

Métodos Multivariados de Análise de Dados*

Draft

Alberson da Silva Miranda

24 de janeiro de 2025

*Código disponível em https://github.com/albersonmiranda/analise_multivariada.

SUMÁRIO

1	INTRODUÇÃO	3
2	REFERENCIAL TEÓRICO	4
3	METODOLOGIA	6
4	RESULTADOS PRELIMINARES	8
5	REFERÊNCIAS	10

1 INTRODUÇÃO

Há diversos trabalhos analisando os determinantes na renda do trabalhador formal, sendo um dos temas mais recorrentes a relação entre a escolaridade e a renda. Neste projeto, busco modificar a abordagem tradicional, que se baseia na elaboração de modelos de regressão log-linear, para uma abordagem probabilística, utilizando logit multinomial para estimar a probabilidade de um trabalhador receber salário mínimo ou menos, até o salário mínimo DIEESE e acima, sob a ótica da escolaridade, raça/cor e sexo.

2 REFERENCIAL TEÓRICO

A primeira abordagem quantitativa, baseada em modelagem estatística, para evidenciar o poder de explicação da escolaridade sobre a renda do trabalhador foi realizada no trabalho seminal de [Mincer \(1974\)](#). O autor propôs um modelo de capital humano, onde a escolaridade é um dos principais determinantes da renda – especificamente, do logaritmo da renda. A especificação que ficou conhecida como *equação de Mincer* é a seguinte:

$$\ln Y = \beta_0 + \beta_1 S + \beta_2 E + \beta_3 S^2 + \beta_4 E^2 + \varepsilon \quad (2.1)$$

em que Y é a renda, S é a escolaridade (medida em anos), E é o tempo de experiência (estimada como a idade subtraída dos anos de escolaridade e da constante 6, considerada a idade para início da formação) e ε é o erro aleatório. O modelo de Mincer foi amplamente utilizado e adaptado em diversos estudos subsequentes, tornando-se uma referência na literatura sobre capital humano.

A partir de dados do censo norte-americano, ele estimou diversas especificações de modelos estatísticos, concluindo que aqueles que utilizaram a escolaridade e tempo de experiência como variáveis dependentes foram capazes de explicar até 33% da variação da renda do trabalhador¹, com o coeficiente de retorno médio de escolaridade (β_1 na Equação 2.1) de até 0.16, o que significa um incremento médio de 17,4% na renda por ano de escolaridade².

Desde então, diversos estudos têm sido realizados para estimar os efeitos da escolaridade sobre a renda do trabalhador, cada um estendendo ou aplicando o modelo de Mincer em diferentes contextos. [Psa-charopoulos e Patrinos * \(2004\)](#) estimam, em média, cerca de 10% de incremento de renda para cada ano de estudo. [Colclough, Kingdon e Patrinos \(2010\)](#) mostram, a partir de dados de 34 países, que até a década de 1990 os retornos a cada nível de escolaridade eram decrescentes, mas que, a partir de então, o retorno a cada ano extra de estudo aumenta a cada incremento no nível de escolaridade. Neste mesmo trabalho, os autores demonstram que os retornos em todos os níveis de escolaridade (primário, secundário e terciário) estão reduzindo ao longo das décadas, mas que a redução é mais acentuada no nível primário.

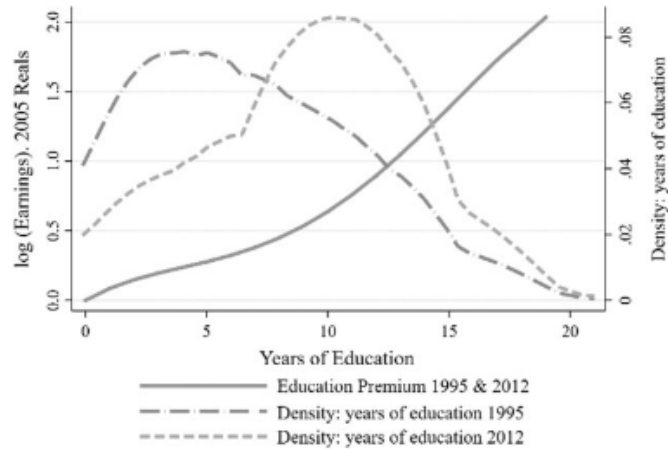
[Ferreira, Firpo e Messina \(2022\)](#) analisam a desigualdade salarial no Brasil no período entre 1995-2012 e mostram que os dois principais fatores que aumentam a desigualdade são 1) a maior disparidade de salários entre diferentes setores da economia, e; 2) o chamado “paradoxo do progresso”, o efeito intensificador da desigualdade quando há aumento da educação da população em uma sociedade em

¹Medidos pelo coeficiente de determinação R^2 .

²Como o modelo utiliza o log da renda, para se computar o efeito em moeda deve-se realizar a operação inversa, resultando em $e^{0.16} = 1.1735$.

que os retornos à educação são convexos, ou seja, aumentam exponencialmente a cada aumento do nível educacional (Figura 2.1). Esse fenômeno é paradoxal no sentido de que se espera que a educação reduza a desigualdade, mas, em sociedades em que os retornos associados aos níveis superiores de educação são muito altos em relação aos menores, a escolaridade tem efeito contrário, de intensificação das desigualdades.

Figura 2.1: Paradoxo do progresso no Brasil.



Fonte: Ferreira et al. (2022).

Nesse sentido, [Altonji, Blom e Meghir \(2012\)](#) aponta que a escolha do curso superior é de grande impacto na determinação da renda, destacando que a diferença no retorno médio entre alguns cursos superiores, como engenharia elétrica e pedagogia, é quase tão grande quanto a diferença média entre indivíduos com ensino médio e ensino superior. [Ophem e Mazza \(2024\)](#) reforçam essa ideia, mostrando que a escolha do curso superior é um dos principais elementos não apenas na renda inicial quanto também na progressão salarial ao longo da carreira.

3 METODOLOGIA

Dois dos caminhos viáveis são via dados da Relação Anual de Informações Sociais (Rais) e a Pesquisa Nacional por Amostra de Domicílios Contínua (PNad Contínua). Ambas bases de dados são disponibilizadas publicamente por órgãos governamentais e estão também tratadas e disponíveis gratuitamente por fontes secundárias, como o *data lake* Base dos Dados (CAVALCANTE, 2022). O acesso, manipulação dos dados e a análise podem ser realizados com o *software* R (R CORE TEAM, 2024).

As variáveis de interesse são:

1. renda média nominal naquele ano
2. ciclo de escolaridade
3. idade
4. raça/cor
5. sexo
6. quantidade de moradores no domicílio, no caso da PNad Contínua, que possibilita o cálculo da renda domiciliar per capita e um novo nível da variável resposta (abaixo do nível da pobreza, que corresponde a R\$ 665 per capita).

Após selecionadas, deve-se aplicar condições às variáveis para obter amostra completa, ou seja, sem valores faltantes, e coerente. Essas condições estão resumidas na tabela a seguir. Elas implicam na restrição às entradas com renda média positiva não nula; na exclusão de entradas sem quaisquer dos campos escolaridade, raça/cor ou sexo preenchidos.

Tabela 3.1: Possíveis valores para as variáveis selecionadas da Rais

Variável	Valores
Sigla UF	ES
Renda Média Nominal	Númericos, não negativos
Ciclo de Escolaridade	Analfabeto, Ensino Fundamental (I/II, completo/incompleto), Ensino Médio (completo/incompleto), Ensino Superior (completo/incompleto), Mestrado ou Doutorado
Idade	Sem restrições
Raça/Cor	Branco, Preto, Pardo, Indígena ou Amarelo
Sexo	Masculino ou Feminino

Fonte: Elaborado pelo Autor (2025)

Além das condições de interesse do pesquisador, é necessário atentar que a Rais trata do mercado de trabalho formal, o que exclui trabalhadores informais e profissionais autônomos.

Seguindo [Mincer \(1974\)](#), mas adaptando à nossa Lei de Diretrizes e Bases da Educação Nacional (LDB), pode-se obter uma aproximação para o tempo de experiência de cada indivíduo a partir da seguinte fórmula:

$$\text{exp} = \text{idade} - \text{anos de formação} - 6 \quad (3.1)$$

Em que supõe-se:

- Idade de 6 anos para ingresso no ensino fundamental
- 5 anos para o fundamental I
- 4 anos para o fundamental II
- 3 anos para o ensino médio
- 4 anos para o superior
- 2 anos para o mestrado
- 4 anos par ao doutorado

Assim, por exemplo, para um indivíduo de 30 anos com mestrado, temos $30 - 18 - 6 = 6$ anos de experiência.

Por fim, a variável resposta (resp). O salário mínimo em 2022 foi de R\$ 1.212,00. O salário mínimo DIEESE, que é o salário mínimo necessário para uma família de quatro pessoas, foi de R\$ 6.647,63, em dezembro. Assim, a variável resposta será dividida em três categorias:

1. Salário mínimo ou menos
2. Acima do salário mínimo, mas abaixo do salário mínimo DIEESE
3. Salário mínimo DIEESE ou mais

A modelagem será realizada por meio de um modelo logit multinomial, que estima a probabilidade de um trabalhador pertencer a uma das categorias da variável resposta, dada a escolaridade, raça/cor e sexo.

4 RESULTADOS PRELIMINARES

Resultados preliminares foram obtidos com os dados da Rais para o ano de 2022, com foco nos trabalhadores do estado do Espírito Santo. A Tabela 4.1 mostra os resultados do modelo logit ordinal estimado com os coeficientes já exponenciados para possibilitar a interpretação direta das razões de chance.

As variáveis de linha de base são:

- **resp**: salário mínimo ou menor
- **grau**: sem nenhum grau de escolaridade
- **raca_cor**: preta
- **sexo**: feminino
- **regiao**: interior

Tomando como exemplo **graudoutorado**, a interpretação do coeficiente é que com razão de chance de 51.03, um indivíduo com doutorado tem 51x mais chances de receber acima do salário mínimo (abaixo ou acima do DIEESE) do que um indivíduo sem escolaridade. Os demais seguem interpretação análoga. Vemos com isso que ter ensino superior eleva em muito as chances de receber maiores remunerações, com a pós-graduação intensificando ainda mais esse efeito.

Além da educação, características do indivíduo escancaram desigualdades. Ser homem triplicam as chances de receber acima do salário mínimo, enquanto ser banco aumentam as chances em 30%. A localidade também exerce influência, com a grande Vitória aumentando as chances em quase 60%.

Em relação ao ponto de corte do salário mínimo do DIEESE, a Tabela 4.2 mostra o sumário sem a exponenciação, incluindo os limites de cada nível. Para alcançar um logit de 8.16 e ser classificado como acima do salário mínimo DIEESE, em média, um indivíduo na linha de base deve ter ensino superior e 40 anos de experiência ($3.23 + 40 \times 0.12 = 8.03$), o que quer dizer que tal remuneração só seria alcançada ao final da carreira. Já homens brancos, localizados na grande Vitória, alcançam essa remuneração com 25 anos de experiência ($3.23 + 0.25 + 1.09 + 0.45 + 25 \times 0.12 = 6.73$).

Adicionando uma camada de curso superior, o trabalho controlaria para a heterogeneidade de remuneração de cada indústria, como evidenciado na revisão de literatura e seria capaz de mostrar como esse tempo de experiência se encurta ou alonga em função da profissão escolhida.

Tabela 4.1: Modelo logit ordinal

<i>Dependent variable:</i>	
resp	
graudoutorado	51.03*** (0.0002)
graufund_I	1.18*** (0.01)
graufund_II	1.41*** (0.01)
graumedio	2.95*** (0.01)
graumestrado	41.55*** (0.001)
grausuperior	25.31*** (0.01)
raca_coramarela	1.39*** (0.003)
raca_corbranca	1.29*** (0.01)
raca_corindigena	1.11*** (0.001)
raca_corparda	1.07*** (0.01)
sexomascuino	2.98*** (0.01)
regiaoqv	1.57*** (0.01)
exp	1.13*** (0.001)
I(exp^2)	1.00*** (0.0000)
Observations	1,060,134

*p<0.1; **p<0.05; ***p<0.01

Fonte: Elaborado pelo Autor (2024).

Tabela 4.2: Sumário do modelo

	Value	Std. Error	t value
graudoutorado	3.93	0.00	20684.80
graufund_I	0.17	0.01	16.22
graufund_II	0.35	0.01	44.95
graumedio	1.08	0.01	201.73
graumestrado	3.73	0.00	5020.24
grausuperior	3.23	0.01	500.46
raca_coramarela	0.33	0.00	118.45
raca_corbranca	0.25	0.01	25.90
raca_corindigena	0.10	0.00	153.77
raca_corparda	0.07	0.01	6.97
sexomascuino	1.09	0.01	176.94
regiaoqv	0.45	0.01	80.91
exp	0.12	0.00	155.30
I(exp^2)	0.00	0.00	-111.90
minimo abaixo_DIEESE	5.39	0.01	503.93
abaixo_DIEESE acima_DIEESE	8.16	0.01	626.57

Fonte: Elaborado pelo Autor (2025)

5 REFERÊNCIAS

ALTONJI, J. G.; BLOM, E.; MEGHIR, C. Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. **Annual Review of Economics**, v. 4, n. 1, p. 185–223, 1 set. 2012. ISSN 1941-1383, 1941-1391. DOI: [10.1146/annurev-economics-080511-110908](https://doi.org/10.1146/annurev-economics-080511-110908). Disponível em: <https://www.annualreviews.org/doi/10.1146/annurev-economics-080511-110908>. Acesso em: 1 mai. 2024. Citado na p. 5.

CAVALCANTE, P. **basedosdados: 'Base Dos Dados' R Client**. [S.l.: s.n.], 2022. Disponível em: <https://CRAN.R-project.org/package=basedosdados>. Citado na p. 6.

COLCLOUGH, C.; KINGDON, G.; PATRINOS, H. The Changing Pattern of Wage Returns to Education and its Implications. **Development Policy Review**, v. 28, n. 6, p. 733–747, 2010. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-7679.2010.00507.x>. ISSN 1467-7679. DOI: [10.1111/j.1467-7679.2010.00507.x](https://doi.org/10.1111/j.1467-7679.2010.00507.x). Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-7679.2010.00507.x>. Acesso em: 27 abr. 2024. Citado na p. 4.

FERREIRA, F. H. G.; FIRPO, S. P.; MESSINA, J. Labor Market Experience and Falling Earnings Inequality in Brazil: 1995–2012. **The World Bank Economic Review**, v. 36, n. 1, p. 37–67, 2 fev. 2022. ISSN 0258-6770. DOI: [10.1093/wber/lhab005](https://doi.org/10.1093/wber/lhab005). Disponível em: <https://doi.org/10.1093/wber/lhab005>. Acesso em: 27 abr. 2024. Citado na p. 4.

MINCER, J. **Schooling, experience, and earnings**. New York: Columbia University Press, 1974. 152 p. (Human behavior and social institutions, 2). ISBN 978-0-87014-265-9. Citado nas pp. 4, 7.

OPHEM, H. van; MAZZA, J. Educational choice, initial wage and wage growth. **Empirical Economics**, 21 mar. 2024. ISSN 1435-8921. DOI: [10.1007/s00181-024-02580-5](https://doi.org/10.1007/s00181-024-02580-5). Disponível em: <https://doi.org/10.1007/s00181-024-02580-5>. Acesso em: 27 abr. 2024. Citado na p. 5.

PSACHAROPOULOS, G.; PATRINOS *, H. A. Returns to investment in education: a further update. **Education Economics**, v. 12, n. 2, p. 111–134, 1 ago. 2004. ISSN 0964-5292, 1469-5782. DOI: [10.1080/0964529042000239140](https://doi.org/10.1080/0964529042000239140). Disponível em: <https://www.tandfonline.com/doi/full/10.1080/0964529042000239140>. Acesso em: 17 out. 2024. Citado na p. 4.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2024. Disponível em: <https://www.R-project.org/>. Citado na p. 6.