

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE MATEMÁTICA  
CURSO DE ESTATÍSTICA

João Flávio Andrade Silva

**RELAÇÕES ENTRE CRESCIMENTO ECONÔMICO E CARACTERÍSTICAS  
SOCIAIS DOS MUNICÍPIOS DO ESTADO DE MINAS GERAIS**

Uberlândia

2015

João Flávio Andrade Silva

**RELAÇÕES ENTRE CRESCIMENTO ECONÔMICO E CARACTERÍSTICAS  
SOCIAIS DOS MUNICÍPIOS DO ESTADO DE MINAS GERAIS**

Trabalho de conclusão de curso de graduação apresentado a Faculdade de Matemática da Universidade Federal de Uberlândia como requisito parcial para a obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Marcelo Tavares

Uberlândia

2015

João Flávio Andrade Silva

**RELAÇÕES ENTRE CRESCIMENTO ECONÔMICO E CARACTERÍSTICAS  
SOCIAIS DOS MUNICÍPIOS DO ESTADO DE MINAS GERAIS**

Trabalho de conclusão de curso de graduação apresentado a Faculdade de Matemática da  
Universidade Federal de Uberlândia como requisito parcial para a obtenção do título de  
Bacharel em Estatística.

Aprovado em: \_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_.

**BANCA EXAMINADORA**

---

Prof. Dr. Ednaldo Carvalho Guimarães – Universidade Federal de Uberlândia

---

Prof. Dr. Rogério de Melo Costa Pinto – Universidade Federal de Uberlândia

---

Prof. Dr. Marcelo Tavares (Orientador) - Universidade Federal de Uberlândia

## RESUMO

Minas Gerais é um estado brasileiro de grande importância, possui a quarta maior área territorial entre os estados, é o segundo maior em tamanho populacional e no ano de 2010 apresentou o terceiro maior produto interno bruto dentre todas as federações. O presente trabalho teve como objetivo geral estabelecer relações entre crescimento econômico e características sociais dos municípios do estado de Minas Gerais por meio da análise de regressão, que tem como finalidade estabelecer relações quantitativas entre variável resposta (dependente) e uma ou mais variáveis preditoras (independentes) a partir do ajuste de um modelo estatístico. Dessa forma a renda per capita foi utilizada como indicador de crescimento econômico e para os aspectos sociais foram utilizados: tamanho populacional, proporção de pessoas empregadas, média de veículos por pessoa, IFDM Educação, IFDM Saúde, proporção de residentes em área urbana e proporção de mulheres em cada município, com base em dados referentes ao ano de 2010. Por meio da análise de regressão constatou-se relação entre as variáveis, sendo possível predição da renda per capita em função: do tamanho da população, do Índice de Empregados, da Média de Veículos, do IFDM Educação, do IFDM Saúde, do Índice de Urbanização e do Índice Feminino. O modelo apresentou bom ajuste, conseguindo captar 74,25% da variabilidade total da renda, o ajuste descreve que há uma relação positiva entre a renda e todas as outras variáveis, e devido o modelo ser do tipo Log-linear as variáveis independentes causam um efeito de crescimento exponencial na variável dependente. A variável de maior impacto na renda, considerando sua variação, é a Índice de Empregados, um incremento de 0,1 nesta é esperada uma taxa de crescimento de 19,71% na renda. Já para as variáveis IFDM Educação, IFDM Saúde, Índice de Urbanização e Índice Feminino um aumento de 0,1 causa uma taxa de crescimento na Renda Per Capita de 4,02%, 3,28%, 2,61% e 10,84%, respectivamente. No caso da População (em milhares) esta mostrou-se pouco influente, um aumento de 10 mil habitantes produz uma taxa de crescimento de 0,31% na renda, e para um aumento de 0,1 na Média de Veículos é esperado um crescimento de 4,89%. O modelo mostra que crescimento e desenvolvimento econômico estão intimamente relacionados, pois como foi indicado pelo modelo: aspectos sociais estão ligados com a renda; um aumento na qualidade das características sociais (variáveis independentes) acarretará em um aumento na renda.

## ABSTRACT

Minas Gerais is a Brazilian state of huge importance, it owns the fourth largest territorial area between the states, it is the second biggest regarding to population amount and at 2010 it presented the third biggest gross national product among all the federations. The current paper had as general purpose to establish relations between the economic growth and social characteristics of the Minas Gerais cities through the regression analysis, which has as goal to set quantitative relations between the response variable (dependent) and one or more predictors variables (independents) from the adjust of a statistical model. In this way the per capita income was used as an economic growth indicator e for the social aspects the following were used: population amount, employed people proportion, average of vehicle per person, FIRJAN municipal development index of education, FIRJAN municipal development index of health, residents in urban area proportion and women proportion in each city, based in data related to 2010. Over the regression analysis it was found relation between the variables, being possible prediction of the per capita income in function of: the population amount, the employed people index, the vehicle average, the FIRJAN municipal development index of education, FIRJAN municipal development index of health, the urbanization index and the female index. The model presented good adjust, able to collect 74,25% of the total variability of income, the adjust describes that there is a positive relation between the income and all the others variables, and due to the Log-Linear type of the model, the independent variables cause an exponential growth effect in the dependent variable. The variable of the biggest income impact, considering its variation, is the employed people index, an increment of 0,1 in this is variable is expected a growth rate of 19,71% in the income. On the other hand, for the FIRJAN municipal development index of education, FIRJAN municipal development index of health, the urbanization index and the female index variables, a growth of 0,1 causes a growth rate in the per capita income of 4,02%, 3,28%, 2,61% and 10,84%, respectively. In the population case (in thousands), it presented low influence, an increase of 10 thousands of habitants produces a growth rate of 0,31% in the income, and for a growth of 0,1 in the vehicle average is expected a growth of 4,89%. The model presents that the growth and the economic development are strongly related, as it was indicated by the model: social aspects are connected to the income; a growth at the quality of the social characteristics (independent variables) will generate an income growth.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>8</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO .....</b>	<b>10</b>
2.1	MINAS GERAIS .....	10
2.2	DESENVOLVIMENTO ECONÔMICO .....	14
2.3	REGRESSÃO LINEAR .....	17
<b>3</b>	<b>MATERIAL E MÉTODOS .....</b>	<b>22</b>
3.1	BASE DE DADOS .....	22
3.2	VARIÁVEIS .....	22
<b>4</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>24</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>39</b>
	<b>REFERÊNCIAS .....</b>	<b>40</b>
	<b>ANEXO A – DEMONSTRAÇÃO TAXA DE CRESCIMENTO .....</b>	<b>42</b>

## LISTA DE FIGURAS

FIGURA 1. ESTADO DE MINAS GERAIS E SUAS MESORREGIÕES .....	10
FIGURA 2. PIRÂMIDE ETÁRIA DE MINAS GERAIS, ANO 2010 .....	12
FIGURA 3. CARTOGRAMA DA RENDA PER CAPITA (R\$) PARA OS MUNICÍPIOS DE MINAS GERAIS	12
FIGURA 4. CARTOGRAMA DO ÍNDICE DE DESENVOLVIMENTO HUMANO MUNICIPAL DE MINAS GERAIS.....	13
FIGURA 5. GRÁFICO NORMAL DE PROBABILIDADES (Q-Q PLOT) PARA O MODELO M1 .....	28
FIGURA 6. DISPERSÃO DOS RESÍDUOS PADRONIZADOS E OS VALORES PREVISTOS PADRONIZADOS .....	29
FIGURA 7. COMPARATIVO DO MODELO LINEAR E O MODELO TRANSFORMADO A PARTIR DOS GRÁFICOS DE DISPERSÃO ENTRE A VARIÁVEL RESPOSTA E AS VARIÁVEIS REGRESSORAS: FEM, SA, VEI, ED. ....	30
FIGURA 8. COMPARATIVO DO MODELO LINEAR E O MODELO TRANSFORMADO A PARTIR DOS GRÁFICOS DE DISPERSÃO ENTRE A VARIÁVEL RESPOSTA E AS VARIÁVEIS REGRESSORAS: EMP, POP E URB .....	31
FIGURA 9. GRÁFICO DE DISPERSÃO DOS RESÍDUOS PADRONIZADOS E OS VALORES PREVISTOS PADRONIZADOS .....	34
FIGURA 10. RESÍDUOS PADRONIZADOS VS VALOR PREVISTO PADRONIZADO MODELO M2 .....	35

## LISTA DE TABELAS

TABELA 1. PIB DE MINAS GERAIS POR MESORREGIÕES – 2010 .....	11
TABELA 2. CORRELAÇÕES DE PEARSON ENTRE VARIÁVEIS DO MODELO M1 E P-VALOR CALCULADO* .....	24
TABELA 3. ANÁLISE DE VARIÂNCIAS MODELO M1 .....	25
TABELA 4. RESUMO DO MODELO M1 .....	25
TABELA 5. COEFICIENTES DO MODELO M1 .....	26
TABELA 6. CORRELAÇÕES DE PEARSON ENTRE VARIÁVEIS DO MODELO M2 E P-VALOR CALCULADO* .....	32
TABELA 7. ANÁLISE DE VARIÂNCIAS MODELO M2 .....	32
TABELA 8. COEFICIENTES DO MODELO M2 .....	33
TABELA 9. CASOS ATÍPICOS DO MODELO M2 .....	35
TABELA 10. ESTATÍSTICA DESCRITIVA DAS OBSERVAÇÕES ATÍPICAS.....	37
TABELA 11. TAXA DE CRESCIMENTO ESPERADO DA RENDA PER CAPITA DE ACORDO O INCREMENTO NAS VARIÁVEIS INDEPENDENTES.....	38



## 1 INTRODUÇÃO

O estado de Minas Gerais configura parcela vultosa do território nacional brasileiro, seja em área territorial, capacidade de produção econômica ou tamanho populacional, tornando-se assim uma parcela representativa do Brasil.

No que diz respeito a crescimento econômico, Minas Gerais apresenta uma grande disparidade entre os municípios que integram sua região, entretanto essa desigualdade está além da capacidade produtiva do estado se estendendo também para aspectos sociais.

De acordo com Carvalho (2008) “O crescimento econômico é importante porque promove o aumento da renda real e a melhoria do padrão de vida. ”. Já segundo Ferreira (2000)

[...] tanto os argumentos teóricos como a evidência empírica mais recente passaram, em grande parte, a sugerir uma relação causal negativa entre desigualdade e crescimento econômico. O principal argumento teórico por trás da hipótese de que a desigualdade reduz o crescimento é a consideração de que, com mercados imperfeitos de capital e sob assimetrias de informação, a desigualdade e a pobreza implicariam na existência de um grupo de agentes sem acesso ao crédito e, portanto, sem a possibilidade de desenvolver projetos cujo valor privado (e social) é positivo.

Dessa forma, a renda per capita, que é o rendimento por pessoa, atrelada ao bem-estar configura um indicador socioeconômico que avalia o grau de desenvolvimento econômico de uma determinada região, o desenvolvimento por sua vez proporciona benefícios a sociedade tornando-a mais igualitária.

O presente trabalho se insere neste contexto com o objetivo geral de estabelecer relações entre crescimento econômico e características sociais dos municípios do estado de Minas Gerais. Uma forma de estabelecer tal relação se faz por meio da utilização da técnica de análise de regressão, que tem como finalidade estabelecer relações quantitativas entre variável resposta (dependente) e uma ou mais variáveis preditoras (independentes) a partir do ajuste de um modelo estatístico.

Para tanto, este trabalho utilizou a renda per capita como indicador de crescimento econômico e para os aspectos sociais foram utilizados: tamanho populacional, proporção de pessoas empregadas, média de veículos por pessoa, IFDM Educação, IFDM Saúde, proporção de residentes em área urbana e proporção de mulheres em cada município, com base em dados referentes ao ano de 2010.

O estudo das relações dessas variáveis tem como objetivo específico tornar mensurável o impacto de características sociais no crescimento econômico, auxiliar na detecção de áreas menos favorecidas socioeconomicamente, podendo ser utilizado para definição de projetos de

fomento regional e política orçamentária municipal, indicando áreas de maior necessidade, para assim proporcionar maior desenvolvimento econômico.

## 2 REFERENCIAL TEÓRICO

### 2.1 MINAS GERAIS

Minas Gerais é um estado brasileiro situado na região sudeste do país, possui a quarta maior área territorial entre os estados e é o segundo maior em tamanho populacional. Sua população em 2010 foi estimada em cerca de 19.597.330, atrás apenas de São Paulo e mais populoso que todos os estados da Região Norte. Possui densidade demográfica de 33,41 habitantes por quilômetro quadrado. Sua capital é Belo horizonte, centro da terceira maior metrópole brasileira, com 5.414.701 residentes na região metropolitana, em 2010. Com base em aspectos sócio econômicos o estado é dividido em 853 municípios o que o torna o estado com mais municípios do brasil, estes municípios fazem parte de 66 microrregiões que por sua vez compõem 12 mesorregiões: Campo das Vertentes, Central Mineira, Jequitinhonha, Metropolitana de Belo Horizonte, Noroeste de Minas, Norte de Minas, Oeste de Minas, Sul e Sudoeste de Minas, Triângulo Mineiro e Alto Paranaíba, Vale do Mucuri, Vale do Rio Doce e Zona da Mata, retratadas na (Figura 1). (IBGE, 2015)



Figura 1. Estado de Minas Gerais e suas mesorregiões  
Fonte: BCB (2013, p.95)

Minas Gerais apresentou o terceiro maior produto interno bruto dentre todas as federações ao final do ano de 2010, contabilizando 351,38 bilhões de reais. Considerando o período 2007 a 2012 e tendo como base o terceiro trimestre de 2007, o PIB mineiro cresceu 14,8%, ante 16,2% do PIB brasileiro, deixando claro o forte efeito que a crise de 2008 e 2009 teve sobre a economia estadual. (BCB, 2013)

Dentre mesorregiões mineiras, a Região Metropolitana de Belo Horizonte concentra 45%, conforme Tabela 1 das atividades econômicas do estado, e é também uma das regiões que apresenta maior crescimento. Apesar de ser composta por 34 municípios, na Região Metropolitana a atividade se concentra em Belo Horizonte, com 43% de participação em Betim, com 23%, e em Contagem, com 15%. Destacam-se também o Triângulo Mineiro e Alto Paranaíba, o Sul e Sudoeste de Minas, Zona da Mata e Vale do Rio Doce que juntas correspondem a mais de 40% do PIB mineiro. As mesorregiões com menor participação no produto mineiro são os Vales do Jequitinhonha e do Mucuri que juntas possuem 2,1% de participação no PIB estadual. (BCB, 2013)

Tabela 1. PIB de Minas Gerais por mesorregiões – 2010

Mesorregiões	PIB		Var. nominal 2010/2006
	R\$ milhões	Distr. %	
RMBH	158207	45	73
Triâng. Mineiro/Alto Paranaíba	52693	15	63,5
Sul/Sudoeste de Minas	41375	11,8	55,2
Zona da Mata	25629	7,3	50,7
Vale do Rio Doce	20981	6	44,1
Norte de Minas	13578	3,9	61,3
Oeste de Minas	13187	3,8	53,6
Campo das Vertentes	6794	1,9	55
Noroeste de Minas	6001	1,7	82,9
Central Mineira	5787	1,6	49,8
Jequitinhonha	4136	1,2	67,4
Vale do Mucuri	3012	0,9	61,5
MG	351381	100	63,6

Fonte: BCB (2013, p.95)

A população mineira era composta em 2010 por 49,2% de homens e 50,8% de mulheres, a expectativa de vida dos mineiros é de 76 anos, sendo que para as mulheres esta expectativa é de 79 anos e para os homens é de aproximadamente 73 anos, na Figura 2 é possível visualizar a distribuição de diferentes faixas etárias da população mineira em 2010. (IBGE, 2015)

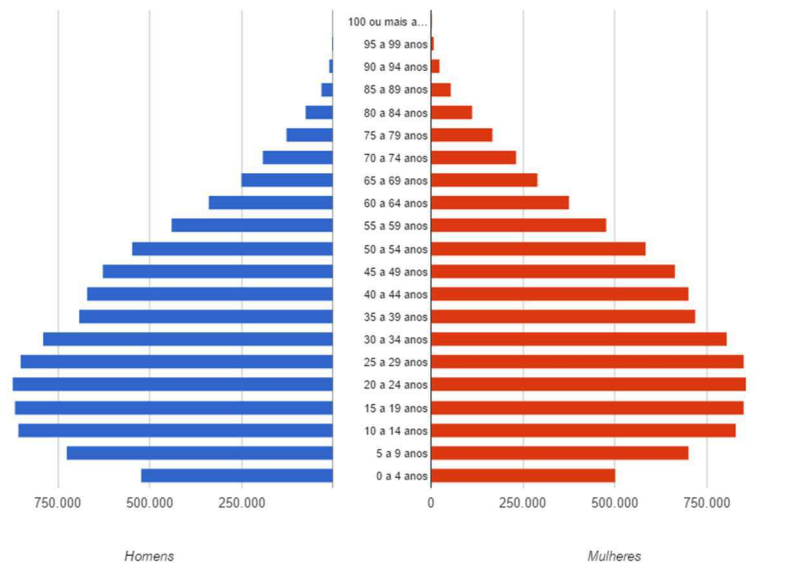


Figura 2. Pirâmide etária de Minas Gerais, ano 2010

Fonte: [www.censo2010.ibge.gov.br/apps/mapa/](http://www.censo2010.ibge.gov.br/apps/mapa/)

O estado apresenta uma grande disparidade entre suas regiões no que se refere ao desenvolvimento econômico e social. As mesorregiões mais ricas do estado, como a Central e o Triângulo Mineiro possuem renda per capita muito acima da das áreas mais pobres localizadas no norte do estado, especialmente nos vales do Jequitinhonha e do Mucuri, de acordo com Figura 3.

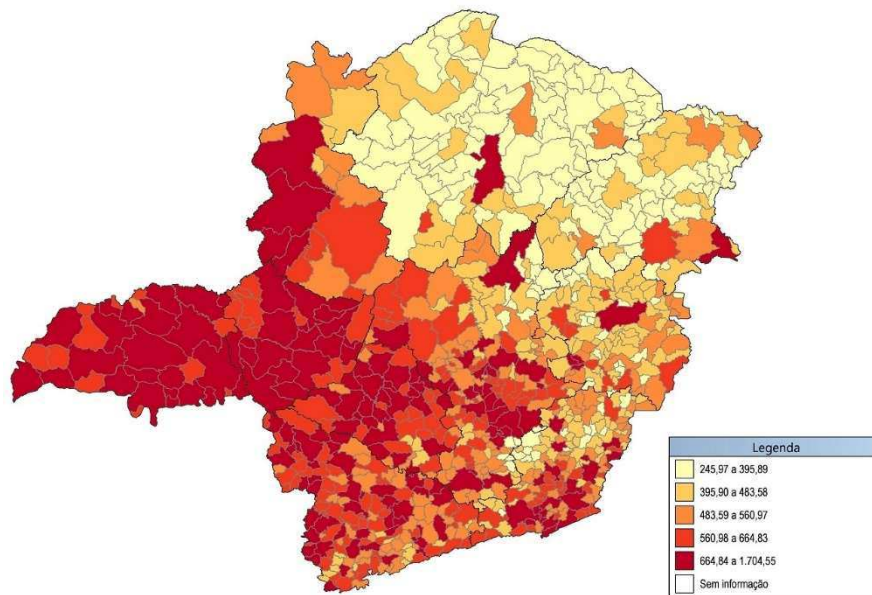


Figura 3. Cartograma da renda per capita (R\$) para os municípios de Minas Gerais

Fonte: Índice Mineiro de Responsabilidade Social 2013

A Figura 4 revela de forma precisa a desigualdade entre as regiões de Minas Gerais. O norte e nordeste do estado concentram a maioria dos municípios com baixo IDH, enquanto a região Sul, Triângulo e Alto Paranaíba que a situação é oposta. A região central do estado, embora seja rica, possui alto índice de desigualdade, ou seja, a renda não é igualmente distribuída entre a população.

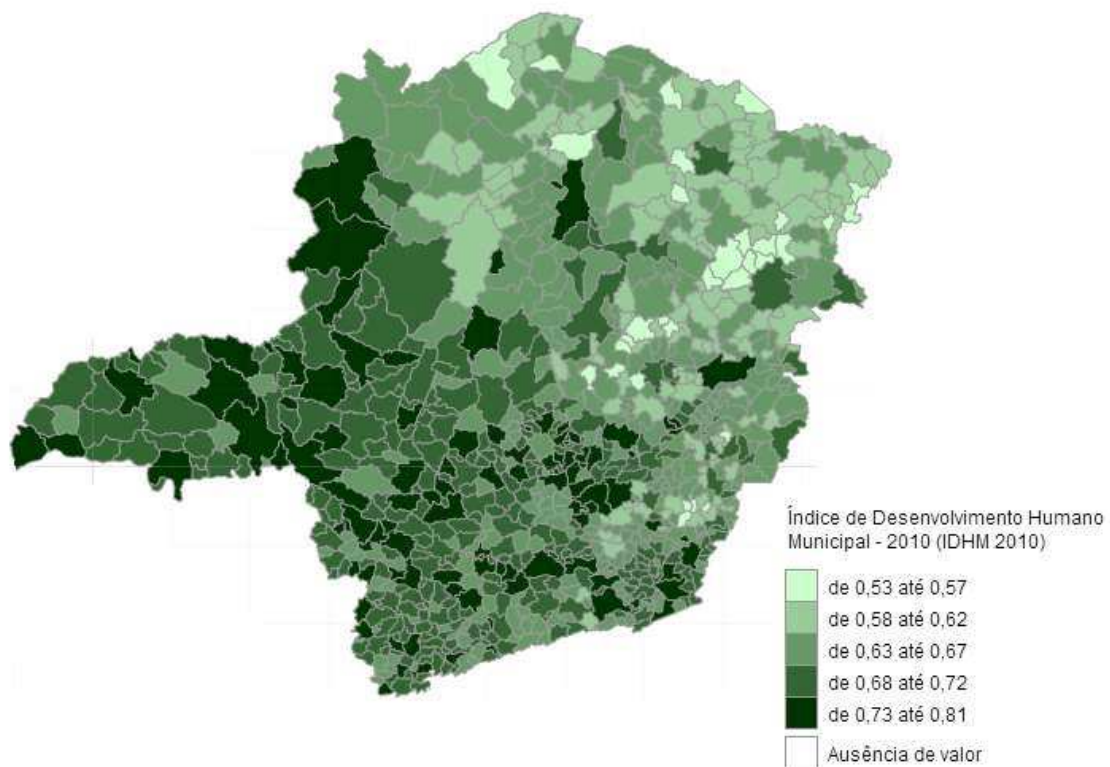


Figura 4. Cartograma do Índice de desenvolvimento Humano Municipal de Minas Gerais.  
Fonte: [www.ibge.gov.br/webcart](http://www.ibge.gov.br/webcart)

O padrão de desigualdade observado na Figura 3 também é visto na Figura 4, há uma separação nítida das áreas onde a renda é mais elevada e regiões em que a renda per capita é extremamente baixa, e a diferença não fica apenas na renda, os municípios com menor renda per capita também possuem uma educação mais precária, saúde deficiente e baixo crescimento econômico.

## 2.2 DESENVOLVIMENTO ECONÔMICO

Os conceitos de desenvolvimento e crescimento econômicos muitas das vezes são confundidos ou acabam sendo tratados como sinônimos. Porém, crescimento e desenvolvimento possuem definições distintas. Segundo a descrição de Sandroni (1999), o crescimento econômico representa o aumento da capacidade de produção da economia e, portanto, da produção de bens e serviços de determinado país ou área econômica. A mensuração do crescimento econômico é realizada com base no índice de crescimento anual do Produto Nacional Bruto (PNB) per capita, que por sua vez, é referente à soma de todos os bens e serviços provenientes da utilização de recursos de uma nação, não levando em consideração o território em que foram produzidos, diferente do Produto Interno Bruto (PIB), que está relacionado com à produção total dos bens e serviços no âmbito do território nacional. Já Vasconcellos (2011), define crescimento econômico de forma sucinta como o crescimento continuado da renda per capita ao longo do tempo. Para este mesmo autor o desenvolvimento econômico aborda um conceito mais qualitativo do que o crescimento econômico, onde as alterações da composição do produto e a alocação dos recursos pelos diferentes setores da economia são incluídas de forma a promover melhorias nos indicadores de bem-estar não somente econômico como também social, ou seja, o desenvolvimento econômico leva em conta o progresso de uma região considerando sua produção e ponderando-a por aspectos sociais como a pobreza, desemprego, desigualdade, condições de saúde, nutrição, educação e moradia. Em suma, o crescimento econômico refere-se ao PIB enquanto que o desenvolvimento econômico possui um conceito mais amplo: além de apresentar crescimento econômico também está relacionado a fatores sociais que proporcionam melhora na qualidade de vida, como é o caso da educação, da saúde e do saneamento básico. Notoriamente, o desenvolvimento é um fenômeno global que atinge toda a estrutura social, política e econômica.

“O crescimento de uma economia é indicado ainda pelo índice de crescimento da força de trabalho, pela proporção da receita nacional poupada e investida e pelo grau de aperfeiçoamento tecnológico.” (SANDRONI, 1999). Neste contexto as diretrizes para elevação da taxa de crescimento têm como foco investimentos em capitais físico, capital humano e tecnologia.

O valor das máquinas, prédios, infraestruturas – tais como: transportes, energia e comunicações – e tecnologia, é dito capital físico. De acordo com Sousa (2006), o capital físico é um dos elementos mais essenciais na criação da riqueza, na medida em que reforça a qualidade e produtividade decorrente do uso de recursos naturais e da capacidade do capital humano.

Embora o capital físico seja muito enfatizado, o capital humano, que consiste no valor do ganho de renda potencial incorporado ao indivíduo, requer uma imprescindível atenção. Esse fator diz respeito às habilidades e talentos inerentes às pessoas, à educação e sucessivamente, às desenvolturas adquiridas.

Sousa (2006), afirma que “Quanto mais alto o nível médio de habilidade e conhecimento, mais fácil será para que os indivíduos em idade produtiva compreendam, apliquem e colham os frutos do progresso técnico e, portanto, mais alto será o padrão de vida do país, de forma que o investimento em educação deve ser o máximo e diversificado a ponto de preparar a população nos mais variados ramos do conhecimento. ”.

Condiz afirmar que o trabalhador médio nos países industrializados é significativamente mais produtivo do que o trabalhador médio nos países em desenvolvimento. Em parte, este aspecto é explicado pela maior quantidade de capital físico, mas também, pelo fato de o trabalhador ser mais qualificado. Em outros termos, houve melhor aquisição de capital humano seja por meios de educação formal, treinamento informal e ou por experiência.

Quanto ao que se refere à tecnologia, segundo KIM (2005), a mesma pode ser definida pela “aplicação prática dos conhecimentos e habilidades para viabilizar o estabelecimento, a operação, melhoria e expansão das condições dessas transformações, assim como do subsequente planejamento e aperfeiçoamento da produção. ”. Em relação à habilidade de criação de novas tecnologias, o autor KIM (2005) relata que

A ‘aptidão de inovar’ consiste nas aptidões de criar e levar adiante novas possibilidades tecnológicas através da prática econômica. [...] A invenção e a inovação são o resultado de atividades tanto formais quanto informais. O termo ‘inovação’ é frequentemente associado à mudança tecnológica na fronteira internacional do conhecimento.

As variações na quantidade e na qualidade de insumos básicos: capital e mão de obra, suscitam o crescimento da produção e da renda. Nesse âmbito, as fontes de crescimento são: o aumento da força de trabalho; o crescimento do estoque de capital, ou da capacidade produtiva; a melhor qualidade da mão-de-obra, por meio de programas de educação, treinamentos e especializações; aperfeiçoamento tecnológico, o qual contribui para uma melhor eficiência na utilização do estoque de capital e a eficiência organizacional, ou seja, eficiência na interação dos insumos. (VASCONCELLOS, 2011)

Sandroni (1999) diz ainda que as mudanças que caracterizam o desenvolvimento econômico consistem no aumento da atividade industrial em comparação com a atividade agrícola, migração da mão-de-obra do campo para as cidades, redução das importações de



produtos industrializados e das exportações de produtos primários e menor dependência de auxílio externo.

Ainda, em relatório, publicado pelo Banco Mundial, intitulado “Relatório Sobre o Desenvolvimento Mundial de 2012”, é exposto que a igualdade de gênero é ideia fundamental e ferramenta de desenvolvimento.

O Banco Mundial, instituição financeira internacional que fomenta o desenvolvimento de países, anualmente produz relatório sobre o desenvolvimento, na edição de 2012 o foco esteve na igualdade de gênero e desenvolvimento, nele é informado que a vida das mulheres em todo o mundo tem melhorado drasticamente, que obtiveram ganhos sem precedentes em direitos, educação, saúde e em acesso a empregos e meios de subsistência. Em muitos países as diferenças de gênero do ensino fundamental vêm sendo superadas, embora em um terço de todos os países há mais meninas no ensino médio que meninos. Em 60 países há mais mulheres jovens frequentando as universidades do que homens. As mulheres estão utilizando sua educação para atuar com maior intensidade na força de trabalho, atualmente, elas representam 40% da força de trabalho global e 43% de todos os agricultores. Além disso, atualmente, em todo o mundo, as mulheres vivem mais do que os homens. (BANCO MUNDIAL, 2015)

Neste documento o Banco Mundial afirma que o aumento da escolaridade das mulheres está relacionado com crescimento econômico e que a igualdade de gênero não é apenas uma questão de justiça social, mas esta traz consigo efeitos econômicos positivos e que os padrões deste progresso e persistência em igualdade de gênero são importantes, tanto para os resultados do desenvolvimento quanto para a formulação de políticas. Eles são relevantes porque a igualdade de gênero constitui por si só de grande valor, em adição, a equidade de gênero proporciona otimização da economia, elevando a produção, potencializando os resultados do desenvolvimento, incluindo perspectivas para a próxima geração e para a qualidade das políticas e instituições da sociedade.

## 2.3 REGRESSÃO LINEAR

Um modelo de regressão é dito linear quando a variável resposta é função linear dos parâmetros, e aos modelos de regressão linear que possuem apenas uma variável preditora é dado o nome de modelo de regressão linear simples (MRLS), e chamamos de modelo de regressão linear múltiplo (MRLM) os casos onde existe mais de uma variável preditora (DRAPER e SMITH, 1998).

O Modelo de Regressão Linear tem definição matricial dada por (DRAPER e SMITH, 1998):

$$Y = X\beta + \varepsilon$$

em que:  $Y$  é o vetor que contém as observações da variável resposta,  $\beta$  é o vetor de parâmetros,  $X$  é a matriz das observações das variáveis preditoras em que o número de colunas é igual à quantidade de parâmetros ( $p$ ) e as linhas correspondem às observações da respectiva variável preditora de cada coluna, exceto a primeira coluna que é uma coluna de elementos iguais a 1, que representa o intercepto, e  $\varepsilon$  é o vetor dos erros que são variáveis aleatórias independentes com distribuição normal com média zero e variância dada por  $\sigma^2 I$ , onde  $I$  é a matriz identidade.

Para construir o modelo é necessário estimar o vetor de parâmetros  $\beta$ . Uma forma de obter tais estimativas é por meio do Método dos Mínimos Quadrados, técnica que consiste em minimizar a soma de quadrados dos erros (resíduos), ou seja, minimizar a função:

$$L = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^t \varepsilon = (Y - X\beta)^t (Y - X\beta)$$

Com  $i = 1, \dots, n$ , e  $n$  é a quantidade de observações. Desta forma o método dos mínimos quadrados consiste em determinar os coeficientes do vetor  $\beta$  que produz a menor soma de quadrados dos erros.

Seja  $L$  uma função convexa, para obter um ponto de mínimo, inicialmente encontramos seus pontos críticos. O ponto crítico da função  $L$  é calculado igualando o gradiente de  $L$  a zero (FRANCO, 2006):

$$\frac{\partial L}{\partial \beta} = 0$$

Trocando  $\beta$  pelo vetor  $\hat{\beta}$ , e isolando este, temos que os estimadores de mínimos quadrados para os parâmetros são dados pelo vetor:

$$\hat{\beta} = (X^t X)^{-1} X^t Y$$

Desde que a matriz  $(X^tX)$  possua inversa.

A partir deste resultado encontramos o modelo linear ajustado.

$$\hat{Y} = X\hat{\beta}$$

Uma medida de qualidade de ajuste do modelo é o coeficiente de determinação ( $R^2$ ) que representa a proporção da variação total explicada pela regressão de  $Y$  por  $X$ , varia entre 0 e 1, sendo que 0 o modelo captura nenhuma variação de  $Y$  e 1 o modelo explica toda a variabilidade. A expressão a seguir mostra como se dá o cálculo do coeficiente de determinação (MILES e SHEVILIN, 2001).

$$R^2 = \frac{SQ_{exp}}{SQ_{tot}}$$

Em que  $SQ_{tot} = \sum_{i=1}^n (Y_i - \bar{Y})^2$  é a soma de quadrada total, representa a variação total de  $Y$  em torno da sua média, e pode ser definida também como:

$$SQ_{tot} = SQ_{exp} + SQ_{erro}$$

$SQ_{exp} = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$  é a soma de quadrado da regressão, representa a variação das esperanças específicas de  $Y$ , dado  $X$ , em torno de sua média.

$SQ_{erro} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$  é a soma de quadrado dos erros, representa a variação de  $Y$  em torno da reta de regressão.

Também, utiliza-se a análise de variância para testar a significância dos parâmetros do modelo, através da comparação de variâncias com base na decomposição da soma de quadrados e nos graus de liberdade.

Utiliza-se da razão  $F$  para testar as seguintes hipóteses:

$$\begin{cases} H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0 \\ H_a: \beta_j \neq 0 \text{ para pelo menos um } j \end{cases}$$

A razão  $F$  de variâncias pode ser calculada através da divisão do quadrado médio do modelo pelo quadrado médio dos resíduos,

$$F_c = \frac{QM_{exp}}{QM_{res}} \sim F_{(\alpha; p-1; n-p)}$$

Em que  $\alpha$  é o nível de significância do teste,  $p$  é a quantidade de parâmetros do modelo e  $n$  é o número de observações.

$$QM_{exp} = \frac{SQ_{exp}}{p-1}, \text{ quadrado médio do modelo;}$$

$$QM_{res} = \frac{SQ_{res}}{n-p}, \text{ quadrado médio dos resíduos;}$$

Por regra de decisão, rejeitamos  $H_0$  caso  $F_c > F_{(\alpha; p-1; n-p)}$ .

O estimador de mínimos quadrados  $\hat{\beta}$  possui distribuição normal,  $p$ -variada, isto é,

$$\hat{\beta} \sim N_p(\beta; \sigma^2(X^t X)^{-1})$$

Temos que para  $\hat{\beta}_j$  elemento de  $\hat{\beta}$ , segue que  $\hat{\beta}_j \sim N(\beta_j; \sigma^2 C_{jj})$  com  $C_{jj}$  sendo o  $j$ -ésimo elemento da diagonal de  $(X^t X)^{-1}$ ,  $j = 1, 2, \dots, p$ .

Logo para testar a significância de qualquer coeficiente de regressão individualmente, ou seja,

$$\begin{cases} H_0: \beta_j = 0 \\ H_a: \beta_j \neq 0 \end{cases}$$

Podemos aplicar o teste t dado por

$$t_c = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t_{(n-p)}$$

Com  $\hat{\sigma}^2 = QM_{res}$ . Por regra de decisão rejeitamos  $H_0$  para os caso em que  $|t_c| > t_{(n-p)}$ .

O intervalo de confiança de  $100(1 - \alpha)\%$  para o coeficiente  $j$  é dado por,

$$IC(\beta_j)_{100(1-\alpha)\%} = \hat{\beta}_j \pm t_{(\alpha/2; n-p)} \sqrt{\hat{\sigma}^2 C_{jj}}$$

Outras análises de ajuste de modelos, o diagnóstico de resíduos e a detecção de pontos atípicos são de grande importância, pois estes pontos podem afetar significativamente o modelo. Ao detectar um ponto atípico, que pode ou não influenciar, deve-se buscar compreender o que faz esta observação diferente das demais. Serão apresentados a seguir algumas medidas que podem auxiliar no diagnóstico.

A medida de *leverage* ( $h_{ii}$ ) mede a distância entre uma observação e todas as outras. Seja a matriz de projeção  $H = X(X^t X)^{-1} X^t$ , o *leverage* é o  $i$ -ésimo elemento da diagonal principal, ou seja,  $h_{ii} = x_i^t (X^t X)^{-1} x_i$ , com o vetor  $x_i^t = (x_{i1}, x_{i2}, \dots, x_{ip})$ . Os valores para essa medida variam de 0 a 1, quanto mais próximo de 1 estiver o valor  $h_{ii}$  maior influência terá a observação  $i$  (Rousseeuw, 1987).

$$\varepsilon \sim N(0; \sigma^2(1 - H))$$

$$\varepsilon_i \sim N(0; \sigma^2(1 - h_{ii}))$$

Observações com valores de *leverage* maiores que  $3p/n$  indicam casos de grande influência (STEVENS, 1992), no qual  $p$  é o número de termos do modelo (incluindo a constante) e  $n$  é o número de observações.

Utiliza-se também a análise dos resíduos estudentizados. O termo estudentizado está ligado ao fato de que estes resíduos, sob normalidade, têm distribuição t de Student com  $(n - p - 1)$  graus de liberdade.

O resíduo estudentizado internamente é dado por

$$rsi_i = \frac{r_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}$$

E o resíduo estudentizado externamente é definido da seguinte forma

$$rse_{(i)} = \frac{r_i}{\hat{\sigma}_{(i)}\sqrt{1 - h_{ii}}}$$

Em que  $r_i$  é o resíduo ordinal, definido por  $r_i = y_i - \hat{\mu}_i$ , e  $\hat{\sigma}_{(i)}^2$  é a estimativa da variância sem a observação  $i$ .

Os resíduos estudentizados externamente permitem a identificação de observações atípicas, sugere-se apurar pontos em que  $|rse_{(i)}| > t_{n-p-1; 1 - \frac{\alpha}{2n}}$ .

O cálculo do  $DFFitS_{(i)}$  mede a alteração provocada no valor ajustado decorrente da exclusão da observação  $i$ . É definido pela expressão

$$DFFitS_{(i)} = \left( \frac{h_{ii}}{1 - h_{ii}} \right)^2 rse_{(i)}$$

Notamos uma relação proporcional positiva do  $DFFitS_{(i)}$  com a medida de *leverage* e os resíduos estudentizados externamente. Um dos critérios adotados para consideração de ponto influente é  $|DFFitS_{(i)}| > 3\sqrt{p/(n - p)}$ .

Já os  $DFBetaS_{(i)}$  mede a alteração provocada no vetor estimado  $\hat{\beta}$  após a retirada da  $i$ -ésima observação, permite avaliar a influência da observação  $i$  sobre o coeficiente de  $X_j$ . Valores absolutos maiores que 1 apontam casos potencialmente influentes para o parâmetro do modelo. É definido por

$$DFBetaS_{j(i)} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{c_{jj}QME_i}}$$

Sendo  $j = 1, 2, \dots, p$ , e  $c_{jj}$  é o  $j$ -ésimo elemento da diagonal da matriz  $(X^t X)^{-1}$ .

A distância de Cook é uma medida de influência global, mede a distância entre os valores ajustados calculados com e sem a  $i$ -ésima observação, ou seja, considera o efeito de um

único caso no modelo como um todo. Valores maiores que 1 merecem atenção (COOK e WEISBERG, 1982).

$$D_{(i)} = \frac{h_{ii} r s_i^2}{p(1 - h_{ii})}$$

No que diz respeito a interpretação do modelo, deve-se observar as relações que as variáveis independentes mantêm com a variável dependente, para o caso linear essa relação é proporcional, porém em alguns casos faz-se necessária a transformação de variáveis e para estes, as relações entre a variável dependente e as variáveis independentes, costumam, não serem proporcionais, como é o caso da transformação Logarítmica, que gera o modelo conhecido como Log-Linear. Dado por

$$\ln Y = X\beta + \varepsilon$$

Para este modelo, alterações nas variáveis independentes causam um efeito de crescimento exponencial na variável dependente. De acordo com a demonstração em anexo (ANEXO A), tem-se:

$$\hat{y}_1 = \hat{y}_0(1 + i)^{\frac{\Delta x}{h}}$$

$$\Delta x = x_1 - x_0;$$

$$i = \frac{y_b - y_a}{y_a};$$

$$h = x_b - x_a;$$

$i$ : taxa de crescimento referente a variação  $h$ ;

$h$ : é a variação de  $x$  que foi utilizado para obter a taxa de crescimento.

### 3 MATERIAL E MÉTODOS

#### 3.1 BASE DE DADOS

A base de dados do presente trabalho é composta por características populacionais, referentes ao ano de 2010, de 848 municípios de Minas Gerais que a Secretaria do Tesouro Nacional mantém dados contábeis (STN, 2015). O acesso a estes dados foi através de links disponibilizados nos sites do Instituto Brasileiro de Geografia e Estatística (IBGE), da Federação das Indústrias do Estado do Rio de Janeiro (FIRJAN), da Fundação João Pinheiro e do Tesouro Nacional.

No site do IBGE foram coletadas informações referentes a quantidade total de: pessoas no município, residentes urbanos, residentes rurais, mulheres, homens, pessoas empregadas, veículos (IBGE, 2015). Através da Fundação João Pinheiro obteve-se dados referentes a Renda Per Capita (FJP, 2015).

A FIRJAN produz o Índice de Desenvolvimento Municipal (IFDM), que tem como objetivo medir o grau de desenvolvimento socioeconômico dos municípios, numa escala de 0 a 1, em três áreas de atuação: Emprego & renda, Educação e Saúde. Neste estudo foram utilizados apenas os índices IFDM – Saúde e IFDM – Educação. IFDM – Emprego & renda não foi utilizado devido a este indicador ser dependente da renda per capita (FIRJAN, 2015).

#### 3.2 VARIÁVEIS

Com o objetivo de reduzir a correlação entre o tamanho populacional e algumas variáveis, foram criados índices e retirado médias para compor novas variáveis, a saber:

- Índice feminino. Representa a proporção, e como tal é definida no intervalo de 0 a 1, de indivíduos do sexo feminino em determinado município;
- Média de veículos. Esta variável é maior ou igual a zero, representa a média de veículos por pessoa de cada município. Obtida tomando a razão da quantidade total de veículos pela quantidade total de pessoas;
- Índice de urbanização. O termo “Urbanização” significa o crescimento da cidade, representa a transformação de área rural em área urbana. Porém no presente estudo a variável “Índice de urbanização” representará apenas o

contraste entre zona urbana e zona rural através da proporção de moradores da área urbana;

- Índice de empregados. Esta variável informa quanto da população total do município está empregada, em uma escala de 0 a 1, obtida tomando a razão da quantidade de pessoas empregadas pela quantidade total de pessoas.

Neste trabalho procurou-se analisar a relação da Renda Per Capita em função do tamanho da população, do Índice de Empregados, da Média de Veículos, do IFDM Educação, do IFDM Saúde, do Índice de Urbanização e do Índice Feminino por meio da análise de regressão. A análise de regressão objetiva estudar a relação entre variável resposta (dependente) e uma ou mais variáveis preditoras (independentes) a fim de estabelecer um modelo estatístico para prever o valor médio da variável resposta (DRAPER e SMITH, 1998).

A análise da qualidade de ajuste do modelo foi realizada através do cálculo do coeficiente de determinação, também foram utilizados gráficos de valores previstos contra resíduos, para avaliar os pontos atípicos foram utilizadas as medidas de *leverage*, resíduos estudentizados externamente, *DFFitS*, *DFBetaS* e distância *Cook*. Todas as análises estatísticas foram realizadas no software IBM SPSS Statistics.



#### 4 RESULTADOS E DISCUSSÃO

Conforme apresentado na metodologia, procurou-se obter um ajuste por intermédio da Regressão Linear Múltipla em que a Renda Per Capita (Rpc) dos municípios de Minas é a variável resposta de interesse e para predizê-la foram selecionadas as variáveis: População (em milhares) (Pop), Índice de Empregados (Emp), Média de Veículos (Veic), IFDM Educação (Ed), IFDM Saúde (Sa), Índice de Urbanização (Urb) e Índice Feminino (Fem), sendo estas as variáveis regressoras candidatas a compor o modelo *M1* apresentado a seguir:

$$\widehat{Rpc} = \hat{\beta}_0 + \hat{\beta}_1 Fem + \hat{\beta}_2 Sa + \hat{\beta}_3 Veic + \hat{\beta}_4 Ed + \hat{\beta}_5 Urb + \hat{\beta}_6 Emp + \hat{\beta}_7 Pop$$

Com  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_7)$ , representando o vetor de estimativas para os parâmetros do modelo proposto.

Para uma análise inicial dos dados procurou-se avaliar as correlações entre as variáveis por meio do coeficiente de correlação linear de Pearson. Os resultados são apresentados na Tabela 2. A Renda Per Capita tem uma moderada correlação positiva com as variáveis IFDM Saúde, Média de Veículos, IFDM Educação, Índice de Urbanização e Índice de Empregados, logo, dentre as 7 variáveis, 5 delas possuem considerável correlação com a variável resposta. Observa-se também que entre as variáveis previsoras não ocorreram fortes correlações, uma indicação de que estas variáveis não são colineares.

Tabela 2. Correlações de Pearson entre variáveis\* do modelo M1 e p-valor calculado\*\*

	Rpc	Fem	Sa	Veic	Ed	Urb	Emp	Pop
Rpc	-	0,229	0,503	0,575	0,531	0,619	0,719	0,386
Fem	0,000	-	-0,005	0,117	0,154	0,392	0,071	0,257
As	0,000	0,437	-	0,385	0,428	0,291	0,462	0,098
Veic	0,000	0,000	0,000	-	0,427	0,365	0,448	0,182
Ed	0,000	0,000	0,000	0,000	-	0,439	0,512	0,112
Urb	0,000	0,000	0,000	0,000	0,000	-	0,646	0,225
Emp	0,000	0,019	0,000	0,000	0,000	0,000	-	0,164
Pop	0,000	0,000	0,002	0,000	0,001	0,000	0,000	-

\*Renda per capita (Rpc), Índice Feminino (Fem), IFDM Saúde (Sa), Média de Veículos (Veic), IFDM Educação (Ed), Índice de Empregados (Emp), População (em milhares) (Pop) e Índice de Urbanização (Urb).

\*\*Acima da diagonal principal estão os valores das correlações e abaixo seus respectivos p-valores.

Para avaliar a significância do modelo na variação dos dados, inicialmente, aplicou-se a Análise de Variância (ANOVA), técnica estatística que tem como objetivo testar a hipótese

de nulidade, em que todos os parâmetros do modelo são iguais a zero, contra a hipótese alternativa, que pelo menos um dos parâmetros é diferente de zero, baseando-se na distribuição  $F$ . A Tabela 3 mostra que a estatística  $F$  obtida de 278,86 gerando um  $p$ -valor abaixo de 0,001, portanto, ao nível de significância de 5%, afirma-se que o modelo  $M1$  é melhor que a média para uma previsão da Renda Per Capita, ou seja, pelo menos um dos parâmetros é diferente de zero.

Tabela 3. Análise de variâncias modelo M1

Fator de Variação	Soma dos Quadrados	Gl	Quadrado Médio	F	Sig.
Regressão	14677819,5	7	2096831,36	278,86	0,000
Residual	6316119,6	840	7519,19		
Total	20993939,1	847			

O ajustamento do modelo com as variáveis propostas forneceu um coeficiente de determinação da ordem de 0,699, indicando que o modelo explica cerca de 69,9% da variação total da variável Renda Per Capita, conforme mostra a Tabela 4.

Tabela 4. Resumo do modelo M1

R	R quadrado	Durbin-Watson
0,836	0,699	1,726

A estatística Durbin-Watson (Tabela 4), utilizada para testar a independência dos resíduos, o valor desta estatística varia entre 0 e 4 e como regra geral, os resíduos são considerados não correlacionados se a estatística é de aproximadamente 2, já um valor próximo de 0 indica forte correlação positiva, enquanto um valor de 4 indica forte correlação negativa. Para o caso modelo M1 a estatística Durbin-Watson apresentou o valor de 1,726, razoavelmente próxima do valor de referência 2, indicando com segurança que os resíduos são não correlacionados, ou seja, podem ser considerados independentes.

Estimativas dos coeficientes de regressão com respectivas significâncias bem como intervalos de confiança são exibidos na Tabela 5.

Tabela 5. Coeficientes do modelo M1

Variáveis do modelo	$\hat{\beta}$	$\hat{\beta}$ padronizado	Sig.	Intervalo de confiança 95,0% para B	
				Limite inferior	Limite superior
(Constante)	-478,299		0,001	-747,587	-209,011
Índice Feminino	647,970	0,051	0,019	106,835	1189,105
IFDM Saúde	156,570	0,145	0,000	108,980	204,160
Média de Veículos	252,064	0,214	0,000	200,092	304,035
IFDM Educação	205,811	0,087	0,000	96,414	315,207
Índice de Urbanização	123,023	0,149	0,000	77,339	168,708
Índice de Empregados	924,651	0,376	0,000	785,930	1063,371
População (em milhares)	0,359	0,215	0,000	0,294	0,425

Os testes t para os estimadores dos parâmetros  $\beta$  indicaram que todos estes são relevantes para o modelo, ao nível de 5% de significância. A mesma informação também pode ser abstraída interpretando os intervalos de confiança da ordem de 95%, como pode-se observar o zero não está contido em nenhum destes intervalos, indicando assim que todas as variáveis são influentes para o modelo. Verificou-se também que todas as variáveis propostas mantêm uma relação positiva com a variável Renda Per Capita, isto quer dizer que ao aumentarmos os valores das variáveis independentes, ou elevar qualquer uma destas variáveis mantendo as demais inalteradas, estas modificações acarretarão em um incremento na Renda Per Capita.

Para averiguar possível existência de dependência entre as variáveis do modelo foram utilizados os valores de FIV (Fator de Inflação de Variância) que ficaram todos abaixo de 10 (MYERS, 1990) (BOWERMAN e O'CONNELL, 1990) e as Tolerâncias que se mantiveram acima de 0,2 (MENARD, 1995), logo podemos concluir que não existe multicolinearidade nos dados, ou seja, as variáveis independentes não são função linear de outras variáveis independentes.

No quadro 1 são apresentados os casos em que os resíduos padronizados foram maiores que 2 ou menores que -2. Considerando a distribuição normal com dois desvios padrões, o que abrange em torno de 95,45% das observações, temos que para uma amostra de tamanho 848, é sensato esperar aproximadamente 39 casos (4,55%) fora dessa região, porém notam-se 30 casos, ou seja, a ocorrência de casos extremos está abaixo do esperado. Ao considerar a região com 2,5 desvios padrões são esperados 9 casos fora desta região, no entanto, encontraram-se 13 casos, ficando agora acima do esperado, o mesmo é visto quando é alterado para 3 desvios padrões, espera-se apenas 1 caso fora da região, contudo, 9 casos foram encontrados. Devemos dar atenção especial para o caso 338 (Nova Lima) em que o resíduo padronizado foi 10,659, vale ressaltar que com 3 desvios padrões tem-se a cobertura de uma área de aproximadamente

99,865% das observações. Estes fatos podem indicar uma maior concentração dos dados no centro da distribuição com maior ocorrência de valores extremos, indicando assim uma distribuição dos resíduos como sendo leptocúrtica.

Quadro 1. Diagnóstico de casos modelo M1

Número de Caso	Residual padrão	Renda Per Capita	Valor previsto	Residual
8	3,339	764,770	475,223	289,547
67	-2,265	377,680	574,101	-196,421
152	2,372	538,980	333,315	205,665
191	3,122	850,810	580,050	270,760
219	2,010	770,840	596,529	174,311
225	2,182	813,830	624,635	189,195
227	2,042	788,590	611,493	177,097
240	2,446	787,190	575,081	212,109
288	3,890	1019,100	681,816	337,284
326	4,054	854,240	502,684	351,556
329	3,559	1027,990	719,343	308,647
338	10,659	1704,550	780,238	924,312
341	2,103	885,550	703,213	182,337
354	-2,741	498,270	735,931	-237,661
389	2,543	617,090	396,616	220,474
499	2,029	754,360	578,444	175,916
506	2,497	773,290	556,737	216,553
520	-2,471	475,180	689,457	-214,277
598	2,897	964,780	713,534	251,246
633	2,013	993,890	819,370	174,520
634	2,050	949,320	771,545	177,775
640	-2,073	508,580	688,350	-179,770
646	3,365	984,140	692,370	291,770
688	2,262	953,900	757,733	196,167
705	3,601	928,210	615,987	312,223
733	3,082	762,510	495,296	267,214
759	2,534	1095,860	876,090	219,770
806	-2,650	617,730	847,503	-229,773
811	-4,862	465,710	887,326	-421,616
845	2,839	961,690	715,537	246,153

Aplicou-se o teste de Kolmogorov-Smirnov com correção de Liliefors para averiguar a pressuposição de normalidade dos resíduos, este teste apresentou o valor de 0,0616 para sua estatística e um p-valor abaixo de 0,0001, sugerindo assim a rejeição da hipótese de normalidade dos resíduos. Utilizou-se o gráfico Q-Q plot para o modelo M1 a fim de explorar possíveis causas para a indicação de não normalidade dos resíduos indicada pelo teste, e que

pode ser observado na Figura 5, onde observa-se a presença de duas observações distantes da reta de normalidade e desvio na cauda direita, um apontamento de que o desvio da normalidade pode ter ocorrido devido a presença de outliers e assimetria.

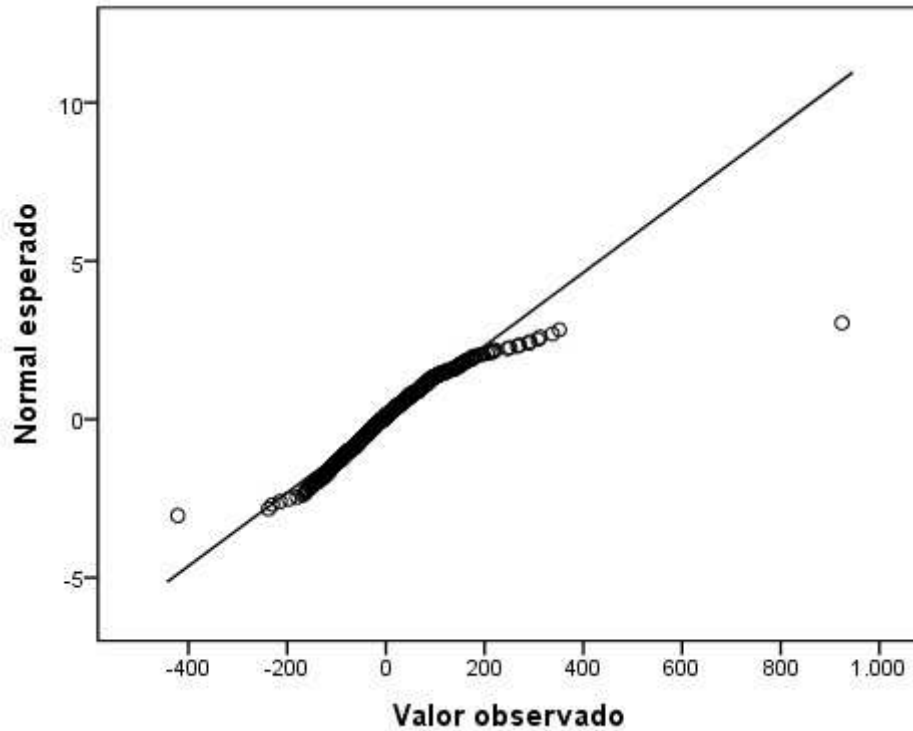


Figura 5. Gráfico normal de probabilidades (Q-Q plot) para o modelo M1

Nota-se na Figura 6 uma alteração na variação dos resíduos, ao passo que o valor previsto aumenta a variação também aumenta, observamos inclusive ocorrência de pontos com grande distanciamento dos demais. Tais observações nos sugerem que os resíduos apresentam heterocedasticidade, ou seja, as variâncias diferem entre os casos.

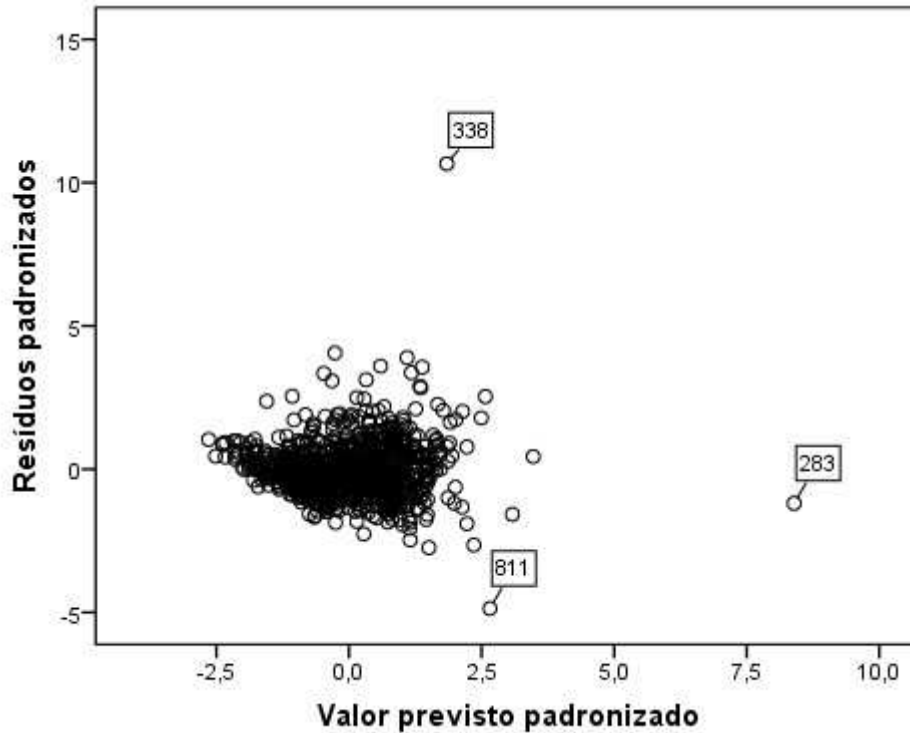


Figura 6. Dispersão dos resíduos padronizados e os valores previstos padronizados

Em virtude do modelo não atender as pressuposições de normalidade e homocedasticidade impostas pelo método, foi proposto uma transformação na variável resposta, selecionamos a transformação logarítmica neperiana devido sua propriedade de estabilizar a variância quando esta tende a crescer à medida que a variável resposta também cresce, há casos em que esta transformação pode também auxiliar na normalização dos dados. A seguir são apresentados os ajustes do modelo linear múltiplo transformado.

A aplicação da transformação logarítmica neperiana é possível neste caso em razão da variável resposta Renda Per Capita possuir valores maiores que zero, não sendo possível a ocorrência desta variável com valor negativo ou igual a zero. Este novo modelo será chamado de *M2* e será expresso da seguinte forma:

$$\ln \widehat{Rcp} = \hat{\beta}_0 + \hat{\beta}_1 Fem + \hat{\beta}_2 Sa + \hat{\beta}_3 Vei + \hat{\beta}_4 Ed + \hat{\beta}_5 Urb + \hat{\beta}_6 Emp + \hat{\beta}_7 Pop$$

Na Figura 7 e Figura 8 temos o comparativo dos gráficos de dispersão das variáveis resposta e variáveis regressoras, entre o modelo transformado e o modelo não transformado.

Observa-se melhora na aproximação de observações muito distantes dos demais pontos após aplicação da transformação logarítmica natural.

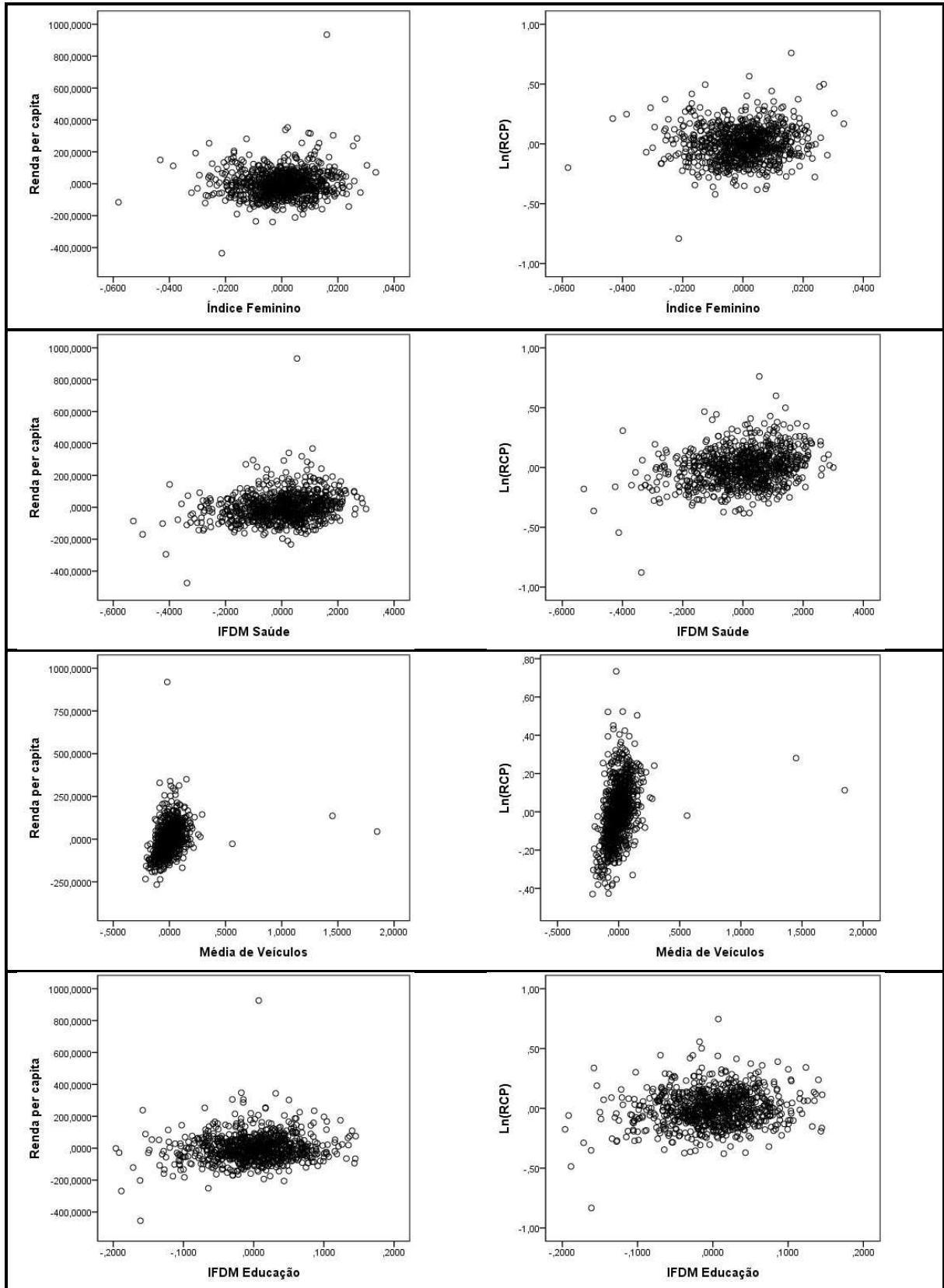


Figura 7. Comparativo do modelo linear e o modelo transformado a partir dos gráficos de dispersão entre a variável resposta e as variáveis regressoras: Fem, Sa, Vei, Ed.

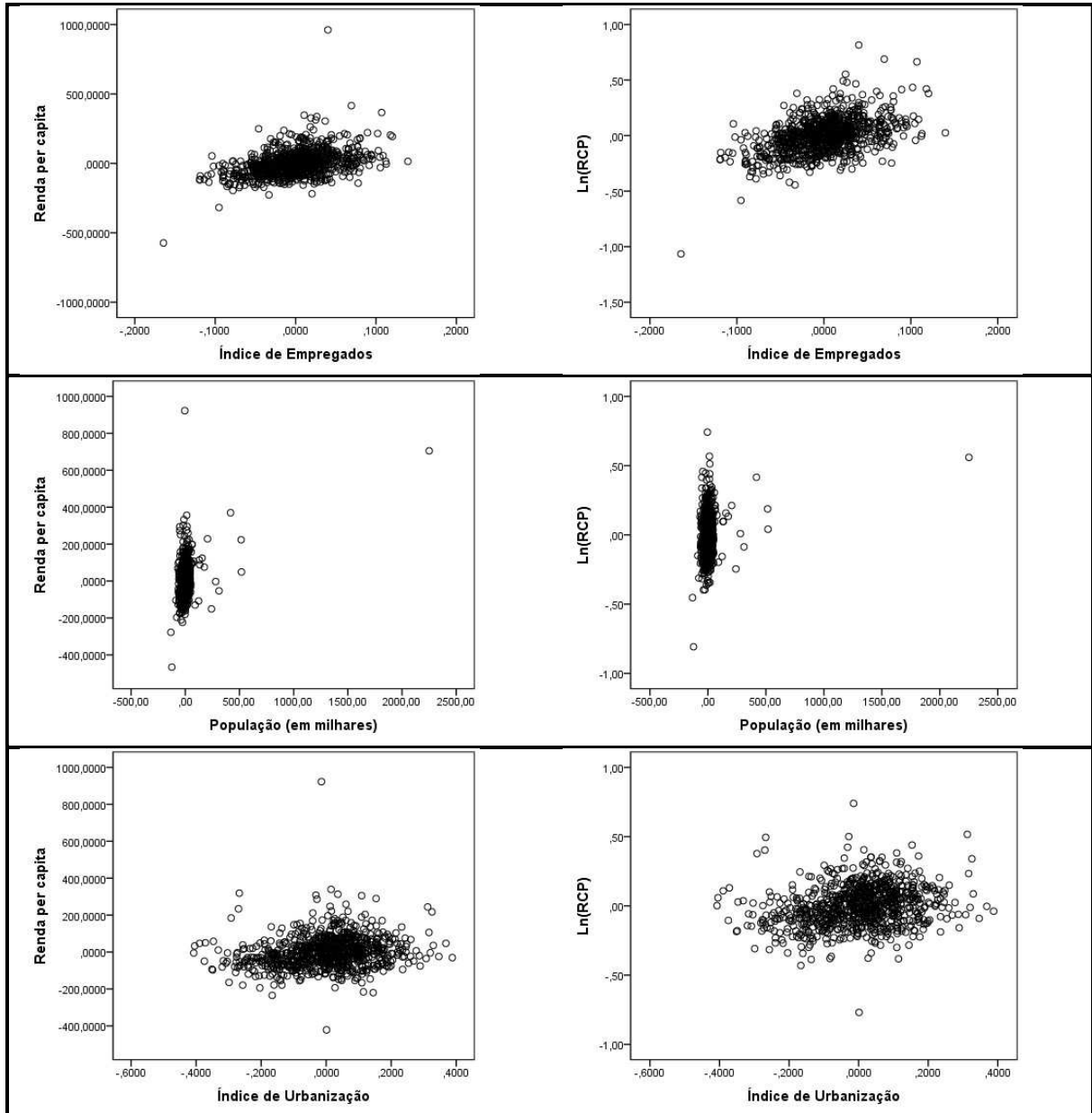


Figura 8. Comparativo do modelo linear e o modelo transformado a partir dos gráficos de dispersão entre a variável resposta e as variáveis regressoras: Emp, Pop e Urb

Após a transformação dos dados obteve-se a correlação da nova variável transformada com as variáveis a compor o  $M2$ , conforme pode-se observar na Tabela 6. As variáveis que tiveram correlação moderada com Renda Per Capita mantiveram este nível com a variável  $\ln(Rpc)$ .



Tabela 6. Correlações de Pearson entre variáveis\* do modelo M2 e p-valor calculado\*\*

	Ln(Rpc)	Fem	Sa	Veic	Ed	Emp	Pop	Urb
Ln(Rpc)	-	0,209	0,543	0,600	0,565	0,766	0,289	0,650
Fem	0,000	-	-0,006	0,117	0,154	0,071	0,257	0,392
As	0,000	0,437	-	0,385	0,428	0,462	0,098	0,291
Veic	0,000	0,000	0,000	-	0,427	0,448	0,182	0,365
Ed	0,000	0,000	0,000	0,000	-	0,512	0,112	0,439
Emp	0,000	0,019	0,000	0,000	0,000	-	0,164	0,646
Pop	0,000	0,000	0,002	0,000	0,001	0,000	-	0,225
Urb	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-

\* Logaritmo neperiano da Renda per capita (Ln(Rpc)), Índice Feminino (Fem), IFDM Saúde (Sa), Média de Veículos (Veic), IFDM Educação (Ed), Índice de Empregados (Emp), População (em milhares) (Pop) e Índice de Urbanização (Urb).

\*\*Acima da diagonal principal estão os valores das correlações e abaixo seus respectivos p-valores.

A avaliação inicial do modelo se deu através da aplicação da análise de variância. A estatística F mostrada na Tabela 7 apresentou um valor 346,117 indica, ao de nível de significância de 5%, que pelo menos um dos parâmetros do modelo é diferente de zero e que o modelo é mais eficiente que a média para prever o  $\ln(Rpc)$ .

Tabela 7. Análise de variâncias modelo M2

	Soma dos Quadrados	Gl	Quadrado Médio	F	Sig.
Regressão	50,197	7	7,171	346,117	0,000
Residual	17,403	840	0,021		
Total	67,601	847			

O modelo obtido com a transformação logarítmica apresentou coeficiente de correlação positivo de 0,862 e consequentemente um coeficiente de determinação de 0,7425, assim o este modelo consegue capturar cerca de 74,25% da variação total dos dados, logo a aplicação da função logarítmica nos dados de Renda Per Capita proporcionou um incremento de 4,34% na captura da variação total em relação ao modelo sem a transformação. A estatística de Durbin-Watson aferida para este modelo foi de 1,6277 mantendo-se próxima do valor 2 de referência, logo pode-se considerar que os resíduos são independentes.

As estimativas dos coeficientes de regressão e também testes t bilaterais para os coeficientes e os intervalos de confiança da ordem de 95% são apresentados na Tabela 8. Pode-se observar que em nenhum dos intervalos existe a presença do zero, logo todas as variáveis do modelo são significativas. Bem como no modelo anterior as variáveis deste também mantem uma relação positiva com a variável Renda Per Capita.

Tabela 8. Coeficientes do modelo M2

Variáveis do modelo	$\hat{\beta}$	$\hat{\beta}$ Padronizado	Sig.	Intervalo de confiança 95,0% para B	
				Limite inferior	Limite superior
(Constante)	4,3772		0,0000	3,9302	4,8242
Índice Feminino	1,0296	0,0455	0,0247	0,1314	1,9279
IFDM Saúde	0,3227	0,1669	0,0000	0,2437	0,4017
Média de Veículos	0,4771	0,2254	0,0000	0,3908	0,5634
IFDM Educação	0,3945	0,0933	0,0000	0,2129	0,5761
Índice de Empregados	1,7993	0,4081	0,0000	1,5691	2,0296
População (em milhares)	0,0003	0,1030	0,0000	0,0002	0,0004
Índice de Urbanização	0,2572	0,1730	0,0000	0,1814	0,3330

Os valores padronizados para os coeficientes, que representam o número de desvios padrão que a variável resposta irá alterar como resultado de uma mudança de um desvio padrão no respectivo previsor, os valores  $\hat{\beta}$  padronizados são mensurados em termos de unidades desvios padrão e são, dessa forma, comparáveis diretamente. Ao compararmos os valores dos coeficientes padronizados constatamos que a variável Índice de Empregados é a de maior impacto no modelo, a cada unidade elevada de desvio padrão, desta variável, causará um aumento de 0,408 desvios padrão no  $\ln(Rpc)$ , Média de Veículos vem logo em seguida provocando um aumento de 0,225 desvios padrão e a variável de menor influência é o Índice Feminino.

Com base nas informações apresentadas na Tabela 8, podemos construir o modelo de regressão para a variável transformada  $\ln(Rpc)$ , o qual é apresentado a seguir:

$$\widehat{\ln Rpc} = 4,377 + 1,03Fem + 0,323Sa + 0,477Vei + 0,394Ed + 0,257Urb + 1,799Emp + 0,00031Pop$$

Para estimativas sobre a Renda Per Capita devemos utilizar o seguinte modelo

$$\widehat{Rpc} = 79,61e^{1,03Fem+0,323Sa + ,477Vei+0,394Ed+0,257Urb+1,799Emp+0,00031Pop}$$

Os valores da estatística FIV (Fator de Inflação de Variância) mantiveram-se todos abaixo de 10 (MYERS, 1990) (BOWERMAN e O'CONNELL, 1990) e as tolerâncias também ficaram acima de 0,2 (MENARD, 1995), logo podemos manter a posição de que não existe multicolinearidade.

O gráfico de normalidade na Figura 9 indica progresso, decorrente da aplicação da transformação logarítmica, na aderência da distribuição dos resíduos a uma distribuição normal. Neste gráfico, observam-se dois pontos 811 (Santa Bárbara do Monte Verde) e 338 (Nova Lima) que se distanciam da reta de referência, indicando que estes municípios podem ser outliers, cabendo análise mais detalhada sobre esses casos.

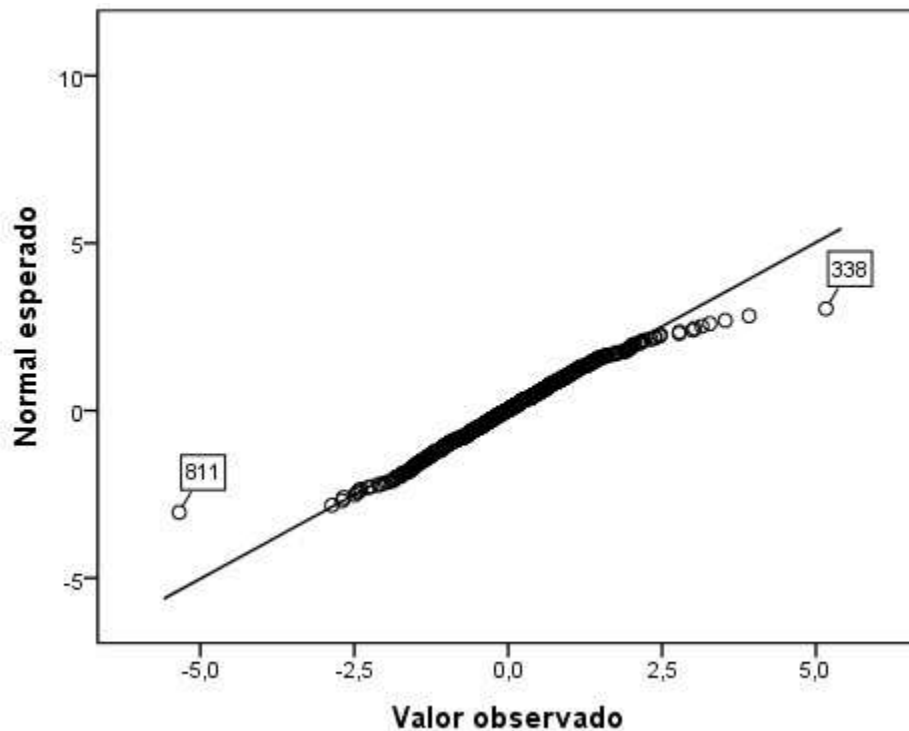


Figura 9. Gráfico de dispersão dos resíduos padronizados e os valores previstos padronizados

O teste Kolmogorov-Smirnov com correção de Liliefors apresentou valor de 0,0289 e p-valor de 0,0908 indicando que não há evidências para rejeitarmos, ao nível de significância de 5%, a hipótese de que os resíduos provem de uma distribuição normal. Figura 10 são apresentados os resíduos padronizados vs valor previsto normalizado. Pode-se verificar que os pontos que antes se apresentavam mais concentrados estão mais dispersos e os mais distantes 338 e 283 se aproximaram dos demais, também não observamos tendência, sugerindo que as suposições de linearidade e homocedasticidade foram atendidas.

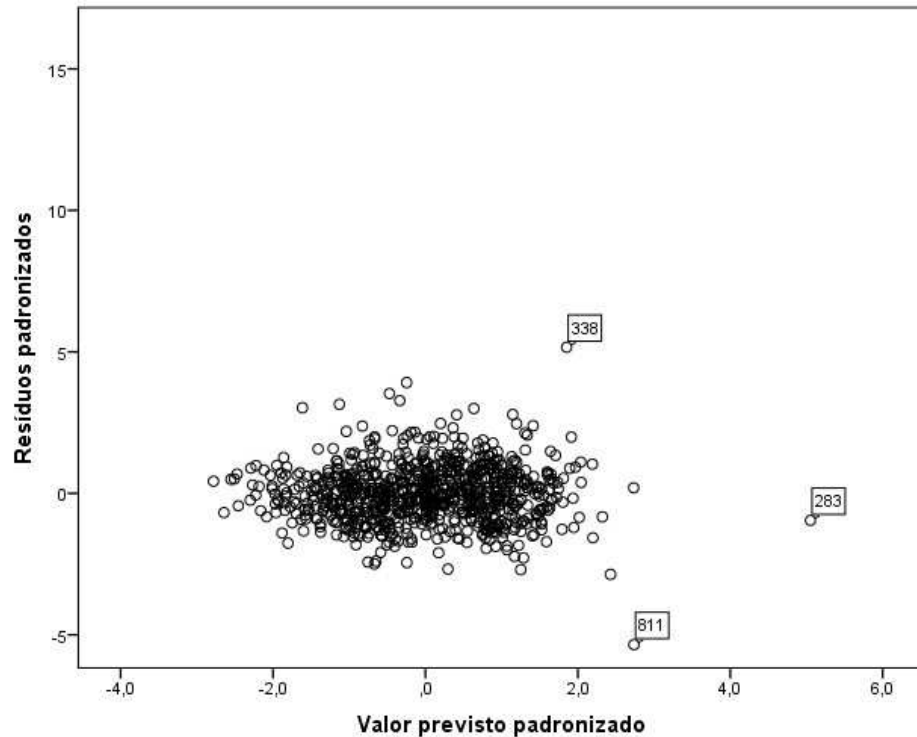


Figura 10. Resíduos padronizados vs Valor previsto padronizado modelo M2

Para auxiliar na avaliação do modelo *M2* sobre as observações atípicas foram consideradas as medidas de *leverage*, *DFFitS*, *DFBetaS*, Distância de *Cook* e Resíduos Studentizados (Tabela 9).

Tabela 9. Casos atípicos do modelo M2

Municípios	Cod	hii	Res(i)	Cook	DFFIT	Vei	Emp	Pop	Pt Atípico
Belo Horizonte	283	<b>0,767</b>	-1,989	<b>1,634</b>	<b>-3,621</b>	0,126	0,073	<b>-3,583</b>	Influente
Braúnas	389	0,012	3,185	0,017	<b>0,369</b>	-0,042	0,058	-0,025	Influente
Campo Florido	191	0,020	2,819	0,022	<b>0,417</b>	0,045	0,085	0,014	Influente
Carmésia	395	<b>0,039</b>	-2,041	0,022	<b>-0,419</b>	-0,356	-0,039	0,063	Influente
Chiador	733	0,014	3,322	0,021	<b>0,412</b>	-0,045	0,292	-0,058	Influente
Contagem	301	<b>0,046</b>	-0,849	0,004	-0,188	0,010	-0,008	-0,174	Influente
Delta	204	<b>0,046</b>	-0,986	0,006	-0,220	-0,060	0,008	-0,015	Influente
Jequitibá	326	0,009	3,968	0,019	<b>0,393</b>	-0,106	0,226	0,020	Influente
José Gonçalves de Minas	138	<b>0,033</b>	-0,065	0,000	-0,012	-0,001	-0,006	-0,001	Influente
Juiz de Fora	759	<b>0,035</b>	2,032	0,020	<b>0,395</b>	-0,039	-0,013	0,333	Influente
Nova Lima	338	0,007	<b>5,266</b>	0,027	<b>0,472</b>	-0,032	0,173	-0,010	Outlier; Inf.
Rio Preto	806	<b>0,203</b>	-3,225	0,329	<b>-1,631</b>	<b>-1,602</b>	0,281	0,185	Influente
Salto da Divisa	152	0,024	3,081	0,030	<b>0,494</b>	-0,084	-0,079	0,020	Influente
Santa Bárbara do Monte Verde	811	<b>0,323</b>	<b>-6,669</b>	<b>2,535</b>	<b>-4,619</b>	<b>-4,580</b>	<b>1,087</b>	0,388	Incons.; Inf.
Uberlândia	244	<b>0,047</b>	0,201	0,000	0,045	0,004	0,000	0,041	Influente
União de Minas	245	0,021	2,028	0,012	<b>0,308</b>	0,018	0,023	0,044	Influente

Ponto Influente é aquele que a sua ausência causa impacto significativo sobre os valores ajustados, a influência pode ser verificada no modelo de forma geral ou apenas nas estimativas de uma ou mais variáveis.

Como critério para seleção de prováveis pontos influentes no modelo utilizamos,  $h_{ii} > 3p/n \cong 0,028$ ,  $|DFFITs_{(i)}| > (3\sqrt{p/(n-p)}) \cong 0,2928$ ,  $|DFBetaS_{(i)}| > 1$  e distância de  $Cook > 1$ . Através do critério DFFitS foram detectadas 12 observações de provável influência para o modelo, a distância de Cook apontou para apenas dois que se distanciam dos demais, já o DFBetaS indicou que em apenas 3 variáveis existem pontos que as influenciam sendo 4 observações as responsáveis. Dentre estas observações destacamos Belo Horizonte, Santa Bárbara do Monte Verde e Rio Preto, pontos de grande influência para o modelo e também para alguns parâmetros.

Pontos inconsistentes são casos que divergem da orientação dos demais. Considerando-se as observações que apresentaram  $|rse_{(i)}| > (t_{n-p-1}; 1 - \frac{\alpha}{2n}) \cong 4,037$  pontos inconsistentes, foram encontrados dois casos, Nova Lima e Santa Bárbara do Monte Verde.

Outliers são observações que não apresentam o mesmo padrão que a maioria das outras observações (ROUSSEUW e VAN ZOMEREN, 1990). São considerados outliers os pontos inconsistentes que apresentam medida de *leverage* pequena, ou seja, apresentaram  $|rse_{(i)}| > (t_{n-p-1}; 1 - \frac{\alpha}{2n}) \cong 4,037$  e  $h_{ii} < 2p/n \cong 0,019$ . Nestas condições apenas um caso foi constatado, Nova Lima.

A Tabela 10 traz as estatísticas descritivas das observações atípicas detectadas pelas medidas de diagnósticos, a estatística descritiva permite compreender com maior clareza o que na realidade fazem destas observações casos atípicos, como é a condição de Belo Horizonte, esta cidade está acima do limite superior em todas as variáveis, isto não seria problema pois o modelo prevê que todas as variáveis contribuem para um aumento na renda per capita, porém o desvio desta cidade em algumas variáveis é de extrema relevância como ocorre com a população, enquanto a média é de 23 mil habitantes, Belo Horizonte possui 2,375 milhões de residentes. Nova Lima é destaque na renda per capita, 1.704,5, mesmo apresentando todos os aspectos acima do limite superior, a renda apresentada não condiz com o observado nos demais casos. Santa Bárbara do Monte Verde é um município pequeno de pouco mais de 2 mil habitantes, todos os seus índices apresentaram valores abaixo do limite inferior, exceto pela média de veículos que é de mais de 2 carros por residente, enquanto que a média é de 0,25.

Tabela 10. Estatística descritiva das observações atípicas

	Rpc	Fem	Sa	Vei	Ed	Emp	Pop	Urb
Média	516,4	0,495	0,671	0,259	0,736	0,305	23,1	0,679
Limite inferior	506,7	0,494	0,661	0,250	0,731	0,300	16,7	0,666
Limite superior	526,3	0,496	0,680	0,268	0,740	0,309	29,4	0,691
Belo Horizonte	1539,3	0,531	0,810	0,564	0,803	0,396	2375,2	1,000
Braúnas	617,1	0,510	0,710	0,153	0,674	0,232	5,0	0,317
Campo Florido	850,8	0,464	0,767	0,292	0,600	0,372	6,9	0,758
Carmésia	502,6	0,502	0,805	0,802	0,638	0,323	2,4	0,536
Chiador	762,5	0,513	0,536	0,171	0,655	0,338	2,8	0,534
Contagem	806,6	0,515	0,824	0,388	0,771	0,392	603,4	0,997
Delta	572,8	0,439	0,681	0,460	0,664	0,397	8,1	0,938
Jequitibá	854,2	0,483	0,788	0,173	0,711	0,331	5,2	0,381
José Gonçalves de Minas	368,6	0,482	0,159	0,191	0,754	0,300	4,6	0,250
Juiz de Fora	1095,9	0,527	0,814	0,346	0,783	0,357	516,2	0,989
Nova Lima	1704,5	0,517	0,835	0,371	0,818	0,413	81,0	0,978
Rio Preto	617,7	0,495	0,475	1,650	0,663	0,295	5,3	0,841
Salto da Divisa	539,0	0,489	0,147	0,031	0,660	0,241	6,9	0,838
Santa Bárbara do Monte Verde	465,7	0,478	0,635	2,052	0,709	0,239	2,8	0,577
Uberlândia	1033,0	0,512	0,893	0,515	0,809	0,399	604,0	0,972
União de Minas	585,6	0,454	0,398	0,218	0,702	0,317	4,4	0,617
Média dos pontos atípicos	738,9	0,495	0,642	0,524	0,713	0,334	264,6	0,720

Devido a transformação logarítmica aplicada no modelo as relações entre as variáveis independentes e a variável dependente não são realizadas de forma proporcional. O modelo ajustado é também conhecido como Log-linear, neste modelo alterações nas variáveis independentes causam um efeito de crescimento exponencial na variável dependente. Na Tabela 11 são apresentadas as taxas de crescimento para a Renda Per Capita conforme aumento de determinada variável dependente, isoladamente, por exemplo: um aumento de 0,1 na variável IFDM Educação, em que todas as outras variáveis permaneçam inalteradas, é esperado um crescimento de 4,02% na Renda Per Capita. As variáveis Índice Feminino, IFDM Saúde, IFDM Educação, Índice de Empregados e Índice de Urbanização, são passíveis de comparação entre elas, devido possuírem a mesma escala (intervalo de 0 a 1), sendo assim, observa-se que Índice de Empregados é a variável que proporcionará maior crescimento na Renda Per Capita, seguido do Índice Feminino. A População (em milhares) mostra-se pouco influente, um aumento de 10 mil habitantes produz uma taxa de crescimento de 0,31% na Rpc, já um aumento de 10% na Média de Veículos proporciona um crescimento de 4,89%.

Tabela 11. Taxa de crescimento esperado da renda per capita de acordo o incremento nas variáveis independentes

	Incremento	Taxa de crescimento (i)
Índice Feminino	0,1	0,1084
IFDM Saúde	0,1	0,0328
Média de Veículos	0,1	0,0489
IFDM Educação	0,1	0,0402
Índice de Empregados	0,1	0,1971
População (em milhares)	10	0,0031
Índice de Urbanização	0,1	0,0261

Neste modelo, ao analisar aumento simultâneo de duas ou mais variáveis as taxas de crescimento serão acumuladas, logo, caso ocorra um aumento simultâneo de 0,1 em Média de Veículos e Índice de Empregados, espera-se uma taxa de crescimento de  $0,2556 [1 + 0,0489) \cdot (1 + 0,1971) - 1]$ .

O ajuste deste modelo mostra-se importante ao identificar características que mais influenciam na Renda Per Capita, sendo assim no crescimento econômico dos municípios mineiros, constituindo dessa forma consulta relevante para definição de políticas de distribuição orçamentária no estado de Minas Gerais que tenham como objetivo proporcionar crescimento e desenvolvimento econômico.

## 5 CONSIDERAÇÕES FINAIS

Por meio da análise de regressão constatou-se que há relação entre as variáveis, sendo possível a variável renda per capita ser descrita em função: do tamanho da população, do Índice de Empregados, da Média de Veículos, do IFDM Educação, do IFDM Saúde, do Índice de Urbanização e do Índice Feminino. O modelo apresentou bom ajuste, conseguindo captar 74,25% da variabilidade total da renda, tal ajuste descreve que há uma relação positiva entre a renda e todas as outras variáveis, e devido o modelo ser do tipo Log-linear as variáveis independentes causam um efeito de crescimento exponencial na variável dependente.

Dentre as variáveis independentes a Índice de Empregados é a variável de maior impacto na predição da renda, considerando sua variação. Para um incremento de 0,1 no Índice de Empregados é esperada uma taxa de crescimento de 19,71% na renda. Já para as variáveis IFDM Educação, IFDM Saúde, Índice de Urbanização e Índice Feminino um aumento de 0,1 causa uma taxa de crescimento na Renda Per Capita de 4,02%, 3,28%, 2,61% e 10,84%, respectivamente. No caso da População (em milhares) esta mostrou-se pouco influente, um aumento de 10 mil habitantes produz uma taxa de crescimento de 0,31% na Rpc, e para um aumento de 0,1 na Média de Veículos proporciona um crescimento de 4,89%.

O fato da variável “Índice Feminino” apresentar-se significativa para o a predição da renda per capita e seu coeficiente ser positivo, confirma o relato do banco mundial, que o desenvolvimento econômico está ligado com a equidade de gênero, o aumento da oportunidade para as mulheres, e apoia a ideia do banco mundial de que políticas focadas em correção de hiatos de gêneros são fundamentais e impactam no desenvolvimento econômico. O estudo confirma, também, a afirmação dos economistas de que a educação é fonte de crescimento.

O modelo mostra que crescimento e desenvolvimento econômico estão intimamente ligados não apenas pela renda, pois como foi possível observar, aspectos sociais estão ligados com a renda; um aumento na qualidade das características sociais (variáveis independentes) acarretará em um aumento na renda.



## REFERÊNCIAS

- BANCO MUNDIAL. Relatório sobre o Desenvolvimento Mundial: Igualdade de Género e Desenvolvimento 2012. **The World bank**, 2015. Disponível em: <<http://siteresources.worldbank.org/INTWDR2012/Resources/7778105-1299699968583/7786210-1315936231894/Overview-Portuguese.pdf>>. Acesso em: 16 novembro 2015.
- BCB. Economia Mineira: estrutura produtiva e desempenho recente. **Banco Central do Brasil**, Janeiro 2013. Disponível em: <<http://www.bcb.gov.br/pec/boletimregional/port/2013/01/br201301b3p.pdf>>. Acesso em: 20 Novembro 2015.
- BOWERMAN, B. L.; O'CONNELL, R. T. **Linear statistical models: an applied approach**. Belmont: Duxbury, 1990.
- CARVALHO, J. L. et al. **Fundamentos de economia**. São Paulo: Cengage Learning, 2008.
- COOK, R. D.; WEISBERG, S. **Residuals influence in regression**. New York: Chapman & Hall, 1982.
- DESENVOLVIMENTO, P. D. N. U. P. O. Programa das Nações Unidas para o Desenvolvimento. **PNUD Brasil**, 2015. Disponível em: <[http://www.pnud.org.br/idh/IDHM.aspx?indiceAccordion=0&li=li\\_IDHM](http://www.pnud.org.br/idh/IDHM.aspx?indiceAccordion=0&li=li_IDHM)>. Acesso em: 19 novembro 2015.
- DRAPER, N. R.; SMITH, H. **Applied Regression Analysis**. New York: Wiley, 1998.
- FERREIRA, F. H. G.; LITCHFIELD, J. Os determinantes da desigualdade de renda no Brasil: luta de classes ou heterogeneidade educacional? In: HENRIQUES, R. **Desigualdade e Pobreza no Brasil**. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada, 2000. p. 49-80.
- FIELD, A. **Descobrimo a Estatística usando o SPSS**. 2. ed. Porto Alegre: Artmed, 2009.
- FIRJAN, F. D. I. D. R. D. J. Índice FIRJAN de Desenvolvimento Municipal. **FIRJAN**, 2015. Disponível em: <<http://www.firjan.com.br/ifdm>>. Acesso em: 20 novembro 2015.
- FJP, F. J. P. Índice Mineiro de Responsabilidade Social (IMRS). **Fundação João Pinheiro**, 2015. Disponível em: <<http://www.fjp.mg.gov.br/>>. Acesso em: 20 novembro 2015.
- FRANCO, N. B. **Cálculo Numérico**. São Paulo: Prentice-Hall, 2006.
- IBGE. **Atlas do censo demográfico 2010**. Rio de Janeiro: IBGE, 2013. ISBN 9788524042812.
- IBGE, I. B. D. G. E. E. Estados. **IBGE**, 2015. Disponível em: <<http://www.ibge.gov.br/estadosat/>>. Acesso em: 20 novembro 2015.
- IBGE, I. B. D. G. E. E. IBGE - Cidades. **IBGE**, 2015. Disponível em: <[www.cidades.ibge.gov.br](http://www.cidades.ibge.gov.br)>. Acesso em: 20 novembro 2015.

KIM, L.; NELSON, R. R. **Tecnologia, aprendizado e inovação**: as experiências das economias de industrialização. Campinas: UNICAMP, 2005.

MENARD, S. **Applied logistic regression analysis**. Thousand Oaks: Sage, 1995.

MILES, J.; SHEVILIN, M. **Applying regression and correlation**: a guide for students and researchers. London: Sage, 2001.

MYERS, R. **Classical and modern regression with applications**. Boston: Duxbury, 1990.

ROUSSEUW, P. J.; LEROY, A. M. **Robust Regression and Outlier Detection**. New York: Wiley, 1987.

ROUSSEUW, P. J.; VAN ZOMEREN, B. C. Unmasking Multivariate Outliers and Leverage. **Journal of the American Statistical Association**, v. 85, n. 411, p. 633-639, 1990.

SANDRONI, P. **Novíssimo Dicionário de Economia**. Best Seller: São Paulo, 1999.

SOUSA, R. B. D. **Cooperativismo de crédito – do básico ao gerencial**. João Pessoa: UFPB, 2006.

STEVENS, J. **Applied multivariate statistics for the social sciences**. Hillsdale: Erlbaum, 1992.

STN, S. D. T. N. Contas Anuais - Artigo Prefeituras e Governos - STN. **Tesouro Nacional**, 2015. Disponível em: <[http://www.tesouro.fazenda.gov.br/pt\\_PT/contas-anuais](http://www.tesouro.fazenda.gov.br/pt_PT/contas-anuais)>. Acesso em: 20 novembro 2015.

VASCONCELLOS, M. A. S. **Economia**: micro e macro. São Paulo: Atlas, 2011.

## ANEXO A – DEMONSTRAÇÃO TAXA DE CRESCIMENTO

$$\ln Y = X\beta + \varepsilon$$

$$\widehat{\ln Y} = X\hat{\beta}$$

$$\hat{Y} = e^{X\hat{\beta}}$$

$$\hat{Y} = e^{\hat{\beta}_0 + \hat{\beta}_1 x}$$

$$\frac{\hat{y}_b}{\hat{y}_a} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_b}}{e^{\hat{\beta}_0 + \hat{\beta}_1 x_a}}$$

$$\therefore \hat{\beta}_1 = \frac{\ln \frac{y_b}{y_a}}{x_b - x_a} = \frac{\ln \frac{y_b}{y_a}}{h} \quad h = x_b - x_a$$

$$\frac{\hat{y}_1}{\hat{y}_0} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1}}{e^{\hat{\beta}_0 + \hat{\beta}_1 x_0}}$$

$$\hat{y}_1 = \hat{y}_0 [e^{\hat{\beta}_1 (x_1 - x_0)}]$$

$$\hat{y}_1 = \hat{y}_0 \left[ e^{\frac{\ln(\frac{y_b}{y_a})}{h} (x_1 - x_0)} \right] \quad \Delta x = x_1 - x_0$$

$$\hat{y}_1 = \hat{y}_0 \left[ e^{\frac{\ln(i+1)}{h} \Delta x} \right] \quad i = \frac{y_b - y_a}{y_a} = \frac{y_b}{y_a} - 1$$

$$\therefore \hat{y}_1 = \hat{y}_0 (1 + i)^{\frac{\Delta x}{h}}$$

*i*: taxa de crescimento referente ao intervalo *h*;

*h*: é o intervalo de *x* que foi utilizado para obter a taxa de crescimento