

4.

METODOLOGIA

O presente capítulo apresenta alguns métodos para a transformação dos dados em resultados relevantes para análise. Iniciamos com a metodologia de extração das escalas, que valida os itens motivacionais do instrumento de coleta de dados do aluno. Este procedimento permite a manipulação de variáveis motivacionais para análises posteriores.

Em seguida, utilizou-se a regressão linear múltipla para verificação das variáveis fortemente associadas à motivação, esta operação permite inferir sobre bons preditores das variáveis motivacionais.

Por fim, a regressão logística binária é utilizada com o objetivo de estudar o desfecho escolar favorável (variável dicotômica) em associação com variáveis demográficas, motivacionais, de práticas culturais e de habilidades com uso de computador.

4.1 Metodologia de extração das escalas

Para a operacionalização do conceito de Motivação empregamos instrumentos utilizados numa pesquisa conduzida por Luc G. Pelletier (2005), no Canadá, com duração de três anos, numa perspectiva longitudinal. Tal instrumento confirmou a existência de escalas de motivação propostas pela Teoria da Autodeterminação. Originalmente os itens para “The Academic Motivation Scale” (escala de motivação acadêmica) foram desenvolvidos por Robert Vallerand e colaboradores (1993).

O estudo de Pelletier investigou as mudanças motivacionais dos alunos durante a passagem de nível de ensino – o equivalente à conclusão do Ensino Fundamental (9º ano) e o ingresso nos dois anos subsequentes do Ensino Médio, numa amostra de 646 alunos. Um dos objetivos explícitos da pesquisa foi a replicação de resultados obtidos em pesquisas anteriores, baseados na Teoria da Autodeterminação, especificando correlações entre as várias escalas de motivação. Uma segunda finalidade da pesquisa foi identificar os padrões de variação das mudanças do perfil motivacional dos alunos durante o período estudado, comparando duas séries subsequentes. Os resultados encontrados não só corroboram o que as pesquisas anteriores já constatarem – à medida que os alunos avançam em sua escolaridade, sua motivação intrínseca diminui –, como também contribuíram para o entendimento dos padrões de mudança nos perfis motivacionais das dimensões da motivação extrínseca e da desmotivação.

Ainda neste artigo, é possível perceber que os autores trabalharam com três formas de regulação da motivação extrínseca, dimensões que foram ajustadas pelos autores da Teoria da Autodeterminação em um artigo em resposta a um debate sobre as convergências e incongruências dos achados das pesquisas e a literatura sobre motivação. Deci, Ryan e Koestner (1999) afirmam em tal artigo que a proposta da microteoria OIT (Teoria das Orientações Causais) é que a motivação extrínseca pode ser concebida por três formas diferentes de regulação: i) por uma regulação externa no caso do comportamento ser controlado por pressões totalmente externas à pessoa; ii) por regulação introjetada, quando o comportamento é controlado por contingências que têm sido internalizadas mas

não integradas e iii) por regulação totalmente integrada (ou assimilada), quando a pessoa se identifica de tal forma com a razão da sua ação, integrando essa forma de regulação, mas mesmo assim, como discutido no capítulo anterior, esse perfil motivacional ainda constitui uma das bases da motivação extrínseca.

Esse novo desenho da concepção das formas de regulação da motivação extrínseca reforça a construção do instrumento de coleta de dados que utilizamos na pesquisa.

Com inspiração no instrumento utilizado por Pelletier et al (2005), o bloco temático que procurou captar o perfil motivacional é apresentado no quadro 2.

Quadro 2. Itens utilizados no instrumento para medição da Motivação

MOTIVAÇÃO INTRÍNSECA	item
1) Porque sinto prazer e satisfação em aprender coisas novas .	23
2) Pelo prazer que sinto quando descubro coisas que eu nunca tinha visto.	16
3) Pelo prazer que sinto em saber mais sobre assuntos que me agradam .	02
4) Porque os estudos me permitem continuar a aprender coisas que me interessam.	07
MOTIVAÇÃO EXTRÍNSECA COM REGULAÇÃO ASSIMILADA	
1) Porque penso que estudar vai me preparar melhor para a carreira que eu escolher.	24
2) Porque eventualmente vai me permitir encontrar um emprego em uma área que eu goste.	01
3) Porque isso vai me ajudar a fazer uma melhor escolha com relação a trabalhos futuros.	11
4) Porque acredito que mais estudo vai aumentar minha competência no mercado de trabalho.	13
MOTIVAÇÃO EXTRÍNSECA COM REGULAÇÃO INTROJETADA	
1) Para mostrar a mim mesmo que sou uma pessoa inteligente.	10
2) Porque quero mostrar a mim mesmo que posso ser bem sucedido no estudo.	12
3) Para provar a mim mesmo que posso terminar o ensino fundamental.	17
4) Porque o fato de ser bem sucedido na escola me faz sentir importante.	21
MOTIVAÇÃO EXTRÍNSECA COM REGULAÇÃO EXTERNA	
1) Porque sem estudo não vou encontrar um emprego que pague bem.	09
2) Para conseguir um emprego melhor no futuro.	18
3) Porque quero levar uma vida mais confortável no futuro.	22
4) Para conseguir, mais tarde, um melhor salário.	03
DESMOTIVAÇÃO	
1) Já tive bons motivos para ir à escola, mas agora me pergunto se vale a pena continuar.	04
2) Tenho dúvidas, não consigo entender o que estou fazendo na escola.	19
3) Não consigo entender porque vou à escola, e francamente, não me interessa.	14
4) Honestamente tenho a impressão que é uma perda de tempo ir à escola.	08

Além dos itens utilizados da pesquisa canadense, a literatura aponta (Finn 1989; Jenkins 1995) que o sentimento de pertencimento ao lugar com o qual se estabelece uma relação mais próxima pode ser uma manifestação da motivação. O

questionário contextual do PISA¹ (Programme for International Student Assessment) para os alunos utiliza alguns itens que medem esse constructo. Na tentativa de captar a motivação com mais essa possível manifestação, decidimos incluir cinco itens relacionados ao sentimento de pertencimento e que estivessem também associados às dimensões ora descritas. O quadro 3 mostra a descrição dos itens.

Quadro 3: Descrição dos itens incorporados do PISA.

NUMERAÇÃO DO ITEM	DIMENSÃO DA MOTIVAÇÃO/DESMOTIVAÇÃO	DESCRIÇÃO DO ITEM
Item 20	Motivação Intrínseca	Porque na escola me sinto à vontade
Item 6	Motivação extrínseca com regulação assimilada	Porque na escola estão os meus melhores amigos.
Item 5	Motivação extrínseca com regulação introjetada	Estudar me deixa entediado
Item 15	Motivação extrínseca com regulação externa	Vou à escola obrigado, me sinto um estranho lá dentro.
Item 25	Desmotivação	Vou à escola, mas sinto que ela não foi feita para mim.

Tendo o aporte da Teoria para a construção das dimensões do conceito Motivação, a tarefa que se segue apresentará os caminhos para a validação dos itens e a escolha das escalas para as modelagens.

Escalas são dispositivos de redução de dados que permitem a construção de uma medida mais exata e abrangente daquilo que se quer medir, tendo em vista que “várias respostas de um respondente podem ser resumidas num único escore, e mesmo assim os detalhes específicos daquelas respostas são mantidos quase que na totalidade” (Babbie 2001, p. 214).

A pesquisa maior na qual este trabalho está ancorado é responsável pelo processo de validação dos itens presentes em todos os instrumentos utilizados (questionário contextual do aluno, do professor e do diretor). Porém, a escolha das escalas para a análise dos modelos poderá ser modificada de acordo com o interesse de cada pesquisador. Por isso, este capítulo apresenta os procedimentos e justificativas para utilização das escalas de motivação.

¹ PISA (Programme for International Student Assessment) é um programa internacional, de iniciativa da OCDE para avaliação de estudantes com 15 anos. Trata-se de um instrumento padronizado para aferir o conhecimento dos estudantes nas áreas de matemática, leitura e ciências (www.pisa.oecd.org).

Para a extração de escalas utilizamos a metodologia de Teoria de Resposta ao Item (TRI), que verifica a relação entre o traço latente do indivíduo respondente (no caso, a motivação) e a probabilidade deste indivíduo marcar uma resposta a um item do questionário (uma dentre as sete opções de concordância). Tal probabilidade está associada a características do indivíduo e a características do item proposto. Além disso, utilizamos a TRI não paramétrica para examinar o quanto a escala construída se afasta da escala perfeita (escala determinística de Guttman), por meio do coeficiente H de Loevinger. Tanto este coeficiente quanto os outros utilizados na TRI não paramétrica são explicados a seguir.

O software utilizado para este procedimento é o MSP (*Mokken Scale Procedure*), que constitui um programa para análise de coeficientes de escalonabilidade e confiabilidade de escalas produzidas por itens politômicos. Além dos coeficientes o programa oferece duas possibilidades de análise:

Procedimento de análise exploratória – método que permite a extração de escalas de modo automático a partir da inserção de vários itens. Deste modo, o próprio programa estabelece as dimensões encontradas. Cabe ao pesquisador a análise dos coeficientes e a interpretação dos arranjos construídos.

Procedimento de análise confirmatória – método que permite verificar a consistência de uma escala a partir de um conjunto de itens selecionados a priori. Normalmente o pesquisador já possui algumas hipóteses de arranjos que necessitam ser validados. A extração das escalas utilizando esse procedimento ainda necessita de verificação dos coeficientes e em alguns casos, novos arranjos, como a exclusão de alguns itens para aumentar a confiabilidade da escala.

Utilizaremos os dois procedimentos para uma análise mais consistente das escalas a serem utilizadas nos modelos utilizando a mesma amostra de dados. Como resposta à questão subsidiária desta pesquisa, os resultados apresentados têm como finalidade verificar se as dimensões existentes no instrumento utilizado para medir a motivação dos alunos são úteis e confiáveis do ponto de vista estatístico.

Para que as escalas sejam consideradas boas medidas de captação do conceito latente, utilizamos alguns coeficientes de ajuste da Teoria da Resposta ao Item não paramétrica e da Teoria Clássica dos testes. Cada coeficiente fornece um escore de escalonabilidade ou confiabilidade para se utilizar a escala. De acordo com os coeficientes encontrados, é possível retirar ou incorporar itens de forma a

ajustar a escala, o que será feito tanto na análise exploratória como na análise confirmatória.

Coeficiente de escalonabilidade de *H de Loewinger*

O *H* de Loewinger é um dos coeficientes extraídos do MSP, portanto, um coeficiente de ajuste da teoria de resposta ao item não paramétrica, e pode ser entendido como uma medida que expressa o quanto a escala se afasta da escala perfeita de Guttman, tendo como referência o modelo de Mokken (1971). O “*H*” de Loewinger mede o grau de associação entre um par de itens e, entre o item e os outros que formam a escala. Cada item possui seu coeficiente e também a escala como um todo. Mokken assume que as escalas podem ser caracterizadas de acordo com o quadro 4.

Quadro 4: Legenda do Coeficiente de Escalonabilidade Rho de Moken.

Coeficientes	Nível de Escalonabilidade
Entre 0,5 e 1	Forte
Entre 0,4 e 0,5	Médio
Entre 0,3 e 0,4	Discreto
Menor que 0,3	Sem escalonabilidade

Fonte: Mokken, 1971.

O coeficiente Rho de Mokken, também extraído por meio do MSP, está relacionado ao número de itens na escala e é associado a sua confiabilidade, isto é, com o quanto de confiança posso atribuir à escala o quanto mede o traço latente. Ou seja, se uma escala possui coeficiente igual a 0,67 por exemplo, isto significa que do total do percentual da variabilidade encontrado nas correlações entre os pares de itens, 67% está captando o traço latente e o restante é ruído na mensuração. Rho de Mokken caracteriza as escalas em fracas, médias ou fortes, de acordo com a segurança com que se podem ordenar os itens em relação ao traço latente que ele se propõe medir. As referências podem ser vistas no quadro 5.

Quadro 5: Legenda dos coeficientes de confiabilidade Rho de Mokken.

Coeficientes	Nível de confiabilidade
Entre 0,7	Forte
Entre 0,6 e 0,7	Média
Menor que 0,6	Fraca

Coeficiente de confiabilidade Alpha de Cronbach

O Alpha de Cronbach é o coeficiente da teoria clássica dos testes que fornece a consistência interna baseada na média de correlação dos itens e é fornecido pelo SPSS (*Statistical Package for the Social Sciences*)². O escore produzido é parecido com o Rho de Mokken porque ambos têm a mesma finalidade e padrões de referência.

² O SPSS é um software de organização de dados e análise estatística para as ciências sociais (www.spss.com.br).

4.2 Regressão Linear Múltipla

A regressão linear é uma ferramenta estatística usada para analisar as relações entre uma variável dependente e várias variáveis preditoras ou independentes (variáveis tidas como previsoras ou potencialmente explicativas). O objetivo da análise de regressão linear é utilizar as variáveis independentes cujos valores são conhecidos para prever os valores da variável dependente. No caso específico da pesquisa, nosso interesse é avaliar que variáveis podem incidir (positiva ou negativamente) nas escalas de motivação, ou seja, nos interessa explicar que fatores, habilidades, condições ou características apreendidas no questionário dos alunos podem explicar, de alguma forma, a motivação. Segundo Hair *et al* (2009) “cada variável independente é ponderada pelo procedimento de regressão para garantir a máxima previsão a partir do conjunto de variáveis independentes” (p. 154). Uma das condições para a realização desta técnica é que a variável dependente seja métrica, ou seja, assuma valores contínuos na base de dados. No caso das variáveis independentes, é possível codificar variáveis nominais e ordinais em variáveis dicotômicas (*dummies*).

Quando a análise de regressão linear envolve somente uma variável explicativa, a técnica é denominada Regressão Linear Simples. Quando são envolvidas duas ou mais variáveis explicativas, emprega-se a expressão Regressão Linear Múltipla.

O modelo de regressão é normalmente descrito da seguinte forma:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \dots + \beta_i x_{ii} + e_i$$

Onde:

Y_i representa o valor da variável dependente;

$\beta_0, \beta_1, \beta_2, \dots, \beta_i$ são os coeficientes de regressão;

$X_1 + X_2 + X_3 \dots X_i +$ são as variáveis independentes;

e_i é o termo de erro³;

$i=1,2,\dots, n$ (número de casos).

³ O termo erro é o padrão de estimativa, isto é, o desvio padrão dos valores da variável dependente em relação à reta de regressão estimada. Pode-se dizer que quanto menor a sua estimativa, melhor é o modelo.

Para este trabalho, um dos modelos de regressão será assim representado:

$$MOT_I_i = \beta_0 + \beta_1 Hab_tec + \beta_2 Uso_edu + \beta_3 ptl_cul + \beta_4 bens + \beta_5 viol_alun + \beta_6 + e_i$$

Onde:

MOT_I_i é a variável dependente, ou seja, a motivação intrínseca do aluno.

$\beta_0, \beta_1, \beta_2, \dots, \beta_k$ são os coeficientes de regressão;

Hab_tec é uma das variáveis independentes – habilidade técnica no uso de computadores;

Uso_edu, **Ptl_cul**, **vio_anul** são outros exemplos de variáveis independentes na base de dados.

e_i é o termo de erro;

$i=1,2,\dots, n$ (número de casos).

Os modelos de regressão foram analisados no programa estatístico SPSS (*Statistical Package for the Social Sciences*).

O primeiro coeficiente a ser determinado é o R^2 (obtido na primeira tabela do *output - Model Sumary*). Sua função é medir a qualidade do ajuste do modelo de regressão ao explicar a variável dependente, ou seja, ele mede a proporção da variação de Y em relação à média que é explicada pela regressão. Em princípio, a qualidade do ajuste será tanto maior quanto mais o R^2 se aproximar de uma unidade. Utilizaremos um exemplo para esclarecer melhor estas idéias. Considerando a tabela 5 e 6, podemos dizer que as duas variáveis independentes (PTL_TELE (práticas de assistir a telejornais e documentários no tempo livre) e USO_EDU (uso educacional do computador) conseguem explicar 65% (R Square) de toda a variabilidade dos dados, considerando a motivação intrínseca como variável dependente.

Tabela 5: Exemplo de Sumário de um Modelo de Regressão

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,255(a)	,065	,065	5,01228

a Predictors: (Constant), PTL_TELE, USO_EDU

Tabela 6: Exemplo de Tabela de Coeficientes de um Modelo de Regressão

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	29,507	,336		87,846	,000
	USO_EDU	,742	,071	,170	10,451	,000
	PTL_TELE	,611	,059	,168	10,378	,000

a Dependent Variable: "Motivação intrínseca do aluno para o aprendizado na escola" (escala soma das variáveis 1.02, 1.07, 1.11, 1.13, 1.16, 1.23 e 1.24)

A tabela 6 (*Coefficients*) é uma tabela de análise dos coeficientes (coluna B) e teste de significância (coluna sig). Os coeficientes estimados aqui dizem respeito à confiabilidade do modelo para ser generalizado. Há duas formas complementares de analisar a confiabilidade do modelo, a primeira se dá pela observação da magnitude do coeficiente B – quanto maior o valor do coeficiente, mais chance de o erro amostral ser pequeno e o valor positivo ser para toda a população. A segunda forma é observando o valor do teste *t* que indica o erro amostral do modelo. Quanto menor o erro amostral, maior a confiança em generalizar. O teste *t* depende principalmente do tamanho da amostra e é dado pela divisão entre o coeficiente B e o erro amostral (Std Error). A significância (sig) nos mostra o quão confiável a variável pode ser. O padrão para a análise pode ser visto no quadro 6:

Quadro 6: Padrão de Significância de Modelos de Regressão

P < 0,001	Significância a 0,1%	Posso generalizar para 99,99% dos casos
P < 0,01	Significância a 1%	Posso generalizar para 99% dos casos
P < 0,05	Significância a 5%	Posso generalizar para 95% dos casos
P < 0,10	Significância a 10%	Posso generalizar para 90% dos casos
P > 0,10	Não significativa	

Considerando o modelo da tabela 6, percebemos que a motivação intrínseca do aluno com uso educacional médio e prática de assistir TV média é de 29, 50 pontos. Essa variável tem significância de 0,000 (ou seja, posso generalizar esse dado para toda a população com confiança de 99,9% de chance de acerto). As variáveis independentes explicam 65,7% da variância (R Square da tabela 10 - Model Sumary) e indicam que o aumento de 1 unidade no escore médio do uso educacional do computador está associado ao aumento de, em média, 0,7 pontos na motivação intrínseca dos alunos, com ótima significância. Em relação à prática de assistir TV no tempo livre (telejornais e documentários), o aumento de 1

unidade nessa prática está associado ao aumento de, em média, 0,6 pontos na motivação intrínseca dos alunos. Este é um exemplo fictício com apenas duas variáveis para compreensão dos coeficientes estimados. As análises que compõem os modelos necessitam de testes e a introdução de mais variáveis para verificação das estimativas.

Todas as variáveis que entraram nos modelos foram padronizadas tendo em vista que as variáveis podem ter escalas e dispersões muito diferentes. A padronização corresponde a uma transformação para a média nula e o desvio padrão unitário de cada variável original.

4.3 Medidas Utilizadas

Para conhecer todas as variáveis que farão parte dos modelos, descrevemos no quadro 7: a sigla utilizada no procedimento, o seu significado, o tipo de variável (dicotômica ou contínua) e uma descrição mais extensa do modo como foi construída a partir dos itens.

Quadro 7: Descrição das Variáveis Utilizadas no Modelo de Regressão

Sigla	Variável	Tipo e Codificação	Descrição
VARIÁVEIS DEPENDENTES			
Mot_i	Motivação Intrínseca	Contínua	Variável de motivação intrínseca com 7 itens da análise exploratória
Mot_e	Motivação Extrínseca	Contínua	Variável de motivação extrínseca com 7 itens da análise exploratória
Mot_d	Desmotivação	Contínua	Variável de desmotivação com 7 itens da análise exploratória
Mi_5	Motivação Intrínseca por análise confirmatória	Contínua	Variável de motivação intrínseca com 4 itens da análise confirmatória
Mi_a5	Motivação Extrínseca por regulação assimilada - análise confirmatória	Contínua	Variável de motivação extrínseca por regulação assimilada com 3 itens da análise confirmatória
Mi_i5	Motivação Extrínseca por regulação introjetada - análise confirmatória	Contínua	Variável de motivação extrínseca por regulação introjetada com 4 itens da análise confirmatória
Mi_e5	Motivação Extrínseca por regulação externa - análise confirmatória	Contínua	Variável de motivação extrínseca por regulação externa com 3 itens da análise confirmatória
Md_5	Desmotivação por análise confirmatória	Contínua	Variável de desmotivação com 4 itens da análise confirmatória

Continuação do Quadro 7: Descrição das Variáveis Utilizadas no Modelo de Regressão

VARIÁVEIS INDEPENDENTES			
FP_d	Estrato escola pequena próxima a favela	dicotômica	Variável <i>dummie</i> representando o estrato escola pequena próxima à favela em relação ao estrato escola grande próxima à favela
nFG_d	Estrato escola grande distante de favela	dicotômica	Variável <i>dummie</i> representando o estrato escola grande distante de favela em relação ao estrato escola grande próxima à favela
nFP_d	Estrato escola pequena distante de favela	dicotômica	Variável <i>dummie</i> representando o estrato escola pequena distante de favela em relação ao estrato escola grande próxima a favela
PM_d	Estrato escola pólo	dicotômica	Variável <i>dummie</i> representando o estrato escola pólo em relação ao estrato escola grande próxima à favela
Fem_d	menina	dicotômica	Variável <i>dummie</i> representando o estudante ser menina em relação ao ser do sexo masculino
Bran_d	branco	dicotômica	Variável <i>dummie</i> representando o estudante ter se autodeclarado branco em relação a ter se autodeclarado pardo.
Amar_d	amarelo	dicotômica	Variável <i>dummie</i> representando o estudante ter se autodeclarado amarelo em relação a ter se autodeclarado pardo.
Pret_d	preto	dicotômica	Variável <i>dummie</i> representando o estudante ter se autodeclarado preto em relação a ter se autodeclarado pardo.
Av1_d	Avançado 1 ano	dicotômica	Variável <i>dummie</i> representando o aluno que está 1 ano avançado na série em relação à idade ideal para o 9º ano.
Av2_d	Avançado 2 anos	dicotômica	Variável <i>dummie</i> representando o aluno que está 2 anos avançado na série em relação à idade ideal para o 9º ano.
At1_d	Atrasado 1 ano	dicotômica	Variável <i>dummie</i> representando o aluno que está 1 ano atrasado na série em relação à idade ideal para o 9º ano.

Continuação do Quadro 7: Descrição das Variáveis Utilizadas no Modelo de Regressão

At2_d	Atrasado 2 anos	dicotômica	Variável <i>dummie</i> representando o aluno que está 2 anos atrasado na série em relação à idade ideal para o 9º ano.
At3_d	Atrasado 3 anos	dicotômica	Variável <i>dummie</i> representando o aluno que está 3 anos atrasado na série em relação à idade ideal para o 9º ano.
M0an_d	Mãe sem escolaridade	dicotômica	Variável <i>dummie</i> representando a escolaridade da mãe com nenhum ano de estudo em relação à mãe com Ensino Médio.
M5an_d	Mãe com cinco anos de escolaridade	dicotômica	Variável <i>dummie</i> representando a escolaridade da mãe com cinco anos de estudo (1º segmento do Ensino Fundamental) em relação à mãe com Ensino Médio.
M9an_d	Mãe com nove anos de escolaridade	dicotômica	Variável <i>dummie</i> representando a escolaridade da mãe com nove anos de estudo (2º segmento do Ensino Fundamental) em relação à mãe com Ensino Médio.
Msup_d	Mãe com ensino superior	dicotômica	Variável <i>dummie</i> representando a escolaridade da mãe com ensino superior em relação à mãe com Ensino Médio.
P0an_d	Pai sem escolaridade	dicotômica	Variável <i>dummie</i> representando a escolaridade do pai com nenhum ano de estudo em relação ao pai com Ensino Médio.
P5an_d	Pai com cinco anos de escolaridade	dicotômica	Variável <i>dummie</i> representando a escolaridade do pai com cinco anos de estudo (1º segmento do Ensino Fundamental) em relação ao pai com Ensino Médio.
P9an_d	Pai com nove anos de escolaridade	dicotômica	Variável <i>dummie</i> representando a escolaridade do pai com nove anos de estudo (2º segmento do Ensino Fundamental) em relação ao pai com Ensino Médio.
Psup_d	Pai com ensino superior	dicotômica	Variável <i>dummie</i> representando a escolaridade do pai com ensino superior em relação ao pai com Ensino Médio.
Usac_d	Uso do computador	dicotômica	Variável <i>dummie</i> do uso do computador em relação ao não uso.

Continuação do Quadro 7: Descrição das Variáveis Utilizadas no Modelo de Regressão

Me1a_d	Usa computador há menos de um ano	dicotômica	Variável <i>dummie</i> do uso do computador há menos de um ano.
Ma1a_d	Usa computador há mais de um ano	dicotômica	Variável <i>dummie</i> do uso do computador há mais de um ano.
Ma3a_d	Usa computador há mais de três anos	dicotômica	Variável <i>dummie</i> do uso do computador há mais de três anos.
Usa_escola	Uso do computador na escola	dicotômica	Variável que indica se o estudante usa computador na escola
Usa_casa	Uso do computador em casa	dicotômica	Variável que indica se o estudante usa computador em casa
Usa_amigos	Uso do computador na casa de amigos	dicotômica	Variável que indica se o estudante usa computador na casa dos amigos
Usa_locpub	Uso do computador em locais públicos	dicotômica	Variável que indica se o estudante usa computador em locais públicos como lan houses, cafés, etc.
Uso_edu	Frequência de uso educacional do computador	contínua	Variável que indica frequência do uso educacional do computador pelos estudantes, como usar programas educacionais, escrever trabalhos para a escola, pesquisar na internet para aprender coisas novas etc..
Uso_tec	Frequência de uso técnico do computador	contínua	Variável que indica frequência do uso técnico do computador pelos estudantes, como postar vídeos digitais, baixar programas pela internet, baixar fotos da câmera digital etc..
Uso_soc	Frequência de uso social do computador	contínua	Variável que indica frequência do uso social do computador pelos estudantes como freqüentar rede social, comunicar-se por skype ou MSN, ler e enviar mensagens de e-mail, etc.
Hab_edu	Habilidade educacional do computador	contínua	Variável que indica a habilidade no uso educacional do computador, como criar uma apresentação multimídia, criar um blog, usar programa de apresentação (Word) etc.

Continuação do Quadro 7: Descrição das Variáveis Utilizadas no Modelo de Regressão

Hab_tec	Habilidade técnica do computador	contínua	Variável que indica a habilidade no uso técnico do computador, como instalar periféricos, editar imagens digitais, mover arquivos de um lugar a outro etc.
Hab_soc	Habilidade social do computador	contínua	Variável que indica a habilidade no uso social do computador como bater papo <i>on line</i> com voz e imagens, escrever em “internetês”, bater papo <i>on line</i> com mensagens escritas, etc.
Ptl_culti	Prática cultural cultivada	contínua	Variável que indica as práticas culturais cultivadas no tempo livre, como ouvir música, ler jornal ou revista, ler livro, ir a museu, ir ao teatro etc.
Ptl_lazer	Prática de lazer	contínua	Variável que indica a participação em atividades de lazer, como ir a festas, ao shopping, à praia, ao cinema, a bares etc.
Ptl_sport	Prática esportiva	contínua	Variável que indica a participação em atividades esportivas, como praticar atividades físicas, assistir a jogos esportivos na televisão, ler a parte esportiva dos jornais etc.
Ptl_celul	Prática do uso do celular	contínua	Variável que indica a frequência do uso do celular no tempo livre, caracterizado em atividades como usar o celular para telefonar, tirar fotos, jogar, enviar mensagens etc.
Ptl_relig	Prática de atividades religiosas	contínua	Variável que indica a frequência de participação em atividades religiosas como ir a missa/culto e ler a bíblia.
Ptl_telev	Frequência em assistir televisão	contínua	Variável que indica a frequência com que costuma assistir a programas na televisão como jornais e documentários.
Disp_midia	Disponibilidade de mídia em casa	contínua	Variável que indica a disponibilidade de recursos de mídia na casa do aluno como televisão, aparelho de DVD, computador etc.
Disp_jorn	Disponibilidade de informação escrita em casa	contínua	Variável que indica a disponibilidade de jornais de informação geral na casa do aluno, como jornais e/ou revistas de informação geral (Veja, Isto é, Época etc.) e assinaturas.

Continuação do Quadro 7: Descrição das Variáveis Utilizadas no Modelo de Regressão

Disp_livro	Disponibilidade de literatura e revistas em casa	contínua	Variável que indica a disponibilidade de revistas de divulgação científica/livros de literatura na casa do aluno.
Bens	Posse de bens	contínua	Variável que indica a posse de bens da família do aluno, como banheiro, telefone fixo, máquina de lavar roupa, geladeira duplex, câmera fotográfica digital, câmera de vídeo, automóvel, motocicleta.
Viol_alun	Violência ao aluno	contínua	Variável que indica a presença de atitudes ou atos de violência em relação aos alunos, tais como agressão verbal, agressão física, roubo etc.
Viol_prof	Violência ao professor	contínua	Variável que indica a presença de atitudes ou atos de violência em relação aos professores, tais como presenciar uso de bebidas alcoólicas, drogas ou portando armas.

4.4 Regressão Logística Binária

Assim como a regressão linear, a Regressão Logística Binária examina a relação entre uma variável dependente (que necessita de explicação) e uma ou mais variáveis independentes. A diferença entre estas técnicas de regressão está no fato de que na regressão logística as variáveis dependentes estão dispostas em categorias (sim/não/alto/baixo/bom/ruim), enquanto na regressão linear estas variáveis são contínuas. Outra diferença está na resposta da regressão logística que é expressa por meio de uma probabilidade de ocorrência ($\text{Exp}(B) = \text{odds ratio}$), enquanto na regressão linear, obtém-se um valor numérico.

Na regressão logística as variáveis independentes (preditoras ou explicativas) podem ser tanto variáveis dicotômicas quanto dados contínuos e as variáveis dependentes poderão estar dispostas em duas ou mais categorias (ordinária e nominal), daí ser denominada Regressão Logística Binária.

Para compreender a leitura dos coeficientes dos *outputs* gerados por este procedimento, vamos exemplificar com uma análise em que a variável independente seja o desfecho, conforme é o foco desta pesquisa e as duas variáveis preditoras são posse de bens e a mãe com cinco anos de escolaridade.

Tabela 7: Exemplo de tabela do output da regressão das variáveis não presentes no modelo.

			Score	df	Sig.
Step 0	Variables	M5AN_D	23,371	1	,000
		DEFASAG	6,627	1	,010
		BENS	19,907	1	,000
		PTL_CELU	,806	1	,369
	Overall Statistics		41,239	4	,000

A última linha da tabela 7 (*overall statistics*) apresenta a estatística qui-quadrado dos resíduos como 41,239 que é totalmente significativa ($p < 0,001$). Esta estatística indica que os coeficientes para as variáveis que não estão no modelo são significativamente diferentes de zero, ou seja, que a adição de uma ou mais destas variáveis ao modelo irá afetar significativamente o seu modelo de previsão; se o p fosse $> 0,05$, isto significaria que nenhuma das variáveis excluídas do

modelo contribuiria de forma significativa para o poder preditivo do modelo e assim, a análise terminaria neste estágio.

A coluna Score oferece o valor da estatística do Escore Eficiente de Roa. Observamos aqui que (M5AN_D – mãe com cinco anos de escolaridade), seguido pela posse de bens apresentam maior escore eficiente de Roa significativos. Defasagem possui um escore mais discreto com significância também discreta.

Tabela 8: Exemplo de tabela do output da regressão das variáveis presentes no modelo.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	M5AN_D	-,656	,177	13,751	1	,000	,519
	DEFASAG	,204	,088	5,375	1	,020	1,226
	BENS	1,127	,317	12,655	1	,000	3,085
	PTL_CELU	-,009	,054	,030	1	,861	,991
	Constant	1,544	,326	22,384	1	,000	4,683

a Variable(s) entered on step 1: M5AN_D, DEFASAG, BENS, PTL_CELU.

A tabela 8 apresenta a estimativa dos coeficientes das variáveis explicativas incluídas no modelo.

Um coeficiente também importante é a estatística de Wald que tem uma distribuição qui-quadrado e indica se o coeficiente B para qualquer das variáveis explicativas difere significativamente de zero. Se isto ocorrer, podemos dizer que o preditor está contribuindo de modo significativo para a explicação do modelo. Deste modo, com relação aos dados do exemplo, podemos dizer que ter mãe com apenas cinco anos de escolaridade é um preditor significativo para um desfecho desfavorável e que a posse de bens (ou nível sócio econômico acima da média dos alunos) representa um fator positivo e importante para um desfecho favorável. A *Exp (b)* é interpretada como uma mudança nas chances: se o valor for maior que 1, ele indica que à medida que o preditor aumenta, aumentam as chances de um desfecho favorável ocorrer; já um valor menor que 1, indica que à medida que o preditor aumenta, as chances de um desfecho favorável ocorrer diminuem. Neste caso, podemos afirmar que as chances do aluno cuja mãe tem cinco anos de escolaridade ter um desfecho favorável reduz em 48,1% (*Exp (B)* de 0,519: 1-0,481) em relação ao aluno cuja mãe tem ensino médio. Já a posse de bens constituiu um fator positivo, aumentando em 1,2 vezes mais a chance de um desfecho favorável em relação ao aluno com posse de bens na média.

Multicolinearidade

A multicolinearidade é observada quando existe uma forte correlação entre duas ou mais variáveis explicativas em um modelo de regressão, o que dificulta a avaliação da importância individual de um preditor. Se ambos preditores são altamente correlacionados e cada um é responsável por uma variância similar dos resultados, como é possível afirmar qual das duas variáveis é a mais importante, uma vez que o modelo pode incluir qualquer uma delas de forma intercambiável?

Um dos mecanismos para se verificar esse dado é fazer um diagnóstico de colinearidade para obtenção de estatísticas, obtidos por meio da tolerância e do Fator de Inflação da Variância (FIV). Para adiantar a verificação da multicolinearidade, executamos o procedimento, gerando os resultados na tabela 9.

Tabela 9: Análise de tolerância e Fator de Inflação da Tolerância da Regressão Logística Binária

Model		Collinearity Statistics	
		Tolerance	VIF
1	USA_ESCO	,966	1,035
	USA_CASA	,457	2,187
	USA_AMIG	,867	1,154
	USA_LOCP	,760	1,316
	USO_EDU	,514	1,947
	USO_TEC	,399	2,504
	USO_SOC	,539	1,854
	HAB_EDU	,377	2,651
	HAB_TEC	,311	3,215
	HAB_SOC	,532	1,881
	PTL_CULT	,640	1,563
	PTL_LAZE	,586	1,706
	PTL_SPOR	,801	1,248
	PTL_CELU	,743	1,346
	PTL_RELI	,864	1,158
	PTL_TELE	,850	1,176
	DISP_MID	,489	2,044
	DISP_JOR	,848	1,179
	DISP_LIV	,862	1,160
	BENS	,774	1,293
	VIOL_ALU	,946	1,057
	V258_A	,957	1,045
	DEFASAG	,954	1,048

a Dependent Variable: desfecho final

Como se verifica, é possível concluir com segurança que não existe colinearidade entre as variáveis preditoras, uma vez que a tolerância (*Tolerance*) e o Fator de Inflação da Variância (*VIF*) das variáveis previsoras são $>0,1$ e <10 , respectivamente. Assim, o segundo previsor tem uma probabilidade maior de ser responsável por uma variância dos resultados do que quando existia um único previsor.