

DAVIDSON JOSÉ MOULEPES

**A FORMAÇÃO DOS FUTUROS PROFESSORES DO ENSINO FUNDAMENTAL
BRASILEIRO: UMA ANÁLISE MULTINÍVEL**

Monografia apresentada como requisito
para obtenção da graduação do Curso
de Ciências Econômicas, do Setor de
Ciências Sociais e Aplicadas da
Universidade Federal do Paraná.

Orientador: Prof^o Flávio de Oliveira
Gonçalves

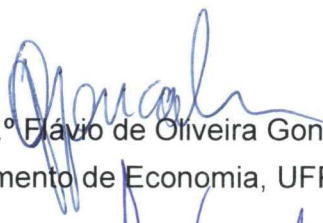
**CURITIBA
2007**

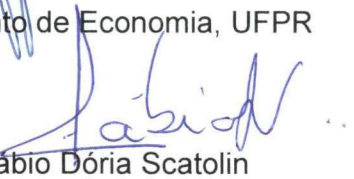
TERMO DE APROVAÇÃO

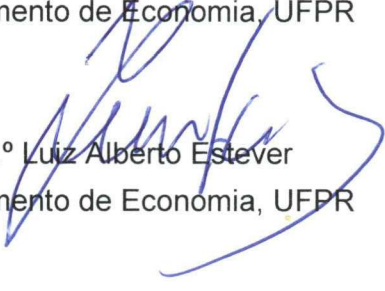
DAVIDSON JOSÉ MOULEPES

A FORMAÇÃO DOS FUTUROS PROFESSORES DO ENSINO FUNDAMENTAL BRASILEIRO: UMA ANÁLISE MULTINÍVEL

Monografia apresentada como requisito parcial para obtenção do grau de Graduação do Curso de Ciências Econômicas, Setor de Ciências Sociais e Aplicadas da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientador: Prof.º Dr.º  Flávio de Oliveira Gonçalves
Departamento de Economia, UFPR

Prof.º Dr.º  Fábio Dória Scatolin
Departamento de Economia, UFPR

Prof.º Dr.º  Luiz Alberto Esteves
Departamento de Economia, UFPR

Curitiba, 29 de Novembro de 2007

***Aos meus amados pais,
Sirley e Hamilton.***

AGRADECIMENTOS

A **DEUS**, por proporcionar a conclusão de mais uma etapa conquistada, que se consuma neste trabalho.

A **FAMÍLIA**, principalmente a minha mãe, que sempre acreditou que eu alcançaria mais esse sonho, e que nunca deixou de amparar e dedicar o seu amor a minha vida.

Aos **AMIGOS** que fizeram parte dessa história e que contribuíram para que a vida acadêmica fosse mais fácil durante os anos que passei na universidade.

Ao professor, orientador e amigo **FLÁVIO DE OLIVEIRA GONÇALVES**, pelo estímulo que me permitiu levar até o fim este trabalho, e mostrar um assunto tão interessante como este, a **EDUCAÇÃO**, além de fornecer o material e todas as ferramentas necessárias para o desenvolvimento do projeto.

Não posso deixar de agradecer ao **MARCO TÚLIO ANICETO FRANÇA**, que teve paciência ao me mostrar o caminho a ser percorrido para chegar aos resultados aqui apresentados, e separar um pouco do seu tempo para me transmitir sua experiência, que foi de grande importância para a conclusão da monografia.

Meu Muito Obrigado a Todos!!!

***“Uma pessoa permanece jovem na medida
em que ainda é capaz de aprender, adquirir novos hábitos
e tolerar contradições.”
(Marie von Ebner-Eschenbach)***

SUMÁRIO

LISTA DE GRÁFICOS.....	vii
LISTA DE TABELAS.....	viii
INTRODUÇÃO.....	1
CAPÍTULO I.....	4
1. METODOLOGIA.....	4
1.1. Modelo Multinível e seus Aspectos Teóricos.....	4
1.2. O modelo de Regressão Hierárquico Linear.....	5
CAPÍTULO II.....	11
2. OS DADOS DO ENADE E A DEFINIÇÃO DAS VARIÁVEIS DO MODELO	11
2.1. Os Dados do ENADE e do Censo da Educação Superior	11
2.2. Definição das Variáveis do Modelo	13
2.2.1. Variáveis do Primeiro Nível: Indivíduo.....	14
2.2.2. Variáveis do Segundo Nível: Instituição	18
CAPÍTULO III.....	23
3. RESULTADOS	23
3.1. Modelo Incondicional.....	33
3.2. Modelo Condicional	35
4. CONSIDERAÇÕES FINAIS	41
5. REFERÊNCIAS BIBLIOGRÁFICAS	46
ANEXO I: Análise Fatorial	47
ANEXO II: Teste de Normalidade dos Resíduos	54

LISTA DE GRÁFICOS

GRÁFICO A.II.1 – Q-Q (Qui-Quadrado) dos Resíduos	55
--	----

LISTA DE TABELAS

TABELA 3.1: Amostra e População do ENADE 2005	13
TABELA 3.2: Variável Renda	15
TABELA 3.3: Variável Sexo	16
TABELA 3.4: Variável Raça	17
TABELA 3.5: Variável Escolaridade do Pai	18
TABELA 3.6: Variável Escolaridade da Mãe	18
TABELA 3.7: Variável Iniciação Científica e Extensão : Nível 1.....	19
TABELA 3.8: Categoria Administrativa	19
TABELA 3.9: Variável Técnicas de Ensino.....	20
TABELA 3.10 Variável Instrumentos de Avaliação	21
TABELA 3.11 Variável Iniciação Científica e Extensão : Nível 2	22
TABELA 4.1: Modelo Incondicional	25
TABELA 4.2: Modelo Condicional 1	28
TABELA 4.3: Modelo Condicional 2 e 3.....	31
TABELA 4.4: Coeficiente : Desvio Padrão	40
TABELA A.1: Análise Fatorial da Variável Infra-Estrutura.....	53
TABELA A.2: Extração	53
TABELA A.3: Matriz do Componente Principal.....	54

A FORMAÇÃO DOS FUTUROS PROFESSORES DO ENSINO FUNDAMENTAL BRASILEIRO: UMA ANÁLISE MULTINÍVEL

Autor: Davidson José Moulepes

Orientador: Prof. Dr. Flávio de Oliveira

Gonçalves

RESUMO

Este trabalho tem por objetivo analisar o impacto das características institucionais e individuais sobre o desempenho dos estudantes formandos dos cursos de Letras, Matemática e Pedagogia no Exame Nacional de Avaliação do Desempenho do Estudante (ENADE) no ano de 2005. Para a análise, foi utilizado o modelo multinível, pois este incorpora a estrutura hierárquica dos dados na estimação. A partir dos resultados, verificou-se que a renda mensal familiar é significativamente relacionada ao desempenho do aluno no exame, assim como a escolaridade do pai e a da mãe. Em relação aos aspectos institucionais, a composição do corpo docente tem efeito positivo sobre o resultado esperado na avaliação. Quanto maior a porcentagem de mestres e doutores dentro da instituição, melhores são os resultados dos formandos, assim como a utilização de aulas práticas e trabalhos em grupo como forma predominante de técnica de ensino utilizada pelos seus professores.

Palavras-chaves: desempenho, modelos lineares hierárquicos, função de produção educacional

INTRODUÇÃO

O sistema de ensino superior brasileiro teve profundas e significativas mudanças em sua composição e funcionamento nos últimos anos. O Censo da Educação Superior, feita anualmente pelo Instituto Nacional de Estudos e Pesquisas Anísio Teixeira - INEP, aponta que em 2005 eram 2.165 Instituições de Ensino Superior. Comparando-se com o ano de 2002, esse número representa um acréscimo de 32,25% de Instituições no período, que totalizavam 1.637.

Diante desses aumentos significativos, é importante analisar a qualidade e a eficiência dessas instituições no que se refere à formação de profissionais capacitados para ingressarem no mercado de trabalho. Não é apenas a preocupação quanto ao ingresso dos estudantes no mercado de trabalho, mas sim, no caso dos cursos de licenciatura oferecidos, como essas instituições estão capacitando os novos professores para o ensino. Questões relevantes nascem desse processo de aceleração quantitativa do ensino superior, e a busca por encontrar quais são os determinantes da eficácia do ensino superior é um dos focos desse trabalho.

O trabalho consiste na construção de um modelo de regressão hierárquico (também conhecido como modelo multinível) capaz de demonstrar quais são as variáveis que explicam a proficiência dos alunos e identificar os determinantes individuais e institucionais que influenciaram os resultados e o desempenho dos estudantes no Exame Nacional de Desempenho do Estudante - ENADE¹ 2005, pelos formandos dos cursos de Letras, Matemática e Pedagogia. Os três cursos foram selecionados em decorrência dos baixos resultados dos alunos da 4ª e 8ª Série na avaliação do SAEB². Trabalhos como o de ALBERNAZ, FERREIRA & FRANCO (2002), apontam que a qualidade do professor, medida principalmente pelo seu nível

¹ ENADE é o Exame Nacional de Desempenho do Estudante que integra o Sistema Nacional de Avaliação do Ensino Superior, realizado pelo INEP. Seu objetivo é de aferir o rendimento dos alunos dos cursos selecionados de graduação, em relação aos conteúdos programáticos, suas habilidades e competências.

² SAEB – Sistema de Avaliação da Educação Básica – é um exame que se propõe a analisar o desempenho dos alunos em algumas séries: 4ª e 8ª Série do Ensino Fundamental e 3º Ano do Ensino Médio – realizado pelo INEP.

de escolaridade contribuem para um melhor desempenho de seus alunos, além de outras características, como nível de renda familiar e a estrutura física da escola. BARBOSA & FERNANDES (2001), apontam que as características associadas ao professor e à sua interação com a turma aparecem com grande poder explicativo no desempenho escolar dos alunos da 4ª Série. É de conhecimento, que não são somente os recursos materiais que tem influência sobre o desempenho dos alunos do ensino fundamental e médio, mas também os professores dessas escolas, que tem sua parcela de influência sobre os resultados, como mostra o trabalho de SOARES (2005). Dessa forma, analisar a performance dos formandos dos cursos de Letras, Matemática e Pedagogia se tornam importante, na medida em que parte desses formandos serão os professores que influenciarão o desempenho dos estudantes do ensino fundamental e médio do amanhã.

Dado a importância de se verificar e quantificar os efeitos dos insumos educacionais no desempenho de estudantes, o objetivo desse trabalho é modelar e estimar uma função de produção educacional que auxilie na compreensão desses determinantes.

Ainda, entre os conjuntos das instituições fundamentais para o desenvolvimento sócio econômico de um país está o seu sistema educacional. A acumulação de capital humano ocupa posição de destaque na linha de pensamento do processo de crescimento econômico de um país, além de também explicar o nível de desigualdade de renda, como já verificados em trabalhos como o de BARROS, HENRIQUE & MENDONÇA (2002).

É de se esperar, portanto que, compreender, identificar e quantificar os efeitos dos determinantes da proficiência dos estudantes brasileiros seja de interesse acadêmico e de principalmente, interesse dos gestores das políticas públicas educacionais. Esses determinantes podem ser tais como escolaridade e experiência dos professores, infra-estrutura acadêmica, e outras variáveis que afetam o desempenho e a aprendizagem dos alunos.

A relação entre os insumos do processo produtivo e o produto resultante é descrita na microeconomia como função de produção. Uma função de produção indica o produto máximo que uma empresa produz para cada combinação específica de um insumo. Na literatura sobre “Educação”, a relação existente entre uma

quantidade de insumos ao processo educacional e seu produto é dita “função de produção educacional”, entendendo-se produto como o desempenho dos alunos, e os insumos como os fatores que determinam o resultado do desempenho do estudante em uma determinada avaliação padronizada.

Sobre o tema, ainda são escassas as análises a nível nacional e internacional, e as que existem ainda são recentes. Entretanto, trabalhos já desenvolvidos apresentaram resultados interessantes no que se refere à eficiência do ensino superior e principalmente do ensino médio-fundamental.

VAN der HULST & JANSEN (2002) trabalharam no sentido de encontrar as características individuais e curriculares sobre o progresso dos alunos do curso de engenharia das Instituições de Ensino Superior dos Países Baixos. Seus resultados mostraram que características individuais, como idade e sexo, têm influência sobre o comportamento dos alunos em relação à duração de seus cursos.

DIAZ (2006) examinou os aspectos individuais e principalmente os institucionais que influenciaram o desempenho dos estudantes no Exame Nacional dos Cursos (Provão) realizado no ano de 2000 para os formandos das áreas de Administração, Direito e Engenharia Civil, e verificou nos aspectos individuais que a situação sócio-econômica do aluno tem um efeito significativo e importante na determinação do rendimento do estudante no exame. No que tange aos aspectos institucionais sobre os graduandos, identificou-se impactos positivos oriundos de maior qualificação e de melhores condições de trabalho do corpo docente, além da dedicação, empenho e pontualidade dos mesmos.

Há que se considerar os trabalhos realizados com o objetivo de procurar identificar características institucionais e individuais que podem influenciar o desempenho escolar dos alunos do ensino fundamental. Dados escolares são estruturados hierarquicamente, alunos, turmas e escolas, e as variáveis representativas dos três níveis podem interagir dentro do mesmo nível, e também com variáveis de outro nível hierárquico. Em decorrência desta natureza, a metodologia utilizada para a análise nesses trabalhos é a utilização dos modelos multinível, descritos por BRYK & RAUDENBUSH (1992).

SOARES (2005) constrói e interpreta um modelo de regressão hierárquica em três níveis (aluno, turma e escola) para explicar a proficiência dos alunos da 4ª Série do ensino fundamental, utilizando a avaliação em língua portuguesa realizada em 2002 pela PROEB³. Seu resultado foi de que no sistema, parte da parcela da variabilidade dos resultados encontrados para a proficiência dos alunos da 4ª série é devido à turma (31,9%) e uma parcela devido à escola (11,5%), incluindo o efeito do professor, e os 56,7% restantes é a variabilidade explicada pelas características do aluno. Logo, estudar os cursos ligados à carreira de magistério se torna importante na medida em que estes poderão qualificar os insumos na explicação dos resultados (baixos) obtidos pelos estudantes do ensino fundamental e médio.

Assim, com o auxílio dos modelos hierárquicos, avaliamos em um primeiro nível, as características individuais dos formandos. Essas características incluem as variáveis relacionadas à etnia, ao gênero, ao nível de renda e outros aspectos condizentes ao ambiente familiar que o estudante está inserido. No segundo nível, as variáveis institucionais, como técnicas de ensino, categoria administrativa, entre outros, serão analisadas com o interesse de se investigar qual é a magnitude do efeito dessas características sobre o rendimento dos estudantes no exame do ENADE.

O trabalho está dividido em três partes, além desta introdução, da conclusão e dos anexos. No primeiro capítulo, apresentamos o modelo e seus aspectos teóricos, descrevendo a metodologia hierárquica linear com o auxílio de um modelo geral com suas hipóteses e, resumidamente, apresentamos os testes de especificação. Em seguida, no segundo capítulo descrevemos a fonte dos dados e o tratamento das variáveis, além das análises descritivas. Por fim, o terceiro capítulo traz os resultados da estimação do modelo hierárquico linear para os estudantes formandos dos cursos selecionados. O última parte do trabalho traz as considerações finais.

³ PROEB é o Programa de Avaliação da Rede Pública de Educação Básica de Minas Gerais feita pelo SIMAVE – Sistema Mineiro de Avaliação da Educação Pública

CAPÍTULO I

O objetivo desse capítulo é detalhar a metodologia hierárquica linear para modelos de regressão que possuem dados agrupados em níveis, como os educacionais. Esse modelo incorpora naturalmente a estrutura hierárquica das variáveis, além de captar as características de cada nível da hierarquia no desempenho dos estudantes e na diferenciação entre as instituições, e ainda, separar a variabilidade nos resultados associada às instituições da variabilidade dentro de cada instituição. Essa apresentação do modelo consiste em inicialmente identificar os aspectos teóricos. Em seguida, elaboramos um modelo multinível geral e apresentamos sinteticamente os testes de especificação do modelo. A partir disso, referendamos a utilização do modelo multinível nesse trabalho.

1. METODOLOGIA

1.1. Modelo Multinível e seus Aspectos Teóricos

Em trabalhos recentes, nas muitas áreas das ciências sociais, uma grande atenção tem sido dada às relações encontradas entre os indivíduos e as características das instituições/grupos às quais eles pertencem. Encontrou-se nesses trabalhos a necessidade de se levar em consideração nas análises desenvolvidas, os efeitos das relações existentes entre os indivíduos e as peculiaridades do ambiente e do universo onde eles vivem. Para se verificar como esses aspectos específicos do ambiente afetam o comportamento e o desenvolvimento dos indivíduos a ela pertencente, há a necessidade de se utilizar uma metodologia que permita a identificação da amplitude desses impactos resultantes das características do contexto ao qual o indivíduo está incluído em relação aos seus aspectos individuais.

A utilização dos modelos multiníveis de regressão para dados educacionais torna possível a determinação e a investigação empírica de questões importantes e

relevantes da Educação como: o que determina a eficácia do Ensino Superior? Quais são os fatores individuais e institucionais que interferem e determinam os resultados obtidos pelos estudantes na avaliação do ENADE? Dado a natureza dos dados selecionados da educação, em que alunos estão agrupados em turmas e as turmas em Instituições, as Instituições em Regiões, e assim por diante, a utilização de modelos clássicos de regressão pode trazer resultados equivocados. A existência de correlação entre indivíduos pertencente a um mesmo grupo, como no caso dos dados educacionais, as estimativas dos erros-padrão dos parâmetros estimados pela análise da regressão clássica através dos Mínimos Quadrados Ordinários, produzem uma estimativa do erro padrão inconsistente. As estimativas produzidas pelos modelos de regressão multinível são geralmente mais conservadoras. O modelo utilizado nesse trabalho incorpora naturalmente na sua construção a estrutura hierárquica dos dados na regressão, resultando dessa maneira, em estatísticas observacionais e inferências mais confiáveis. Além disso, o modelo permite que se isolem os efeitos das variáveis do modelo segundo seu agrupamento, considerando a estrutura aninhada das observações. Um modelo de análise que não reconheça essa estrutura de agrupamento das variáveis tornará os resultados viesados e com probabilidade de problemas de especificação do modelo.

1.2. O modelo de Regressão Hierárquico Linear

Para melhor entendermos a estrutura e a forma de como se dá a estimação da regressão utilizando as técnicas do modelo multinível, consideremos as seguintes informações para se criar um modelo simplificado:

- Dois Níveis – Formandos e Instituições;
- O que se pretende explicar é o desempenho dos formandos através de um conjunto de variáveis explicativas: uma variável medida no nível do formando e outra variável medida no nível da instituição;

- Os formandos são identificados pelo índice i (que varia de 1 a n_k , sendo n_k o número de formandos pertencentes à instituição k), e as instituições identificadas pelo índice k (que varia de 1 a K).
- A proficiência do formando i da Instituição k é a variável explicada ou dependente do modelo, e a variável X_{ik} é a variável explicativa (independente) associada ao formando i da instituição k .

Dado as informações acima, o modelo de regressão clássico especifica a seguinte relação às variáveis:

$$Y_{ik} = \beta_0 + \beta_1 X_{ik} + \pi_{ik}$$

Onde:

Y_{ik} É a proficiência do formando i da instituição k ;

β_0 É o intercepto e pode ser entendido como o valor esperado da proficiência do formando no caso da variável X_{ik} ser nulo;

β_1 São os coeficientes de inclinação e representam o impacto da variável X_{ik} na proficiência do formando. Para cada acréscimo de uma unidade de X_{ik} , a média do desempenho do aluno irá variar β_1 unidades;

π_{ik} É o erro do modelo. São os efeitos individuais que são explicados por outras variáveis não incluídas no modelo. O erro possui distribuição normal com média zero, variância σ_π^2 , constante e homogênea entre os grupos e não correlacionado, isto é, $\pi_{ik} \sim N(0, \sigma_\pi^2)$.

O modelo de dois níveis considera formando como nível 1 e instituição como nível 2. Para facilitar a compreensão, consideremos a relação entre uma variável explicativa, X_{ik} , onde i representa o formando e k a Instituição a que pertence, e uma variável dependente, Y_{ik} . Vamos supor que a variável X_{ik} represente a situação sócio econômica de cada formando i , e a variável Y_{ik} a proficiência, obtida como o resultado da avaliação do ENADE em uma amostra de K Instituições. O modelo de dois níveis pode ser representado da seguinte maneira, a

partir da elaboração de um modelo dito incondicional, ou modelo nulo, que consiste no modelo em que a regressão estimada contém o menor número possível de regressores.

$$(1) Y_{ik} = \beta_{0k} + \beta_{1k}X_{ik} + \pi_{ik}$$

Onde,

$$(2) \begin{cases} \beta_{0k} = \gamma_{00} + \mu_{0k} \\ \beta_{1k} = \gamma_{10} \end{cases}$$

Reformulando o modelo, e substituindo (2) em (1) temos:

$$(3) Y_{ik} = \gamma_{00} + \gamma_{10}X_{ik} + \pi_{ik} + \mu_{0k}$$

Na equação 3, têm-se as variáveis aleatórias π_{ik} e μ_{0k} , que são mutuamente independentes com distribuição normal e média zero. A variância de π_{ik} é σ_{π}^2 , que representa a variância entre os formandos e σ_{μ}^2 é a variância de μ_{0k} , que representa a variância entre as instituições. Outra hipótese assumida para o modelo multinível é de que os regressores do nível 1 e 2 são não correlacionados com os resíduos e não há correlação entre os termos de resíduo aos dois níveis.

A equação apresentada é um modelo mais simples, de apenas dois níveis, onde os formandos estão agrupados no nível 1 e as instituições no nível 2, representado por apenas um termo aleatório. Entretanto, podemos generalizar para se construir um modelo que considere mais interceptos aleatórios, com mais variáveis explicativas associadas a cada um dos vários níveis. A nova estrutura generalizada fica da seguinte maneira:

$$(4) Y_{ik} = \beta_{0k} + \beta_{1k}X_{1ik} + \dots + \beta_{tk}X_{tik} + \pi_{ik} \text{ - onde } t \text{ são todas as variáveis } X_{ik} \text{ (} t=1, \dots, t \text{)}$$

$$(5) \beta_{hk} = \gamma_{h0} + \gamma_{h1}Z_{1k} + \dots + \gamma_{hq}Z_{qk} + \mu_{hk} \text{ - onde } q \text{ são todas as variáveis } Z_k, \text{ explicativas no nível 2: (} h=1, \dots, h \text{)}$$

O modelo acima incorpora intercepto e parâmetros de inclinação aleatórios para todas as t variáveis X_{ik} , sendo explicados pelas q variáveis Z_k . Substituindo a equação (5) em (4), temos o seguinte modelo generalizado:

$$(6) Y_{ik} = \gamma_{(0)} + \sum_{h=1}^l \gamma_{h0} X_{hik} + \sum_{t=1}^b \gamma_{0t} Z_{ht} + \sum_{j=1}^q \sum_{h=1}^l \gamma_{hj} Z_{jt} X_{hik} + \mu_{(0)} + \sum_{h=1}^l \mu_{hk} X_{hik} + \pi_{ik}$$

As hipóteses básicas para o modelo representado na equação 6, são as seguintes:

- i. $\pi_{ik} \sim N(0, \sigma_{\pi}^2)$: os erros no nível 1 são independentes e normalmente distribuídos com média 0 e variância σ_{π}^2 , para todos os indivíduos do nível 1 dentro de cada unidade do nível 2.
- ii. $\mu_{hk} \sim N(0, \sigma_{\mu}^2)$: os erros no nível 2 possuem distribuição normal com média 0 e variância σ_{μ}^2 ;
- iii. $\text{cov}(X_{ik}, \pi_{ik}) = 0$: os regressores do nível 1 são independentes dos resíduos π_{ik} ;
- iv. $\text{cov}(Z_{ht}, \mu_{hk}) = 0$: os preditores do nível 2 são não correlacionados com os resíduos μ_{hk} ;
- v. $\text{cov}(\pi_{ik}, \mu_{hk}) = 0$: inexistência de correlação entre os termos de resíduos dos dois níveis.

Para estimar os parâmetros do modelo, os parâmetros fixos, γ_{hk} , as variâncias e as covariâncias - $\sigma_{\pi, \mu}^2$, τ_{π}^2 e τ_{μ}^2 - levando em consideração a hipótese de normalidade dos termos π_{ik} e μ_{ik} , existem na literatura algumas alternativas, entre elas:

- ❑ Método da Máxima Verossimilhança;
- ❑ Método da Máxima Verossimilhança Restrita;
- ❑ Método Bayesiano;

As estimativas que serão encontradas nesse trabalho serão produzidas através do procedimento dos Mínimos Quadrados Generalizados Iterativos (MQGI). Dado as hipóteses iniciais do modelo de que os erros têm distribuição normal, os estimadores têm as propriedades estatísticas semelhantes às encontradas pelo método da máxima verossimilhança. O método consiste na estimação dos parâmetros fixos do modelo pelo Método dos Mínimos Quadrados Ordinários. Com os resultados, obtêm-se as estimativas dos resíduos que permitem que se estime a matriz de variâncias e covariâncias do modelo. Com a matriz, estimam-se novamente os parâmetros fixos utilizando o Método dos Mínimos Quadrados Generalizados. Novamente, com os resultados dos resíduos, re-estima a matriz de variância e covariância do modelo. Utiliza-se esse método até que os resultados converjam. Entretanto, todo esse procedimento será realizado pelo software HLM. As referências estatísticas clássicas para estes modelos estão indicados na literatura utilizada nesse trabalho.

Assim como nas análises dos modelos clássicos de regressões, é necessário aplicar testes de especificação no modelo hierárquico linear. O objetivo desses testes é a verificação do ajuste das variáveis a equação estimada. Segundo BRYK & RAUDENBUSH (1992), uma má especificação do modelo pode viesar as estimativas, e conseqüentemente, levar o pesquisador a inferências e análise estatísticas das variáveis equivocadas.

A aplicação do teste de especificação do modelo consiste em utilizar o mesmo conjunto de regressores nas equações de todos os níveis que se está estimando. Para BRYK & RAUDENBUSH (1992) esse procedimento torna as inferências dos coeficientes de segundo nível independentes dos coeficientes estimados no modelo nulo.

A do teste t é utilizado para a escolha dos regressores, através da análise investigativa de regressão simples dos resíduos *bayesianos* provenientes do segundo nível da regressão, (nesse trabalho estamos trabalhando com apenas dois níveis, mas é necessária a análise dos resíduos no terceiro nível quando for o caso).

Ainda, o modelo hierárquico assume a hipótese de homogeneidade da variância no nível 1 e de normalidade dos resíduos. Caso essa hipótese de homogeneidade seja violada, as estimativas serão ineficientes e os desvios padrão

do segundo nível se tornarão viesados. A propriedade de normalidade no nível 1 é relevante na especificação dos níveis inferiores. A violação dessa hipótese interfere nos cálculos dos intervalos de confiança e conseqüentemente, pode tornar as estatísticas testes inválidas, já que estes interferem nos testes de hipóteses. Como esse trabalho não tem o foco na demonstração dos cálculos, mas sim a compreensão dos resultados, indicamos o artigo de BRYN & RAUDENBUSH, intitulado "*Hierarchical linear models: applications and data analysis methods*" (1992) para melhor entender os procedimentos dos testes de homogeneidade de variância no nível 1 e da suposição da normalidade dos resíduos, além de uma demonstração mais detalhada dos modelos multinível.

CAPÍTULO II

O objetivo desse capítulo é apresentar os dados do ENADE e fazer uma análise descritiva das variáveis utilizadas nesse trabalho. Inicialmente detalhamos o procedimento utilizado pelo INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira na construção da base de dados, além de mostrar a amostra e a população pertinente ao trabalho. Em seguida, identificamos as variáveis no nível 1, dos estudantes, e do nível 2, das instituições.

2. OS DADOS DO ENADE E A DEFINIÇÃO DAS VARIÁVEIS DO MODELO

2.1. Os Dados do ENADE e do Censo da Educação Superior

Os dados no ENADE (Exame Nacional de Desempenho de Estudantes), parte integrante do Sistema Nacional de Avaliação da Educação Superior (SINAES) utilizados no trabalho referem-se ao exame aplicado no ano de 2005. Segundo o INEP (Instituto Nacional de Estudos e Pesquisas Educacionais), autarquia vinculada ao Ministério da Educação, o ENADE tem por objetivo “... aferir o desempenho dos estudantes em relação aos conteúdos programáticos previstos nas diretrizes curriculares dos respectivos cursos de graduação, suas habilidades... e suas competências...”.

Como parte integrante da avaliação do sistema, o ENADE é complementado com 105 perguntas de um questionário sócio-econômico para os estudantes, um questionário para os coordenadores dos cursos, as questões de avaliação da prova e os dados do Censo da Educação Superior.

O exame é aplicado periodicamente, através de amostras de estudantes de diversas áreas do conhecimento. Os estudantes selecionados são os que ingressaram no curso e os que estão se formando. O resultado da avaliação é ordenado em uma escala de cinco níveis.

No plano de amostragem do ENADE 2005 foram contempladas às áreas de: Arquitetura e Urbanismo, Biologia, Ciências Sociais, Computação, Engenharia,

Filosofia, Física, Geografia, História, Letras, Matemática, Pedagogia e Química, totalizando 20 grupos distintos. Foram geradas diferentes distribuições de amostragem. Os parâmetros variáveis na simulação foram os tamanhos da amostra e os erros de amostragem. O esquema adotado foi a amostragem estratificada com seleção aleatória simples em cada estrato. Os cursos correspondem aos estratos, e os alunos, às unidades de seleção.

No exame de 2005, participaram da avaliação 5.511 cursos que foram agrupados em 20 áreas de conhecimento, num total de 277.476 estudantes, dentre eles 147.600 ingressantes e 129.876 concluintes, em um universo de 525.685 estudantes, 292.008 ingressantes e 233.677 concluintes, o que mostra que 52,8% do total de estudantes dos cursos selecionados foram avaliados. Neste trabalho, o que se procura avaliar é o desempenho dos estudantes formandos, dos cursos de Pedagogia, Matemática e Letras. Nesses, 891 cursos de Pedagogia, 714 de Letras e 457 de Matemática foram avaliados. Nos cursos, o total de alunos participantes do exame foi de 58.292 estudantes concluintes, sendo 22.870 de Letras, 9.243 de Matemática e de 26.179 formandos em Pedagogia. A Tabela 2.1 mostra o plano amostrai deste trabalho.

TABELA 2.1 – AMOSTRA E POPULAÇÃO DO ENADE 2005

Área	Amostra			População		
	Ingressante	Concluinte	TOTAL	Ingressante	Concluinte	TOTAL
Letras	23.330	22.870	46.200	39.271	36.368	75.639
Matemática	10.229	9.243	19.472	19.006	13.581	32.587
Pedagogia	23.318	26.179	49.497	59.243	67.890	127.133
TOTAL	56.877	58.292	115.169	117.520	117.839	235.359

FONTE: MEC/INEP/SINAES

NOTA: Números extraídos do Banco de Dados do ENADE 2005

É importante destacar, que durante a análise dos dados, muitos dos estudantes inscritos não compareceram a prova, não responderam ou deixaram de entregar o questionário sobre os dados sócio econômicos. Dessa forma, a amostra da população pode não coincidir com dados utilizados nas regressões, já que aquele

está considerando todos os estudantes selecionados e inscritos para o exame. Quando estamos estimando a regressão, os dados faltantes são excluídos da amostra, podendo nesse caso variar em relação aos dados da tabela 2.1.

O Censo da Educação Superior é um conjunto de informações sobre as instituições de ensino superior, incluindo os cursos de graduação presenciais e à distância. Os últimos dados disponíveis do Censo são de 2005. Entre as informações disponíveis, está a composição do corpo docente, a carga horária dos professores e outras variáveis que tem impacto sobre o desempenho dos estudantes no exame do ENADE. Os dados de 2005 mostram que foram oferecidas 2.435.987 vagas pelas 2.165 Instituições de Ensino Superior, sendo 97 Universidades Federais, 75 Estaduais, 59 Municipais e 1.934 Instituições de Ensino Superior Privado, o que representa 89,3% do total das Entidades.

2.2. Definição das Variáveis do Modelo

Essa parte do trabalho destina-se a apresentação das variáveis utilizadas no modelo, descrevendo e identificando os níveis a que elas pertencem. Os dados foram obtidos das informações fornecidas pelo INEP, através do Banco de Dados do ENADE e do Censo da Educação Superior de 2005. Em particular, utilizamos a amostra sobre o desempenho dos estudantes concluintes dos cursos de Letras, Matemática e Pedagogia, num total de 58.292 estudantes pertencentes a 625 Instituições. Esse banco de dados fornece informações que vão desde o desempenho⁴ dos estudantes, como resultado da nota obtida no exame, até uma série de outras características dos estudantes, dos professores e da instituição na qual estudam. Entre os bancos de dados utilizados, encontram-se as informações referentes ao ENADE e o Censo da Educação Superior 2005, ambos fornecidos pelo INEP.

⁴ A avaliação do desempenho dos estudantes de cada curso que participa do ENADE é expressa por meio de conceitos, ordenados em uma escala de cinco níveis. Neste trabalho, utilizamos a nota geral obtida pelos estudantes, disponível no banco de dados do ENADE. O que nos interessa, é o quanto os “futuros professores” são capacitados, e não o quanto ele evoluiu dentro da instituição.

2.2.1. Variáveis do Primeiro Nível: Indivíduo

A partir das informações originárias no questionário preenchido na ocasião da realização das provas do ENADE e disponível no banco de dados, construímos as seguintes variáveis:

➤ **Nota Geral** – Variável endógena do nosso modelo. É a nota média da formação geral e do componente específico dos alunos concluintes na prova do ENADE⁵.

➤ **Renda** – É a variável da resposta do estudante sobre a faixa de renda mensal de sua família. Essa variável requer que ela seja construída como uma variável binária. Assim, a variável renda será definida da seguinte maneira:

TABELA 3.2 – VARIÁVEL RENDA

Variável	Descrição	Frequência	%
Renda_1	1: Renda até 3 Salários Mínimos 0: Caso Contrário	15.960	34,91
Renda_2	1: Renda de 3 até 10 Salários Mínimos 0: Caso Contrário	24.850	54,35
Renda_3	1: Renda de 11 a 20 salários mínimos 0: Caso Contrário	3.754	8,21
Renda_4	1: Renda de 21 a 30 salários mínimos 0: Caso Contrário	782	1,71
Renda_5	1: Renda acima de 30 salários mínimos 0: Caso Contrário	374	0,82
Total		45.720	100,00

FONTE: INEP

NOTA: Os dados para a construção da tabela foram extraídos do banco de dados do ENADE

A análise dos dados mostra que, do percentual válido, ou seja, descontados os alunos que não responderam a questionário (27,72 % do total),

⁵ A nota publicada para as instituições de Ensino Superior são reflexo da diferença entre as notas obtidas pelos alunos concluintes e ingressantes. Neste trabalho, optamos por utilizar apenas as notas dos concluintes, dado que não desejamos avaliar as instituições de ensino e sim o seu produto final, que em última instância é a qualidade dos professores que colocam no mercado de trabalho.

34,91% dos concluintes responderam que a faixa de renda mensal da família é de até 3 salários-mínimos. Ainda, 54,35% se encontram na cada de 3 a 10 salários, e apenas 0,82% dos estudantes possuem uma renda mensal familiar superior a 30 salários mínimos.

É importante explicar que, dado que o estudante de nível superior tem consciência do nível de renda de sua família, não foi necessário construir uma variável a partir de uma análise fatorial⁶ de várias variáveis para compor o nível sócio econômico desse aluno.

➤ **Sexo** – Variável que associa se o estudante é do sexo masculino ou feminino.

TABELA 3.3 – VARIÁVEL SEXO

Sexo	TOTAL	Percentual	Letras	Matemática	Pedagogia
Masculino	11.287	17.84%	4.213	4.747	2.327
Feminino	51.968	82.16%	20.680	5.563	25.725
Total	63.255	100.00	24.893	10.310	28.052

Fonte: INEP

Os dados mostram que 82,16% dos estudantes concluintes dos cursos de Letras, Matemática e Pedagogia são do sexo feminino, mostrando um interesse maior das mulheres para a Licenciatura e para a profissão, principalmente para o curso de pedagogia, onde existem aproximadamente 11 mulheres formandos para cada homem.

⁶ Esse método foi utilizado em vários trabalhos, como em SOARES (2005), por acreditar que o nível sócio econômico representa um conjunto de informações, e não pode ser apreendido num único item do questionário, mas sim esse conjunto. Neste contexto, insere-se a análise fatorial, a fim de construir uma variável que represente o nível de renda dos alunos. Entretanto, nesse trabalho, optou-se por não utilizar a análise fatorial para a variável renda,

➤ **Escola Pública** - Essa variável quer captar o fato do estudante ter estudado o ensino médio - 2º Grau – em escola pública ou privada. A partir das 5 respostas possíveis dos estudantes, criou-se uma variável binária, com valores que podem ser 0 ou 1, sendo 1 para quem cursou o ensino médio todo ou a maior parte na rede pública de educação ou 0 para o aluno que tenha estudado em escola particular. Os dados revelam que 23,8% dos alunos estudaram em escola particular no ensino médio.

➤ **Raça** – Variável construída a partir da resposta do questionário dos alunos de como o estudante se considera. Variáveis binárias, compostas dos seguintes elementos:

TABELA 3.4 – VARIÁVEL RAÇA

Variável	Descrição	Frequência - %
Branco	1: Se considera Branco 0: Caso Contrário	28.609 – 62,4%
Negro	1: Se considera Negro 0: Caso Contrário	2.813 – 6,13%
Pardo	1: Se considera Pardo 0: Caso Contrário	13.328 – 29%
Oriental	1: Se considera Oriental 0: Caso Contrário	622 – 1,36%
Indígena	1: Se considera Indígena 0: Caso Contrário	510 – 1,11%
Total		45.882 – 100%

Fonte: INEP

Pela tabela, verifica-se que a maioria dos estudantes são brancos, correspondendo a mais de 60% do total da amostra, seguidos dos pardos (29%), negros (6.13%), oriental (1.36%) e por ultimo, os indígenas, que representam 1.11% dos formandos.

➤ **Escolaridade do Pai** – Variável que considera o efeito da escolaridade do pai sobre o rendimento do aluno. Leva em consideração apenas se o pai tem

formação superior ou não. Variável binária descrita como: 1 = pai com formação superior, 0=caso contrário.

TABELA 3.5 – VARIÁVEL ESCOLARIDADE DO PAI

Valor	Descrição	TOTAL	Percentual	Letras	Matemática	Pedagogia
1	Formação Superior	3.984	8.71%	1.935	416	1.633
0	Caso Contrário	41.772	91.29%	18.152	4.345	19.275
Total		45.756	100.00%	20.087	4.761	20.908

Fonte: INEP

É importante verificar que apenas 8,71% dos pais do total dos alunos nos cursos selecionados possuem curso superior.

➤ **Escolaridade da Mãe** – Idêntico a variável da escolaridade do pai. Leva em consideração qual o efeito da escolaridade da mãe (se formação superior ou não) no desempenho do estudante.

TABELA 3.6 – ESCOLARIDADE DA MÃE

Valor	Descrição	TOTAL	Percentual	Letras	Matemática	Pedagogia
1	Formação Superior	4.730	10.30%	2.424	502	1.804
0	Caso Contrário	41.210	89.70%	17.749	4.268	19.193
Total		45.940	100.00%	20.173	4.770	20.997

Fonte: INEP

Dos cursos selecionados e das respostas válidas, 10,30% das mães dos formandos possuem curso superior, conforme informa a tabela 3.6.

➤ **Iniciação Científica e Extensão** - Variável que identifica o efeito do estudante ter participado de algum programa de iniciação científica ou de extensão durante sua formação.

TABELA 3.7 – VARIÁVEL INICIAÇÃO CIENTÍFICA E EXTENSÃO – NÍVEL 1

Programas	Sim – (%)	Não – (%)	TOTAL
Iniciação Científica	28.527 – (62.41%)	17.181 – (37.59%)	45.708
Extensão	27.662 – (60.50%)	18.057 – (39.50%)	45.719

Fonte: INEP

A Tabela 3.7 mostra que aproximadamente 60% dos alunos freqüentou algum programa de iniciação científica ou de extensão.

2.2.2. Variáveis do Segundo Nível: Instituição

➤ **Categoria Administrativa da Instituição:** Essa variável corresponde a categoria da Instituição a qual pertence. É uma variável binária, que pode assumir o valor de 0 ou 1, como mostra a tabela a seguir:

TABELA 3.8 – CATEGORIA ADMINISTRATIVA

Ordem	Categoria	Freqüência	Percentual
1	Federal	9.585	15,2
2	Estadual	12.277	19,4
3	Municipal	2.179	3,4
4	Particular	39.214	62,0
Total		63.255	100,0

Fonte: INEP

Pela tabela, é possível notar que as instituições particulares são responsáveis por 62% dos formandos nas carreiras analisadas. Essa variável vai captar o efeito que a categoria de Instituição tem sobre o rendimento do aluno. Quer

saber e quantificar o ganho ou a perda que o formando irá ter por estudar em uma Instituição Federal, Estadual, Municipal ou Particular.

➤ **Estrutura Física** – A influência das instalações das Instituições utilizadas pelos alunos durante o curso. Essa variável foi construída através da análise fatorial – ver anexo – por acreditar que nas respostas do questionário do ENADE 2005, não há apenas uma única variável que represente a situação da infraestrutura da Instituição, mas sim um conjunto de informações que juntas, fornecem com maior precisão o nível dessa variável. O valor da variável se encontra no intervalo de 1,5 à -3,20.

➤ **Técnicas de Ensino** – Variável no nível da instituição que quer quantificar o efeito das técnicas de ensino utilizadas pelos professores no desempenho dos estudantes. As técnicas de ensino aqui estudadas referem-se ao procedimento que a maioria dos professores adota predominantemente nas aulas. Entre as metodologias adotadas, estão as aulas expositivas, aulas práticas, trabalhos em grupo desenvolvidos em sala de aulas, aulas expositivas com participação dos estudantes e outras técnicas não incluídas.

A tabela a seguir mostra a frequência das técnicas utilizadas predominantemente nas instituições.

TABELA 3.9 – VARIÁVEL TÉCNICAS DE ENSINO

Técnica de Ensino	Variável	TOTAL	Percentual	Letras	Matemática	Pedagogia
Aulas Expositivas	Tec_Ens1	5.278	11.52%	2.544	1.520	1.214
AE – Participação dos Estudantes	Tec_Ens2	26.835	58.54%	12.356	2.240	12.239
Aulas Práticas	Tec_Ens3	1.271	2.77%	620	196	455
Trabalhos de Grupo	Tec_Ens4	11.440	24.96%	4.233	629	6.578
Outras	Tec_Ens5	1.010	2.21%	390	170	450
Total		45.834	100.00%	20.143	4.755	20.936

Fonte: INEP

Pelas respostas dos alunos ao questionário, percebe-se que a maioria dos professores, cerca de 60%, utiliza-se das aulas expositivas com participação dos estudantes como metodologia e técnica de ensino. O que se pretende verificar com essa variável é qual o ganho ou a perda (na média) que o estudante formando tem, ao ter aprendido através dos procedimentos utilizados predominantemente pelos seus professores.

Essa variável foi construída como uma variável binária, sendo 1 para cada técnica utilizada, e 0 quando a resposta for negativa.

➤ **Instrumentos de Avaliação** - Essa variável tem o mesmo sentido da variável “**Técnicas de Ensino**”. Pretende-se verificar qual o instrumento de avaliação utilizado predominantemente pelos professores da instituição ao qual o formando pertence que tem um maior impacto positivo sobre o desempenho do aluno. Entre esses instrumentos, estão:

- Provas Escritas Discursivas
- Testes Objetivos
- Trabalhos de Grupo
- Trabalhos Individuais
- Provas Práticas.

TABELA 3.10 – VARIÁVEL INSTRUMENTOS DE AVALIAÇÃO

Instrumento Avaliação	Variável	Freqüência	Percentual
Provas Escritas Discursivas	Ins_Ava1	23.822	52,14%
Testes Objetivos	Ins_Ava2	1.958	4,28%
Trabalhos de Grupo	Ins_Ava3	16.283	35,63%
Trabalhos Individuais	Ins_Ava4	2.010	4,40%
Provas Práticas	Ins_Ava5	1.619	3,55%
Total		45.692	100,00%

Fonte: INEP

Os instrumentos de avaliação amplamente utilizados pelos professores são as Provas Escritas Discursivas e os Trabalhos em Grupo, correspondendo juntos a 87,77% do total. O menos utilizado pelo corpo docente são as provas escritas, com apenas 3,55%.

➤ **Iniciação Científica** - Essa variável tem por objetivo captar o efeito que o oferecimento de programas de iniciação científica tem sobre o desempenho dos estudantes. Quer se verificar se existe algum diferencial entre as instituições que oferecem esses programas e as que não possuem projetos iniciação científica.

➤ **Extensão** – Idem a variável “**Iniciação Científica**”.

TABELA 3.11 – VARIÁVEL INICIAÇÃO CIENTÍFICA E EXTENSÃO – NÍVEL 2

Programas	Sím – (%)	Não – (%)	TOTAL
Iniciação Científica	503 – (80,22%)	124 – (19,78%)	627
Extensão	459 – (73,21%)	168 – (26,79%)	627

Fonte: INEP

Pela tabela, verifica-se que aproximadamente 80% das instituições analisadas oferecem programas de Iniciação Científica e Projetos de Pesquisa e Extensão.

➤ **Composição do Corpo Docente** - Essa variável representa a formação do corpo docente da instituição. Verifica se o fato de uma instituição possuir dentro do quadro funcional, uma maior parcela de professores doutores, por exemplo, tem algum efeito sobre o desempenho do estudante.

TABELA 3.12 – COMPOSIÇÃO DO CORPO DOCENTE

Docentes	TOTAL	Percentual
Sem Graduação	27	0,00%
Graduação	31.524	13,39%
Especialização	65.400	27,78%
Mestrado	86.922	36,92%
Doutorado	51.566	21,91%
Total	235.439	100,00%

Fonte: INEP

Pela tabela 3.12 verifica-se que 21,91% dos professores das instituições de nível superior possuem curso de doutorado. Dos 235.439, apenas 13,39% deles possuem apenas curso de graduação, enquanto que 36,92% cursaram o mestrado e 65.400 deles cursaram algum curso de especialização.

Os dados apresentados na tabela referem-se a todas as instituições brasileiras que participaram do Censo da Educação Superior. Não há como separar os professores da instituição e identificar a composição do corpo docente para todos os cursos selecionado no trabalho. Na regressão, os valores variam de 0 a 1, pois foram utilizados o percentual de cada categoria de docente.

CAPÍTULO III

O foco deste capítulo consiste na elaboração do modelo hierárquico linear utilizando os dados dos formandos dos cursos de Letras, Matemática e Pedagogia com o intuito de verificar os principais determinantes da proficiência dos alunos no Exame Nacional de Desempenho dos Estudantes no ano de 2005, e identificar em que variáveis a qualidade na formação de futuros professores é evidenciada.

3. RESULTADOS

3.1. Modelo Incondicional

Antes de analisar as estimativas, é importante lembrar que os resultados foram obtidos através do software HLM versão 6.04. Ainda, o modelo estimado é um modelo cuja estrutura hierárquica possui apenas dois níveis: 1º nível – alunos, e 2º nível – instituições.

Inicialmente, estimaremos o modelo nulo, ou modelo incondicional. A lógica desse primeiro modelo na análise hierárquica é estimá-lo sem incorporar qualquer variável explicativa, além disso, esse primeiro modelo fornece informações relevantes ao considerar a variabilidade da proficiência dentro dos dois níveis. O modelo incondicional tem a seguinte especificação:

$$(7) \quad NT_GER_{ij} = \beta_{0j} + \pi_{ij}$$

$$(8) \quad \beta_{0j} = \gamma_{00} + \mu_{0j}$$

Onde NT_GER_{ij} é a nota geral obtida no ENADE pelo aluno i da instituição j . A variável β_{0j} deve ser interpretada como a média do aluno formando selecionado aleatoriamente em uma instituição também selecionada aleatoriamente. Dito de outra forma, o modelo incondicional está representando o desempenho do

aluno como a soma do desempenho médio da escola, β_{0j} ; e de um resíduo π_{ij} , que depende do estudante e da Instituição de Ensino Superior – IES. Assim como dito no capítulo de apresentação do modelo, pressupõe-se que este resíduo tenha distribuição normal com média zero e variância constante σ^2 . Os resultados do modelo são mostrados na tabela 4.1:

TABELA 4.1 – MODELO INCONDICIONAL

Efeitos Fixos	Coefficiente	Desvio Padrão	p-valor	
β_{0j}	43,42	0,28	0,000	
Efeitos Aleatórios	Variância	G.L	X²	p-valor
Média Instituições μ_{0j}	42,31	623	6.023,49	0,000
Efeito do Aluno π_{ij}	210,78			
Deviance = 372.751,38				

Na tabela 4.1 temos os valores dos efeitos fixos e dos efeitos aleatórios. Os efeitos fixos referem-se às estimativas dos coeficientes das variáveis explicativas, sejam para nível 1 ou 2. Os efeitos aleatórios referem-se aos cálculos da variância e covariância. Em nosso modelo, o valor obtido de 43,42 para o coeficiente β_{0j} sugere que a média da nota geral dos alunos formandos nos Cursos de Letras, Matemática e Pedagogia no exame do ENADE foram de 43,42. É uma nota considerada baixa, já que a escala de nota varia de 0 a 100.

No modelo nulo, ainda verificamos que a estimativa da variância entre os alunos, ou seja, variância no nível 1 foi de 210,78 enquanto que a variância entre as instituições, variância no nível 2 foi de 42,31, ambas estaticamente significativas. Com as variâncias estimadas, pode-se calcular a correlação intra-classe. Esta é a correlação entre os valores da variável NT_GER_{ij} , referentes a dois formandos, retirados aleatoriamente do grupo e também selecionados aleatoriamente entre as instituições. A estatística da correlação intra-classe é calculada conforme equação (9):

$$(9) \rho(NT_GER_{ij}, NT_GER_{i'j'}) = \frac{\tau_2^0}{\tau_0^2 + \sigma^2}$$

A decomposição da variância total, conforme resultados dos efeitos aleatórios da tabela 4.1 mostra que 16,72% da variância total é explicada pelas características relacionadas à instituição, e 83,28% referem-se as características dos formandos. Dessa forma, podemos concluir que existe uma diferença de 83,28% no desempenho de formandos no exame do ENADE 2005 dentro da mesma instituição, ou seja, a grande parte da variância deve-se à variação intra-instituição e não a variação entre - instituições, que correspondem a somente 16,72% do total. Com esses resultados, a utilização dos Modelos Hierárquicos Lineares na estimação da função de produção educacional brasileira para os formandos dos cursos de Letras, Matemática e Pedagogia é recomendável.

A confirmação da significância estatística dos efeitos aleatórios é mostrada pela estatística $\chi^2 = 6.023,49$ com 623 graus de liberdade. A estatística p – valor de 0,000 mostra que a variância no nível da instituição é estatisticamente significativa a menos de 1%. Dessa forma, acredita-se que as características institucionais afetam de forma distinta o desempenho dos estudantes. Com esse resultado, conclui-se pela rejeição da hipótese de que a variância do parâmetro β_{0j} é zero.

Os próximos modelos apresentados buscam entender quais as características dos futuros professores e as institucionais que tornam as IES mais eficazes e equitativas.

3.2. Modelo Condicional

A partir dos resultados preliminares com o modelo incondicional, podemos estimar modelos cujas características dos alunos e as características das instituições são incorporadas a fim de estudar e identificar os determinantes da proficiência no ensino superior.

No primeiro modelo a ser estimado incluímos três variáveis de controle no nível dos formandos: variável renda, sexo e raça.

A variável renda (nível sócio econômico) tem por objetivo quantificar o efeito de se ter maior ou menor renda sobre a nota geral obtida no exame. A inclusão das outras duas variáveis, sexo e raça, controlam possíveis efeitos de gênero e da etnia sobre o desempenho.

Os efeitos fixos e aleatórios para o modelo de verificação da proficiência dos formandos, juntamente com os resultados, encontram-se na tabela 4.2. As equações (10) e (11) representam o modelo estimado.

Primeiro Modelo:

$$(10) \quad NT_GER_{ij} = \beta_{0j} + \beta_{1j}(Sex_masc) + \beta_{2j}(Negro) + \beta_{3j}(Pardo) + \beta_{4j}(Indígena) + \\ + \beta_{5j}(Re_nda_2) + \beta_{6j}(Re_nda_3) + \beta_{7j}(Re_nda_4) + \pi_{ij}$$

$$(11) \quad \begin{aligned} \beta_{0j} &= \gamma_{00} + \mu_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} \\ \beta_{3j} &= \gamma_{30} \\ \beta_{4j} &= \gamma_{40} \\ \beta_{5j} &= \gamma_{50} \\ \beta_{6j} &= \gamma_{60} \\ \beta_{7j} &= \gamma_{70} \end{aligned}$$

Os resultados estão na tabela 4.2:

TABELA 4.2 – MODELO CONDICIONAL 1

Efeitos Fixos	Coeficiente	Desvio Padrão	<i>p</i> -valor	
Instituição	43,65	0,29	0,000	
Sexo.Masculino	-1,89	0,40	0,000	
Negro	-1,21	0,47	0,010	
Pardo	-0,38	0,25	0,037	
Indígena	-1,15	0,99	0,049	
Renda_2	1,25	0,22	0,000	
Renda_3	1,78	0,36	0,000	
Renda_4	3,33	0,73	0,000	
Efeitos Aleatórios	Variância	G.L	X ²	<i>p</i> -valor
Média das Instituições	38,12	623	5.823,50	0,000
Efeito do Aluno	208,26			
Deviance = 372.560,55				

A partir dos resultados da tabela 4.2 verifica-se que a estimativa para o desempenho médio das instituições foi de 43,65. Ao incluir algumas variáveis ao nível do formando, podemos considerar resultados importantes. O primeiro deles refere-se ao sinal negativo do coeficiente **sexo.masculino**. Controlando os demais efeitos individuais, as mulheres têm um resultado superior na média, em 1,89 pontos do que os homens.

Outro ponto a ser destacado, diz respeito às diferenças de etnia nos desempenho dos estudantes. Comparando-se negros, indígenas e pardos com os estudantes brancos, todos eles têm um resultado esperado na média pior, sendo que os formandos considerados negros obtiveram um resultado inferior em 1,21, os indígenas 1,15 e os pardos 0,38. Esse resultado pode servir de alarme para aqueles interessados em reduzir as desigualdades étnicas no acesso à educação superior. Não só é necessário dar mais oportunidades para todos os estudantes, sejam eles da raça que for, mas sim, fazer com que todos tenham desempenhos regulares e equitativos.

O último resultado a ser analisado por esse modelo consiste nas estimativas dos coeficientes da variável renda. Nosso grupo de referência são os formandos com renda mensal familiar de até 3 salários-mínimos. Entre os quatro

níveis de renda analisados, a estatística dos estudantes com nível de renda superior a 30 salários mínimos não se mostrou significativa. Ainda sim, os estudantes cujas famílias possuem uma renda na faixa de 21 a 30 salários mínimos obtêm uma média superior em 3,33 pontos, e os formandos do grupo **Renda_3** (entre 11 a 20 salários mínimos) com 1,78 a mais na média em relação ao nosso grupo de referência.

O segundo e o terceiro modelo condicional será apresentado, incluindo-se nele variáveis que controlam os efeitos institucionais além de outras variáveis no nível do estudante.

Segundo Modelo:

Equação (12):

$$NT_GER_{ij} = \beta_{0j} + \beta_{1j}(Sex_masc) + \beta_{2j}(Negro) + \beta_{3j}(Pardo) + \beta_{4j}(Indígena) + \beta_{5j}(Renda_2) + \beta_{6j}(Renda_3) + \beta_{7j}(Renda_4) + \beta_{8j}(Pub_Ens) + \beta_{9j}(Esc_pai) + \beta_{10j}(Esc_Mãe) + \beta_{11j}(Ini_Cien) + \beta_{12j}(Extensão) + \pi_{ij}$$

Equação (13):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Tec_Ens3) + \gamma_{02}*(Tec_Ens4) + \gamma_{03}*(Tec_Ens5) + \gamma_{04}*(Ins_Ava3) + \gamma_{05}*(Ins_Ava4) + \gamma_{06}*(Ins_Ava5) + \gamma_{07}*(Infra_Estrutura) + \gamma_{08}*(Federal) + \gamma_{09}*(Estadual) + \gamma_{010}*(Municipal) + \gamma_{011}*(Doe_Mest) + \gamma_{012}*(Doe_Dout) + \mu_{ij}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(Tec_Ens3) + \gamma_{22}(Ins_Ava3)$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}(Tec_Ens3) + \gamma_{32}(Ins_Ava3)$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}(Tec_Ens3) + \gamma_{42}(Ins_Ava3)$$

$$\beta_{5j} = \gamma_{50}$$

$$\beta_{6j} = \gamma_{60}$$

$$\beta_{7j} = \gamma_{70}$$

$$\beta_{8j} = \gamma_{80}$$

$$\beta_{9j} = \gamma_{90}$$

$$\beta_{10j} = \gamma_{100}$$

$$\beta_{11j} = \gamma_{110}$$

$$\beta_{12j} = \gamma_{120}$$

Na equação (12), do nível 1, foram acrescentado as variáveis **Pub_Ens**, **Esc_Pai**, **Esc_Mãe**, **Ini_Cien1** e **Extensão1**. No nível 2, representado pela equação (13), foram incorporados as características da Instituição.

Terceiro Modelo:

Equação (14):

$$NT_GER_{ij} = \beta_{0j} + \beta_{1j}(Sex_masc) + \beta_{2j}(Negro) + \beta_{3j}(Pardo) + \beta_{4j}(Indígena) + \beta_{5j}(Re_nda_2) + \beta_{6j}(Re_nda_3) + \beta_{7j}(Re_nda_4) + \beta_{8j}(Pub_Ens) + \beta_{9j}(Esc_pai) + \beta_{10j}(Esc_Mãe) + \beta_{11j}(Ini_Cien) + \beta_{12j}(Extensão) + \pi_{ij}$$

Equação (15):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(Tec_Ens3) + \gamma_{02}*(Tec_Ens4) + \gamma_{03}*(Tec_Ens5) + \gamma_{04}*(Ins_Ava3) + \gamma_{05}*(Ins_Ava4) + \gamma_{06}*(Ins_Ava5) + \gamma_{07}*(Infra_Estrutura) + \gamma_{08}*(Federal) + \gamma_{09}*(Estadual) + \gamma_{010}*(Municipal) + \gamma_{011}*(Doe_Mest) + \gamma_{012}*(Doe_Dout) + \mu_{ij}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40}$$

$$\beta_{5j} = \gamma_{50}$$

$$\beta_{6j} = \gamma_{60}$$

$$\beta_{7j} = \gamma_{70}$$

$$\beta_{8j} = \gamma_{80} + \gamma_{81}(Doc_Mest) + \gamma_{82}(Doc_Dout) + \gamma_{83}(Infra_Estrutura)$$

$$\beta_{9j} = \gamma_{90}$$

$$\beta_{10j} = \gamma_{100}$$

$$\beta_{11j} = \gamma_{110}$$

$$\beta_{12j} = \gamma_{120}$$

Os resultados dos modelos 2 e 3 de verificação de eficácia na formação de futuros professores encontram-se na tabela 4.3.

TABELA 4.3 - MODELOS CONDICIONAIS

continua

Efeitos Fixos		Modelo 1		Modelo 2	
	β_{0j}	Coefficiente	P-valor	Coefficiente	p-valor
Instituição		44,81	0,003	44,35	0,000
Tec_Ens3	γ_{01}	3,49	0,056	3,51	0,041
Tec_Ens4	γ_{02}	2,53	0,042	2,52	0,038
Tec_Ens5	γ_{03}	-1,12	0,015	-1,09	0,009
Ins_Ava3	γ_{04}	2,17	0,001	2,12	0,001
Ins_Ava4	γ_{05}	2,21	0,000	2,17	0,000
Ins_Ava5	γ_{06}	-1,16	0,001	-1,16	0,000
Estrutura	γ_{07}	2,65	0,058	2,71	0,006
Federal	γ_{08}	4,35	0,005	4,36	0,005
Estadual	γ_{09}	4,47	0,011	4,47	0,017
Municipal	γ_{010}	-2,43	0,017	-2,40	0,017
Doc_Mest	γ_{011}	1,11	0,073	1,11	0,071
Doc_Dout	γ_{012}	1,72	0,001	1,73	0,002
Sexo Masculino	β_{1j}, γ_{10}	-1,94	0,000	1,39	0,000
Negro	β_{2j}, γ_{20}			-2,40	0,000
Intercpto	γ_{20}	-2,01	0,000		
Tec_Ens3	γ_{21}	0,12	0,064		
Ins_Ava3	γ_{22}	1,42	0,044		
Pardo	β_{3j}, γ_{30}			-0,71	0,008
Intercpto	γ_{30}	-0,64	0,000		
Tec_Ens3	γ_{31}	0,19	0,010		
Ins_Ava3	γ_{32}	0,22	0,420		
Indígena	β_{4j}, γ_{40}			-2,40	0,000
Intercpto	γ_{40}	-2,42	0,050		
Tec_Ens3	γ_{41}	0,67	0,051		

TABELA 4.3 - MODELOS CONDICIONAIS

conclusão

Efeitos Fixos		MODELO 2		MODELO 3	
		Coefficiente	p-valor	Coefficiente	p-valor
Ins_Ava3	γ_{42}	1,12	0,083		
Renda_2	β_{5j}, γ_{50}	1,31	0,000	1,51	0,000
Renda_3	β_{6j}, γ_{60}	1,85	0,000	2,08	0,000
Renda_4	β_{7j}, γ_{70}	3,26	0,000	3,07	0,000
Pub_Ens	β_{8j}, γ_{80}	2,32	0,000		
Intercepto	γ_{80}			1,99	0,009
Doc_Mest	γ_{81}			0,03	0,091
Doc_Dout	γ_{82}			0,07	0,088
Estrutura	γ_{83}			1,10	0,019
Esc_Pai	β_{9j}, γ_{90}	1,01	0,023	1,14	0,017
Esc_Mãe	$\beta_{10j}, \gamma_{100}$	0,89	0,007	1,08	0,068
Ini_Cien	$\beta_{11j}, \gamma_{110}$	1,53	0,059	1,52	0,054
Extensão	$\beta_{12j}, \gamma_{120}$	0,61	0,007	-0,25	0,001
Efeitos Aleatórios		Variância	GL	X ²	p-valor
Modelo 2					
Média das Instituições		19,72	547	3418,85	0,000
Efeito do Aluno		198,68			
Deviance = 322.501,027					
Modelo 3					
Média das Instituições		19,81	547	3436,84	0,000
Efeito do Aluno		198,65			
Deviance = 322.475,76					

Os resultados relacionados à etnia mostraram-se estaticamente significativos para os dois modelos, e devem ser analisados comparativamente. No segundo modelo, levou-se em consideração o fato do estudante ter estudado em instituição onde a maioria dos professores utilizou como forma predominante de técnica de ensino as aulas práticas, concatenadas com os trabalhos em grupo como método de avaliação do aprendizado. Caso essa condição seja satisfeita, o impacto da etnia do estudante sobre a desempenho médio do aluno no exame é menor, ou

seja, um estudante negro tem um resultado médio esperado inferior ao grupo de referência (branco) em 2,4 pontos, (se todos os outros coeficientes forem iguais a zero), mas se ele estudou numa IES onde as técnicas de ensino e instrumentos de avaliação empregados pelo corpo docente potencializam o resultado dos seus alunos, o desempenho médio esperado é inferior em apenas 0,47 pontos, reduzindo-se em aproximadamente 80% o diferencial atribuído ao fato do estudante ser negro. O mesmo raciocínio deve ser feito para os estudantes considerados indígenas e pardos. A tabela mostra que as três diferentes etnias analisadas mostraram redução do impacto no desempenho esperado, evidenciando que os professores das IES devem utilizar essas metodologias de ensinos para reduzir o diferencial relacionado a diferentes raças dos estudantes brasileiros.

A variável relacionada ao sexo mostrou que as estudantes têm um desempenho esperado superior em 1,94 pontos em relação aos estudantes homens. Em relação à condição sócio econômica, assim como encontrado por DIAZ (2006), foi observado um padrão distinto para os coeficientes da variável renda. A análise mostra que os estudantes na faixa de renda de 3 até 30 salários mínimos, ou seja, as variáveis **Renda_2**, **Renda_3** e **Renda_4**, obtiveram um desempenho superior ao grupo de referência, quer seja, estudantes com nível de renda de até 3 salários mínimos. Outro ponto a ser destacado na análise é resultado verificado para os alunos com renda familiar superior a 30 salários mínimos. O coeficiente não se mostrou significativo, evidenciando que existe um limite de renda que permite que o estudante formando obtenha um resultado esperado estatisticamente significativo melhor do que o grupo de referência. Nesse trabalho, poderíamos dizer, vulgarmente, que o fato do estudante ser muito rico não tem qualquer efeito sobre a nota obtida no exame. Entretanto, se o estudante pertencer à classe média, verificasse que ele tem um rendimento superior em 10% em relação ao estudante considerado de classe baixa.

Os coeficientes das variáveis **Esc_Pai**, **Esc_Mae**, e **Pub_Ens** devem ser analisados conjuntamente. Os resultados obtidos se mostraram estatisticamente significativos á 10%. A tabela mostra que o estudante formando com pai e mãe com ensino superior tem um acréscimo na media de 1,90 pontos em relação aos alunos que disseram que os pais não possuem ensino superior. Esse coeficiente mostra a

evidência de que o ambiente familiar ao qual o aluno está inserido produz efeitos positivos sobre o seu desempenho no exame, indicando que os alunos que estão se formando, adicionado o curso superior dos pais, tendem a apresentar um rendimento relativamente melhor do que os formandos que apresentam um histórico de educação não superior dos seus pais.

Entre as análises até aqui feitas, um resultado interessante, foi o coeficiente positivo da variável **Pub_Ens**, que tem como objetivo captar o efeito do formando ter estudado o ensino médio em escola pública ou privada. Neste trabalho, o resultado mostrou que o aluno que estudou o ensino médio em escola pública tem um desempenho consideravelmente superior a aqueles alunos que estudaram em escola privada. Essa diferença foi de 2.32 pontos. O resultado corrobora com **DIAZ (2006)** evidenciando inicialmente que o resultado negativo para a escola privada indicaria que ao estudar o segundo grau nessas escolas os alunos não têm garantia de melhor desempenho, pelo menos ao final do curso superior. Ainda, ela diz: "... muito mais provável, parece ser hipótese de existência de um problema de seleção, na medida em que, aqueles alunos oriundos de uma escola secundária pública que lograram concluir o ensino superior seriam os mais competentes e habilidosos." Entretanto, esses resultados devem ser analisados com cautela, na medida em que os problemas de seleção dos estudantes não foram adequadamente tratados. O outro exercício feito para a variável **Pub_Ens** está mostrada no terceiro modelo, onde incorporamos a variável **Doc_Mest**, **Doc_Dout** e **Infra-Estrutura**. Nesse modelo, verificamos que o aluno que frequentou escola pública no ensino médio não se comporta de maneira aleatória entre as instituições. Os resultados mostraram que o estudante tem o efeito potencializado quando este estuda numa IES que possui o corpo docente altamente especializado (doutores e mestres) e numa instituição onde a infra-estrutura é de qualidade. Esse resultado permite entender que o aluno de escola pública, mais esforçado e dedicado, é beneficiado ao final do seu curso, pois a qualidade na formação dos futuros professores é evidenciada com esses alunos.

Em relação às práticas didáticas dos docentes, analisando as técnicas de ensino utilizadas em salas de aula e os métodos de avaliação empregados durante o curso, verificou-se que o professor que se utiliza de trabalhos em grupo e aulas práticas para ensinar os alunos, tem efeito positivo sobre o resultado dos formandos

no ENADE. No que diz respeito aos instrumentos de avaliação, os trabalhos em grupo e os trabalhos individuais também influenciam positivamente o desempenho dos formandos dos cursos selecionados. Um resultado que contradiz com a prática de ensino, é o resultado negativo de -1.16 pontos para os estudantes que tiveram seus professores utilizando como forma predominante de avaliação a prova prática como meio de averiguar o aprendizado do aluno. Diz que esse resultado se contradiz, pois como técnica de ensino, as aulas práticas mostraram efeito positivo sobre o desempenho, ao passo que a prova pratica se mostrou “ineficiente” como instrumento de avaliação, resultado esse mostrado pelo coeficiente negativo da variável. É também um resultado para ser analisado delicadamente, pois nos cursos selecionados, (Letras, Matemática e Pedagogia) tem uma boa parte da grade curricular voltada à licenciatura, o que faz com que as “Aulas Práticas” sejam constantes no decorrer desses cursos, não sendo uma metodologia propriamente dita de ensino, mas sim, uma obrigatoriedade para a formação superior. É importante destacar, que não somente melhores práticas de docência aumentam o desempenho esperado de seus alunos, mas também, diminuem as diferenças relacionadas à etnia, ao dar mais igualdade de aprendizado aos alunos, como mostrado nos resultados do segundo modelo.

A possibilidade da existência de endogeneidade que DIAZ (2006) destaca ao analisar os efeitos da participação de alunos nas atividades de pesquisa e extensão, e de engajamento nos programas de iniciação científica, avaliada pela possibilidade de haver características não observáveis entre os dois grupos de alunos que poderiam explicar as diferenças no desempenho dos estudantes, ao invés do efeito casual da participação desses alunos nos programas e projetos de extensão, fez com que esse trabalho analisasse o efeito também pelo nível institucional, tentando verificar se há diferenças no desempenho dos estudantes das instituições que oferecem programas de iniciação científica e atividades de pesquisa/extensão, em contraposição as instituições que não oferecem esses itens. Entretanto, os resultados obtidos dos coeficientes das variáveis **Ini_Cien2** e **Extensão2**, além de ter um efeito pequeno sobre o desempenho do formando,

mostrou-se não significativo a 10%⁷. A conclusão que se extrai desse resultado, e que se parece com a realidade, é o fato de que estudar ou não em uma universidade que ofereça programas de iniciação científica ou projetos de pesquisa/extensão não têm efeito qualquer sobre o desempenho médio do estudante. O que realmente conta e afeta o resultado do aluno, é o fato do estudante participar ou não das atividades que a instituição possa oferecer, resultado esse já encontrado em trabalhos anteriores, exposto mesmo no começo desta análise. Para a comprovação do mencionado, pela tabela, verifica-se nos futuros professores que durante a sua formação, freqüentaram programas de iniciação científica ou projetos de pesquisa e extensão obtiveram um resultado esperado na média melhor em 2,14 pontos. Esse resultado, como já encontrados em outros trabalhos, abre uma discussão importante. Em próximos trabalhos, seria interessante verificar como que o desenvolvimento dessas atividades por parte dos estudantes reduz o impacto positivo de maior renda média familiar, ou dito de outra forma, como as atividades de pesquisa/extensão e iniciação científica geram o efeito atenuador sobre o impacto negativo resultante da condição sócio-econômica desfavorável para os alunos mais carentes. Não somente o nível sócio econômico, mas outras características que afetam negativamente o rendimento médio dos alunos, de forma a preparar não somente professores mais “iguais”, mas melhorar o rendimento de todos, na medida em que o sistema educacional brasileiro de ensino superior tem formado professores cujas médias não ultrapassam a 50% da nota, ou seja, estamos formando professores “REPROVADOS”.

Ainda, a variável que representa o efeito de melhor infra-estrutura da IES no desempenho médio do aluno mostrou-se estaticamente significativa a 5%. A análise revela que alunos que estudam em instituições que oferecem melhores condições de infra-estrutura (melhores equipamentos, melhores bibliotecas, acervos atualizados, entre outros) apresentam um resultado esperado médio superior em 0,85 pontos. Apesar de o acréscimo ser marginal isoladamente, essa variável potencializa o resultado esperado médio do estudante que provêm de escola pública, como mostrada em análise anterior.

⁷ Os dados foram retirados da Tabela 4.3 pois numericamente, não acrescentam de forma significativa a análise.

Igualmente aos resultados encontrados nos trabalhos de análise da educação primária e de nível médio, a variável que capta o efeito da infra-estrutura da instituição sobre o rendimento dos alunos mostrou significativa. Logo, podemos dizer que existe correlação entre o desempenho do estudante e a infra-estrutura da instituição a qual pertence. O melhor ambiente estrutural pode influenciar positivamente a performance do aluno da 4ª Serie em aproximadamente 15% em relação ao aluno que estuda numa escola onde a estrutura é considerada precária, conforme resultado obtido por FRANÇA (2006). Entretanto, para o nível secundário, analisado através do exame do SAEB, o mesmo autor encontra na variável de infra-estrutura um redutor na nota do aluno. Pondera o resultado indicando que apesar dos investimentos em infra-estrutura de equipamento diminuir a performance do estudante na escola, a segurança do ambiente, e melhoria do acesso do estudante a escola, acesso esse refletido no transporte, contribui para um melhor resultado do aluno do ensino médio. O resultado encontrado para o ensino superior, apesar do efeito positivo reduzido sobre o rendimento do formando, deve estar relacionado mais a capacidade e a dedicação individual de cada estudante para uma melhor nota no ENADE, e a infra-estrutura não representa uma variável de proporções significantes para explicar as diferenças de notas entre os alunos de diferentes instituições de ensino superior brasileiro. O acesso a biblioteca, a disponibilização de periódicos, instalações para a leitura e outras variáveis relacionadas à infra-estrutura não pode influenciar o aluno marginalmente, mas a utilização dessas ferramentas fornecidas pela IES é que faz com que o aluno alcance um resultado esperado melhor em comparação com os outros estudantes.

As variáveis referentes à categoria administrativa das instituições apresentaram coeficientes positivos e negativos. Essa variável teve o objetivo de captar o efeito de se estudar em uma instituição pública ou privada. Ainda, as instituições públicas foram divididas em Federal, Estadual e Municipal. A tabela 4.3 mostra que os alunos das instituições estaduais e federais têm seus resultados aumentado em aproximadamente 4 pontos a mais na média que os alunos das instituições particulares, conforme mostrados pelos modelos 2 e 3. Em seguida, verifica-se que as IES municipais reduzem em 2.4 pontos a nota média dos seus alunos. Esses resultados poderiam estar refletindo as diferenças no gerenciamento

de recursos, mascarados por outras variáveis. A partir desse princípio, poderíamos dizer que as faculdades administradas pelos governos estaduais são mais eficazes no que diz respeito ao modo como elas estão gerindo os recursos disponíveis, ao passo que as instituições vinculadas aos municípios, estão direcionando de maneira errônea, se pensarmos que as diferenças que esses coeficientes estão mostrando estão diretamente relacionados a administração de recursos, sejam eles para remunerar os professores, para melhorar o ambiente de trabalho, condicionado ao investimento na capacitação do corpo docente da instituição. Esse ponto de vista é baseado trabalho de DIAZ (2006), que encontrou evidência dos impactos efetivos positivos do corpo docente e de suas características sobre o rendimento dos alunos dos cursos analisados. Acredita-se que não somente a composição do corpo docente (quantidade de mestre e doutores da instituição), mas também melhores condições de trabalho, refletidas na redução de carga horária trabalhada, salários melhores, políticas de valorização e capacitação dos professores são variáveis responsáveis pelo melhor desempenho dos estudantes. Esse resultado é evidenciado pelas variáveis que captam os efeitos da qualificação dos professores sobre os alunos. A tabela mostra que nas instituições onde o quadro do corpo docente é na sua maioria composta de professores doutores e mestres, elevam a proficiência dos estudantes formandos, enquanto que não existe nenhuma correlação dos professores graduados e especialistas com o resultado alcançado pelos alunos.

É importante notar ainda que a variância entre as instituições (nível 2) reduziu-se no Terceiro Modelo (19,81) comparativamente ao Modelo Nulo (42,31). Assim, controlado os aspectos individuais e institucionais, a correlação existente entre dois alunos da mesma instituição, ou seja, a correlação intra-classe reduziu-se para 9,07%. De outra forma, dizemos que ao incluir as variáveis no terceiro modelo, explicamos 45,75% da variância existente entre as instituições. Utilizamos o terceiro modelo como referência, pois foi este que apresentou a menor **deviance**⁸ entre os modelos, mostrando que é o melhor modelo especificado.

⁸ Segundo FRANÇA (2006) a máxima verossimilhança produz a estatística de *deviance* que corresponde ao grau de ajustamento das observações em relação a reta de regressão. O significado da *deviance* é que quanto menor o valor correspondente, maior o ajustamento dos dados.

Um último exercício para incrementar a análise dos dados é feita a partir dos coeficientes estimados e do desvio padrão da variável dependente do modelo. Esse exercício é feito dividindo-se o coeficiente estimado pelo desvio padrão da variável explicada do modelo original, ou seja, da amostra utilizada. Essa estatística se distribui como uma normal padronizada, e podemos verificar para qual percentil o estudante se desloca, satisfeita a hipótese do coeficiente analisado. Assim, a Tabela 4.4 mostra as estatísticas encontradas para os coeficiente do terceiro modelo.

TABELA 4.4 - COEFICIENTE : DESVIO PADRÃO

Variável	Coeficiente	Devio Padrão Y	Estatística	Normalizada	Percentil
Sexo Masculino	-1,94	17,97	-0,1080	-0,044	0,456
Negro	-2,4	17,97	-0,1336	-0,052	0,448
Pardo	-0,71	17,97	-0,0395	-0,052	0,448
Indígena	-2,4	17,97	-0,1336	-0,056	0,444
Renda_2	1,51	17,97	0,0840	0,032	0,532
Renda_3	2,08	17,97	0,1157	0,044	0,544
Renda_4	3,07	17,97	0,1708	0,071	0,571
Pub_Ens					
Intercepto	1,99	17,97	0,1107	0,044	0,544
* Doc_Mest	0,03	17,97	0,0017	0,000	0,500
* Doc_Dout	0,07	17,97	0,0039	0,000	0,500
* Infra_Estrutura	1,1	17,97	0,0612	0,024	0,524
Esc_Pai	1,14	17,97	0,0634	0,024	0,524
Esc_Mãe	1,08	17,97	0,0601	0,024	0,524
Ini_Cient1	1,52	17,97	0,0846	0,032	0,532
Extensão1	0,61	17,97	0,0339	0,012	0,512
Doc_Mest	1,11	17,97	0,0618	0,024	0,524
Doc_Dout	1,73	17,97	0,0963	0,036	0,536
Tec_Ens3	3,51	17,97	0,1953	0,079	0,579
Tec_Ens4	2,52	17,97	0,1402	0,056	0,556
Tec_Ens5	-1,09	17,97	-0,0607	-0,024	0,476
Ins_Ava3	2,12	17,97	0,1180	0,048	0,548
Ins_Ava4	2,17	17,97	0,1208	0,048	0,548
Ins_Ava5	-1,16	17,97	-0,0646	-0,024	0,476
Infra_Estrutura	2,71	17,97	0,1508	0,060	0,560
Federal	4,36	17,97	0,2426	0,095	0,595
Estadual	4,47	17,97	0,2448	0,095	0,595
Municipal	-2,4	17,97	0,1352	-0,064	0,436

Fonte: Elaboração Própria

A análise que deve ser feita, consiste em verificar para qual percentil desloca o estudante na presença de determinada característica. Pela tabela, vemos que o estudante do sexo masculino se encontra no 45,60% percentil da distribuição normal, o que quer dizer, que o estudante formando masculino está 0,6% abaixo da média nacional. No que se refere às etnias analisadas, o negros possuem um desempenho 5,2% abaixo da nota média, e assim por diante é feita as análises

pelos percentis demonstrados na tabela. Entre os resultados que mais se destacaram, encontram-se as variáveis relacionadas ao nível sócio econômico. Os formandos com renda entre 20 e 30 salários mínimos apresentam um resultado 7,1% acima da média. Em relação a categoria administrativa da instituição, as estaduais e federais foram as que mais apresentaram resultados relativamente melhores, com um desempenho em aproximadamente 10% maior.

4. CONSIDERAÇÕES FINAIS

O objetivo desse trabalho foi modelar e estimar uma função de produção educacional que auxiliasse na compreensão dos determinantes quantitativos do desempenho educacional dos estudantes formandos dos cursos de Letras, Matemática e Pedagogia. A qualidade educacional do ensino superior nesse trabalho foi entendida como o desempenho dos alunos no Exame Nacional de Avaliação de Desempenho – ENADE no ano de 2005. A demonstração do conhecimento ou as habilidades cognitivas apresentadas através do desempenho do estudante torna-se uma variável quantitativa da qualidade da educação.

Assim, a utilização do modelo hierárquico linear foi aplicada aos dados do ENADE 2005 a fim de verificar quais são as variáveis que afetam o desempenho dos futuros professores. Os estimadores multinível, como também são conhecidos, é apropriado para modelos com variáveis aleatórias, obtidas a partir de um plano amostral distribuído não identicamente e independente, mas sim aninhados de forma não aleatórias em subgrupos relevantes. Os resultados mostram que a proficiência dos alunos é compreendida em função de dois níveis distintos: as características individuais e familiares dos formandos, e as características da instituição e do corpo docente que o aluno está inserido. Dito isso, utiliza-se a proficiência do aluno como variável dependente do primeiro nível, e suas características familiares e individuais como variáveis independentes. O segundo nível tem como variável explicada os coeficientes estimados no primeiro nível e como variáveis explicativas, as características das instituições e dos professores.

Nesse sentido, o segundo capítulo do trabalho se propôs a apresentar o mecanismo metodológico para o modelo multinível. A natureza dos dados, onde os alunos estão agrupados em turmas, as turmas em instituições, as instituições em estados, e assim por diante, estão estruturados de forma hierárquica, de modo que os modelos de regressão clássicos se tornam inconsistentes na estimação dos parâmetros do modelo. O modelo multinível incorpora naturalmente a estrutura do agrupamento dos dados, permitindo uma melhor análise e compreensão dos resultados obtidos.

O terceiro capítulo apresentou os dados do ENADE, identificando as variáveis e o nível de cada uma delas. Ainda, no mesmo capítulo analisamos descritivamente as variáveis, mostrando dados importantes do ensino superior brasileiro.

No quarto e último capítulo do trabalho verificamos os resultados obtidos através do software HLM. Foram estimados três modelos diferentes, além do modelo nulo, que é parte do procedimento metodológico da análise.

Inicialmente, a partir do modelo nulo estimado, verificou-se que a Nota Geral Média dos alunos formandos nos cursos selecionados foi de 43.42 pontos, numa nota cuja escala varia de 0 a 100. A decomposição da variância mostrou que 16,72% da variância total é explicada por características relacionadas à Instituição, enquanto que as características individuais e familiares correspondem por 83,28% da variabilidade na proficiência dos estudantes. Com esses dados, referendamos a utilização do modelo hierárquico para os dados educacionais relacionados ao Exame Nacional de Avaliação de Desempenho. ALBERNAZ (2002) mostra que essa variância inter-instituição é de aproximadamente 28% para os alunos da 4ª Série. Ou seja, a variância inter-instituição é mais evidenciada em alunos do ensino fundamental. Dito de outra forma, as características ligadas a escola afetam mais o desempenho desses alunos em comparado com os efeitos das IES sobre os futuros professores.

Os dois últimos modelos estimados apresentaram resultados importantes. Ao incluir variáveis explicativas no nível 1 e 2, verificou-se que a condição econômica dos alunos formandos é muito importante como determinante do seu rendimento. O fato relevante dessa análise, é mostrado pelo padrão não regular dos efeitos positivos dos três extratos de níveis de renda do modelo. O grupo de referência da análise são os alunos que possuem uma renda familiar mensal de até 3 salários mínimos. Esse padrão é visto através das diferenças quantitativas dos efeitos dos três níveis de renda, sendo que os estudantes pertencentes ao quarto grupo analisado, ou seja, os formandos com renda familiar mensal entre 21 e 30 salários mínimos tem o maior efeito positivo. O ponto que merece destaque está ligado ao grupo com renda familiar mensal superior a 30 salários mínimos. A estimativa do coeficiente não se mostrou significativa, evidenciando que não existe

uma relação robusta com a proficiência desse estudante. A conclusão, pelo menos a nível estatístico, é de que os estudantes das classes média e média alta, predominantemente os estudantes do quarto grupo, são os alunos que mais se beneficiam do nível sócio econômico familiar. Em suma, o que parece existir é um limite do efeito propulsor do nível de renda sobre o desempenho dos estudantes.

Além disso, foi possível verificar que o ambiente em que o formando vive contribui de maneira significativa para o seu rendimento no exame. Esse efeito foi captado também através das variáveis relacionadas à escolaridade dos pais, assim como dos níveis de renda. Os alunos cujos pais possuem ensino superior tendem a obter aproximadamente 2,00 pontos na média a mais do que os alunos com pais sem formação universitária, ou seja, a qualidade do ensino é evidenciada pelos formandos filhos de pais com diploma superior. Esse ponto da análise pode ser explicado pelo fato de que o ambiente familiar criado pelos pais com formação superior poderia desenvolver o progresso intelectual dos filhos. Esse ambiente seria favorecido pelo acesso a computadores, livros, informação, enfim, cria-se um padrão de vida que é mais comum por classes de nível sócio econômico mais elevado. Essa melhor renda familiar, propiciada pela escolaridade superior dos pais, já que este é um determinante daquele, favorece o rendimento do aluno.

Ainda no nível 1, os resultados indicam que todas as raças analisadas em comparação com os estudantes brancos têm em média um desempenho esperado inferior na avaliação do INEP, com exceção do oriental que não mostrou coeficiente significativo. Os homens também tendem a obter uma média 1.88 pontos a menos que as mulheres. Se for homem e de origem negro, por exemplo, existe uma redução de rendimento esperado de mais de 4,34 pontos (modelo 3). Das características individuais com efeito positivo importante no desempenho do aluno, além do nível de renda, os formandos que freqüentaram o ensino médio em escola pública apresentam uma performance melhor 2,00 pontos na média do que os formandos vindos de escolas particulares. Merece um parêntese na análise, o fato de se ter encontrado nas aulas práticas como técnicas de ensino e trabalhos em grupo como instrumentos de avaliação um atenuador no efeito negativo relacionado à etnia. Dessa forma, as IES cujo corpo docente utilizada como ferramenta de ensino e avaliação aquilo que diminui o diferencial entre as diferentes raças de

estudantes contribui para a formação de futuros professores com mais equidade, levando-se em conta apenas as características étnicas. Não podemos confundir nesse ponto equidade com qualidade, já que mesmo assim, os resultados obtidos na avaliação do ENADE pelos estudantes foram insuficientes, ao não atingir pelo menos 50% da nota da prova.

No terceiro modelo, verifica-se que a composição do corpo docente e a infra-estrutura da IES impulsionam-potencializam de forma positiva o resultado esperado do aluno que estudou em escola pública.

As análises do nível institucional, revelam efeitos sobre o desempenho dos alunos oriundos das técnicas de ensino e dos instrumentos de avaliação adotados pelo corpo docente da instituição. O efeito positivo é proveniente de aulas praticas e trabalhos em grupo como forma predominante de ensino utilizado pelos professores, além de trabalhos individuais e em grupos como ferramenta de averiguação do aprendizado obtido pelos alunos. Ainda, não foi possível comprovar a existência de algum impacto em decorrência do formando estudar em instituições que ofereçam programas de iniciação científica e projetos de pesquisa e extensão. Entretanto, o resultado encontrado ao nível do estudante evidencia que o efeito é do aluno, ou seja, o estudante que desenvolveu alguma atividade de pesquisa durante a sua formação ou participou de algum projeto de iniciação científica tem na média um desempenho esperado melhor em comparação com outros estudantes.

A variável infra-estrutura mostrou significativa nos modelos estimados. O coeficiente mostrou que uma melhor estrutura da IES aumenta em aproximadamente 3% a nota média esperada dos alunos selecionados.

Verificou-se uma diferença positiva entre o desempenho dos alunos de instituições públicas federais e estaduais comparativamente ao verificado nas instituições privadas, quando controlado os demais aspectos. Entretanto, os estudantes das instituições públicas municipais apresentaram um efeito negativo de 2,40 em relação às instituições particulares.

As estimativas encontradas neste trabalho levaram em consideração apenas um modelo com dois níveis. Entretanto, para futuros trabalhos acerca da Educação que utilizam a mesma metodologia adotada por este, seria interessante

controlar um terceiro nível por estados, tentando verificar se existe alguma variabilidade na nota esperada do ENADE por regiões. Ainda, não foi possível distinguir nos cursos selecionados quais são os alunos de licenciatura dos bacharelados. No banco de dados fornecido pelo INEP não é possível fazer essa diferenciação. Por último, estimamos o modelo com os três cursos conjuntamente. Uma análise separada de cada curso pode ser o foco de um próximo trabalho.

Neste trabalho, procurou-se explorar os dados gerados pelo Exame Nacional de Avaliação de Desempenho, com o objetivo de identificar as variáveis que influenciam a proficiência dos futuros professores da rede de ensino brasileiras. Não foi possível incluir todas as variáveis desejadas no modelo, mas, mesmo assim, verificamos que ainda existem diferenças significativas na qualidade da educação no ensino superior. Parece conclusivo, que para melhorar o rendimento dos estudantes do ensino fundamental e médio, é necessário pensar em melhorias no rendimento dos alunos dos cursos destinado a licenciatura, como o caso dos cursos de Letras, Matemática e Pedagogia. Os modelos mostraram que ainda existem diferenças entre de resultados dos alunos de mesmas IES, mas esses podem ser contornados quando dentro da instituição, há políticas afirmativas que beneficiem alunos esforçados e interessados na carreira de magistério. É de se esperar, portanto, com a conclusão desse trabalho, que melhorias na infra-estrutura das instituições, ligados com práticas didáticas mais eficazes são variáveis que contribuem para a queda do diferencial existente entre alunos, sejam eles relacionados à renda mensal familiar ou a etnia. As políticas públicas educacionais não devem apenas se preocupar com a variância intra-instituições, fazendo com que as IES sejam mais equitativas, mas sim, melhorar em muito a formação de futuros professores, já que estes ainda estão aquém do esperado para que estejam capacitados a ensinar,

5. REFERÊNCIAS BIBLIOGRÁFICAS

ALBERNAZ, A., FERREIRA, F. & FRANCO, C. (2002): *Qualidade e equidade na educação fundamental brasileira*. **Pesquisa e Planejamento Econômico**, v.32, n.3, dez. 2002.

BARRETO, M. L., DOURADO, M. I. C., FERREIRA, L. D. A., OLIVEIRA, N. F. & SANTOS, C. A. S. (2000): *Modelagem Multinível*. **Sitientibus**, n.22, p.89-98, jan./jun.2000

BARBOSA, M.E.F. & FERNANDES, C. (2001): "A escola Brasileira faz diferença? Uma investigação dos efeitos da escola na proficiência em Matemática dos alunos da 4ª Série." **Programas de Seminários Acadêmicos**, nº 09.2001. Instituto de Pesquisas Econômicas/USP.

BARROS, R. P., HENRIQUE, R. & MENDONÇA, R. (2002): *Inequality and poverty in Brazil: Portrait of the unacceptable stability*. *Brazilian Review Of Social Science*, v.2, 2002.

BELTRÃO, K. I. , FERRÃO, M. E. & SANTOS, D. P. (2002): *Modelo de Regressão Multinível: Aplicação ao Estudo do Impacto da Política de Não-Repetência no Desempenho Escolar dos Alunos da 4ª Série*. **Pesquisa e Planejamento Econômico**, v. 32, nº 03, dez. 2002

BRYK, S. & RAUDEMESH, W. (1986): *A Hierarchical model for studying school effects*. **Sociology of education**, v.59: pp.1-17.

BRYK, S. & RAUDEMESH, W. (1992): *Hierarchical linear models: application and data analysis methods*. London: **Sage publication**. 265p.

DIAZ, M. D. M. (2006): *Efetividade no Ensino Superior Brasileiro: Aplicação de Modelos Multinível á Análise dos Resultados do Exame Nacional de Cursos*. Artigo aprovado para Publicação em Economia – Revista da ANPEC. Disponível < <http://www.anpec.org.br/revista/aprovados.htm>>

FRANÇA, M. T. A.: **Transmissão de Desigualdade Intergeracional e Qualidade Educacional: Uma Investigação Microeconômica a partir do SAEB 2003**. Curitiba, 2006. 136 f. Dissertação (Mestre em Desenvolvimento Econômico) – Setor de Ciências Sociais e Aplicadas, Universidade Federal do Paraná.

GOLDSTEIN H. (1995): *Multilevel Statistical Models*. 2ª edition. **Kendall's Library of Statistics** 3. Edward Arnold.

HAIR, J. F., ANDERSON, R. E., TATHAM, R. L. & BLACK, W. C.: **Análise multivariada de dados**. 5 Ed. Porto Alegre: Bookman, 2005.

HAIJ, Z.,S.: **Evolução e Desempenho dos Bancos durante o plano real**. São Paulo, 2005. Dissertação – Departametro de Contabilidade e Atuária Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo.

MACHADO, A. M., MARTINS, L., MORO, S. & RIOS, J.: "*Qualidade do ensino em Matemática: determinantes do desempenho de alunos em escolas públicas estaduais mineiras*". Disponível < <http://www.anpec.org.br/encontro2006/artigos/A06A091.pdf> > Acesso: 10 de out. 2007.

SOARES, T.M. (2005): *Modelo de Três Níveis Hierárquicos para a Proficiência dos Alunos de 4ª Série Avaliados no Teste de Língua Portuguesa do Simave/Proeb-2002*. **Revista Brasileira de Educação**, n. 029, p. 73-87.

SOARES, T.M. & MENDONÇA, M.C.M. (2003): *Construção de um modelo de Regressão Hierárquico para os Dados do Simave-2000*. **Pesquisa Operacional**, v. 23, n. 3, p.421-441

VAN der HULST, M. & JANSEN E. (2002) *Effects of Curriculum Organisation on Study Progress in Engineering Studies*. **Higher Educacion**, v. 43, p.489-506

Microdados do ENADE 2005 – *Exame Nacional de Avaliação do Desempenho do Estudante* – (2006): Brasil. INEP.

ANEXO I: Análise Fatorial

Os estudos que utilizam uma função de produção educacional para qualificar e quantificar as variáveis que influenciam os resultados de estudantes em exames como ENADE, para alunos do ensino superior, ou o SAEB⁹, para alunos do ensino fundamental e médio, demonstram através da análise fatorial, que não há apenas um único item a qual pode explicar as características da infra-estrutura da entidade ou mesmo o nível sócio econômico do aluno. Nos dados do SAEB, por exemplo, através do seu questionário, não se encontra uma única pergunta que possa refletir perfeitamente o nível sócio econômico, mas sim, há questões sobre a quantidade de televisão em cores, automóveis, computadores e outras que em conjunto, compõe a situação econômica de cada aluno.

A análise fatorial é uma classe de métodos de análises estatísticas multivariadas¹⁰ cujo principal objetivo é definir a estrutura subjacente em uma matriz de dados. Em suma, análise fatorial aborda o problema de analisar a estrutura das correlações, ou inter-relações, entre um grande número de variáveis (como por exemplo, a composição do nível sócio econômico do aluno através do banco de dados do SAEB), definindo um conjunto de fatores comuns. Entende-se que o fator é uma variável que explica as variáveis originais. Segundo Soares (2006, p.68), "... é a combinação linear das variáveis originais e representa os constructos que resumem ou explicam o conjunto original de variáveis observadas.". Ainda, segundo Hajj (2005, p.93):

Análise fatorial é um técnica multivariada de interdependência em que todas as variáveis são simultaneamente consideradas, cada uma relacionada com as demais, a fim de estudar as inter-relações existentes entre elas, buscando a redução dos dados. Ela pode ser utilizada para se examinar as variáveis latentes em função dos relacionamentos entre as variáveis manifestas.

⁹ SAEB - Sistema de Avaliação da Educação Básica é um exame que se propõe a analisar o desempenho dos alunos em algumas séries: 4ª e 8ª Série do Ensino Fundamental e 3º do Ensino Médio – realizado pelo INEP.

¹⁰ A Análise Multivariada corresponde a um grande número de métodos e técnicas que utilizam simultaneamente todas as variáveis na interpretação teórica do conjunto de dados obtidos.

Com a análise fatorial, podem-se descrever os dados em um número menor de conceitos do que as variáveis individuais originais e melhor interpretar os dados, além de facilitar a compreensão do fenômeno ao qual se está estudando. O objetivo é resumir um grande número de variáveis iniciais num diminuto número de variáveis de entendimento dos dados.

Para a realização da análise fatorial, existem basicamente quatro passos fundamentais: Entrada de Dados, Cálculo das Correlações, Extração de Fatores iniciais e a rotação.

A etapa da entrada de dados consiste basicamente em identificar uma similaridade entre as variáveis para que estas sejam passíveis de análise fatorial. A segunda etapa consiste no cálculo da matriz de correlação das variáveis para verificação do seu grau de associação. Aqui, o que se quer verificar, é a adequação da aplicação da análise fatorial para as variáveis em questão. A primeira análise dessa etapa consiste em se observar o KMO – *Kaiser-Meyer-Olkin*, onde se verifica se a correlação de cada par de variáveis pode ser explicada pelas demais variáveis incluídas na análise. Seu valor pode variar de 0 a 1, onde se pode analisar da seguinte maneira:

- KMO entre 0,9 – 1,0: Grau de ajuste à Análise Fatorial Ótima;
- KMO entre 0,8 – 0,9: Grau de ajuste a Análise Fatorial Boa;
- KMO entre 0,7 – 0,8: Grau de Ajuste a Análise Fatorial Médio;
- KMO entre 0,6 – 0,7: Grau de Ajuste a Análise Fatorial Razoável;
- KMO entre 0,5 – 0,6: Grau de Ajuste à Análise Fatorial Ruim;
- KMO < 0,5: Grau de Ajuste à Análise Fatorial Inaceitável.

O teste de esfericidade BTS ou *Barlett's Test os Sphericity* é baseado na distribuição estatística de “*qui-quadrado*”, e testa se a matriz de correlação é idêntica à matriz identidade, procurando identificar a correlação existente entre as variáveis. Em outras palavras, verifica a significância da correlação. A hipótese nula para o teste é de que a matriz correlação é uma matriz identidade, isto é, de que não há

correlação existente entre as variáveis. Se o nível de significância for pequeno (procura-se uma nível de significância $< 0,05$), a aplicação da Análise Fatorial é adequada.

A terceira etapa da análise fatorial descreve-se como a extração dos fatores mais significativos que representarão os dados, por meio do método adequado. O que se pretende, é saber o quanto o modelo representa os dados originais. O método utilizado nesse trabalho, e amplamente utilizado em outros trabalhos de assunto correlato, é o das componentes principais. O objetivo dessa etapa é encontrar um conjunto de fatores que formem uma combinação linear das variáveis originais. Assim, se as n variáveis são correlacionadas entre si, elas serão reformuladas e combinadas para formar um novo fator, uma nova variável. Podemos definir a combinação linear entre as variáveis da seguinte maneira:

$$Y_j = C_{1j}X_1 + C_{2j}X_2 + \dots + C_{nj}X_n$$

Onde Y_j é uma combinação linear das variáveis $X_1, X_2, \dots, X_n$ e é denominado o componente principal. Segundo HAIR (2005, p.89):

“O método das componentes principais para a análise fatorial envolve a procura de um conjunto de valores de linear C_{nj} que forme uma combinação linear que explique mais a variância da matriz de correlação que qualquer outro conjunto de valores para C_{nj} . É chamado de primeiro fator principal. A seguir, a variância explicada pelo primeiro fator é subtraída da matriz de correlações original, resultando-se assim, as matrizes residuais, adotando-se o mesmo procedimento anterior, obtém-se o segundo fator principal, e assim, sucessivamente todos os fatores principais, até que uma variância muito pequena permaneça sem explicação. A natureza deste procedimento permite extrair fatores que não são correlacionados ou que tenham correlação muito pequenas uns com os outros...”

A quarta etapa para da análise fatorial baseia-se num procedimento chamado rotação da matriz. Para auxiliar na interpretação dos fatores extraídos na

etapa anterior, a solução inicial deve ser rotada. Existem duas formas de procedimento, sendo uma delas a rotação **varimax**¹¹, que mantém os valores não correlacionados, e a rotação oblíqua, que torna as variáveis correlacionadas entre si. A idéia do procedimento é identificar as variáveis que possuem alta correlação e outras com baixa correlação.

Nesse trabalho, utilizou-se a análise fatorial para se construir a variável correspondente a infra-estrutura das instituições de ensino superior. Do questionário do ENADE, seis questões referem-se a infra estrutura da instituição. São elas:

1 – Como são as instalações físicas (salas de aula, laboratórios, ambientes de trabalho/estudo) utilizadas no seu curso?

2 – Como são os equipamentos do laboratório utilizados pelo seu curso?

3 – Como você avalia o acervo da biblioteca, quanto à atualização, face às necessidades curriculares do seu curso?

4 – Com relação aos livros mais usados no curso, o numero de exemplares disponíveis na biblioteca atenda ao alunado?

5 – Como você avalia o acervo de periódicos científico-acadêmicos disponíveis na biblioteca, quanto a atualização?

6 – Como são as instalações de leitura e estudo?

A partir do SPSS, obtemos os seguintes resultados:

¹¹ A partir do sistema SPSS é feito o cálculo dessa rotação, e a complexidade do procedimento do cálculo não é o foco deste trabalho.

TABELA A.1 – ANÁLISE FATORIAL DA VARIÁVEL INFRA- ESTRUTURA

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		,842
Bartlett's Test Of Sphericity	Approx. Chi-Square	79366,312
	Df	15
	Sig.	,000

Ao analisar a medida de adequacidade da amostra através do índice de Kaiser-Meyer-Olkin (KMO), verifica-se que a análise fatorial é adequada. O valor obtido de 0,842 é satisfatório, e pode-se prosseguir com a análise.

O teste de esfericidade de Barlett dá um valor de $\chi^2 = 79366,312$ com 15 graus de liberdade. Dessa forma, verifica-se que $\chi^2 > 0,995$, logo, rejeita-se a hipótese nula de que a matriz de correlação é uma matriz identidade, ou seja, as variáveis são correlacionadas.

Em seguida, os quadros mostram o total da variância original das variáveis explicadas após a extração da primeira componente principal. O autovalor é a medida que avalia a contribuição do fator ao modelo construído pela Análise Fatorial, sendo que um pequeno valor, sugere pequena contribuição do fator na explicação das variações das variáveis originais.

TABELA A.2 - EXTRAÇÃO

Component	Initial Eigenvalues			Extraction Sums Ops Square Loadings		
	Total	% Of Variance	Cumulative %	Total	% Of Variance	Cumulative %
1	3,065	51,085	51,085	3,065	51,085	51,085
2	0,820	13,667	64,752			
3	0,701	11,691	76,443			
4	0,561	9,358	85,801			
5	0,507	8,452	94,253			
6	0,345	5,747	100,000			

Pela tabela, extraiu-se apenas o primeiro componente principal, que explica 51,085% da variância total das 6 variáveis.

A tabela a seguir, mostra as matrizes de comunalidades ou cargas fatoriais. A comunalidade é a medida de quanto da variância de uma variável é explicada pelo

fator derivado pela Análise Fatorial. Ela avalia o peso de uma variável original com todas as outras variáveis incluídas no modelo.

TABELA A.3 – MATRIZ DO COMPONENTE PRINCIPAL

Variável	Componente
	1
Acervo da Biblioteca	0,816
Quantidade de Livro Disponível	0,798
Instalação para Leitura	0,728
Disponibilidade de Periódicos	0,705
Instalações Físicas	0,609
Equipamento de Laboratório	0,603

Na interpretação dos fatores, grande importância deve ser dada as cargas fatoriais. O valor da carga ao quadrado representa a quantidade de variância total da variável expressa pelo fator. Assim, uma carga de 0,816 significa que 66,58% da variância original da variável é explicada pelo fator observado. De forma que, a construção da variável infra estrutura da instituição para este trabalho é: $Infra_{ij} = 0,816 \text{ (Acervo)} + 0,798 \text{ (Livro Disponível)} + 0,728 \text{ (Instalações de Leitura)} + 0,705 \text{ (Disponibilidade de Periódicos)} + 0,609 \text{ (Instalações Físicas)} + 0,603 \text{ (Equipamentos de Laboratório)}$.

ANEXO II: Teste de Normalidade dos Resíduos

Verificou-se a hipótese de normalidade dos resíduos através do exame do gráfico qui-quadrado. Ressaltamos que o gráfico foi obtido através do software SPSS.

GRÁFICO A.II.1: Q-Q (QUI-QUADRADO) DOS RESÍDUOS

