

## CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ  
Rio de Janeiro - RJ - Brasil  
ISSN 1413-9022 / v. 28, p. 47 - 59, 2010

# ESTUDO E APLICAÇÃO DA ANÁLISE MULTIVARIADA DE DADOS PARA A VALIDAÇÃO DAS ESTRUTURAS FÍSICAS E PEDAGÓGICAS DE ESCOLAS PÚBLICAS DO SUL DO PARANÁ

Antônio Carlos Machnicki  
PPGMNE – Universidade Federal do Paraná (UFPR)  
acmachnicki@seed.pr.gov.br  
machnicki@uol.com.br

Anselmo Chaves Neto  
PPGMNE – Universidade Federal do Paraná (UFPR)  
anselmo@ufpr.br

Liliana Madalena Gramani  
PPGMNE – Universidade Federal do Paraná (UFPR)  
l.gramani@gmail.com

### 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 técnicas 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.

## 1. Introdução

Na busca da qualidade educacional devemos encontrar caminhos que auxiliem professores, alunos e gestores escolares, na tomada de decisões, a fim de que avancemos em direção a uma escola pública de qualidade. Segundo De Souza (2001), a qualidade dos serviços educacionais depende, em grande parte, da adequação de suas instalações físicas. Segundo ele, o ambiente de uma escola também educa e pode tornar a formação de seus alunos não apenas mais eficiente como também muito agradável.

Quantificar como a tecnologia, o material didático, a sala de aula, as condições de estudo, o trabalho do professor, as condições físicas dos estabelecimentos de ensino, o ambiente escolar e o ambiente social influenciam no rendimento escolar é o objetivo principal do trabalho. Para esse fim, são investigadas variáveis relacionadas ao ensino fundamental e médio, em escolas públicas, com a finalidade de identificar as variáveis que melhor se correlacionam, com o desempenho dos alunos e com a eficiência das escolas, fornecendo subsídios e propondo formas de avanços tanto na estrutura física, de equipamentos e pedagógica de modo que o reflexo seja a qualidade educacional oferecida pelas instituições de ensino, com ganho imediato nos índices educacionais brasileiros, na formação da pessoa e na qualidade do cidadão formado.

## 2. Revisão Bibliográfica

A análise multivariada compreende todos os métodos estatísticos que podem ser aplicados na análise simultânea de múltiplas medidas sobre cada indivíduo ou objeto que se está investigando. Segundo Crespo (2006, p.13), a estatística é uma ciência matemática que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões.

A variável estatística é composta por combinações lineares de variáveis com pesos empiricamente determinados. Por exemplo, uma variável estatística de  $n$  variáveis ponderadas  $(x_1, x_2, \dots, x_i)$  pode ser escrita como:  $y = w_1x_1 + w_2x_2 + \dots + w_ix_i$  cujo  $x_i$  com  $i = 1, 2, 3, \dots, n$ , é a variável observada e  $w_i$ , com  $i = 1, 2, 3, \dots, n$ , é o peso que a técnica multivariada determina. O resultado apresenta um valor único que mostra a combinação do conjunto de variáveis que atinge o objetivo específico com a técnica multivariada adotada. Para a representação de dados adota-se a notação matricial  $[x_{jk}]$

que indica um valor particular da  $k$ -ésima variável mensurada na  $j$ -ésima unidade amostral.

## 2.1 Análise fatorial

A análise fatorial é uma forma de redução de dados aonde obtém-se, com a utilização da técnica, uma forma menor de descrever os dados que inicialmente eram em um número muito maior e, por conseguinte, mais difíceis de serem analisados e interpretados. Cada dimensão é representada por um fator que por sua vez se associa a um grupo de variáveis e tem por objetivo aumentar o poder explicativo considerando o conjunto inteiro de variáveis.

### 2.1.1 Suposições na análise fatorial

Segundo Hair Jr. *et al.* (2005, p.98), se a inspeção da matriz de dados não revelar um número substancial de correlações maiores que 0,30, então a análise fatorial não é recomendada. Podemos usar a medida de adequação da amostra (MSA) para quantificar o grau de inter-correlações entre as variáveis e a possível aplicação da análise fatorial. Para Hair Jr. *et al.* (2005), tal medida deve ser interpretada com as seguintes orientações e seguindo os valores do MSA (0,80 ou acima admirável e abaixo de 0,50 inaceitável). Devemos procurar a homogeneidade da amostra com relação a estrutura fatorial dominante, ou seja, grupos muito distintos de conjuntos de variáveis devem ser analisados separadamente, pois se duas ou mais sub-amostras muito distintas são combinadas as correlações são fracas resultando uma estrutura única de cada grupo.

## 2.2 Análise de regressão

Para Smailes e McGrane (2002, p.115), quando existem propósitos para a tomada de decisões, é útil identificar se existem relações lineares entre duas ou mais variáveis e, se existir, quantificar sua força. Na visão de Hair Jr. *et al.* (2005, p.136), 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. Afirma que, a análise de regressão é a técnica que vai ponderar a variável independente garantindo a máxima previsão. Segundo Downing e Clark (2006), na regressão múltipla a relação entre a variável dependente  $y$  e as  $m-1$  variáveis independentes  $x_1, x_2, \dots, x_{m-1}$  é dada pela equação:  $y = B_1x_1 + B_2x_2 + \dots + B_{m-1}x_{m-1} + B_m + e$  Os verdadeiros valores para os

parâmetros  $B_1, B_2, B_{m-1}, B_m$  são desconhecidos, precisam ser estimados.  $B_1$  representa o efeito que  $x_1$  tem sobre  $y$ , admitindo que  $x_2$  permaneça constante. Da mesma forma,  $B_2$  representa o efeito de  $x_2$  sobre  $y$  admitindo que  $x_1$  fique constante. “ $e$ ” é uma variável aleatória conhecida como termo erro. Admite-se que “ $e$ ” tem distribuição normal, com esperança zero e variância ( $\sigma^2$ ) desconhecida.

### 3. Metodologia

O trabalho investigou 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.

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 com análise fatorial e análise de regressão.

#### 3.1 Transformando dados dicotômicos

As técnicas estatísticas multivariadas dependem de variáveis cujos valores são métricos. As variáveis pesquisadas apresentavam características não métricas e sua transformação foi feita pela atribuição de pesos para cada um dos conceitos, em uma escala de 10 a 3. Com isso produziu-se uma matriz vetor que padronizou cada variável. Exemplificando: considerando os conceitos ótimo, bom, regular, ruim, insatisfatório, não tem, fez-se a padronização pela matriz

$$\text{variável} = [10 \ 9 \ 7 \ 5 \ 4 \ 3]^* \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix} : (w_1 + w_2 + w_3 + w_4 + w_5 + w_6), \text{ onde:}$$

$w_1, w_2, w_3, w_4, w_5$  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:

	<i>variáveis</i>												
<i>escolas</i>	$E_1$	$E_2$	.	.	$E_n$	.	$P_1$	$P_2$	.	.	$P_n$	.	$NA$
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	$E_{22}$	$E_{22}$	.	.	$E_{22n}$	.	$P_{22}$	$P_{22}$	.	.	$P_{22n}$	.	$NA_{22}$

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.

#### 4. 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.

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.

##### 4.1 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 correlações, entre as variáveis, maiores do que 0,30 e matriz de resíduos que indicam aleatoriedade. As comunalidades, entre as variáveis e os fatores, têm a menor o valor 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.

Tabela 1- pesos estimados- variáveis estruturais

Var.	F1	F2	F3	F4	F5	Com.	Var.esp.
1	0,1873	-0,0601	<b>0,6718</b>	-0,4501	0,2904	0,78	0,22
2	0,2089	-0,2651	-0,0733	<b>-0,7985</b>	0,0555	0,76	0,24
3	-0,4424	-0,1892	0,4090	0,1134	<b>-0,6177</b>	0,79	0,21
4	0,2282	0,1839	<b>0,7998</b>	0,1737	-0,1579	0,78	0,22
5	0,5026	0,2900	0,0735	<b>-0,7199</b>	-0,0444	0,86	0,14
6	-0,1226	0,0215	0,0136	<b>-0,8117</b>	-0,1485	0,70	0,30
7	-0,1748	<b>-0,8672</b>	-0,0662	0,1402	0,0725	0,81	0,19
8	0,0377	-0,0282	-0,0604	-0,0855	<b>-0,9330</b>	0,88	0,12
9	<b>0,8124</b>	-0,2115	0,3393	-0,0727	0,0787	0,83	0,17
10	0,2033	<b>-0,7481</b>	-0,0260	-0,3218	-0,3511	0,83	0,17
11	<b>0,8403</b>	-0,0505	0,0716	-0,1131	-0,0044	0,73	0,27
12	<b>0,8214</b>	0,3081	0,0365	-0,1325	0,0172	0,79	0,21
Autovalor	3,53	2,06	1,53	1,37	1,07		
Prop.Acum.	22,57	36,59	48,25	66,81	79,50		

Fonte: dados da pesquisa. Programa computacional matlab 6.5 (função fator).

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

Tabela 2 – número de fatores com descrição das variáveis estruturais

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: dados da pesquisa

Para a análise de regressão temos o modelo constituído a partir de onze das doze variáveis estruturais, apresentando um poder de explicação da variabilidade dos dados de 59,5% e uma estimativa média sobre a linha de regressão de 0,468874. A regressão é feita considerando a equação  $NA=13,2 - 0,388E1 - 0,740E2 + 0,0420E3 - 0,0611E4 + 0,528E5 - 0,182E6 - 0,181E8 + 0,189E9 + 0,274E10 - 0,092E11 + 0,084E12$ , onde NA é a variável nível de aprendizagem e E1 a E12 são as variáveis estruturais. Com a aplicação da equação de regressão, acima, identificamos o conjunto ideal de estruturas físicas o conjunto com erro quadrático de 0,0000, atingindo o extremo superior com erro quadrático de 0,3721.

## 4.2 Resultados para as variáveis pedagógicas

A análise fatorial identificou um conjunto com cinco fatores. Quanto aos métodos de adequação da amostra, a mesma apresenta um  $Q^2$  de 187,6132 e MSA de 0,5420, possuindo 43,94% das correlações maiores do que 0,30 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.

Tabela 3- pesos estimados- variáveis pedagógicas

Var.	F1	F2	F3	F4	F5	Com.	Var. esp.
1	-0,0689	0,0566	<b>-0,9239</b>	-0,2924	-0,1129	0,96	0,04
2	-0,2310	0,0781	<b>-0,8826</b>	-0,1241	-0,0932	0,86	0,14
3	-0,4036	0,0887	<b>-0,8189</b>	0,0088	0,0036	0,84	0,16
4	-0,1145	0,1373	-0,1115	0,0701	<b>-0,9336</b>	0,92	0,08
5	<b>-0,7813</b>	0,0158	-0,4039	-0,0289	-0,2350	0,83	0,17
6	<b>-0,8596</b>	-0,0741	-0,0483	-0,2796	-0,2263	0,88	0,12
7	<b>-0,7410</b>	-0,1005	-0,2973	-0,4877	-0,2665	0,96	0,04
8	<b>-0,7069</b>	0,0309	-0,6232	0,2110	0,0544	0,94	0,06
9	-0,1332	0,0901	-0,1846	<b>-0,9044</b>	0,1158	0,89	0,11
10	-0,0663	<b>0,9005</b>	-0,0524	0,0672	-0,0512	0,83	0,17
11	0,0919	<b>0,7260</b>	-0,1610	-0,3591	-0,1969	0,73	0,27
12	<b>0,8284</b>	-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: dados da pesquisa. Programa computacional matlab 6.5 (função fator)

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

Tabela 4- número de fatores com descrição de variáveis pedagógicas

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

Fonte: dados da pesquisa



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”. A melhor estrutura escolar é aquela que apresenta menor diferença (erro de regressão) entre o nível desejado de aprendizagem dos alunos (coletado na pesquisa) e o nível calculado a partir da equação  $NA=15,8 - 0,844P2 - 0,429P4 + 0,251P5 - 1,07P6 + 1,51P7 + 0,543P8 - 1,23P9 + 0,170P10 + 0,290P11 + 0,257P12$ . Constatou-se que, a melhor estrutura pedagógica é aquela que apresenta um erro de regressão ao quadrado de 0,0025, sendo o limite superior atingido com um erro quadrático de 0,2704. Tal modelo tem um percentual explicativo da variabilidade dos dados de 62,40% e uma variância média sobre a linha de regressão de 0,431368.

### 4.3 Resultados para as variáveis auxiliares

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. Constam da análise fatorial vinte e uma variáveis auxiliares com 55,71% de correlações acima de 0,30. Avaliou-se as intercorrelações 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 segue.

Tabela 5 – fatores com descrição das variáveis auxiliares

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; televisão; 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

Fonte: dados da pesquisa

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. Aplicando a equação de regressão,  $NA = 69,7 + 0,0319AP1 + 0,0357AP3 - 0,0253AP5 - 0,302AP7 + 0,0534AP8 + 0,171AE3 - 0,180AE4 - 0,0382AE5 + 0,0777 AE6 - 0,140AE7 - 0,169AE8 + 0,132AE9 + 0,0360AE11 - 0,0444AE12 + 0,0164AE13 + 0,0133AE14 + 0,0737AE15 - 0,278AP10 - 0,0300AE17$ , determina-se o conjunto de variáveis auxiliares que melhor ganho trás para a aprendizagem.

Na equação de regressão os coeficientes das variáveis representam o quanto 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.

## 5. Análise dos resultados

A análise fatorial revelou o conjunto mais homogêneo de variáveis, dentro de cada fator. 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 estrutura única para possíveis investimentos. 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. 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.

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 apresentam erros maiores ou iguais a 0,178525. Com relação 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. 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.

## 6 Considerações Finais

Partindo das primícias de que para uma escola pública de qualidade são necessários recursos estruturais, pedagógicos e de características classificadas, no referido artigo, como auxiliares é que avaliou-se, segundo pesquisa, e com o uso das

técnicas estatísticas da análise fatorial e da análise de regressão as vinte e duas instituições escolas/colégios da região sul do Paraná.

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. 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 sócio-político, família, trabalho, assim como, características sociais e culturais do público escolar.

## Referências

- CRESPO, A. A. **Estatística Fácil**. São Paulo: Saraiva, 2002.
- DE SOUZA, P. N. P. **LDB e Educação Superior**. Thomson Learning, 2001.
- DOWNING, D.; CLARK, J. **Estatística Aplicada**. São Paulo: Saraiva, 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, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W.C. **Análise Multivariada de Dados**. Porto Alegre: Bookman, 2005.
- SAMPAIO, B.; GUIMARÃES, J. Diferenças de eficiência entre ensino público e privado no Brasil. **Revista Economia Aplicada**. Volume 13. Ribeirão Preto, 2009.
- SMAILES, J.; MCGRANE, A. **Estatística Aplicada à Administração com Excel**. São Paulo: Atlas, 2002.

## **STUDY AND APPLICATION OF MULTIVARIATE DATA ANALYSIS FOR VALIDATION OF PHYSICAL AND TEACHING STRUCTURES OF PUBLIC SCHOOLS OF SOUTHERN PARANA**

### **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 public 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.*

**Key-words:** Educational; School; Multivariate Analysis.