as desigualdades entre as instituições de ensino pertencentes a um mesmo núcleo educacional ou a uma mesma região de forma a equiparar o ensino diminuindo suas diferenças.

2 Metodologia

O trabalho investigou “n” variáveis relacionadas ao ensino fundamental e médio em escolas públicas pertencentes ao núcleo regional de educação de União da Vitória (PR), identificando as variáveis que melhor se correlacionam, com o desempenho dos alunos e com a eficiência das escolas, fornecendo subsídios para avanços na estrutura física, de equipamentos e pedagógica, tendo reflexo na qualidade educacional oferecida, pelas instituições de ensino, e na formação do cidadão.

Segundo Sampaio e Guimarães (2009), a importância da acumulação de capital humano é destacado como aspecto crucial para o desenvolvimento econômico de um país, influencia na redução do crescimento populacional, na queda da mortalidade infantil e no aumento da expectativa de vida.

Foram coletadas informações em vinte e duas instituições de ensino de oito municípios, num total de mil, oitocentos e vinte e seis alunos pesquisados ou 15,77% do total. As pesquisas foram feitas através de questionários produzidos para esse fim e envolvem variáveis estruturais (relacionadas a estrutura física), pedagógicas (relacionadas as questões didáticas e administrativas) e auxiliares (relacionadas as questões sociais e econômicas).

As técnicas de análises são a estatística multivariada de dados, mais especificamente, utilizou-se a análise fatorial que, segundo Hair Jr et al (2005), seus objetivos são os de fornecer um processo para a redução de dados identificando as estruturas comuns que podem ser usadas no acúmulo de variáveis. Junto com a análise fatorial foi feita a análise de componentes principais que, para Johnson e Wichern (1988), tem como fim tanto a redução dos dados como a interpretação que pode ser feita pela visualização do percentual explicativo da variabilidade dos dados de cada uma das componentes. Outra ferramenta multivariada estudada é a análise de regressão múltipla que, para Moita Neto (2004), a principal aplicação da regressão múltipla, após encontrar a relação matemática, é produzir valores para a variável dependente quando se tem as variáveis independentes, isto é, ela pode ser usada na predição dos resultados. Para Hair Jr. et al (2005) a análise de regressão múltipla tem como objetivo o de trabalhar com as variáveis independentes cujos valores são conhecidos para prever os valores de variáveis dependentes selecionadas pelo pesquisador.

Como meio de cálculo dos resultados utiliza-se os programas matlab 6.5, minitab 15 e excel. Tais softwares são aplicados nos conjuntos de variáveis estruturais, pedagógicas e auxiliares.

2.1 Transformando dados dicotômicos

Variáveis dicotômicas são aquelas oriundas de bancos de dados com valores não-métricos. As técnicas estatísticas multivariadas dependem de variáveis cujos valores são métricos, para tanto, devemos encontrar uma forma adequada de transformação de dados quando esses são não-métricos. O meio utilizado para a transformação dos dados não-métricos, constante nos questionários, para dados métricos, foi pela atribuição de pesos para cada um dos conceitos, de ótimo a não tem, por exemplo, em uma escala de 10 a 3. Com isso produzimos uma matriz vetor que padroniza cada variável. Exemplificando: se considerar os conceitos ótimo, bom, regular, ruim, insatisfatório, não tem, posso fazer a padronização pela matriz

$$\text{var}\,iável = \begin{bmatrix}10 & 9 & 7 & 5 & 4 & 3\end{bmatrix} * \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix} : \left(w_1 + w_2 + w_3 + w_4 + w_5 + w_6\right), \text{ onde:}$$

w_1 ,w_2 ,w_3 ,w_4 ,w_5 e w_6 são os valores totais das respostas de cada colégio aos conceitos ótimo, bom, regular, ruim, insatisfatório, não tem, respectivamente.

Os dados da variável estatística E_1 ficam padronizados conforme equação:

$$E_1 = \frac{10w_1 + 9w_2 + 7w_3 + 5w_4 + 4w_5 + 3w_6}{w_1 + w_2 + w_3 + w_4 + w_5 + w_6}$$

Genericamente, a variável P_n é padronizada como:

$$P_n = \frac{10w_1 + 9w_2 + 7w_3 + 5w_4 + 4w_5 + 3w_6}{w_1 + w_2 + w_3 + w_4 + w_5 + w_6}$$

Os dados padronizados, numericamente, constituem a matriz:

variáveis

escolas	E_1	E_2	.	.	E_n	.	P_1	P_2	.	.	P_n	.	NA
1	E_{11}	E_{12}	.	.	E_{1n}	.	P_{11}	P_{12}	.	.	P_{1n}	.	NA_{11}
2	E_{21}	E_{22}	.	.	E_{2n}	.	P_{21}	P_{22}	.	.	P_{2n}	.	NA_{21}
.	.	.			.		.	.			.		.
.	.	.			.		.	.			.		.
.	.	.			.		.	.			.		.
j	E_{j1}	E_{j2}	.	.	E_{jn}	.	P_{j1}	P_{j2}	.	.	P_{jn}	.	NA_{j1}
.	.	.			.		.	.			.		.
.	.	.			.		.	.			.		.
.	.	..			.		.	..			.		.
22	E_{221}	E_{222}	.	.	E_{22n}	.	P_{221}	P_{222}	.	.	P_{22n}	.	NA_{221}

onde:

$E_1; E_2; E_n$ são as variáveis estruturais.

$P_1; P_2; P_n$ são as variáveis pedagógicas.

NA é a variável nível de aprendizagem.

j é a e-ésima escola.

E_{jn}, P_{jn}, NA_{j1} são valores numéricos obtidos após aplicação do cálculo com o respectivo peso.

3 Resultados

Os resultados estão organizados da seguinte forma:

a) Resultados gerados a partir da matriz de variáveis estruturais;

b) Resultados gerados a partir da matriz de variáveis pedagógicas;

c) Resultados gerados a partir da matriz de variáveis auxiliares.

Consideramos, neste artigo, os resultados gerados da análise fatorial e da regressão multivariada.

A análise fatorial tem como fim reduzir o número de variáveis e de indicar os conjuntos de variáveis, distribuídos nos fatores, que são mais homogêneos.

Com a análise de regressão qualifica-se o melhor modelo de estrutura física, pedagógica ou de caráter auxiliar. Para a qualificação da melhor estrutura utiliza-se tabelas que indicam:

a) O nível de aprendizagem identificado por cada escola e coletado na pesquisa;

b) A previsão do nível de aprendizagem pela equação de regressão multivariada que é tomada considerando-se o melhor modelo selecionado a partir do programa minitab 15 função **stat>regression>Best subsets**, com resposta para a variável dependente "NA" nível de aprendizagem;

c) O erro entre o nível de aprendizagem, item a, e a previsão, item b;

d) O erro quadrático.

3.1 Resultados para as variáveis estruturais

A análise fatorial revelou uma estrutura com cinco fatores e com poder explicativo de 79,50% da variabilidade total do modelo. Apresenta bom número de correlações, entre as variáveis, maiores do que 0,30 e matriz de resíduos que indicam aleatoriedade. As comunalidades entre as variáveis e os fatores são boas, sendo a menor 0,70 e a maior 0,88 estando dentro do limite inferior aceitável para que possamos considerar os resultados oriundos da análise fatorial. O teste de adequacidade da amostra revela um MSA de 0,5696 com Q^2 de 94,4802.

Matriz de fatores com pesos estimados e comunalidades

Var.	*F1*	*F2*	*F3*	*F4*	*F5*	*Com.*	*Var.esp.*
1	0,1873	-0,0601	**0,6718**	-0,4501	0,2904	0,78	0,22
2	0,2089	-0,2651	-0,0733	**-0,7985**	0.0555	0,76	0,24
3	-0,4424	-0,1892	0,4090	0,1134	**-0,6177**	0,79	0,21
4	0,2282	0,1839	**0,7998**	0,1737	-0,1579	0,78	0,22
5	0,5026	0,2900	0,0735	**-0,7199**	-0,0444	0,86	0,14
6	-0,1226	0,0215	0,0136	**-0,8117**	-0,1485	0,70	0,30
7	-0,1748	**-0,8672**	-0,0662	0,1402	0,0725	0,81	0,19
8	0,0377	-0,0282	-0,0604	-0,0855	**-0,9330**	0,88	0,12
9	**0,8124**	-0,2115	0,3393	-0,0727	0,0787	0,83	0,17
10	0,2033	**-0,7481**	-0,0260	-0,3218	-0,3511	0,83	0,17
11	**0,8403**	-0,0505	0,0716	-0,1131	-0,0044	0,73	0,27
12	**0,8214**	0,3081	0,0365	-0,1325	0,0172	0,79	0,21

Autovalor	*3,53*	*2,06*	*1,53*	*1,37*	*1,07*
Prop. acumulada	22,57	36,59	48,25	66,81	79,50

Fonte: matlab 6.5 (função fator).
Tabela 1- pesos estimados- variáveis estruturais

A partir de cada peso estimado com rotação da matriz, rotação varimax, obtém-se os fatores e as respectivas variáveis em cada fator.

Fator	*Variáveis*	*Descrição das variáveis*
F1	E9, E11, E12	Sala de aula, quadro, estado das carteiras
F2	E7, E10	Quadra de esportes, áudio e vídeo
F3	E1, E4	Biblioteca, lab. de informática
F4	E2, E5, E6	Serv.gerais, banheiro, pátio
F5	E3, E8	Lab. de ciências, refeitório

Fonte: matlab 6.5
Tabela 2 – número de fatores com descrição das variáveis estruturais

A análise de regressão multivariada, para as variáveis estruturais, tem como fim, identificar, pelo menor erro quadrático, aquela estrutura que melhor serve como modelo em resposta ao anseio das escolas com relação ao nível de aprendizagem desejável de seus alunos. Para esse fim utilizou-se a equação de regressão multivariada, constante do quadro abaixo, comparando o número pesquisado com a previsão da equação de regressão múltipla, calculando o erro de previsão e o erro de previsão quadrado.

Variável estatística de regressão
$NA = b_0 + b_1E1 + b_2E2 + b_3E3 + b_4E4 + b_5E5 + b_6E6 + b_7E8 + b_8E9 + b_9E10 + b_{10}E11 + b_{11}E12$
Equação de previsão NA=13,2 - 0,388E1 – 0,740E2 +0,0420E3 – 0,0611E4 + 0,528E5 – 0,182E6 – 0,181E8 + 0,189E9 + 0,274E10 – 0,092E11 + 0,084E12

Tabela 3- equação de regressão- variáveis estruturais.

O modelo de regressão é construído a partir de onze das doze variáveis estruturais, tem um poder de explicação da variabilidade de 59,6% e apresenta uma estimativa média sobre a linha de regressão de 0,468874.

3.2 Resultados para as variáveis pedagógicas

Consideramos como variáveis pedagógicas as variáveis ligadas diretamente aos processos de ensino. Tais variáveis são influenciadas pelas pessoas que ocupam determinado cargo ou função desde, professor, professor pedagogo, diretor, secretario, servidor administrativo, servidor de serviços gerais e estudante. Diferentemente das variáveis estruturais cuja melhoria, muitas vezes, requer recursos financeiros, as variáveis pedagógicas precisam de trabalho em equipe, do conjunto escola família, professor aluno, diretor administração, estando sujeitas as emoções e aos jogos de interesse.

A análise fatorial identificou um conjunto com cinco fatores. A distribuição das variáveis dentro de cada fator obedeceu o critério do maior carregamento após a rotação da matriz pela rotação varimax. Quanto aos métodos de adequação da amostra, a mesma apresenta um Q^2 de 187,6132 e MSA de 0,5420, possui um médio índice de correlações entre as variáveis maiores do

que 0,30, sendo 43,94% do total de correlações e a matriz de resíduos sugere que eles foram gerados de maneira aleatória, sem tendência de formação. As comunalidades, entre as variáveis e os fatores, estão acima da comunalidade mínima aceitável, sendo, a menor delas 0,73 e a maior 0,96. O percentual explicativo do modelo atinge 87,31% de explicação da variabilidade dos dados.

Matriz dos fatores, com pesos estimados e comunalidades.

Var.	*F1*	*F2*	*F3*	*F4*	*F5*	*Com.*	*Var. esp.*
1	-0,0689	0,0566	**-0,9239**	-0,2924	-0,1129	0,96	0,04
2	-0,2310	0,0781	**-0,8826**	-0,1241	-0,0932	0,86	0,14
3	-0,4036	0,0887	**-0,8189**	0,0088	0,0036	0,84	0,16
4	-0,1145	0,1373	-0,1115	0,0701	**-0,9336**	0,92	0,08
5	**-0,7813**	0,0158	-0,4039	-0,0289	-0,2350	0,83	0,17
6	**-0,8596**	-0,0741	-0,0483	-0,2796	-0,2263	0,88	0,12
7	**-0,7410**	-0,1005	-0,2973	-0,4877	-0,2665	0,96	0,04
8	**-0,7069**	0,0309	-0,6232	0,2110	0,0544	0,94	0,06
9	-0,1332	0.0901	-0,1846	**-0,9044**	0,1158	0,89	0,11
10	-0,0663	**0,9005**	-0,0524	0,0672	-0,0512	0,83	0,17
11	0,0919	**0,7260**	-0,1610	-0,3591	-0,1969	0,73	0,27
12	**0,8284**	-0,1739	0,1691	-0,0648	-0,3146	0,85	0,15
Autovalor	5,30		1,71	1,25	1,18		1,04
Prop.acum.	27,91		39,82	65,23	77,09		87,31

Fonte: matlab 6.5 função fator

Tabela 4- pesos estimados- variáveis pedagógicas

Para as variáveis pedagógicas, o conjunto de fatores e as respectivas variáveis fica composto segundo a tabela abaixo.

Fator	*Variáveis*	*Descrição das variáveis*
F1	P5, P6, P7,P8,P12	Cond. de estudo; mat. did. para o prof.; mat. did. para o aluno; grau de compr. dos prof. com relação a escola; de que maneira gostaria de ser avaliado.
F2	P10, P11	Pais ou responsáveis vão a escola com que frequência; quanto tempo você estuda, em média, por dia, fora da escola.
F3	P1, P2, P3	Equipe pedagógica; direção; secretaria.
F4	P9	Sua relação com os colegas de sala
F5	P4	Limpeza

Tabela 5- número de fatores com descrição de variáveis pedagógicas

Destacamos que, o favorecimento da análise fatorial torna possível que os colégios demandem esforços sobre os conjuntos de variáveis carregadas nos fatores e não pela análise de aprimoramento em separado de cada uma delas. Tal incremento minimiza o trabalho, demanda menos tempo e facilita a tomada de decisão.

A análise de regressão multivariada, no conjunto de variáveis pedagógicas, faz-se pelo melhor modelo, ou seja, aquele que apresenta maior R^2 e menor S. A variável dependente é a variável NA "nível de aprendizagem". O melhor modelo é aquele que apresenta menor diferença (erro de regressão) entre o nível desejado de aprendizagem dos alunos e o nível calculado a partir da equação de regressão multivariada.

Modelo de regressão pedagógico:

Variável estatística de regressão:
$NA = b_0 + b_1P2 + b_2P4 + b_3P5 + b_4P6 + b_5P7 + b_6P8 + b_7P9 + b_8P10 + b_9P11 + b_{10}P12$
Equação de regressão: NA=15,8 – 0,844P2 – 0,429P4 + 0,251P5 – 1,07P6 + 1,51P7 + 0,543P8 – 1,23P9 + 0,170P10 + 0,290P11 + 0,257P12

fonte: minitab>stat>regression>regression.

Tabela 6- variáveis pedagógicas- equação de regressão

O modelo é formado por dez das doze variáveis pedagógicas e apresenta um percentual explicativo da variabilidade dos dados de 62,40%.

3.3 Resultados para as variáveis auxiliares.

Os conjuntos de variáveis auxiliares têm um peso que é determinante no ritmo da escola e no nível de aprendizagem dos educandos, influenciado pelo acréscimo de estímulo, que o fator positivo de cada variável, pode proporcionar. A soma de fatores notoriamente favoráveis a aprendizagem como, por exemplo, boa condição de vida econômica, boa estrutura de bens na família, moradia própria, estudo em cursos paralelos as aulas escolares, faz com que a vida escolar melhore facilitando o trabalho dos gestores, professores e funcionários, pois, demanda que, o tempo gasto, por esses profissionais, com auxílios e serviços aos estudantes, seja gasto exclusivamente com o saber e o ensinar. Nesse caminho, há de se julgar e ponderar cada estrutura posta a disposição das escolas de ensino juntamente com o meio onde elas estão alocadas e em que condições, aqui chamadas de auxiliares, chega toda a sua clientela.

Constam da análise fatorial vinte e uma variáveis auxiliares com 55,71% de correlações acima de 0,30, baixo p valor e bom Q^2. Avaliou-se as inter-correlações e a matriz de resíduos. As comunalidades estão acima de 0,70 e o percentual explicativo do modelo é de 84,11% da variabilidade total dos dados.

Usando os pesos estimados da matriz rotacionada os seis fatores gerados na análise fatorial ficaram compostos conforme a tabela:

Fator	*Variáveis*	*Descrição das variáveis*
F1	AP4; AE5; AE6; AE7; AE9; AE10;AE11;AE12;AE14;AE17	Ajuda programas sociais; vídeo; computador; banheiro dentro de casa; aparelho de som; televisor; DVD; internet; asfalto na rua; renda mensal.
F2	AP3; AE3; AE15	Prática de esportes; casa própria; trabalho.
F3	AP1; AE8	Língua estrangeira; água encanada.
F4	AE13; AP10	Luz; importância da escola.
F5	AP7; AE4	Uso de drogas; geladeira.
F6	AP5; AP8	Curso superior; bebida alcoólica

Tabela 7 – fatores com descrição das variáveis auxiliares.

No modelo de análise fatorial os fatores representam grupos de variáveis. As variáveis são agrupadas nos fatores segundo sua similaridade. O percentual explicativo individual dos fatores decresce a medida que se aumenta o número de fatores, haja visto, a melhor combinação de variáveis estar no primeiro fator, depois no segundo e assim sucessivamente até termos toda a variância dos dados dissipada.

A equação multivariada de regressão para as variáveis auxiliares foi selecionada a partir da função stat>regression>Best subsets do programa minitab 15. O programa apresenta como melhor modelo o modelo que possui um poder explicativo da variabilidade R^2 de 95,0 e uma estimativa da variância média sobre a linha de regressão S de 0,36953. A equação de regressão para as variáveis auxiliares é composta por dezenove das vinte e três variáveis auxiliares que favorecem a aprendizagem e tem como resposta a variável dependente **NA** (nível de aprendizagem).

Variável estatística de regressão:
$NA = b_0 + b_1 AP1 + b_2 AP3 + b_3 AP5 + b_4 AP7 + b_5 AP8 + b_6 AE3 + b_7 AE4 + b_8 AE5 + b_9 AE6 + b_{10} AE7 + b_{11} AE8 + b_{12} AE9 + b_{13} AE11 + b_{14} AE12 + b_{15} AE13 + b_{16} AE14 + b_{17} AE15 + b_{18} AP10 + b_{19} AE17$
Equação de regressão: NA= 69,7+ 0,0319AP1 + 0,0357AP3 – 0,0253AP5 – 0,302AP7 + 0,0534AP8 + 0,171AE3 – 0,180AE4 – 0,0382AE5 + 0,0777 AE6 – 0,140AE7 – 0,169AE8 + 0,132AE9 + 0,0360AE11 – 0,0444AE12 + 0,0164AE13 + 0,0133AE14 + 0,0737AE15 – 0,278AP10 – 0,0300AE17

Fonte: minitab stat>regression>regression

Tabela 8 – variáveis auxiliares – equação de regressão.

Na equação de regressão os coeficientes das variáveis representam o quando a variável contribui para o nível de aprendizagem quando todas as demais variáveis estão fixas, dentro do modelo. As variáveis que contribuem positivamente para o aumento do nível de aprendizagem são as variáveis AP1 (língua estrangeira); AP3 (prática de esportes); AP8 (bebida alcoólica- não injeção); AE3 (casa própria); AE6 (computador); AE9 (banheiro dentro de casa); AE11 (DVD); AE13 (luz); AE14 (asfalto na rua) e AE15 (trabalho). Melhora nos coeficientes dessas variáveis aumenta o nível de aprendizagem, logo, espera-se que tais coeficientes sejam os maiores possíveis. As demais variáveis conduzem a uma perda de nível de aprendizagem, logo, espera-se que os valores de tais variáveis sejam os menores possíveis.

4 Análise dos resultados

A amostra inicial de alunos é de mil, oitocentos e vinte e seis sendo desse total 49,56 % do sexo masculino e 50,44% do sexo feminino distribuídos entre os vinte e dois colégios pesquisados e que estão entre as faixas de idade de dez a dezoito anos.

Com respeito a região de moradia 31,87% residem na zona rural e 68,13% dos alunos residem na zona urbana. O nível de escolaridade da mãe revela que 3,23% delas não possuem nenhuma escolaridade e que 36,80% sequer completaram o ensino fundamental. Já o nível de escolaridade dos pais é de 3,12% sem escolaridade alguma e de 32,37 % que não completou o ensino fundamental. Esses dados apontam para uma baixa escolaridade dos pais, ou seja, em média, aproximadamente, 37 % dos pais estudaram menos de oito anos escolares, dado esse tomado sem levar em conta o percentual de respondentes que desconhecem o nível de escolaridade de seus pais, que chega a 15,66% no caso da mãe e, 22,34% para o pai.

Com relação a renda para o suprimento das necessidades familiares constata-se que 34,17% das famílias suprem suas necessidades com rendimentos de até um salário mínimo e que 1,86 % das famílias têm uma renda entre onze a quinze salários mínimos. Vinte e dois vírgula setenta e oito porcento das famílias recebem ajuda de programas sociais sendo desse total 84,62% que recebem bolsa família, 19,71% participam do programa leite das crianças, e, 4,33% das famílias recebem ajuda dos dois programas.

A análise fatorial revelou o conjunto mais homogêneo de variáveis, dentro de cada grupo, e tem como fim revelar a estrutura dominante de variáveis, ou seja, revelar os conjuntos que melhor explicam a variabilidade do modelo. Para as **variáveis estruturais** o conjunto sala de aula, quadro negro e estado das carteiras alcançou o melhor poder explicativo, 22,57% e pode ser pensado como uma única estrutura. Dependendo da quantidade de investimentos sugere-se que os mesmos sejam aplicados em tal conjunto de variáveis, pois, maiores ganhos estruturais se alcança na sua correção, melhora ou conserto. Esse conjunto reflete melhor o anseio dos estudantes com relação a estrutura física das escolas pesquisadas. As **variáveis pedagógicas** creditam o maior poder explicativo as variáveis condições de estudo, material didático para o professor, material didático para o aluno, grau de comprometimento dos professores com relação a escola que lecionam e a variável de que maneira o aluno gostaria de ser avaliado, atingindo o índice de 27,91% e podem ser pensadas como estrutura única em investimentos pedagógicos, pois, refletem melhor a variabilidade do modelo dentro da expectativa apontada pelos estudantes.

Olhando para as **variáveis auxiliares que favorecem a aprendizagem** o modelo é melhor explicado pelas variáveis que consideram que o estudante ou que a sua família não recebe ajuda de programas sociais, possui vídeo cassete, DVD, aparelho de som, televisor, têm computador e é conectado com a internet, o banheiro é dentro de casa, têm asfalto na rua em que mora e que a renda mensal da família é acima de três salários mínimos. Tais variáveis são altamente correlacionadas e explicam 34% da variabilidade de dados do modelo, assim sendo, a melhora em tais estruturas auxiliares melhora a expectativa do conjunto todo, ou seja, teremos ganhos mais consideráveis se os investimentos forem direcionados para essas variáveis do que, por exemplo, se investirmos em um conjunto de variáveis com um poder explicativo menor.

Como meio de identificar, dentre as vinte e duas instituições de ensino, aquelas que melhor representam as estruturas físicas, pedagógicas e de bens auxiliares para que sirvam de modelo e de parâmetro para as demais instituições, não pesquisadas, foi feita a análise de regressão. A **regressão para as variáveis estruturais** aponta a variação de erro no intervalo de 0,0000 até 0,3721. Vinte e cinco por cento das instituições apresentam erros menores ou iguais a 0,011575. Cinquenta por cento das escolas possuem erro menor ou igual a 0,05085 e vinte e cinco por cento das vinte e duas escolas apresentam erros maiores ou iguais a 0,178525.

As escolas melhor classificadas quanto as variáveis estruturais apresentam erro de previsão, com relação ao nível de aprendizagem, no máximo de 0,0121, para mais ou para menos e, podem ser tomadas como modelos de estruturas físicas que servem ao nível de aprendizagem desejado e sugerido dentro de suas particularidades.

Com respeito as **variáveis pedagógicas** os resultados da análise de regressão mostram um erro quadrado que varia entre 0,0025 e 0,2704, sendo estas a melhor e a pior previsão, respectivamente, obtida pela equação de regressão dentro das expectativas de nível de aprendizagem pesquisado. O primeiro quartil, das variáveis pedagógicas, é de 0,030675; a mediana é de 0,073 e o terceiro quartil é de 0,146325. Esses dados apontam que, em torno de seis escolas, apresentam erros de regressão, iguais ou inferiores a 0,030675, entre o nível de aprendizagem desejável e o previsto pelo modelo de regressão. Tais escolas podem ser tomadas como exemplos de estruturas pedagógicas se obedecidas as suas peculiaridades individuais. É possível, ainda, prever aonde as instituições com maiores erros, aquelas que apresentam erros quadrados acima de 0,146325 devem investir para que se aproximem da linha de regressão alcançando índice melhor de satisfação entre o desejável o real.

No tocante as **variáveis auxiliares**, a equação de regressão multivariada selecionada como modelo apresenta erros quadrados de regressão que variam entre 0,0001 e 2,0736. Vinte e cinco por cento das escolas apresentam erros superiores ou iguais a 0,6455, estando entre os piores índices, avaliados na pesquisa, no que diz respeito a estrutura de variáveis auxiliares, propostas neste artigo, e comparadas com os índices de aprendizagens desejáveis pelas escolas. Já, as vinte e cinco por cento, melhores instituições, em torno de seis, apresentam erros quadrados iguais ou inferiores a 0,024825; e onze escolas, cinquenta por cento do total, apresentam erros quadrados iguais ou inferiores a 0,14825.

5 Conclusão

Tendo como objetivo encontrar caminhos positivos que auxiliem professores, alunos e gestores escolares na tomada de decisões a fim de que possamos avançar em direção a uma escola pública de qualidade é que desenvolveu-se o trabalho.

A pesquisa aponta para a melhora de rendimento escolar associado a melhora de vida, portanto, é necessário medirmos como as famílias investem seus ganhos, relacionando-os aos bens que possuem e ao tipo de educação que recebem e que estão sujeitos. É de censo comum que a qualidade da educação depende de investimentos dos poderes públicos e também de investimentos e incentivos por parte das famílias, mas, nesta análise fica claro a necessidade de termos, como um todo, uma melhor partição da renda, por parte dos empregadores, de forma que um percentual considerável das famílias possa gerir melhor as condições de seus filhos quanto a qualidade de vida oferecendo melhores condições de sustento e de bens materiais fundamentais

ao crescimento físico e intelectual, havendo, dessa forma, um desafogar do sistema de ensino que muitas vezes assume tais compromissos.

As diretrizes curriculares da educação básica da secretaria de Estado da Educação do Paraná (2008), na dimensão do conhecimento estende, ao estudante, que esse tenha uma formação necessária para o enfrentamento com vistas a transformação da realidade social, econômica e política de seu tempo.

Nesse sentido, o trabalho oferece mais um modelo para a avaliação das instituições de ensino a partir de dados que medem a estrutura física e pedagógica das escolas/colégios. Inova na maneira de tratar dados dicotômicos, na escolha das variáveis, que e feita com prévia separação entre estrutural, pedagógica e auxiliares e na utilização da análise de regressão linear multivariada.

Tal avaliação pode servir de meio ou ajuda, no que é tratado na escola, por meio dos conteúdos das disciplinas, pois, oferece base para a análise estrutural, pedagógica e social, da clientela e das instituições, escolas/colégios de ensino, tendo como fim uma escola que se forma por fatores internos e por fatores ditos externos como os determinados pelo regime socio-político, família, trabalho, assim como, características sociais e culturais do público escolar.

6 Referências

ARANHA, Maria Lúcia de Arruda. **História da educação**. São Paulo: moderna, 1997.

BARBOSA, S. G. **Avaliação das escolas do núcleo regional de educação de Paranavaí através de data envolopment analysis, análise de regressão e correlação**. 2007, 216 f. Dissertação (métodos numéricos em engenharia). Universidade Federal do Paraná, Curitiba, 2007.

BOCK, R. D.; AITIKIN, M. **Marginal Maximun Likelihood estimation of the parameters: application of em algorithm.** Psychometrika, 1981.

CRESPO, A. Arnot. **Estatística fácil**. São Paulo: Saraiva, 2002.

DA FONSECA, J. Simon; MARTINS, G. de Andrade; TOLEDO, G. Luciano. **Estatística aplicada.** Atlas, 1978.

DE SOUZA, Paulo N. Pereira. **LDB e educação superior.** Thomson Learning, 2001.

DOWNING, Douglas; CLARK, Jeffrey. **Estatística aplicada.** São Paulo: Saraiva, 2006.

ESQUARCINI, Rogério; MARQUES, Jair Mendes. **Classificação dos municípios paranaenses segundo suas políticas setoriais pela análise multivariada**. Revista FAE, v.9, Nº 1. Curitiba, 2006.

GOVERNO DO ESTADO DO PARANÁ; SECRETARIA DE ESTADO DA EDUCAÇÃO; SUPERINTENDÊNCIA DA EDUCAÇÃO. **Diretrizes curriculares de matemática para as séries finais do ensino fundamental e para o ensino médio**. Curitiba: 2008.

HAIR Jr, Joseph F. et al. **Análise Multivariada de Dados**. Tradução de Adonai S. Sant'anna e Anselmo C. Neto. Porto Alegre: Bookman, 2005.

JOHNSON, Richard A.; WICHERN, Dean W. **Aplied Multivariate Statistical Analysis.** Englewood Clifs: Prentice-Hall, 1988

LEVINE, David M.; BERENSON, Mark; STEPHAN, David. **Estatística: teoria e aplicações.** LTC, 2000.

LOPES, Luís Felipe Dias; ZANELLA, Andréia. **Identificação de fatores que influenciam na qualidade do ensino de matemática, através da análise fatorial**. Revista eletrônica Sistemas & Gestão. 2007.

MARQUES, A. F. **Aplicação da análise multivariada na infra-estrutura e desempenho das escolas públicas do ensino fundamental e médio pertencentes ao núcleo regional de educação de Paranavaí**. 2007, 124 f. Dissertação (métodos numéricos em engenharia). Universidade Federal do Paraná, Curitiba, 2007.

MOITA NETO, JM. **Estatística multivariada. Uma visão didática-metodológica.** Filosofia da ciência. 2004.

PAIVA BELLO, José Luiz de. **História da educação no Brasil.** Rio de Janeiro, 1998.

SACRISTÁN, J. Gimeno; GÓMEZ, A. I. Pérez. **Compreender e transformar o ensino.** Artmed, 2000.

SAMPAIO, Breno; GUIMARÃES, Juliana. **Diferenças de eficiência entre ensino público e privado no Brasil.** Revista economia aplicada. Volume 13. Ribeirão Preto, 2009.

SMAILES, Joanne; MCGRANE, Ângela. **Estatística aplicada à administração com excel**. Tradução de Christiane Brito. Bázan tecnologia e linguística Ltda. São Paulo: Atlas, 2002.

SOARES, Tufi Machado. **Utilização da teoria da resposta ao item na produção de indicadores socioeconômicos.** Minas Gerais: Universidade federal de Juiz de Fora.

TAVARES Jr., Fernando. **Planejamento e produção de dados- parte I.** Governo de Minas. Centro de referência virtual do professor.