

**UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE CIÊNCIAS NATURAIS E EXATAS  
CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA E MODELAGEM  
QUANTITATIVA**

**AVALIAÇÃO DE FATORES DE RISCO  
RELACIONADOS À OBTENÇÃO DA META DO IDEB**

**MONOGRAFIA DE ESPECIALIZAÇÃO**

**Gabriela Granzotto Fillipin**

**Santa Maria, RS, Brasil**

**2015**

# **AVALIAÇÃO DE FATORES DE RISCO RELACIONADOS À OBTENÇÃO DA META DO IDEB**

**Gabriela Granzotto Fillipin**

Monografia apresentada ao Curso de Especialização em Estatística e Modelagem Quantitativa da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Especialista em Estatística e Modelagem Quantitativa**

**Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Luciane Flores Jacobi**

**Santa Maria, RS, Brasil**

**2015**

**Universidade Federal de Santa Maria  
Centro de Ciências da Saúde  
Curso de Especialização em Estatística e Modelagem Quantitativa**

**A Comissão Examinadora, abaixo assinada,  
aprova a Monografia de Especialização**

**AVALIAÇÃO DE FATORES DE RISCO RELACIONADOS À  
OBTENÇÃO DA META DO IDEB**

elaborada por  
**Gabriela Granzotto Phillipin**

como requisito parcial para obtenção do grau de  
**Especialista em Estatística e Modelagem Quantitativa**

COMISSÃO EXAMINADORA:

  
**Luciane Flores Jacobi, Dra.**  
(Presidente/Orientadora)

  
**Roselaine Ruviano Zanini, Dra. (UFSM)**

  
**Anaelma Bragança de Moraes, Dra. (UFSM)**

Santa Maria, 17 de julho de 2015

## AGRADECIMENTOS

Agradeço a Deus primeiramente, pela sabedoria e pela coragem, por me dar forças para suportar as infundáveis madrugadas de estudos, por não me deixar perder a lucidez e tampouco o ânimo, e por me ajudar a perseverar até o fim.

À minha família, pelo carinho, apoio, compreensão, e incentivo, e por acreditarem na minha capacidade e desejarem sempre o melhor.

À minha mãe e meu pai, pelo carinho de sempre, pelo apoio, pelas orações, e por acreditarem no meu potencial.

Ao meu marido Jader, pela paciência, carinho e dedicação em todos os momentos.

À minha amada Lola, pelas incansáveis noites que esteve ao meu lado me fazendo companhia, e por tantas outras vezes em que ficava à porta me esperando. Amor eterno.

Aos amigos, pela paciência e carinho ao saber que estaria ausente em alguns momentos, e por sempre me apoiarem nesta caminhada.

À Prof.<sup>a</sup> Dr.<sup>a</sup> Luciane Flores Jacobi, pela maravilhosa orientação, por toda sua paciência, dedicação, carinho e compreensão, e principalmente por ter contribuído muito com meu crescimento pessoal e profissional.

À Prof.<sup>a</sup> Dr.<sup>a</sup> Roselaine Zanini, por sua dedicação não somente nas disciplinas ministradas, mas também no seu companheirismo, amizade e alegria com que conduzia o curso, e por todas as vezes que nos deu suporte.

À Prof.<sup>a</sup> Anaelena pelas maravilhosas aulas de inferência.

Ao Prof.<sup>o</sup> Augusto, por ouvir nossos desabafos durante as aulas.

Aos demais professores que fizeram parte deste crescimento, por todas as contribuições enriquecedoras.

Aos queridos colegas, “sobreviventes”, pela amizade, carinho, apoio, pelas alegrias e tristezas compartilhadas nesta jornada. Jamais esquecerei pessoas tão queridas.

## RESUMO

Monografia de Especialização  
Curso de Especialização em Estatística e Modelagem Quantitativa  
Universidade Federal de Santa Maria

### **AVALIAÇÃO DE FATORES DE RISCO RELACIONADOS À OBTENÇÃO DA META DO IDEB**

AUTORA: GABRIELA GRANZOTTO FILLIPIN

ORIENTADORA: LUCIANE FLORES JACOBI

Local e data da defesa: Santa Maria, 17 de julho de 2015.

O objetivo deste estudo é identificar quais os fatores que influenciam no baixo rendimento escolar e são refletidos nas notas do Índice de Desenvolvimento da Educação Básica (IDEB). Foram coletados dados das escolas municipais, estaduais e federais, num total de 62 escolas da cidade de Santa Maria, de 2011 e 2013, do sistema de ensino fundamental, disponibilizados no site Qedu. Fez-se uso de variáveis explicativas como localização das escolas, número de funcionários, número de alunos matriculados, número de alunos reprovados, a existência de biblioteca, sala de leitura, laboratório de informática e de ciências, quadra de esportes, refeitório e acessibilidade. Para a análise dos respectivos dados, foi utilizado um modelo de regressão logística múltiplo. A variável dependente para o modelo foi se a escola tinha ou não atingido a meta estabelecida para o IDEB naquele ano. Como o IDEB do ensino fundamental é avaliado em duas etapas: anos iniciais (5º ano) e anos finais (9º ano), obteve-se um total de dois modelos. Os resultados do ajuste do modelo indicam que os preditos significativos ( $p \leq 0,05$ ) para o 5º ano de 2011 foram o número de alunos reprovados (OR = 1,063) e sala de leitura (OR = 8,286), enquanto que em 2013 as variáveis inclusas foram laboratório de ciências (OR = 0,159) e número de alunos reprovados (OR = 1,079). Para o 9º ano, o ajuste do modelo não apresentou nenhuma variável significativa no ano de 2011 e também de 2013, não gerando portanto um modelo. Destaca-se a variável número de reprovações que se fez presente em todos os modelos ajustados, sendo que sua contribuição é de forma negativa, pois quanto maior o seu valor, menor a chance de se atingir a meta do IDEB. Logo, já que ela contribui significativamente para o alcance da meta, deve-se ter maior atenção com esta variável. Com este estudo, pode-se identificar algumas das variáveis que influenciam significativamente no resultado das notas do IDEB, podendo usar estes resultados como auxílio às escolas para a tomada de decisões e para o planejamento das políticas públicas a fim de melhorar a educação e elevar o índice de qualidade.

**Palavras-chave:** IDEB; Regressão Logística; *Odds Ratio*. Qualidade na Educação.

## **ABSTRACT**

Specialization Monograph  
Specialization Course in Statistics and Quantitative Modeling  
Federal University of Santa Maria

### **EVALUATION OF RISK FACTORS RELATED TO THE ATTAINMENT OF THE IDEB TARGET**

ADVISOR: GABRIELA GRANZOTTO FILLIPIN

ADVISER: LUCIANE FLORES JACOBI

Defense: July 17<sup>th</sup>, 2015. Santa Maria

The objective of this study is to identify what factors influence the poor academic performance and are reflected in the scores of the Basic Education Development Index (IDEB). Data were collected from municipal, state and federal schools from the elementary school system, a total of 62 schools in Santa Maria, from 2011 to 2013, available in the Qedu website. Explanatory variables were used, such as location of schools, number of employees, number of students enrolled, number of students who failed, the existence of a library, reading room, computer lab and science lab, sports court, cafeteria and accessibility. For the analysis the data, a multiple logistic regression model was used. The dependent variable for the model was whether school had or had not reached the target set for IDEB that year. As the IDEB elementary school is evaluated in two stages, initial years (5th year) and final years (9th year), a total of two models was obtained. The results of the model adjustment indicate that the significant predictors ( $p \leq 0.05$ ) for the 5th year of 2011 were the number of students who failed (OR = 1.063) and reading room (OR = 8,286), while in 2013 the included variables were science laboratory (OR = 0.159) and number of students who failed (OR = 1.079). For the 9th year, the model adjustment did not show any significant variable neither in 2011 nor in 2013, therefore not generating a model. As for the year 2013, the only variable included was the number of students who failed (OR = 1.079). The variable number of failures, which was present in all the adjusted models, stands out due to its negative contribution, because the higher its value, the smaller the chance of attaining the IDEB target. Thus, since it significantly influences the achievement of the target, this variable should receive more attention. With this study, some of the variables that significantly influence the results of IDEB scores can be identified, which can be used as an aid to schools for decision making and planning of public policies to improve education and raise the quality score.

**Keywords:** IDEB, Logistic Regression, *Odds Ratio*, Quality in Education.

## LISTA DE FIGURAS E QUADROS

Figura 1 - Gráfico comparativo entre o número de escolas que atingiram ou não a meta do IDEB estabelecida, por série e ano .....	29
Quadro 1 - Distribuição do número de escolas da Educação Básica do município de Santa Maria, por nível e sistema de ensino .....	25

## LISTA DE TABELAS

Tabela 1 - Relação da estrutura física das escolas, por série, em função da conquista ou não da meta estabelecida do Ideb no ano de 2011 .....	30
Tabela 2 - Relação da estrutura física das escolas, por série, em função da conquista ou não da meta estabelecida do Ideb no ano de 2013 .....	31
Tabela 3 - Resultados do ajuste da regressão logística univariada do modelo para a meta M5° do ano de 2011 .....	33
Tabela 4 - Resultados do ajuste da regressão logística univariada do modelo para a meta M9° do ano de 2011 .....	33
Tabela 5 - Resultados do ajuste da regressão logística univariada do modelo para a meta M5° do ano de 2013 .....	34
Tabela 6 - Resultados do ajuste da regressão logística univariada do modelo para a meta M9° do ano de 2013 .....	34
Tabela 7 - Resultados do ajuste da regressão logística multivariada do modelo para a meta M5° do ano de 2011 .....	35
Tabela 8 - Resultados do ajuste da regressão logística multivariada do modelo para a meta M9° do ano de 2011 .....	35
Tabela 9 - Resultados do ajuste da regressão logística multivariada do modelo para a meta M5° do ano de 2013 .....	36
Tabela 10 - Resultados do ajuste da regressão logística multivariada do modelo para a meta M9° do ano de 2013 .....	37

## LISTA DE ABREVIATURAS E SIGLAS

A. Matriculados	Número de alunos matriculados
A. Reprovados	Número de alunos reprovados
Enem	Exame Nacional do Ensino Médio
IBGE	Instituto Brasileiro de Geografia e Estatística
IDEB	Índice de Desenvolvimento da Educação Básica
Inep	Instituto Nacional de Estudos e Pesquisas
M5°	Alcançou ou não a meta no 5° ano
M9°	Alcançou ou não a meta no 9° ano
MEC	Ministério da Educação
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
OR	<i>Odds ratio</i> (Razão de chances)
PDE	Plano de Desenvolvimento da Educação
PISA	Programa Internacional de Avaliação dos Estudantes
QEdU	Qualidade Educacional
Saeb	Sistema de Avaliação da Educação Básica
VL	Verossimilhança-log

# SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>10</b>
<b>1.1 Objetivos .....</b>	<b>11</b>
1.1.1 Objetivo geral .....	11
1.1.2 Objetivos específicos .....	11
<b>1.2 Justificativa .....</b>	<b>12</b>
<b>1.3 Estrutura .....</b>	<b>12</b>
<b>2. REVISÃO DE LITERATURA .....</b>	<b>13</b>
2.1 IDEB – Índice de Desenvolvimento da Educação Básica .....	13
2.2 A Regressão Logística .....	17
<b>3. MATERIAIS E MÉTODOS .....</b>	<b>25</b>
3.1 Estrutura do banco de dados .....	26
3.2 Variável dependente .....	27
3.3 Variáveis independentes ou covariáveis .....	27
3.4 Modelo de Regressão Logística .....	27
<b>4. RESULTADOS E DISCUSSÕES .....</b>	<b>29</b>
<b>5. CONSIDERAÇÕES FINAIS .....</b>	<b>40</b>
<b>6. REFERÊNCIAS .....</b>	<b>41</b>

# 1. INTRODUÇÃO

A educação no Brasil vem sendo, frequentemente, discutida por educadores, gestores, formuladores de políticas, especialistas, no que tange à qualidade. No entanto, não há uma definição para qualidade, uma vez que cada pessoa pode adotar critérios diferentes para determiná-la, baseados em suas perspectivas e realidade social na qual está inserida, ou seja, seu conceito é subjetivo.

Os argumentos de Rios (2001) para a qualidade são:

O conceito de qualidade é totalizante, abrangente, multidimensional. É social e historicamente determinado porque emerge em uma realidade específica de um contexto concreto. Portanto, uma análise crítica da qualidade deverá considerar todos esses aspectos, articulando aqueles de ordem técnica e pedagógica aos de caráter político-ideológico (Rios 2001, p.64).

Para os autores Dourado, Oliveira e Santos (2007), a qualidade é repleta de muitos significados e interpretações, ressaltando que, para a qualidade escolar se materializar, não basta direcionar os esforços a um único enfoque, e sim em ações que devem se realizar dentro e fora da escola, enfrentando situações como a violência, abandono, acesso à cultura e aos esportes, saúde e emprego, através de políticas públicas, ou seja, a qualidade educacional está ligada aos fatores sociais, culturais e econômicos da sociedade.

Deste modo, observa-se que a qualidade da educação está ligada a diversas questões que compõem um ambiente escolar, como formação docente, estrutura física, prática pedagógica, gestão escolar, aplicação de recursos, perfil socioeconômico e cultural das famílias, desigualdade social, dentre outros.

Pinto (2008) também salienta outros itens que têm forte impacto na qualidade da educação, como: tamanho das escolas, equipe profissional estável, equipe de direção presente na escola, participação da comunidade escolar na gestão, presença de grêmios e outras instâncias mediadoras, os quais possuem aspectos que devem ser buscados dentro e fora do ambiente escolar, uma vez que a qualidade na educação ultrapassa os intramuros da escola.

Em consequência destas discussões, vem a necessidade de estabelecer critérios e mecanismos para determinar a qualidade do ensino. Os critérios valorizados que norteiam esta qualidade são eficiência, eficácia e produtividade, e os mecanismos utilizados são os indicadores de qualidade, como o Índice de Desenvolvimento da Educação Básica (IDEB) que será objeto de estudo deste trabalho, no âmbito do Ensino Fundamental.

O IDEB, segundo Fernandes (2007), é utilizado para monitorar o desempenho dos alunos e detectar possíveis falhas no sistema de ensino. Este indicador de qualidade é formado a partir do desempenho em exames padronizados e informações de indicadores de fluxo.

Os exames padronizados são a Prova Brasil, Sistema de Avaliação do Ensino Básico (Saeb) e Exame Nacional do Ensino Médio (Enem), que avaliam os estudantes nas etapas iniciais e finais do ensino fundamental, 5º e 9º ano, e 3ª ano do ensino médio.

Entretanto, estes indicadores de qualidade consideram apenas o resultado obtido na aplicação dos exames padronizados, e não o contexto social em que a escola está inserida.

Para Pinto (2008), não existe qualidade na escola se não há recursos mínimos suficientes para atendê-la, como infraestrutura, equipamentos, bibliotecas, professores remunerados, laboratórios de informática, etc.

Estas, dentre tantas outras variáveis, que não são consideradas no momento de fazer a avaliação dos estudantes, interferem no rendimento destes alunos, fazendo com que seus índices de desempenho não tenham os resultados esperados nas notas do IDEB, conforme salienta Freitas (2007): “diante deste quadro, escolher apenas uma variável, desempenho do aluno, para analisar a educação básica brasileira, como o IDEB faz, é certamente temerário em face deste complexo de variáveis.”

Após o levantamento destas questões, decidiu-se realizar este estudo a fim de identificar as variáveis que influenciam nos resultados, mas são deixadas de lado na hora de realizar as avaliações da educação escolar.

## **1.1 Objetivos**

### 1.1.1 Objetivo geral

Este trabalho tem por objetivo identificar quais os fatores que influenciam no baixo rendimento escolar e são refletidos nas notas do IDEB.

### 1.1.2 Objetivos específicos

- Realizar uma análise descritiva das variáveis em estudo;
- Modelar uma regressão logística múltipla para determinar a probabilidade de atingir as metas estabelecidas no IDEB, obtendo-se os fatores de risco a ela associados;
- Identificar variáveis relevantes ao alcance da meta do IDEB, a fim de auxiliar as escolas na tomada de decisões para melhorar o índice de qualidade.

## **1.2 Justificativa**

A qualidade da educação no Brasil é medida por meio de índices de desempenho, como o IDEB, que considera, além da média de desempenho nos testes padronizados, o fluxo escolar.

Além dos índices, existem outros fatores que contribuem, significativamente, para a qualidade na educação, mas que não são considerados no momento da avaliação, como por exemplo, a quantidade de alunos matriculados, o número de reprovações, o número de funcionários, a existência de biblioteca, sala de leitura, refeitório, quadra de esportes, laboratório de informática e laboratório de ciências, na escola.

Baseado nestes fatores, decidiu-se estudar estas variáveis a fim de identificar sua significância, de modo que as escolas possam adotar medidas e estabelecer metas para modificar esta realidade, fazendo com que a qualidade da educação aumente e, conseqüentemente, o seu índice de desempenho.

## **1.3 Estrutura do trabalho**

Este trabalho está estruturado da seguinte forma: no Capítulo 1 apresenta-se a Introdução, contendo os Objetivos, a Justificativa e a Metodologia; no Capítulo 2, a Revisão de Literatura sobre o IDEB, a Regressão Logística e os Métodos Estatísticos utilizados; no Capítulo 3, apresentam-se os Materiais e Métodos utilizados na pesquisa; o Capítulo 4 traz os Resultados e Discussões; no Capítulo 5, a Conclusão e, por fim; no Capítulo 6, as Referências Bibliográficas.

## **2 REVISÃO DE LITERATURA**

### **2.1 Índice de Desenvolvimento da Educação Básica**

O Ministério da Educação (MEC), sob o eixo do Plano de Desenvolvimento da Educação (PDE), no que trata da educação básica, fixou metas para melhorar os indicadores educacionais, por meio do Termo de Adesão ao Compromisso Todos pela Educação.

O Termo de Adesão é objeto de compromisso dos municípios para, em parceria com o MEC e o governo estadual, junto com toda a comunidade e suas lideranças, profissionais da educação, pais e alunos, meios políticos e administrativos, e os recursos municipais disponíveis, em que a escola está inserida, melhorar os indicadores educacionais. O indicador objetivo para a verificação do cumprimento do Termo de Adesão é o IDEB.

O IDEB é um indicador de qualidade criado pelo MEC e lançado em 2007, com o propósito de detectar escolas e/ou redes de ensino cujos alunos apresentam baixo rendimento escolar e também monitorar a evolução temporal do desempenho desses alunos.

Segundo Fernandes (2007), o IDEB é um indicador educacional composto por outros dois indicadores: rendimento escolar (taxa média de aprovação na etapa de ensino) e desempenho (pontuação média dos estudantes nos testes padronizados: Prova Brasil e Saeb). É um índice fácil de compreender, simples de calcular, e aplicável às escolas. A combinação destes indicadores do IDEB, pode variar entre 0 e 10, expressando o andamento do sistema de ensino.

A meta nacional, estabelecida pelo MEC, é alcançar a média 6,0 em 2021, tendo em vista a simbologia do Bicentenário da Independência no ano seguinte. A definição desta meta direciona o cálculo das metas intermediárias individuais do IDEB para o Brasil, estados, municípios e escolas, que deverão unir esforços para que esse objetivo seja alcançado. Com o IDEB, criado em 2007, foram calculadas as metas para cada uma das esferas, de dois em dois anos.

Esta meta foi determinada a partir da observação da média dos países desenvolvidos (países membros da Organização para a Cooperação e Desenvolvimento Econômico (OCDE)), considerando o nível de qualidade educacional em termos de proficiência e rendimento, nos anos iniciais do ensino fundamental. Esta comparação se deu por meio da

distribuição das proficiências observadas no Programa Internacional de Avaliação dos Estudantes (PISA) e no Saeb.

Fernandes (2007) salienta que, tanto para o Brasil quanto para os estados, municípios e escolas, para projetar as trajetórias esperadas para o IDEB ao longo do tempo, devem-se considerar as seguintes situações:

- “parte-se do princípio de que essas trajetórias têm o comportamento de uma função logística”;

- “para o Brasil alcançar a meta estipulada no tempo adequado, a contribuição em esforço de cada rede de ensino e escola deve partir de metas individuais diferenciadas”;

- “o esforço a ser empregado por cada esfera deve objetivar, além do alcance das metas intermediárias de curto prazo, a convergência dos IDEB atingidos pelas demais redes a médio e longo prazos, ou seja, os esforços de cada rede devem também contribuir para a redução das desigualdades em termos de qualidade educacional”.

Baseados nestas premissas, para se calcular as projeções do Brasil, é necessário ter conhecimento das seguintes informações: valor inicial do IDEB, a meta para o Brasil, o tempo para esta meta ser alcançada e o esforço necessário para chegar à meta.

Para o valor inicial do IDEB foi utilizado o do ano de 2005, ou seja,  $t = 0$ , enquanto que o tempo é definido de acordo com cada fase de ensino. O esforço necessário ( $\gamma_i$ ) foi calculado por meio da equação (1), descrita por Fernandes (2007), e serve para garantir que o Brasil atinja a meta no tempo esperado.

$$IDEB_{it} = \frac{1}{1 + e^{-\left(\ln\left(\frac{IDEB_{i0}}{10 - IDEB_{i0}}\right) + \gamma_i \cdot t\right)}} \quad (1)$$

$$t = \begin{cases} 0, \dots, 16 & \text{para metas da 1ª fase do ensino fundamental} \\ 0, \dots, 20 & \text{para metas da 2ª fase do ensino fundamental} \\ 0, \dots, 23 & \text{para metas do ensino médio} \end{cases} \\ \text{(o tempo é dado em anos, desde o ano do IDEB inicial)}$$

$$\begin{cases} i: & \text{município, UF, Brasil, rede de ensino ou escola;} \\ IDEB_{it}: & \text{valor do IDEB no ano } t \text{ para determinado } i \\ IDEB_{i0}: & \text{IDEB inicial (} t=0 \text{) para determinado } i \\ \gamma_i: & \text{esforço individual} \end{cases}$$

Com o esforço calculado, ele foi mantido constante, para que fosse possível determinar quantos anos o Brasil levaria para se aproximar do valor de convergência do IDEB, estipulado em 9,9, sendo que se tem o IDEB variando de 0 a 10.

O processo para se calcular as metas intermediárias dos estados, municípios e escolas, é semelhante ao que foi usado para o cálculo das metas do Brasil. Porém, os objetivos do Brasil como um todo são de curto prazo, enquanto que os demais são de longo prazo, para que se obtenha a convergência entre as redes, tendo em vista que cada uma delas seguirá uma trajetória diferente ao longo dos anos, devido às suas características individuais, mas almejando o mesmo propósito, ou seja, que o País atinja, em 2022 o desejável nível de qualidade da educação.

Segundo Fernandes (2007), o IDEB é o resultado do produto entre o desempenho e rendimento escolar (ou o inverso do tempo médio de conclusão de uma série).

O IDEB é calculado a partir da fórmula geral dada:

$$IDEB_j = N_{ji}P_{ji} \quad (2)$$

$$0 \leq N_j \leq 10; \quad 0 \leq P_j \leq 1; \quad 0 \leq IDEB_j \leq 10$$

Em que,

$IDEB_j$  = Índice de desenvolvimento da educação básica da unidade j (escola, rede de ensino, município, etc.);

$N_{ji}$  = Média padronizada da proficiência em Língua Portuguesa e Matemática, obtida em determinado exame padronizado, para estudantes da unidade j ao final da etapa de ensino considerada;

$P_{ji}$  = Indicador de rendimento baseado na taxa de aprovação da etapa de ensino dos alunos da unidade j.

$i$  = Ano do exame (Saeb e Prova Brasil) e do Censo Escolar.

A média de proficiência determinada em (2) por  $N_{ji}$  é padronizada para que esteja, assim como o IDEB, entre 0 e 10, e é obtida através da fórmula (3) e (4):

$$N_{ji} = \frac{n_{ji}^{lp} + n_{ji}^{mat}}{2} \quad (3)$$

$$n_{ji}^{\alpha} = \frac{S_{ji}^{\alpha} - S_{inf}^{\alpha}}{S_{sup}^{\alpha} - S_{inf}^{\alpha}} \times 10 \quad (4)$$

Onde,

$n_{ji}^{\alpha}$  = proficiência na disciplina  $\alpha$ , obtida pela unidade j, no ano i, padronizada para valores entre 0 e 10;

$\alpha$  = disciplina (matemática ou língua portuguesa);

$S_{ji}^\alpha$  = proficiência média na disciplina, não padronizada, dos alunos da unidade j, obtida no exame do ano i;

$S_{inf}^\alpha$  = limite inferior da média de proficiência (matemática ou língua portuguesa) do Saeb 1997;

$S_{sup}^\alpha$  = limite superior da média de proficiência (matemática ou língua portuguesa) do Saeb 1997.

A proficiência média obedece a seguinte regra, para as unidades escolares: se  $S_{ji}^\alpha < S_{inf}^\alpha$  o desempenho é fixado em  $S_{inf}^\alpha$ ; e de forma análoga se  $S_{ji}^\alpha > S_{sup}^\alpha$  o desempenho é fixado em  $S_{sup}^\alpha$ .

Para o cálculo destas proficiências, utiliza-se a média e o desvio padrão obtidos no Saeb de 1997, ano em que foi definida a escala do Saeb, para cada etapa de ensino, de acordo com as fórmulas descritas em (5) e (6).

$$S_{inf}^\alpha = \bar{X}_\alpha - (3 \times DP) \quad (5)$$

$$S_{sup}^\alpha = \bar{X}_\alpha + (3 \times DP) \quad (6)$$

O indicador de rendimento  $P_j$  é calculado a partir da fórmula descrita em (7):

$$P_{ji} = \frac{1}{T_{ji}} \quad (7)$$

Sendo que  $T_{ji}$  é o tempo médio para a conclusão de uma etapa dos estudantes de uma unidade j, e é calculado como mostra a fórmula (8):

$$T_{ji} = \sum_{r=1}^n \frac{1}{p^r} \quad (8)$$

Onde

$p^r$  = proporção de aprovados em cada uma das séries consideradas (é a taxa de aprovação da r-ésima série da etapa educacional considerada);

$r = 1, 2, 3, \dots, n$ , onde n é o número de séries com taxa de aprovação positiva;

Sendo P o inverso do tempo médio para conclusão de uma série, então, pode-se calcular o IDEB através da forma descrita em (9):

$$IDEB_{ji} = \frac{N_{ji}}{T_{ji}} \quad (9)$$

## 2.2. A Regressão Logística

A regressão logística múltipla é similar à regressão linear múltipla, porém difere desta pela natureza da variável dependente (resposta) que é classificada como categórica, podendo ser dicotômica, sendo as variáveis predictoras contínuas ou categóricas, conforme aponta Field (2009).

Uma variável é dicotômica ou binária, quando assume somente dois valores, podendo ser 0 ou 1, sendo estes mutuamente excludentes, ou seja, a ocorrência de um implica na não ocorrência do outro. O valor 0 (zero) é usado quando o evento de interesse ocorre, ou seja quando se obtém um sucesso, enquanto que o valor 1 é o seu complementar, ou seja, fracasso.

Quando se tem variáveis respostas dicotômicas, algumas propriedades da regressão linear múltipla perdem sua eficácia e até mesmo sua validade, gerando assim a necessidade de ajustes para que se possa modelar esta resposta adequadamente, e estes ajustes se concretizam na regressão logística.

Por isso, é relevante discutir, mesmo que brevemente, a regressão linear múltipla para que se possa inserir a regressão logística de maneira simples e de fácil compreensão.

Na regressão linear simples, a variável de saída (dependente)  $Y$  é determinada através da equação linear abaixo:

$$Y = b_0 + b_1X + \varepsilon \quad (10)$$

Na equação (10) tem-se representado o coeficiente  $b_0$  como o intercepto, o  $b_1$  é a inclinação, e  $\varepsilon$  o erro do modelo de regressão. O intercepto  $b_0$  representa o valor da variável resposta  $Y$ , quando a variável predictor  $X$  for zero. Já o coeficiente  $b_1$  representa a variação ocorrida na variável de saída quando a variável predictor for alterada. Logo o sinal deste coeficiente  $b_1$  indica que, se o sinal for positivo, o acréscimo na variável predictor implica em um acréscimo em  $Y$ , e se o sinal for negativo, o acréscimo em  $X$  gera um decréscimo na variável de saída.

Uma equação semelhante é apresentada na regressão múltipla, porém existem vários previsores cada qual com seu próprio coeficiente. Assim  $Y$  poderá ser previsto através da equação a seguir:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon \quad (11)$$

Surge, portanto, uma importante diferença entre a regressão logística e a linear. A quantidade a ser modelada em qualquer problema de regressão é o valor médio da variável resposta, dado os valores das variáveis previsoras, também conhecido como valor esperado,  $E(Y|X = x)$ . No entanto, este valor esperado é uma probabilidade para a regressão logística e por isso assume valores entre 0 e 1, o que não acontece na regressão linear onde essa média pode variar de  $-\infty$  a  $+\infty$ .

Outra grande diferença entre as regressões é na distribuição de probabilidade dos erros. Enquanto para a linear eles seguem uma distribuição Normal com média 0 e variância constante, na logística, eles seguem uma distribuição Binomial com média 0 e variância  $\pi(x)[1 - \pi(x)]$ . Para contornar este problema, é necessário realizar uma transformação logarítmica, que possibilitará expressar um relacionamento não-linear de uma forma linear (FIELD, 2009).

Assim, a equação logarítmica, após realizada sua transformação chamada de logit, aplicada sob a equação (10), está representada na equação (12). O mesmo procedimento pode ser feito em relação a equação (11), e está representada pela equação (13).

$$\text{logit}(Y) = b_0 + b_1X \quad (12)$$

$$\text{logit}(Y) = b_0 + b_1X + b_2X_2 + \dots + b_nX_n \quad (13)$$

Hair (2009) salienta que a variável dependente de natureza binária tem algumas propriedades que violam as suposições da regressão linear múltipla, como visto anteriormente. Por isso, a distribuição logística de probabilidade impera nas escolhas para realizar a modelagem dos dados que tem resposta binária, por ser de fácil utilização e possuir extrema flexibilidade, assim como por proporcionar interpretações ricas em significados práticos.

Na regressão logística, segundo Field (2009), em vez de se prever o valor da variável  $Y$  a partir de um predictor  $X$ , como na regressão linear, ou diversas variáveis previsoras ( $Xs$ ),

como ocorre na regressão múltipla, é previsto a probabilidade de  $Y$  ocorrer conhecidos os valores de  $X$  ou  $X_s$ .

Como foi visto, a regressão logística tem várias semelhanças com as equações de regressão recém descritas. O primeiro modelo apresentado se assemelha ao da regressão simples, pois existe um único preditor  $X$  para a probabilidade  $Y$  ocorrer, denominado regressão logística univariada.

Sua equação é dada por:

$$P(Y) = \frac{1}{1+e^{-(b_0+b_1X_1+\varepsilon)}} \quad (14)$$

Onde  $e$  é a base dos logaritmos naturais e os demais coeficientes que se encontram entre parênteses na equação formam uma combinação linear idêntica à da regressão simples.

De forma análoga, apresenta-se a seguir o modelo de regressão logística múltipla, que inclui a equação de regressão múltipla, em que se pode inserir qualquer número de preditores.

$$P(Y) = \frac{1}{1+e^{-(b_0+b_1X_1+b_2X_2+\dots+b_nX_n+\varepsilon)}} \quad (15)$$

A equação dada para  $P(Y)$  está baseada no princípio de que ela expressa uma equação de regressão múltipla em termos logarítmicos para resolver o problema da linearidade dos dados, que fica violada quando a variável de saída é dicotômica (FIELD 2009).

É importante ressaltar que, como se trata de uma probabilidade, seus valores não podem ultrapassar o intervalo característico:  $0 \leq P(Y) \leq 1$ . Se obtiver um valor próximo de zero, significa que é bastante improvável a ocorrência de  $Y$ , ao passo que se o valor for bem próximo de 1, a ocorrência de  $Y$  é bem provável.

Quando a análise é executada, é preciso estimar o valor dos coeficientes para que a equação seja utilizada. A estimação destes parâmetros é realizada pelo ajustamento do modelo, por meio do método da máxima verossimilhança, no qual uma das principais medidas de avaliação da regressão logística é o valor da verossimilhança-log, descrita por Field (2009).

Assim como na regressão múltipla utiliza-se o  $R^2$  (Correlação de Pearson ao quadrado) para avaliar se um modelo adere aos dados, por meio da comparação dos valores observados e dos previstos, de forma análoga, é utilizado na regressão logística a

verossimilhança-log (VL) para avaliar a aderência do modelo. Sua equação, apresentada a seguir, é baseada na soma das probabilidades associadas com a saída real e a prevista.

Assim, pode-se dizer que a verossimilhança-log indica quanta informação não explicada ainda existe no modelo, após ele ter sido ajustado. Sendo assim, quanto maior for o valor encontrado para esta estatística, mais observações não explicadas existirão, fazendo com que o modelo tenha uma aderência pobre (Field, 2009).

$$\text{verossimilhança} - \log = \sum_{i=1}^N \{Y_i \ln(P(Y_i)) + (1 - Y_i) \ln[1 - P(Y_i)]\} \quad (16)$$

Logo, para medir o ajuste da estimação do modelo, deve-se multiplicar o logaritmo do valor da verossimilhança por -2 (chamado de -2VL), que segue a distribuição qui-quadrado. Assim, para se ter o ajuste perfeito do modelo, seu valor deve ser 0; podendo, portanto, concluir que quanto menor o seu valor, maior o poder preditivo do modelo.

Contudo, pode-se calcular uma versão mais adequada para o ajuste do modelo na regressão logística, chamado de estatística R, que é a correlação parcial entre a variável de saída e cada uma das variáveis previsoras, podendo variar de -1 a 1. Quanto maior o valor de R, maior a sua contribuição para o modelo. A equação a seguir fornece o cálculo de R, conforme descreve Field (2009).

$$R = \pm \sqrt{\frac{\text{wald} - (2 \times gl)}{-2VL(\text{original})}} \quad (17)$$

Porém, como o cálculo de R depende da estatística de Wald, ele não é uma medida precisa, devendo, portanto, ser tratado com cautela.

Outras medidas foram descritas com valores semelhantes ao do  $R^2$  da regressão linear, de modo que pudesse ser utilizado na regressão logística.

Hosmer e Lemeshow (1989) descreveram esta medida apresentada na equação 18, baseada na verossimilhança-log do valor original e antes de adicionar um predictor qualquer. O  $R_V^2$  pode variar de 0 a 1, onde o valor zero indica que os previsores não contribuem, e 1 indica que o modelo prevê totalmente a variável de saída.

$$R_V^2 = \frac{-2VL(\text{Modelo})}{-2VL(\text{Original})} \quad (18)$$

Outra medida é o  $R_{CS}^2$  de Cox e Snell, que também se baseia na verossimilhança-log do modelo novo e do original, juntamente com o tamanho da amostra.

$$R_{CS}^2 = 1 - e^{\left[-\frac{2}{n}(VL(novo) - VL(basico))\right]} \quad (19)$$

Nagelkerke (1991) sugeriu uma correção na estatística de Cox e Snell, pelo fato de que esta estatística nunca alcança o seu valor teórico máximo. Esta correção está representada na equação seguinte.

$$R_N^2 = \frac{R_{CS}^2}{1 - e^{\left[-\frac{2(VL(basico))}{n}\right]}} \quad (20)$$

Todas estas medidas descritas apresentam diferenças, não somente na forma de cálculo, como também em seus resultados. Porém, de um modo geral, é possível considerá-las praticamente iguais, já que existe controvérsia sobre qual seria o valor mais adequado, pois em termos de interpretação, todas se assemelham ao  $R^2$  da regressão linear, uma vez que fornecem o grau de aderência do modelo.

Field (2009) salienta que é preciso verificar, além do ajuste do modelo aos dados, a contribuição individual de cada um dos previsores, assim como é feito na regressão linear. A estatística de Wald apresenta uma distribuição qui-quadrado, e assim como o teste t da regressão linear, ela informa se o coeficiente b para cada predictor é significativamente diferente de zero, ou seja, se contribui significativamente para a previsão de Y.

A equação do teste, descrita por Hosmer e Lemeshow (1989), diz que o teste é obtido por meio da comparação da estimativa de máxima verossimilhança do parâmetro ( $b_i$ ) com a estimativa do seu erro padrão ( $EP_{b_i}$ ), onde tem-se  $H_0: b_i = 0$  e  $H_1: b_i \neq 0$ . E, por seguir uma distribuição normal padrão, compara-se esta razão com um valor tabelado de Z (bilateral) para a significância  $\alpha$  desejada.

A seguir tem-se a equação de Wald.

$$Wald = \frac{b_i}{EP_{b_i}} \quad (21)$$

Apesar de simples, deve-se ter cautela ao usar a estatística de Wald, pois se o coeficiente (b) for grande, ele tende a inflacionar o erro padrão, fazendo com que a estatística fique subestimada, ou seja, esta inflação aumenta a probabilidade de se rejeitar o predictor, quando na verdade ele contribui para o modelo.

Além de avaliar a contribuição individual de cada um dos previsores, a avaliação geral do modelo também deve ser verificada. Para tal, o teste de Hosmer e Lemeshow utiliza a comparação entre as frequências observadas e as esperadas da variável dependente, baseia-se na estatística da distribuição qui-quadrado com  $(g - 2)$  graus de liberdade (Hosmer e Lemeshow, 1989).

Field (2009) diz que: “essa estatística testa a hipótese de que os dados observados são significativamente diferentes dos valores previstos pelo modelo”. Logo, o que se espera do modelo é que apresente um valor não-significativo para este teste, o que indica que o modelo irá apresentar boa previsão dos valores.

Hair (2009) diz que o teste apresenta uma ampla medida da precisão preditiva do modelo no qual se baseia na real previsão da variável resposta, e não no valor da verossimilhança. O autor ressalta que o uso correto do teste depende de uma amostra com no mínimo 50 casos, já que consiste em realizar divisões no número de observações em aproximadamente 10 classes iguais para depois comparar as frequências. Também ressalta que a estatística qui-quadrado é sensível ao tamanho da amostra, logo quando se tem amostras grandes pode apresentar diferenças muito pequenas, porém significativas.

O *Odds Ratio* (OR), ou seja, razão de chances, é um indicador da mudança nas chances resultantes da mudança de uma unidade no predictor, conforme explica Field (2009). Ele ainda diz que o OR é semelhante aos coeficientes  $b$  da regressão logística, porém, mais fácil de compreender e de explicar quando a variável preditora é categórica, por não ser necessária uma transformação logarítmica.

Para calcular o OR é necessário primeiro calcular a chance de o evento ocorrer e de o evento não ocorrer, e depois a razão entre as duas. Para isso, deve-se utilizar as fórmulas a seguir:

$$\text{Chance} = \frac{P(\text{evento ocorrer})}{P(\text{evento não ocorrer})} \quad (22)$$

Sendo que para encontrar a probabilidade de um evento ocorrer é usada a equação 23:

$$P(\text{evento } Y) = \frac{1}{1 + e^{-(b_0 + b_1 X + \dots + b_n X_n)}} \quad (23)$$

E para a probabilidade do evento não ocorrer se usa a equação 24:

$$P(\text{evento } Y \text{ não ocorrer}) = 1 - P(\text{evento } Y \text{ ocorrer}) \quad (24)$$

Logo, deve-se utilizar o mesmo procedimento após a variável previsor ter mudado em uma unidade. Assim, tem-se as chances antes e depois da mudança de uma unidade na variável previsor, que são necessárias para se obter o OR, definido pela fórmula:

$$\text{OR} = \frac{\text{probabilidade após a mudança de uma unidade no previsor}}{\text{probabilidade original}} \quad (25)$$

Calculada essa proporção, pode-se interpretar o OR em relação a mudança na chance, com base na relação abaixo:

- OR > 1: quando aumenta o previsor, aumentam as chances da saída ocorrer (é considerado fator de risco);
- OR < 1: quando aumenta o previsor, diminuem as chances da saída ocorrer (é considerado fator de proteção);
- OR = 1: não há associação entre o previsor e a saída (não é considerado fator de risco).

Hosmer e Lemeshow (1989) defendem que antes de aceitar definitivamente que um determinado modelo se encaixa aos dados, uma análise de resíduos deve ser realizada. Field (2009) diz que “o exame dos resíduos, depois de qualquer análise, é extremamente importante”, pois através dele pode-se verificar a qualidade do ajuste do modelo aos dados observados, e descreve os resíduos como sendo a diferença entre os valores previstos pelo modelo e os observados na amostra. Se os resíduos são pequenos, é porque o modelo se ajusta bem aos dados da amostra, e o contrário ocorre se os resíduos forem grandes.

Field(2009) ainda apresenta os principais objetivos da análise de resíduos na regressão logística que são: “isolar pontos em que o modelo tem pouca aderência” (1), e “isolar pontos que exercem uma influência indevida no modelo” (2).

Para avaliar o primeiro objetivo, devem-se examinar os resíduos estudentizados, padronizados e as estatísticas de desvio, onde todos estes têm a mesma propriedade: 95% dos casos devem apresentar valores entre  $\pm 1,96$ , e 99% entre  $\pm 2,58$ .

Para o segundo objetivo, devem-se utilizar as estatísticas de influência: distância de Cook, que deve ter valor menor do que 1; DFBeta que é uma versão padronizada da anterior, onde valores maiores do que 1 indicam casos potencialmente influentes; Leverage que é o valor da influência esperada, a qual deve estar entre 0 e 1, em que o valor 0 indica sem nenhuma influência e o valor 1 indica completa influência. O Leverage é calculado pela fórmula  $(k + 1)/N$ , onde k é o número de previsores e N é o tamanho da amostra.

### 3 MATERIAIS E MÉTODOS

A cidade de Santa Maria está localizada no “Coração do Rio Grande”, ou seja, na região central do Rio Grande do Sul, e possui 274.838 habitantes, conforme dados estimados, em 2014, pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Possui uma área territorial que se aproxima de 1.782 Km<sup>2</sup> e fica distante da capital a aproximadamente 300 Km. É considerada como cidade universitária, por ter instalada no município uma das maiores universidades públicas do Brasil, a Universidade Federal de Santa Maria, e também por ter diversas instituições de ensino sendo também chamada de “Cidade Cultura”.

Segundo os dados educacionais do censo 2012 também divulgados pelo IBGE, Santa Maria possui 79 pré-escolas, tendo 50 escolas públicas, 106 escolas de ensino fundamental, das quais 86 são públicas, e 36 escolas de ensino médio, com 26 delas públicas, conforme é apresentado no Quadro 1:

Quadro 1 – Distribuição do número de escolas da Educação Básica do município de Santa Maria, por nível e sistema de ensino

Descrição	Quantidade
Educação Infantil – Escola Pública Municipal	46
Educação Infantil – Escola Pública Estadual	03
Educação Infantil – Escola Pública Federal	01
Educação Infantil – Escola Particular	29
Ensino Fundamental – Escola Pública Municipal	52
Ensino Fundamental – Escola Pública Estadual	33
Ensino Fundamental – Escola Pública Federal	01
Ensino Fundamental – Escola Particular	20
Ensino Médio – Escola Pública Municipal	00
Ensino Médio – Escola Pública Estadual	23
Ensino Médio – Escola Pública Federal	03
Ensino Médio – Escola Particular	10

Fonte: IBGE (2012).

Para o desenvolvimento do presente trabalho, foram escolhidas as escolas públicas por estarem localizadas em diversas regiões do município, incluindo escolas localizadas na zona rural e em bairros com alto índice de violência e criminalidade.

### **3.1 Estrutura do banco de dados**

Utilizou-se o site Qedu e o Portal do Ideb/Inep para a obtenção dos dados das escolas públicas de Santa Maria/RS, participantes do IDEB. O IDEB é estabelecido nas etapas iniciais e finais do ensino fundamental, e os dados coletados para a pesquisa referem-se aos anos de 2011 e 2013, logo, os resultados são apresentados em quatro grupos: 5º e 9º ano de 2011, 5º e 9º ano de 2013.

As variáveis de interesse foram a localização das escolas, o número de funcionários, o número de alunos matriculados e reprovados em cada série, assim como dados sobre a estrutura das escolas, ou seja, se a escola possuía ou não biblioteca, sala de leitura, laboratório de informática, laboratório de ciências, quadra de esportes, refeitório e acessibilidade. Por meio do Portal do IDEB, foi verificado se as escolas atingiram ou não a meta estabelecida para os anos em estudo.

Com o banco de dados coletados, foram realizadas as análises descritivas com o uso do software Estatística 9.1, além da modelagem de regressão logística, apresentada por Hosmer e Lemeshow e por Field (2009) e Hair (2009), para descobrir quais fatores podem ser decisivos para a obtenção das notas do IDEB, por meio do software SPSS 13.0.

É importante ressaltar que algumas escolas foram excluídas do banco de dados por apresentarem dados insuficientes para as análises em questão. Sendo assim, o número de escolas analisadas em cada etapa e em seu respectivo ano foi de: 51 em 2011 e de 43 em 2013 para o 5º ano, enquanto que para o 9º ano em 2011 foi de 52 e 37 em 2013.

Destaca-se a covariável refeitório, em 2011, apenas uma escola não possuía, já em 2013 todas as escolas pesquisadas tinham, e a covariável acessibilidade, que nenhuma das escolas possuía, nos dois anos analisados. Já a covariável laboratório de informática, apenas uma escola, em 2011, não possuía. E, portanto, as que não apresentaram resultados foram excluídas das análises.

Para a análise dos resultados encontrados, foram tomadas como base as últimas duas avaliações realizadas pelo MEC e, por referência, os Indicadores da Qualidade na Educação que servem para direcionar as políticas públicas de melhoria na educação.

### 3.2 Variável dependente

Utilizou-se como variável dependente (Y) para este estudo se a escola atingiu ou não a meta do IDEB estabelecida para cada escola, em cada etapa de avaliação e para o ano em questão. Esta variável só pode assumir dois valores, o que faz dela uma variável dicotômica, que foi assim definida: 1 para a escola que não alcançou a meta ou 0 (zero) para a escola que alcançou a meta.

### 3.3 Variáveis independentes ou covariáveis

As covariáveis utilizadas foram àquelas disponíveis no site Qedu e as informações relacionadas ao censo escolar, que foram categorizadas em a escola possuir (0) ou não (1) a determinada característica.

As covariáveis: número de alunos matriculados e reprovados, que foram coletados para cada série, foram organizadas em dois grupos: séries iniciais (5º ano) e séries finais (9º ano), por possuírem algumas séries com número de alunos muito baixo ou zero.

Foram calculadas algumas medidas descritivas para estas covariáveis a fim de auxiliar na avaliação e também a análise de regressão logística univariada. Os dados resultantes deste ajuste possibilitaram identificar as variáveis com resultado significativo e que são favoráveis para entrar no modelo múltiplo.

### 3.4 Modelo estatístico

Para o processo de modelagem, foram estabelecidos alguns passos, conforme a literatura orienta, sendo eles: o teste de significância para as covariáveis, a seleção destas variáveis no modelo logístico e a avaliação do ajuste do modelo.

Inicialmente, realizou-se a análise da regressão logística univariada entre cada uma das covariáveis com a variável dependente. Foram selecionadas ao modelo de regressão múltiplo as covariáveis que apresentaram  $p \leq 0,25$  no teste univariado, ou seja,

apresentaram evidências de associação com a variável dependente Y. As demais covariáveis foram desconsideradas para o modelo múltiplo.

A análise da regressão univariada possibilitou ainda determinar o valor do OR bruto e os respectivos intervalos de confiança de 95%.

Para a análise de regressão logística múltipla, utilizou-se o método “enter”, no qual todas as covariáveis independentes significativas, obtidas do modelo univariado, foram colocadas ao mesmo tempo no modelo múltiplo.

Após esta análise, todas as covariáveis não significativas, ou seja, com  $p > 0,05$ , foram retiradas do modelo múltiplo, uma a uma. A cada covariável retirada, verificou-se o efeito da saída nas que permaneciam no modelo.

No próximo capítulo, são apresentados os resultados obtidos a partir da análise destes dados.

## 4 RESULTADOS E DISCUSSÃO

Das 86 escolas públicas de ensino fundamental existentes na cidade de Santa Maria, 62 (72,09%) escolas tiveram seus dados disponibilizados no site, em virtude das exigências que as escolas têm de cumprir para participarem do programa: ter um número mínimo de 20 alunos matriculados nas séries avaliadas e realizar a Prova Brasil, entre outros.

É mostrado na Figura 1 o número de escolas que atingiram ou não o objetivo estabelecido nas respectivas séries e anos.

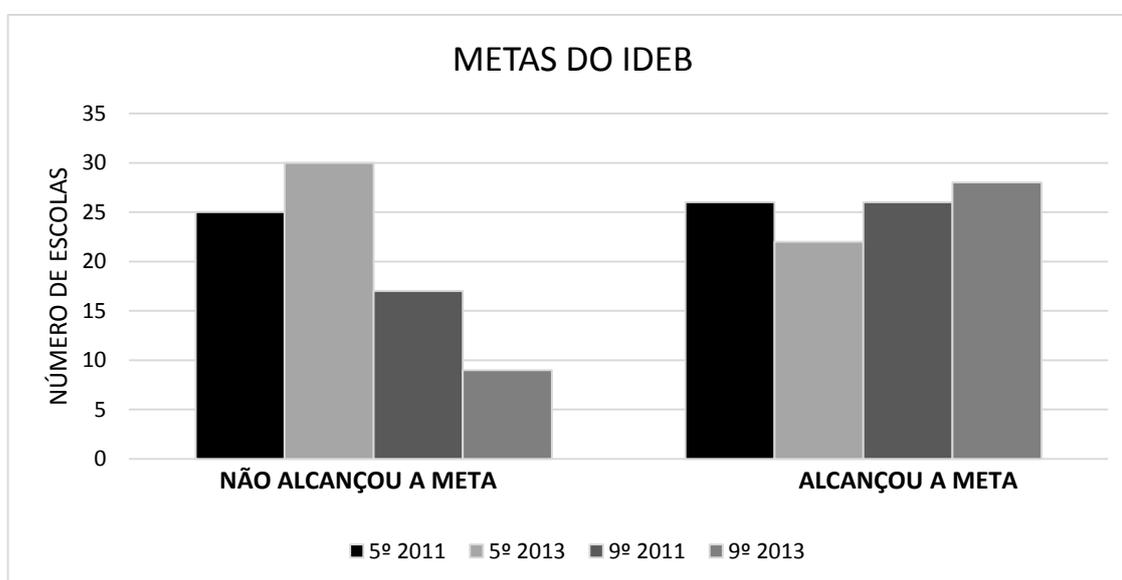


Figura 1: Gráfico comparativo entre o número de escolas que atingiram ou não a meta do IDEB estabelecida, por série e ano.

Observa-se, a partir da Figura 1, que apenas em 2013 no 5º ano, o número de escolas que alcançaram a meta superou as que não alcançaram. Em 2011, também no 5º ano, houve uma diferença muito pequena, apenas uma escola a mais, ou seja, 25 escolas não atingiram e 26 escolas atingiram a meta. O mesmo não ocorreu no 9º ano de 2013, em que esta diferença foi maior, sendo que o número de escolas que atingiram o objetivo ultrapassou em 19 as que não atingiram.

Comparando o 5º ano de 2011 com o 5º ano de 2013, houve uma melhora, pois em 2011 49% (25/51) das escolas alcançaram a meta, enquanto que em 2013 58% (30/52), mas mesmo assim, não houve diferença significativa ( $p = 0,4936$ ) pelo teste do Qui-Quadrado.

Ao contrário do que se viu com o 5º ano, no 9º ano houve uma diminuição, também não significativa ( $p = 0,2267$ ) pelo teste do Qui-Quadrado, no percentual de escolas que alcançaram a meta estabelecida, ou seja, em 2011 39% (17/43) enquanto que em 2013 este percentual foi de apenas 24% (9/37) das escolas que alcançaram a meta.

Apesar da maioria das escolas ter atingido a meta do IDEB, isto não significa que a escola teve uma melhoria no sistema de ensino. Muitas delas tiveram um aumento nas notas do IDEB quando comparados os anos de 2011 para 2013, porém este aumento ainda não corresponde ao desejado.

Bello (2012) aponta que a mídia de um modo geral, descreve a causa do insucesso nos resultados do IDEB ao processo de formação de professores, em particular, aos em exercício na rede pública de ensino, apontando até para fatores como baixos salários e falta de infraestrutura.

Entretanto, a partir de estudos realizados por outros autores, incluindo o de Chirinéa (2010), pode-se verificar que existem diversos fatores que podem explicar os “maus” resultados do IDEB, como por exemplo, o número de reprovações que foi destaque no trabalho. Já no estudo realizado por Gouveia, Souza e Tavares (2009), as diferentes condições sociais na região em questão foi o destaque para os resultados obtidos no IDEB.

Ao realizar uma análise básica dos dados coletados, referente às informações obtidas sobre a estrutura física das escolas, foram obtidos os resultados que são apresentados, na Tabela 1 referentes a 2011, e a Tabela 2 os dados referentes a 2013.

Tabela 1 - Relação da estrutura física das escolas, por série, em função da conquista ou não da meta estabelecida do IDEB no ano de 2011.

Variáveis	Atingiu a meta				
	5º ano		9º ano		
	Sim	Não	Sim	Não	
Biblioteca	Sim	8 15,68%)	13(25,49%)	8 (18,60%)	15 (34,88%)
	Não	17 33,33%)	13 (25,49%)	9 (20,93%)	11 (25,58%)
Sala de leitura	Sim	16 31,37%)	8 (15,68%)	11 (25,58%)	9 (20,93%)
	Não	9 (17,65%)	18 (35,29%)	6 (13,95%)	17 (39,53%)
Laboratório de ciências	Sim	9 (17,65%)	12 (23,53%)	9 (20,93%)	14 (32,56%)
	Não	16 31,37%)	14 (27,45%)	8 (18,60%)	12 (27,91%)
Quadra de esportes	Sim	15 (29,41%)	19 (37,25%)	14 (32,56%)	20 (46,51%)
	Não	10 (19,61%)	7 (13,72%)	3 (6,97%)	6 (13,95%)

Percebe-se que, do total de escolas observadas em 2011, 41,18% delas possuía bibliotecas no 5º ano, sendo que, em 2013, este número passou para 53,85%. Destaca-se também que, apenas 15,68% delas alcançaram a meta estabelecida em 2011, e 23,07% em 2013. Já para o 9º ano, 53,49% das escolas em 2011 possuíam biblioteca, enquanto que em 2013 o número é de 70,27%.

Tabela 2 - Relação da estrutura física das escolas, por série, em função da conquista ou não da meta estabelecida do Ideb no ano de 2013.

Variáveis	Atingiu a meta				
	5º ano		9º ano		
	Sim	Não	Sim	Não	
Biblioteca	Sim	12 (23,07%)	16 (30,77%)	5 (15,51%)	21 (56,76%)
	Não	18 (34,61%)	6 (11,54%)	4 (10,81%)	7 (18,92%)
Sala de leitura	Sim	18 (34,61%)	9 (17,31%)	7 (18,92%)	13 (35,13%)
	Não	12 (23,07%)	13 (25%)	2 (5,40%)	15 (40,54%)
Laboratório de ciências	Sim	5 (9,61%)	14 (26,92%)	5 (15,51%)	14 (37,84%)
	Não	25 (48,07%)	8 (15,38%)	4 (10,81%)	14 (37,84%)
Quadra de esportes	Sim	22 (42,31%)	17 (32,69%)	6 (16,22%)	26 (70,27%)
	Não	8 (15,38%)	5 (9,61%)	3 (8,11%)	2 (5,40%)

Das escolas que possuem sala de leitura, 16 (31,37%) obtiveram êxito em 2011 e 18 (34,61%) em 2013 para o 5º ano, e apenas 7 (18,92%) em 2013 para o 9º ano. Em 2011, 14 (32,56%) escolas com quadra de esportes atingiram a meta no 9º ano, e o mesmo não se repete em 2013, quando apenas 6 (16,22%) atingiram a meta.

Das escolas que possuíam laboratório de ciências, 31,37% delas atingiram o objetivo estabelecido em 2011 para o 5º ano. Esta porcentagem corresponde ao número de 16 escolas que, mesmo não possuindo o laboratório, atingiram a meta. E este número aumenta em 2013 para 25 (48,07%).

Seguindo as análises básicas, pode-se destacar que, em 2011, para cada escola, houve em média 223,2 alunos matriculados nos anos iniciais do ensino fundamental, e 233,6 nos anos finais. Já em 2013, estes números foram de 204,33 e de 211,97, respectivamente.

O número de alunos reprovados em cada escola analisada apresentou uma média de 23,37 e desvio padrão de 15,26 para o 5º ano, e 54,14 e 35,3, respectivamente, para o 9º ano, ambos de 2011. Já para o ano de 2013, estes números foram de 16,96 de média e 12,82 de desvio padrão para o 5º ano, e 48,11 e 27,37, respectivamente, para o 9º ano.

Foi possível verificar, ainda, que o número de funcionários em cada escola foi em média 52,82 em 2011 e 56,16 em 2013.

Soares (2004) diz que parte considerável dos resultados provenientes de pesquisas escolares podem ser explicados por características individuais dos alunos. E dentro destas características estão as condições sociais, as de gênero e a etnia dos estudantes. No entanto, estas condições não foram analisadas na pesquisa pelo fato dela ter sido limitada aos dados divulgados pelo Governo Federal e Municipal, e não por pesquisa de campo.

Também se deve atentar para as diferenças sociais, das regiões onde as escolas estão inseridas, as quais também se expressam na desigualdade do IDEB, conforme salienta Gouveia, Souza e Tavares (2009).

Já Vidal e Vieira (2011) destacam que a melhoria do IDEB está fortemente ligada à evolução das taxas de aprovação. Conseqüentemente, uma piora nas notas do IDEB, está ligada as taxas de reprovação, o que segundo Pessoa (20--?), apesar de todos os esforços aplicados, ainda continua alta no âmbito do ensino fundamental.

Bello (2012) diz ainda que, a melhoria dos resultados nas avaliações são traduzidas como melhorias de desempenho dos alunos e também das práticas pedagógicas realizadas por professores, instituições - escolas, universidades, enfim, pelo sistema de ensino como um todo.

Entretanto, muitas dessas melhorias estão ligadas às exigências que as avaliações externas cobram, que é o desempenho dos estudantes, conforme diz Chirinéa (2010). Em sua pesquisa de campo, a autora verificou que o dado mais significativo foi a estratégia que as escolas utilizaram como meio de atingir estas exigências, ou seja, treinaram seus alunos para a realização dos testes padronizados, adaptando o ambiente escolar e os conteúdos curriculares, com o intuito de receber méritos e não sanções, uma vez que a política instalada vincula o resultado das escolas com bônus para o professor.

Em 2011, para o 5º ano, as covariáveis testadas e candidatas a entrar no modelo múltiplo foram: biblioteca ( $p = 0,195$ ), sala de leitura ( $p = 0,020$ ) e número de alunos reprovados ( $p = 0,106$ ), conforme se pode verificar na Tabela 3.

Tabela 3 - Resultados do ajuste da regressão logística univariada do modelo para a meta M5º do ano de 2011

Variáveis	B	p-valor	OR bruto	IC 95% para OR	
				Inferior	Superior
Bairro	0,021	0,432	1,021	0,969	1,076
Funcionários	0,007	0,449	1,007	0,988	1,027
Biblioteca	-0,754	0,195	0,471	0,151	1,470
Sala de leitura	1,386	0,020	4,000	1,246	12,842
Lab. Ciências	-0,421	0,462	0,656	0,213	2,018
Quadra de esportes	-0,593	0,324	0,553	0,170	1,798
A. matriculados	0,000	0,884	1,000	0,995	1,005
A. reprovados	0,034	0,106	1,034	0,993	1,077

B:estimativas dos parâmetros da regressão; p-valor: significância ao nível de 25% da estatística de Wald; OR bruto

Para o 9º ano, obtiveram-se as covariáveis: sala de leitura ( $p = 0,057$ ) e número de alunos matriculados ( $p = 0,230$ ) como favoráveis a entrar no modelo múltiplo, que podem ser verificados na Tabelas 4.

Tabela 4 - Resultados do ajuste da regressão logística univariada do modelo para a meta M9º do ano de 2011

Variáveis	B	p-valor	OR bruto	IC 95% para OR	
				Inferior	Superior
Bairro	-0,022	0,443	0,979	0,926	1,034
Funcionários	-0,005	0,585	0,995	0,979	1,012
Biblioteca	-0,428	0,495	0,652	0,191	2,230
Sala de leitura	1,242	0,057	3,463	0,961	12,474
Lab. ciências	-0,036	0,954	0,964	0,283	3,284
Quadra de esportes	0,336	0,670	1,400	0,299	6,564
A. matriculados	-0,004	0,230	0,996	0,991	1,002
A. reprovados	0,008	0,415	1,008	0,989	1,027

B:estimativas dos parâmetros da regressão; p-valor: significância ao nível de 25% da estatística de Wald; OR bruto

Em 2013, para o 5º ano, nota-se que as covariáveis selecionadas para a regressão múltipla foram: números funcionários ( $p = 0,026$ ), biblioteca ( $p = 0,022$ ), sala de leitura ( $p$

= 0,176), laboratório de ciências ( $p = 0,001$ ), número de alunos matriculados ( $p = 0,205$ ) e número de alunos reprovados ( $p = 0,006$ ), conforme Tabela 5.

Tabela 5 - Resultados do ajuste da regressão logística univariada do modelo para a meta M5º do ano de 2013

Variáveis	B	p-valor	OR bruto	IC 95% para OR	
				Inferior	Superior
Bairro	-0,15	0,555	0,985	0,937	1,035
Funcionários	0,024	0,026	1,025	1,003	1,047
Biblioteca	-1,386	0,022	0,250	0,076	0,821
Sala de leitura	0,773	0,176	2,167	0,706	6,645
Lab. ciências	-2,169	0,001	0,114	0,031	0,413
Quadra de esportes	-0,212	0,746	0,809	0,224	2,921
A. matriculados	0,003	0,205	1,003	0,998	1,009
A. reprovados	0,092	0,006	1,096	1,027	1,169

B:estimativas dos parâmetros da regressão; p-valor: significância ao nível de 25% da estatística de Wald; OR bruto

Na Tabela 6, são mostradas as covariáveis correspondente ao 9º ano de 2013, e que são candidatas ao modelo múltiplo. Verificou-se que se apresentaram significativas as covariáveis: sala de leitura ( $p = 0,083$ ), quadra de esportes ( $p = 0,066$ ) e número de alunos reprovados ( $p = 0,019$ ).

Tabela 6 - Resultados do ajuste da regressão logística univariada do modelo para a meta M9º do ano de 2013

Variáveis	B	p-valor	OR bruto	IC 95% para OR	
				Inferior	Superior
Bairro	0,022	0,537	1,023	0,953	1,098
Funcionários	-0,001	0,878	0,999	0,983	1,015
Biblioteca	-0,875	0,274	0,417	0,087	2,000
Sala de leitura	1,540	0,083	4,667	0,819	26,604
Lab. ciências	0,223	0,772	1,250	0,276	5,653
Quadra de esportes	-1,872	0,066	0,154	0,021	1,134
A. matriculados	0,002	0,623	1,002	0,974	1,011
A. Reprovados	0,076	0,019	1,079	1,012	1,150

B:estimativas dos parâmetros da regressão; p-valor: significância ao nível de 25% da estatística de Wald; OR: razão de chances

As covariáveis significativas da Tabela 3, referente ao 5º ano de 2011, testadas no modelo múltiplo tem seus resultados apresentados na Tabela 7.

Tabela 7 - Resultados do ajuste da regressão logística multivariada do modelo para a meta M5º do ano de 2011

Variáveis	B	p-valor	OR ajustado	IC 95% para OR	
				Inferior	Superior
Sala de leitura	2,115	0,004	8,286	1,950	35,206
A. reprovados	0,061	0,017	1,063	1,011	1,117
Constante	-2,508	0,006	0,081	---	---

B:estimativas dos parâmetros da regressão; p-valor: significância ao nível de 25% da estatística de Wald; OR: razão de chances.

Os resultados mostraram que o número de reprovações e ter sala de leitura na escola interferem significativamente no alcance da meta do IDEB. Como a significância destas duas covariáveis estão abaixo de 5%, verifica-se que elas possuem relação com o rendimento escolar.

O modelo múltiplo selecionado está representado pela equação 26 (Modelo 1), cujas covariáveis indicam a chance de se atingir a meta do IDEB para o ano de 2011, no 5º ano.

$$\text{Logit}(Y_1) = -2,508 + 2,115*(\text{sala de leitura}) + 0,061*(\text{A. reprovados}) \quad (26)$$

A mesma análise foi feita em relação aos resultados obtidos para o 9º ano de 2011, (Tabela 4) e os resultados do ajuste são mostrados na Tabela 8.

Tabela 8 - Resultados do ajuste da regressão logística multivariada do modelo para a meta M9º do ano de 2011

Variáveis	B	p-valor	OR ajustado	IC 95% para OR ajustado	
				Inferior	Superior
Sala de leitura	1,241	0,062	3,460	0,938	12,761
A. matriculados	-0,004	0,255	0,996	0,990	1,003
Constante	0,638	0,453	1,892	---	---

B:estimativas dos parâmetros da regressão; p-valor: significância ao nível de 25% da estatística de Wald; OR: razão de chances

Logo, pode-se concluir que, apesar das covariáveis sala de leitura e número de alunos matriculados terem sido candidatas ao modelo com  $\alpha$  inferior a 25%, não foi possível encontrar um modelo logístico com covariáveis significativas.

Usando os dados da Tabela 5, foram incluídas, no modelo múltiplo, as variáveis que se apresentaram favoráveis ao alcance da meta estabelecida para o ano de 2013, quando apenas duas delas obtiveram a significância esperada ( $p \leq 0,05$ ). A Tabela 9 mostra o resultado final da regressão logística múltipla para esta etapa.

Tabela 9 - Resultados do ajuste da regressão logística multivariada do modelo para a meta M5° do ano de 2013

Variáveis	B	p-valor	OR ajustado	IC 95% para OR ajustado	
				Inferior	Superior
Lab. Ciências	-2,029	0,006	0,132	0,031	0,560
A. reprovados	0,097	0,014	1,101	1,020	1,190
Constante	-0,803	0,347	0,448	---	---

B:estimativas dos parâmetros da regressão; p-valor: significância ao nível de 25% da estatística de Wald; OR: razão de chances.

Tem-se que as duas covariáveis resultantes possuem a significância desejada, o que confirma a existência de relação com o objetivo das escolas, mostrando que interferem significativamente no alcance da meta estabelecida para o ano de 2013.

Apresenta-se assim, na equação 27 (Modelo 2), o resultado do modelo múltiplo, onde as covariáveis mostram a chance de se atingir a meta estabelecida para o 5° ano em 2013.

$$\text{Logit}(Y_2) = - 0,803 - 2,029*(\text{lab. ciências}) + 0,097*(\text{A. reprovados}) \quad (27)$$

Ao analisar os dados da Tabela 6, observou-se que apenas uma das covariáveis teve valor significativo, podendo este resultado ser verificado através da Tabela 10.

Tabela 10 - Resultados do ajuste da regressão logística multivariada do modelo para a meta M9º do ano de 2013

Variáveis	B	p-valor	OR ajustado	IC 95% para OR ajustado	
				Inferior	Superior
A. reprovados	0,076	0,019	1,079	1,012	1,150
Constante	-1,746	0,120	0,174	---	---

B:estimativas dos parâmetros da regressão; p-valor: significância ao nível de 25% da estatística de Wald; OR: razão de chances.

Sendo assim, o último modelo apresentado na equação 28 (Modelo 3), é constituído apenas da covariável número de alunos reprovados e da constante, gerando um modelo linear.

$$\text{Logit}(Y_3) = -1,746 + 0,076*(A. reprovados) \quad (28)$$

É possível observar que a covariável número de alunos reprovados apareceu em todos os modelos, apresentando valores próximos. Logo é notável que se deve ter um cuidado especial com esta covariável, pois ela contribui negativamente para o alcance da meta.

A avaliação geral dos modelos foi verificada por meio do teste de Hosmer e Lemeshow, em que apenas um dos modelos apresentou a significância esperada, ou seja, um valor não-significativo, indicando que não há diferença significativa entre os valores observados na amostra e os valores preditos dos modelos. Sendo assim o Modelo 1 apresentou p-valor = 0,802, o que mostra que o modelo é suficientemente apropriado para os dados em estudo.

Para avaliar a aderência de cada modelo, foi observado o valor de  $R_N^2$  de Nagelkerke. O Modelo 1 apresentou o valor de 0,299 o que mostra um ajuste pobre deste modelo.  $R_N^2 = 0,478$  foi o valor obtido para o Modelo 2, apresentando um ajuste superior aos demais, já que o Modelo 3 teve valor de 0,358. Sendo assim, o Modelo 2 teve 0,478 de sua variação explicada.

Na análise dos resíduos, os Modelos 1 e 2 apresentaram todos os valores da distância de Cook menores do que 1, o que indica que não existem casos influentes afetando o modelo. Já o Modelo 3 apresentou um valor acima de 1, porém, como as demais estatísticas apresentaram-se bem, isto não deve ser motivo de preocupação.

O valor do DFBeta também informa se há casos potencialmente influentes, já que é uma versão padronizada da distância de Cook, sendo que os três modelos analisados apresentaram valores dentro do esperado, ou seja, valores menores do que 1.

Para todos os modelos analisados, o valor do Leverage apresentou-se dentro do previsto, entre 0 e 1, o que indica que não existem casos exercendo influência indevida no modelo.

Ao avaliar os resíduos padronizados, observou-se que o Modelo 2 e o Modelo 3 apresentaram um valor acima do esperado, valor maior que  $\pm 3$ , indicando que devem ser analisados com cuidado. No Modelo 2, fez-se a retirada deste valor problemático e realizou-se novamente a regressão logística, resultando em um novo modelo com todas as estatísticas dentro dos valores esperados.

Já para o Modelo 3, ao fazer a retirada dos valores discrepantes, o novo modelo apresentou as estatísticas esperadas, porém a única variável, que era significativa, teve p-valor = 0,072. Ou seja, embora todos os dados tenham se ajustado adequadamente para as estatísticas da análise dos resíduos, não foi possível encontrar um modelo que explicasse o desempenho das escolas no IDEB.

Analisando o OR das equações apresentadas, de acordo com os dados obtidos na Tabela 7, verifica-se que não ter sala de leitura aumenta as chances de não alcançar a meta em 8,286 vezes. Para a covariável número de alunos reprovados, a chance é de 1,063 vezes maior de não atingir a meta do IDEB, o que mostra claramente o cuidado que se deve ter com o aumento do número de reprovações.

Os resultados do OR, na Tabela 9, indicam que o fato da escola possuir laboratório de ciências resulta, em 0,868 vezes, ou seja, 86,8% a mais de chance de se alcançar a meta, do que se a escola não possuísse. De modo semelhante, pode ser feita a análise para o número de alunos reprovados, o que dá uma chance de 1,101 vezes maior de não atingir a meta, do que se não houvesse reprovações.

Novamente, a covariável número de reprovações se mostra importante para o alcance da meta, reforçando a ideia de que se deve ter um cuidado especial com esta covariável.

Utilizando a equação 13, e o Modelo 1, considerando não ter sala de leitura e ter alunos reprovados, tem-se:  $\text{logit}(Y) = -0,393$ , portanto a probabilidade P (equação 15) de alcançar a meta do IDEB é igual a 40,3%.

Ainda, usando a equação 13 e o Modelo 2, considerando a situação de não ter laboratório de ciências e ter aluno reprovado, obtém-se  $\text{logit}(Y) = -2,832$ , que substituindo na equação 15, resulta na probabilidade (P) de 5,56% de alcançar a meta do IDEB. Em uma segunda situação, supondo que o número de alunos reprovados seja 15, e que a escola

possua laboratório de ciências, obtém-se  $\text{logit}(Y) = 0,652$ , o que resulta numa probabilidade de 65,75% de alcançar a meta.

Com a análise dos modelos, verifica-se que sendo o número de alunos reprovados maior do que zero, a probabilidade diminui, ou seja, quanto maior o índice de reprovação, menor a probabilidade de atingir a meta estabelecida para o IDEB. Portanto, deve-se ter atenção redobrada com esta covariável.

Os modelos logísticos encontrados têm semelhança com o que descreve Chirinéa (2010), em seu estudo realizado em duas cidades do estado de São Paulo, quando salienta que índice baixo de aprovação compromete o desempenho no IDEB, o que foi comprovado nos resultados encontrados. Logo, escolas que não apresentam reprovação, apesar de serem uma minoria, têm altas probabilidades de atingirem as metas.

## 5 CONSIDERAÇÕES FINAIS

Ao término deste estudo, pode-se concluir que foram atingidos os objetivos propostos, pois o desenvolvimento desta pesquisa possibilitou identificar algumas das variáveis que influenciam no baixo rendimento escolar e são refletidos nas notas do IDEB, e também para a compreensão dos procedimentos de cálculos envolvidos tanto na técnica de análise de regressão logística, como no cálculo do IDEB.

No que diz respeito ao desenvolvimento do ajuste dos modelos, para os dados das escolas investigadas do município de Santa Maria, nos anos de 2011 e 2013 no ensino fundamental, foi identificada a covariável número de reprovações como a principal causadora dos baixos índices do IDEB, o que corrobora com os resultados encontrados por outros autores, visto que, em sua síntese, o IDEB é um indicador de qualidade que considera rendimento escolar e desempenho.

Como se pode verificar, o presente trabalho mostrou que o número de reprovações é o mais importante, mas não o único fator responsável pelo rendimento escolar. As covariáveis sala de leitura e laboratório de ciências também contribuíram significativamente.

Com isso, acredita-se que estes resultados podem contribuir para o planejamento das políticas públicas da educação, e auxiliar as escolas na tomada de decisões a fim de melhorar o índice de qualidade.

Para finalizar, deixa-se como sugestão para o desenvolvimento de novas pesquisas; uma vez que existem diversos aspectos que não foram estudados, e que têm ligação com a qualidade da educação escolar; avaliar as variáveis: condições socioeconômicas, etnia e idade dos alunos, escolaridade dos pais, situação dos professores (efetivos e/ou temporários), práticas pedagógicas, entre outras, sendo necessário realizar uma pesquisa de campo para se obter tais informações que não são disponibilizadas pelos meios digitais.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

BELLO, S. E. L. **As práticas curriculares em Matemática que se produzem pelo governo do Ideb.** Horizontes. Itatiba, v. 30, n. 2, p. 19-30, jul./dez. 2012.

CHIRINÉA, A. M. **O índice de desenvolvimento da educação básica (IDEB) e as dimensões associadas à qualidade da educação na escola pública municipal.** 2010. 121p. Dissertação (Mestrado em Educação) – Universidade Estadual Paulista, Marília, 2010.

DOURADO, L. F.; OLIVEIRA, J. F.; SANTOS, C.A. **A qualidade da educação: conceitos e definições.** Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2007.

FERNANDES, R. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP, Ministério da Educação – MEC. **Índice de desenvolvimento da educação básica (Ideb): metas intermediárias para a sua trajetória no Brasil, Estados, Municípios e Escolas.** Brasília. [2007?a].

\_\_\_\_\_. **Índice de desenvolvimento da educação básica (Ideb).** Instituto Nacional de Estudos e Pesquisas Educacionais ‘Anísio Teixeira’ – INEP, Ministério da Educação – MEC. (2007b).

FIELD, A. **Descobrimo a ESTATÍSTICA usando o SPSS.** 2.ed. Porto Alegre: Artmed, 2009.

FREITAS, L. C. de. **Eliminação adiada: o caso das classes populares no interior da escola e a ocultação da (má) qualidade do ensino.** Educação & Sociedade. Campinas, v. 28, n. 100, p. 965-987, abr. 2007.

GOUVEIA, A. B.; SOUZA, A. R.; TAVARES, T. M. **O Ideb e as políticas educacionais na região metropolitana de Curitiba.** Est. Aval. Educ. São Paulo, v. 20, n. 42, p. 45-48, jan./abr. 2009.

HAIR JR, J.F. et al. **Análise Multivariada de Dados.** 6 ed. Porto Alegre: Bookman, 2009.

HOSMER, Jr. D.W.; LEMESHOW, S. **Applied Logistic Regression.** New York: John Wiley, 1989.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA. **Cidades**. Disponível em: <<http://www.cidades.ibge.gov.br>>. Acesso em 06 mai. 2015.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Portal Ideb**. Brasília, 2015. Disponível em: <<http://portal.inep.gov.br/web/portal-ideb>>. Acesso em 12 mar. 2015.

PESSOA, H. C. **O desempenho dos alunos do ensino fundamental dos municípios de Rondônia**: uma abordagem a partir dos sistemas de avaliações interna e externa. In: TRIBUNAL de Contas do Estado de Rondônia. [Rondônia], 20--?. Disponível em: <<http://http://www.tce.ro.gov.br/nova/publicacoes/>>. Acesso em: 12 mar. 2015.

PINTO, J. M. R. **O custo de uma educação de qualidade**. In: CORREA, B. C; GARCIA, T. O. Políticas educacionais do trabalho da escola. São Paulo: Xamã, 2008.

QUEDU. **Ideb**. Disponível em: <<http://www.qedu.org.br>>. Acesso em 12 mar. 2015.

RIOS, T. A. **Compreender e ensinar**: por uma docência da melhor qualidade. 2 ed. São Paulo: Cortez, 2001.

SOARES, J. F. **Qualidade e equidade na educação básica brasileira**: a evidencia do Saeb. Arquivos Analíticos de Políticas educativas, v. 12, n. 38, p. 1-24, 2004.

VIDAL, E. M; VIEIRA, S. L. **Gestão educacional e resultados no Ideb**: um estudo de caso em dez municípios cearenses. Est. Aval. Educ. São Paulo, v. 22, n. 50, p. 419-434, set./dez. 2011.