

A escolaridade em 1991 permite prever a renda em 2010? Uma análise de regressão usando dados municipais dos censos demográficos do IBGE

Leda Grasielle Oliveira¹,

ledagrasielle@gmail.com, UNIFAL-MG, Brasil

Patrícia de Siqueira Ramos, Dra.

siqueirapaty@gmail.com, UNIFAL-MG, Brasil

Lincoln Frias, Dr.

lincolnfrias@gmail.com, UNIFAL-MG, Brasil

Resumo

Há um consenso em torno da ideia de que a educação é o principal mecanismo de desenvolvimento econômico de um país e também de mobilidade social individual. O objetivo desta pesquisa foi tomar esse consenso como hipótese e buscar evidências empíricas que pudessem confirmar sua veracidade, pois é fundamental que os municípios monitorem a situação educacional de seus cidadãos e planejem políticas públicas de acordo com a situação identificada. Além da análise do coeficiente de correlação, foi utilizada a análise de regressão múltipla para mensurar a associação entre escolaridade e renda. A análise, feita por meio da linguagem R, investigou o estado de Minas Gerais e o Brasil como um todo. A variável explicada – renda – foi avaliada a partir do censo demográfico de 2010, enquanto as variáveis explicativas – indicadores educacionais – foram avaliadas a partir do censo de 1991. Os resultados confirmaram a hipótese de que a escolaridade no ano de 1991 influencia a renda em 2010, denotando, inclusive, que o estado de Minas Gerais pode ser considerado um retrato do Brasil. A confirmação de que os indicadores educacionais são capazes de prever a renda aponta para a importância do investimento em políticas públicas de educação.

Palavras-Chave: Educação. Renda. Regressão.

¹ Bolsista de iniciação científica do CNPq durante a realização desta pesquisa.

Abstract

There is a consensus around the idea that education is the main mechanism of economic development of a country and also of individual social mobility. The objective of this research was to take this consensus as a hypothesis and to search empirical evidence that could confirm its veracity, since it is fundamental that municipalities monitor the educational situation of their citizens and plan public policies according to the identified situation. In addition to the correlation coefficient analysis, multiple regression analysis was used to measure the association between schooling and income. The analysis, made using the R language, investigated the state of Minas Gerais and Brazil as a whole. The explained variable - income - was evaluated from the 2010 demographic census, while the explanatory variables - educational indicators - were evaluated from the 1991 census. The results confirmed the hypothesis that schooling in 1991 influences income in 2010, denoting, even, that the state of Minas Gerais can be considered a portrait of Brazil. The confirmation that educational indicators are capable of predicting income points to the importance of investing in public education policies.

Keywords: Education. Income. Regression.

1. Introdução

Como se pode facilmente observar, há um consenso em torno da ideia de que a educação é o principal mecanismo de desenvolvimento econômico de um país e também de mobilidade social individual. O objetivo desta pesquisa é tomar esse consenso como uma hipótese e buscar evidências empíricas que possam confirmar sua veracidade. É fundamental que os municípios monitorem a situação educacional de seus cidadãos e planejem políticas públicas de acordo com a situação identificada.

Na literatura especializada, há um debate acirrado sobre se, e como, a educação pode estar relacionada ao desenvolvimento econômico. Resende e Wyllie (2006), por exemplo, defendem que existe essa relação, enquanto Souza e Carvalhaes (2014) não encontram evidências dessa relação no caso da recente queda da desigualdade brasileira, apesar de realmente haver uma associação entre escolaridade e rendimento (ROCHA et al., 2012). Por sua vez, o influente livro *The race between education and technology* (GOLDIN; KATZ, 2009) documenta o fato de que o impacto da educação sobre o desenvolvimento é mediado pelas inovações tecnológicas.

De qualquer maneira, se a educação é realmente um pressuposto essencial da melhoria das condições de vida e do progresso econômico, é de se esperar, então, que uma maior escolaridade esteja associada a uma maior renda (MINCER, 1974). Isto é, a educação contribuiria para o desenvolvimento econômico sendo um determinante da renda por causa do fenômeno chamado "retorno à educação", a retribuição financeira marginal para o investimento em cada ano adicional de educação. E se isso for verdade no nível individual, provavelmente também seja verdadeiro no nível municipal. Sendo assim, esta pesquisa toma a renda per capita municipal como uma medida de desenvolvimento econômico e investiga se variações na escolaridade média estão associadas a variações no desenvolvimento econômico (isto é, à renda per capita).

Além disso, é usada uma defasagem temporal, pois o objetivo aqui proposto é avaliar a relação entre educação e renda, portanto não é adequado analisá-las na mesma unidade temporal porque a hipótese que está em jogo é de que indivíduos mais escolarizados têm um melhor desempenho ocupacional. Quando avaliadas na mesma unidade temporal, a escolaridade acaba sendo mais consequência do que causa do nível de renda. Portanto, é preciso avaliar a escolaridade em uma unidade temporal anterior, quando os indivíduos responsáveis pelo nível de renda no período posterior já adquiriram sua escolaridade e já iniciaram sua vida laboral.

Por isso, a variável dependente, a renda, é avaliada a partir do Censo Demográfico do ano de 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE), e as variáveis explicativas, os indicadores educacionais, são retirados do Censo Demográfico do ano de 1991.

Portanto, o objetivo da pesquisa aqui proposta é utilizar a análise de regressão múltipla para mensurar a associação entre escolaridade e renda. O restante deste trabalho consiste na revisão de literatura sobre a relação entre educação e rendimentos, na metodologia, na descrição e análise dos resultados e na apresentação das considerações finais.

2. Revisão da literatura: a relação entre educação e renda

Como foi dito, há um consenso quase unânime de que a educação é capaz de trazer altos retornos para a população de um país como, por exemplo, maior desenvolvimento econômico, maior renda (devido a melhores oportunidades de emprego), maior chance de participação na democracia, maior possibilidade de ascensão de classe social etc. Um exemplo da importância que é dada ao tema é o slogan do governo federal anunciado no discurso de posse da presidenta Dilma Rousseff em sua reeleição: “Brasil, pátria educadora”. Além disso, em época de campanhas políticas, os candidatos geralmente têm como um dos focos principais os investimentos que pretendem fazer em educação. A população também tem esse pensamento, pois, em geral, exige e é a favor de mais investimentos no setor.

A educação, além de proporcionar maiores ganhos em rendimentos futuros, por ajudar na busca de um emprego melhor, também possui impacto em diversos outros fatores como, por exemplo, a diminuição da mortalidade infantil observada em Taiwan após uma lei que obrigava os pais a mandarem os filhos à escola até certa idade (BANERJEE; DUFLO, 2012). Além disso, no Malawi e no Quênia, as garotas que não desistiam dos estudos tinham menores chances de engravidar (BANERJEE; DUFLO, 2012).

Há ainda a questão do empoderamento feminino por meio da educação. As mulheres escolarizadas são mais propensas a desfrutar de um trabalho satisfatório, a cuidar de sua própria saúde, de sua família, a dar opiniões em debates públicos etc. A educação das mulheres também tem importância fundamental para o crescimento econômico e para a saúde das crianças, promovendo a capacidade de absorver novas informações sobre saúde, nutrição e higiene e para estimular e facilitar a aprendizagem das crianças (SEN, 2000).

Ainda nessa linha do empoderamento feminino, Maciel, Campêlo e Raposo (2001) analisam as mudanças na distribuição salarial das mulheres no Brasil nos anos de 1992 e 1999, dando destaque ao retorno em educação. Utilizando o método de regressão quantílica concluíram que houve uma crescente participação da mulher na força de trabalho no período estudado e que, de 1992 para 1999, houve um decréscimo no retorno em educação para as mulheres que possuíam até o ensino fundamental, indicando que em 1999 a escolaridade passou a ser mais significativa. Já para as mulheres que possuíam ensino médio e ensino superior, o retorno em educação foi superior em 1999, denotando uma maior valorização da educação no salário.

Filho e Pessoa (2008) estudaram os retornos da educação através da Taxa Interna de Retorno (TIR), uma taxa que iguala o valor presente dos custos de um ano a mais de educação ao valor presente dos benefícios deste ano adicional investigando se existe um prêmio salarial por completar um ano, ou um ciclo (primário, fundamental, médio e superior), a mais de educação. No trabalho é dado destaque ao chamado “efeito diploma”, refletido em um aumento na média de salários para os anos de educação em que se completam os ciclos quando comparados aos anos anteriores dos mesmos. Esse efeito pode ser observado para todos os ciclos, nos anos 4 (fundamental I), 8 (fundamental II), 11 (ensino médio) e 15 (ensino superior).

Além disso, o período pré-escolar pode ser de extrema importância na formação dos indivíduos. Filho e Pessoa (2008) ressaltam ainda que a aplicação da TIR leva em consideração o fato de que crianças que frequentam a pré-escola possuem uma probabilidade maior de completar o próximo ciclo escolar do que aquelas que não a frequentam. Um resultado importante é que a menor taxa de retorno média ocorre no primeiro ano de estudo (9,2%). Todas as outras taxas se encontram acima desse nível, o que aponta para a importância do investimento em educação no Brasil. Já as taxas de retorno médias por ciclo de educação são bastante elevadas devido, possivelmente, ao já mencionado efeito diploma, sendo de 13,7% para o ginásio e 19,1% para o ensino médio. Sendo assim, Filho e Pessoa (2008) confirmam a ideia de que a educação é um investimento altamente rentável tanto para os indivíduos, quanto para o país como um todo, devido às altas taxas de retorno observadas para o país.

Porém, para Resende e Wyllie (2006), estudos não econométricos, como o citado acima, nos quais são identificados os custos de educação e rendimentos do trabalho para o cálculo da taxa interna de retorno pecam pela dificuldade em efetuar o controle para outras variáveis que também podem ser determinantes dos rendimentos, conforme seria mais facilmente alcançável em uma análise econométrica. Em seu trabalho, Resende e Wyllie (2006) investigam os retornos para educação no Brasil e seus resultados apontam para retornos importantes para a escolaridade formal. Os retornos encontrados foram de 12,6% para mulheres e 15,9% para homens, ao contrário dos resultados de Silva e Kassouf (2000), que encontram retornos maiores para as mulheres em relação aos homens.

Silva e Kassouf (2000), com base nas equações de rendimentos estimadas calculam os retornos não somente à escolaridade, mas também à experiência, para homens e mulheres no mercado formal e

informal. Em relação ao retorno à escolaridade, esse variou de 4,58%, para homens no mercado informal, a 14,61%, para os homens no mercado formal. Já para as mulheres, as taxas de retorno à escolaridade variaram de 8,22%, no mercado informal, a 17,71% no mercado formal, ou seja, os retornos à escolaridade são maiores para as mulheres nos dois mercados de trabalho. É importante destacar ainda que os retornos à escolaridade foram maiores do que os retornos à experiência, com exceção para homens que estão no mercado de trabalho informal.

Os retornos à escolaridade obtidos no mercado de trabalho formal para o homem e para a mulher são de aproximadamente 3 e 2 vezes, respectivamente, os obtidos no mercado informal. Silva e Kassouf (2000) investigaram ainda a desigualdade salarial devido à discriminação por gênero, verificando que, no mercado de trabalho formal, os diferenciais de salário que existem entre homens e mulheres são inteiramente devidos à discriminação, enquanto no mercado de trabalho informal essa questão continua num patamar bastante grave, pois as diferenças nas características individuais explicam apenas 3% do diferencial de rendimentos e o restante (97%) também é determinado pela discriminação (SILVA; KASSOUF, 2000). Dessa forma, pode-se dizer que a discriminação salarial por gênero ocorre em ambos os mercados, formal e informal, sendo um pouco maior no mercado de trabalho formal.

Analisando a relação entre educação e desenvolvimento sob outro aspecto, Curi e Filho (2010) examinam os determinantes dos gastos com educação no Brasil estimando modelos em que as famílias decidem matricular ou não seus filhos em uma escola privada e modelos para analisar o nível de gastos dessas famílias com educação. A principal conclusão é que os principais determinantes da decisão das famílias de matricular os filhos nas escolas privadas são a educação da mãe, a renda familiar, a oferta de escolas públicas, o custo da educação no estado e a região de moradia. Em relação ao valor gasto com as mensalidades, o maior determinante, como já era de se esperar, é a renda familiar per capita.

Soares e Gonzaga (1997) investigam a existência de dualidade no mercado de trabalho brasileiro. A existência de dualidade seria se trabalhadores igualmente qualificados estivessem empregados em setores distintos do mercado, sendo um dos setores com altas perspectivas de ascensão profissional, altos retornos à educação e à experiência, estabilidade do emprego e boas condições de trabalho, o setor primário, e outro com pequenas perspectivas de evolução no emprego, pequenos retornos à educação e à experiência, instabilidade do emprego e condições ruins de trabalho, o setor secundário. Soares e Gonzaga (1997) verificam que não existem evidências de dualidade no mercado de trabalho brasileiro, uma vez que o ajuste de um modelo dual para os dados brasileiros leva à estimação de dois setores formados por trabalhadores com níveis de educação bastante distintos. Para os autores, a educação sozinha é capaz de determinar o setor ao qual o indivíduo faz parte, ou seja, a educação consegue determinar se o indivíduo faz parte do setor com postos ruins de trabalho ou do setor com bons postos de trabalho. Sendo assim, descrevem que não se pode afirmar que existe dualidade dentro do mercado de trabalho brasileiro.

3. Metodologia

Na presente pesquisa foi utilizada a análise de regressão múltipla para mensurar a associação entre indicadores municipais de educação em 1991 e renda per capita em 2010. A variável explicada – renda per capita (rdpc) – é avaliada a partir do Censo Demográfico do ano de 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE), enquanto as variáveis explicativas – indicadores educacionais – são avaliadas a partir do censo do ano de 1991. Essa análise é feita tanto para o estado de Minas Gerais quanto para o Brasil, com o intuito de investigar se ambos seguem um mesmo padrão. Os dados foram retirados do Atlas de Desenvolvimento Humano no Brasil 2013 (disponível em www.atlasbrasil.org.br), desenvolvido em parceria do Programa das Nações Unidas para o Desenvolvimento (PNUD) com a Fundação João Pinheiro (FJP) e o Instituto de Pesquisa Econômica Aplicada (IPEA), o qual utiliza os dados dos Censos Demográficos realizados pelo IBGE em 1991, 2000 e 2010 para calcular cerca de 230 variáveis para os 5.565 municípios brasileiros. A vantagem de se utilizar o Atlas Brasil é o fato de que os dados já estão perfeitamente tabulados em formato .xls.

A pesquisa começou com a análise descritiva das variáveis por meio de tabelas, diagramas de dispersão e medidas de correlação que indicam quais variáveis educacionais estão mais associadas à renda per capita. Em seguida, são ajustados os modelos de regressão múltipla com a intenção de identificar se, e quais, variáveis educacionais de 1991 são capazes de prever a variabilidade da renda em 2010.

A análise de regressão foi feita através da linguagem R (R DEVELOPMENT CORE TEAM, 2014), por intermédio da interface RStudio, cuja função é tornar o uso do R mais intuitivo (RSTUDIO, 2015). Ambos são softwares livres que podem ser utilizados sem custos de licença e com uma extensa coleção de pacotes adicionais também gratuitos.

3.1. Regressão Múltipla

A análise de regressão é um método para investigar a associação entre duas ou mais variáveis. Essa associação é descrita por meio de um modelo que conecta a variável resposta ou dependente, Y , a uma ou mais variáveis explicativas ou predictoras, X_1, X_2, \dots, X_p . O modelo de regressão linear é dado por

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon,$$

em que $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ são os coeficientes do modelo, seus parâmetros, e são constantes desconhecidas que devem ser estimadas a partir dos dados e ε é o erro aleatório do modelo. O erro aleatório é responsável pela variabilidade em Y que não pode ser explicada pelo efeito linear das p variáveis predictoras (CHATTERJEE; HADI, 2006; DRAPER; SMITH, 1998).

Em um conjunto de dados com n observações, o modelo apresentado acima pode ser escrito como

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

em que y_i representa o i -ésimo valor da variável resposta Y , x_{i1}, \dots, x_{ip} representam os valores da variável preditora na observação i e ε_i representa o erro na aproximação de y_i .

Para estimar os coeficientes $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ se utiliza o método dos mínimos quadrados para minimizar a soma de quadrados dos erros (CHATTERJEE; HADI, 2006; DRAPER; SMITH, 1998). A estimativa $\hat{\beta}_0$ é denominada intercepto e $\hat{\beta}_j$ é a estimativa do coeficiente de regressão (parcial) da variável preditora X_j .

Utilizando os coeficientes de regressão estimados, a equação de regressão ajustada pode ser escrita como:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_p X_p$$

do $X_1 = X_2 = \dots = X_p = 0$ e o coeficiente de regressão $\beta_j, j = 1, 2, \dots, p$, representa uma estimativa da mudança em Y correspondente ao acréscimo de uma unidade em X_j quando todas as outras variáveis regressoras se mantêm constantes (SWEENEY et al., 2013). O coeficiente de regressão β_j é também chamado de coeficiente de regressão parcial porque representa a contribuição de X_j para a variável resposta Y após ter sido ajustada para as outras variáveis regressoras (CHATTERJEE; HADI, 2006).

Há algumas suposições sobre a parcela de erro do modelo (1) que são necessárias para que este seja válido. Considera-se que ε é uma variável aleatória com média igual a zero; sua variância, designada por σ^2 , é idêntica para todos os valores das variáveis X_j ; seus valores são independentes e ela é normalmente distribuída (CHATTERJEE; HADI, 2006).

Após ajustar um modelo de regressão múltipla, algum tipo de avaliação sobre sua qualidade deve ser feita. O coeficiente de determinação R^2 é utilizado para este fim e é calculado elevando-se ao quadrado o coeficiente de correlação R :

$$R^2 = \left(\frac{\sum (y_i - \bar{y}) (\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum (y_i - \bar{y})^2 (\hat{y}_i - \bar{\hat{y}})^2}} \right)^2$$

em que \bar{y} é a média da variável resposta Y e $\bar{\hat{y}}$ é a média dos valores ajustados. Interpreta-se R^2 como a proporção da variabilidade total em Y que pode ser explicada pelas variáveis explicativas X_1, \dots, X_p . Em regressão múltipla, R é denominado coeficiente de correlação múltipla porque mede a relação entre Y e um conjunto de variáveis X_1, \dots, X_p (CHATTERJEE; HADI, 2006).

O coeficiente de determinação também pode ser escrito como:

$$R^2 = \frac{SQReg}{SQTotal}$$

em que $SQReg$ é a soma de quadrados da regressão e $SQTotal$ a soma de quadrados total. Mais detalhes podem ser consultados em Chatterjee e Hadi (2006).

O teste F é utilizado para determinar se existe uma relação significativa entre a variável Y e o conjunto das variáveis explicativas $X_j, j = 1, \dots, p$, sendo chamado de teste de significância global. Se o teste F exibir uma significância global, um teste t é usado para determinar se cada uma das variáveis explicativas individuais é significativa, sendo denominado teste de significância individual. As hipóteses do teste F são $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$ versus H_1 : um ou mais parâmetros não são iguais a zero. A estatística do teste F é $F = QMReg = QMRes$, em que $QMReg = SQReg/p$ e $QMRes = SQRes = (n - p - 1)$.

Se, a partir do teste F , concluir-se que a relação de regressão múltipla é significativa, testes t podem ser realizados para determinar a significância de cada um dos parâmetros individuais. A estatística do teste t para testar $H_0: \beta_j = 0$ versus $H_1: \beta_j \neq 0$ é:

$$t_j = \frac{\hat{\beta}_j}{S_{\hat{\beta}_j}}$$

em que possui uma distribuição t de Student com $n - p - 1$ graus de liberdade. O termo no denominador é o erro padrão de β_j . Deve-se comparar o valor-p do teste com o nível de significância preestabelecido α e rejeitar H_0 se $p(|t_j|) \leq \alpha$. Caso H_0 seja rejeitada, então j é diferente de zero e a variável explicativa X_j é uma preditora significativa de Y . Além disso, intervalos de confiança para cada β_j também podem ser obtidos (CHATTERJEE; HADI, 2006).

4. Descrição e análise dos resultados

Em um primeiro momento, foi criado um arquivo utilizando a extensão RData com os dados completos do Censo Demográfico do IBGE dos anos de 1991 – indicadores educacionais – e 2010 – renda per capita. Como já dito anteriormente, nesta pesquisa, a renda per capita municipal é tomada como uma medida de desenvolvimento econômico. Por isso, foi investigada a relação entre as variáveis de escolaridade e a renda per capita para que fosse mensurada a associação entre escolaridade e renda.

Uma primeira maneira de mensurar essa associação é examinar os seus coeficientes de correlação, pois observando a correlação entre as variáveis é facilitado o processo de escolha das variáveis explicativas que irão compor o modelo de regressão linear múltipla. As tabelas 1 e 2 (anexo) mostram os casos em que as correlações são maiores do que 0.6 entre os indicadores educacionais e a renda per capita. O coeficiente de correlação de Pearson mede a relação entre duas variáveis, tanto positivamente quanto negativamente. Um coeficiente de valor negativo significa que as variáveis têm uma relação inversa, ou seja, enquanto uma cresce, a outra tende a decrescer. O contrário vale para uma correlação positiva, se uma cresce, a outra também tende a crescer ou se uma decresce, a outra tende a decrescer.

É importante ressaltar que as variáveis foram escolhidas com base nos valores das correlações, mas não somente neles, pois é importante analisar se, teoricamente, as variáveis fazem sentido para o modelo. Portanto, além do critério de observar os valores das correlações, foram escolhidas aquelas que não meçam aspectos semelhantes. Em um modelo de regressão é interessante escolher poucas variáveis que possam resumir as demais para evitar redundâncias. Sendo assim, os seguintes indicadores educacionais foram escolhidos para compor o modelo:

Quadro 1 – Variáveis utilizadas. Fonte: elaboração própria. Dados: Atlas Brasil, 2013.

Sigla	Significado
t_analf18a24	taxa de analfabetismo de 18 a 24 anos
t_analf11a14	taxa de analfabetismo de 11 a 14 anos
t_fund11a13	% de pessoas de 11 a 13 anos completando ou com fundamental completo
t_atraso_0_fund	% de pessoas de 6 a 14 anos no fundamental sem atraso
t_fund18m	% de pessoas de 18 anos ou mais com fundamental completo
t_flmed	taxa de frequência líquida ao ensino médio

No gráfico 1.1 pode ser observada uma correlação negativa entre as variáveis renda per capita e taxa de analfabetismo de 18 a 24 anos, o que significa que quanto maior a taxa de analfabetismo em 1991, menor a renda per capita em 2010. O valor do coeficiente de correlação foi -0,67. Este resultado está de acordo com a hipótese de que a escolaridade aumenta o desenvolvimento econômico, sendo de se esperar uma correlação negativa entre essas variáveis, dada a suposição de que quanto maior a taxa de analfabetismo, menor a renda de determinada população.

O gráfico 1.2 apresenta a correlação entre renda per capita e taxa de frequência líquida ao ensino médio. Nesse caso, pode-se observar uma correlação positiva entre as variáveis, ou seja, quanto maior a taxa de frequência líquida ao ensino médio em 1991, maior a renda per capita em 2010. O valor do coeficiente de correlação foi 0,63. Essa relação positiva entre um indicador educacional e a renda per capita também é de se esperar, dada a suposição de que uma maior educação esteja correlacionada a um maior nível de renda.

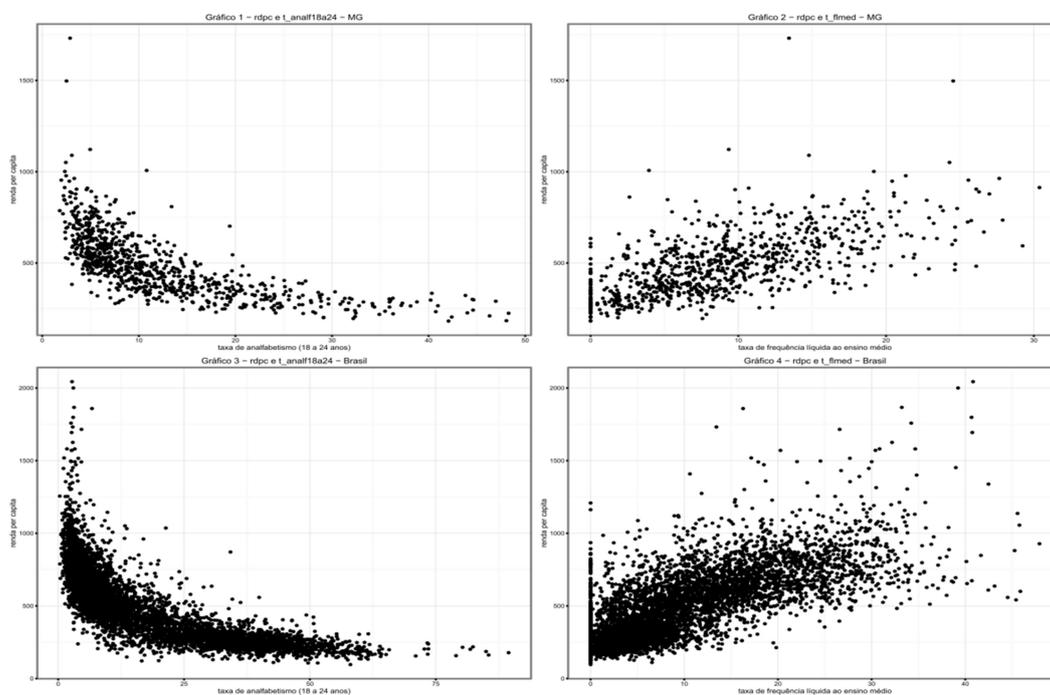


Figura 1 – A relação entre renda per capita e indicadores educacionais. Fonte: elaboração própria. Dados: Atlas Brasil, 2013.

Como já dito anteriormente, a mesma análise foi feita para o Brasil, com o intuito de investigar se a relação entre escolaridade e renda para o país segue o mesmo padrão da relação para o estado de Minas Gerais. Sendo assim, o gráfico 1.3 apresenta a relação entre a renda per capita e a taxa de analfabetismo de 18 a 24 anos para o Brasil. Novamente pode ser observada uma alta correlação negativa entre as variáveis. O valor dessa correlação foi $-0,76$. E, por último, o gráfico 1.4 apresenta a relação entre renda per capita e taxa de frequência líquida ao ensino médio. Assim como em Minas Gerais, novamente, observa-se uma alta correlação positiva entre as variáveis. O coeficiente de correlação nesse caso foi $0,72$. Portanto, em relação ao comportamento dessas variáveis, Minas Gerais pode ser considerada uma miniatura do Brasil.

Sendo assim, observa-se que essa primeira análise confirmou a hipótese de que o nível de escolaridade pode influenciar no nível de renda. É interessante, portanto, utilizar a análise de regressão linear múltipla para avaliar se essa relação é estatisticamente significativa ou se pode ser apenas fruto do acaso.

Tabela 1 – Modelo de regressão para o estado de Minas Gerais. Fonte: elaboração própria. Dados: Atlas Brasil, 2013.

Variável dependente: renda per capita (rdpc)			
t_analf18a24	-2,536*	t_fund18m	11,445***
	-1,309		-0,852
t_analf11a14	-1,57	t_flmed	0,194
	-1,158		-0,876
t_fund11a13	1,173**	constant	240,300***
	-0,521		-23,224
t_atraso_0_fund	2,656***		
	-0,604		
Observações			853
R ²			0,681
R ² ajustado			0,679
Estatística F			301,011***

*valor-p<0,1; **valor-p<0,05; ***valor-p<0,01

A tabela 1 apresenta os resultados do modelo de regressão para o estado de Minas Gerais. A partir do modelo pode-se observar que o valor do coeficiente de correlação, R² ajustado, foi relativamente alto (0,68), o que significa que a variação das variáveis explicativas consegue explicar cerca de 68% da variação na variável explicada (renda per capita). Outro fator interessante são os coeficientes dos indicadores educacionais. Percebe-se que as variáveis que medem o analfabetismo possuem coeficientes negativos, o que aponta para uma relação negativa com a variável explicada, enquanto o contrário ocorre com as variáveis que indicam maior nível de educação.

Ainda analisando a Tabela 1, pode-se observar que duas das variáveis educacionais possuem relação estatisticamente significativa com a renda per capita a um nível 0,01 de significância, já que o seu valor-p foi menor do que 0,01 para duas das variáveis (t_atraso_0_fund, t_fund18m). Uma das variáveis foi significativa a um nível de significância de 0,05 (t_fund11a13), outra se mostrou significativa a um nível de significância de 0,1 (t_analf18a24), e, por fim, uma variável não se mostrou estatisticamente significativa (t_analf11a14).

A tabela 2 apresenta os resultados do modelo de regressão múltipla para o Brasil. Pode-se observar, a partir do modelo, que o valor do R² ajustado também foi relativamente alto (0,75), o que significa que a variação das variáveis explicativas consegue explicar cerca de 75% da variação na variável explicada (renda per capita).

O valor do R² ajustado do modelo de regressão para o Brasil foi maior em comparação ao estado de Minas Gerais, indicando que o modelo pode ser mais adequado para o Brasil. Novamente pode-se observar que as variáveis sobre analfabetismo, ou seja, que indicam menor nível educacional apresentam coeficientes negativos, apontando para uma relação negativa com a renda per capita,

enquanto as variáveis que indicam maiores níveis educacionais possuem coeficientes positivos, apontando para uma relação positiva com a renda.

Tabela 2 – Modelo de regressão para o Brasil. Fonte: elaboração própria. Dados: Atlas Brasil, 2013.

Variável dependente: renda per capita (rdpc)			
t_analf18a24	-1,814***	t_fund18m	8,705***
	-0,471		-0,308
t_analf11a14	-1,401***	t_flmed	0,605*
	-0,386		-0,354
t_fund11a13	2,361***	constant	223,278***
	-0,247		-9,642
t_atraso_0_fund	3,445***		
	-0,257		
Observações			5565
R ²			0,759
R ² ajustado			0,758
Estatística F			2912,867***

*valor-p<0,1; **valor-p<0,05; ***valor-p<0,01

Observa-se ainda, na tabela 2, que todas as variáveis se mostraram estatisticamente significativas, sendo cinco delas significativas a um nível 0,01 de significância, já que o seu valor-p foi menor do que 0,01 para cinco das variáveis (t_analf18a24, t_analf11a14, t_fund11a13, t_atraso_0_fund, t_fund18m) e apenas uma se mostrou significativa a um nível de significância menor, de 0.01 (t_flmed).

4. Considerações Finais

Pode-se dizer que a hipótese de que a escolaridade em 1991 é capaz de prever a variação da renda per capita em 2010 foi confirmada através das análises de regressão apresentadas. A confirmação de que os indicadores educacionais são capazes de prever a renda aponta para a importância de se investir em políticas públicas de educação, pois além de proporcionar maiores ganhos em rendimentos futuros, como demonstrado no presente trabalho, a educação traz altos retornos para a população de um país, como maior desenvolvimento econômico, maior chance de participação na democracia, maior possibilidade de ascensão de classe social etc.

As análises se mostraram interessantes na medida em que foi possível a obtenção de variáveis estatisticamente significativas para o modelo de regressão linear múltipla e de um coeficiente de determinação, R², relativamente alto, tanto para o estado de Minas Gerais quanto para o Brasil. O estado de Minas Gerais parece ser um pequeno retrato do Brasil, pois as análises para o país e para o estado foram análogas, conforme análise apresentada na seção anterior, inclusive com os gráficos de dispersão apresentando formatos e tendências bastante semelhantes.

Referências bibliográficas

BANERJEE, A. V.; DUFLO, E. A economia dos pobres: repensar de modo radical a luta contra a pobreza global. Portugal: Círculo de leitores. 2012.

CHATTERJEE, S.; HADI, A. S. Regression analysis by example. New Jersey: John Wiley & sons, 2006.

CURI, A. Z.; FILHO, N. A. M. Determinantes dos gastos com educação no Brasil. Pesquisa e planejamento econômico, vol. 40, n. 1, 2010.

DRAPER, N. R.; SMITH, H. Applied regression analysis. 3.ed. Nova York: John Wiley & Sons, Inc., 1998

FILHO, F. H. B.; PESSÔA, S. Retorno da educação no Brasil. Pesquisa e Planejamento econômico, vol. 38, n. 1, 2008.

GOLDIN, C.; KATZ, L. The race between education and technology. New York: Harvard University Press, 2009.

IPEA; PNUD; FJP. Atlas do desenvolvimento humano no Brasil, 2013.

MACIEL, M. C; CAMPELO, A. K.; RAPOSO, M. C. F. A dinâmica das mudanças na distribuição salarial e no retorno em educação para mulheres: uma aplicação de regressão quantílica. Anais do XXIX Encontro Nacional de Economia, 2001.

MINCER, J. B. Schooling, experience and earnings. New York: Columbia University Press, 1974.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2015. Disponível em: <http://www.rproject.org/>. Acesso em: 15 nov. 2015.

RESENDE, M.; WYLLIE, R. Retornos para educação no Brasil: evidências empíricas adicionais. Revista de Economia Aplicada, vol. 10, n. 3, pág. 349-365, 2006.

ROCHA, F. et al. Retornos à educação e discriminação no mercado de trabalho brasileiro: evidências por regressões quantílicas em categorias ocupacionais. Anais do 40º Encontro Nacional de Economia, ANPEC, 2012.

RSTUDIO. Rstudio: integrated development for R 2015. Disponível em: <http://rstudio.org/>. Acesso em: 22 ago 2015.

SEN, A. K. Desenvolvimento como liberdade. São Paulo: Companhia das Letras, 2000.

SILVA, N. D. V.; KASSOUF, A. L. Mercados de trabalho formal e informal: uma análise da discriminação e da segmentação. Revista nova economia, vol. 10, n. 1, 2000.

SOARES, R. R.; GONZAGA, G. Determinação de salários no Brasil: dualidade ou não-linearidade no retorno à educação?. Texto para discussão n. 380, Pontifícia Universidade Católica de Rio de Janeiro, Departamento de Economia, 1997.

SOUZA, P.; CARVALHAES, F. Estrutura de classes, educação e queda da desigualdade de renda (2002-2011). Dados-Revista de Ciências Sociais, v. 57, n. 1, p. 101-128, 2014.

Anexos

Tabela A1 – Correlações entre variáveis de escolaridade e renda per capita (>0.6), Minas Gerais.

Variáveis	Correlação	Variáveis	Correlação
t_fund18m	0,78	t_atraso_0_fund	0,66
t_fund25m	0,78	t_atraso_0_basico	0,65
t_med25m	0,76	e_anosestudo	0,64
t_med18m	0,76	t_analf11a14	-0,64
t_analf25m	-0,76	t_fund16a18	0,64
t_analf18m	-0,76	t_fbmed	0,64
t_analf15m	-0,75	t_analf15a17	-0,64
t_fund11a13	0,72	t_atraso_2_basico	-0,63
t_fund12a14	0,71	t_flmed	0,63
t_analf25a29	-0,69	t_fund15a17	0,62
t_fund18a24	0,68	t_med18a24	0,61
t_analf18a24	-0,68	t_fbbas	0,60
t_atraso_2_fund	-0,67		

Tabela A2 – Correlações entre variáveis de escolaridade e renda per capita (> 0.6), Brasil.

Variáveis	Correlação	Variáveis	Correlação
t_analf25m	-0,84	t_analf15a17	-0,74
t_analf18m	-0,83	t_fund25m	0,74
t_analf15m	-0,83	t_fund18a24	0,73
t_fund11a13	0,82	t_flmed	0,72
t_fund12a14	0,81	t_med18m	0,72
t_atraso_0_fund	0,80	t_med25m	0,69
t_atraso_0_basico	0,80	t_med18a24	0,68
t_atraso_2_fund	-0,79	t_fbmed	0,66
t_atraso_2_basico	-0,79	t_med18a20	0,65
t_analf25a29	-0,79	t_freq6a14	0,64
t_analf18a24	-0,77	t_flfund	0,64
t_analf11a14	-0,76	t_med19a21	0,63
t_fund18m	0,76	t_fbssuper	0,62
t_fund16a18	0,76	t_flssuper	0,61
t_fund15a17	0,75	t_freq6a17	0,60
e_anosestudo	0,74	t_flbss	0,60

Variáveis	Correlação	Variáveis	Correlação
t_analf25m	-0,84	t_analf15a17	-0,74
t_analf18m	-0,83	t_fund25m	0,74
t_analf15m	-0,83	t_fund18a24	0,73
t_fund11a13	0,82	t_flmed	0,72
t_fund12a14	0,81	t_med18m	0,72
t_atraso_0_fund	0,80	t_med25m	0,69
t_atraso_0_basico	0,80	t_med18a24	0,68
t_atraso_2_fund	-0,79	t_fbmed	0,66
t_atraso_2_basico	-0,79	t_med18a20	0,65
t_analf25a29	-0,79	t_freq6a14	0,64
t_analf18a24	-0,77	t_ffund	0,64
t_analf11a14	-0,76	t_med19a21	0,63
t_fund18m	0,76	t_fbsuper	0,62
t_fund16a18	0,76	t_flsuper	0,61
t_fund15a17	0,75	t_freq6a17	0,60
e_anosestudo	0,74	t_flbas	0,60