

MODELAGEM DOS EFEITOS DE VARIÁVEIS SOCIOECONÔMICAS NA FORMA DE REALIZAR AS ATIVIDADES DIÁRIAS – REMOTA, HÍBRIDA OU PRESENCIAL

Ana Paula Teixeira da Silva¹ (IC), Josiane Palma Lima (PQ)¹

¹Universidade Federal de Itajubá

Palavras-chave: Mobilidade. Multicolinearidade. Regressão. Variável dependente. Variável independente.

Introdução

A recente pandemia por Covid-19 culminou em diversas mudanças comportamentais, entre elas, as alterações na mobilidade dos cidadãos devido à necessidade de isolamento social como forma de contenção à propagação do novo coronavírus. A vida cotidiana e a participação pessoal nas atividades foram interrompidas, devido às medidas de bloqueio e distanciamento social aplicadas pelo governo, acarretando em uma drástica redução nas viagens diárias. Por outro lado, verificou-se a utilização e expansão de ferramentas tecnológicas como formas de estudo ou trabalho, configurando o e-learning e o home office, por exemplo, que foram empregadas para substituir as atividades presenciais (MOURATIDIS e PAPAGIANNAKIS, 2021).

A partir dessas mudanças nos modelos de trabalho e na mobilidade, importantes impulsos e oportunidades podem surgir para, a longo prazo, alcançar uma mobilidade mais sustentável. Provavelmente, a mudança mais duradoura é o home office, o que reduz o número de deslocamentos e altera a escolha do meio de transporte, sendo importante analisar o que acontecerá com uma sociedade diferenciada, que não pode mais se mover como de costume (RASCHKE *et al.*, 2021).

Devido às alterações nos deslocamentos, a realização de atividades remotas também pode impactar, a longo prazo, no uso do solo, onde as residências e escritórios podem se ajustar ao home office e às compras online, alterando, também, o comportamento de viagem. As mudanças na mobilidade urbana dependerão do grau de adoção de atividades online, dos sistemas de transporte, adoção de mobilidade compartilhada, avanços nas tecnologias de transporte, forma urbana, políticas e regulamentações urbanas e fatores sociais (MOURATIDIS e PAPAGIANNAKIS, 2021).

Porém, a decisão de se realizar o home-office e o e-learning está relacionada com as características socioeconômicas e comportamentais das pessoas. Desta forma, este trabalho irá abordar questões relativas à forma de realização das atividades diárias, remoto, híbrido ou presencial, e sua dependência com variáveis socioeconômicas. Entender melhor os fatores que influenciam na realização das atividades remotas é uma

forma efetiva de apoio à tomada de decisão para o planejamento de transportes.

Para a realização da pesquisa, foi utilizada uma abordagem quantitativa, por intermédio do método de modelagem da regressão logística multinomial. Primeiramente, foi aplicado um questionário para a população de Itajubá, Minas Gerais, a fim de coletar dados acerca das características socioeconômicas da população. Em seguida, foi realizada uma análise descritiva da frequência das variáveis e as mesmas foram submetidas à análise de regressão, verificando a multicolinearidade e acurácia do modelo, além da influência das variáveis independentes na variável dependente.

Metodologia

Com o objetivo de analisar as variáveis socioeconômicas que influenciam no modo de realização das atividades principais, foi aplicado um questionário com os moradores do município de Itajubá, localizado no Sul de Minas Gerais, envolvendo perguntas sobre o perfil socioeconômico dos moradores da cidade e sobre suas atividades diárias e o modo como são realizadas.

A coleta de dados foi realizada durante dois meses, de 29 de abril de 2022 a 29 de junho de 2022, por meio da plataforma *Google Forms* com divulgação nas redes sociais.

A fim de verificar a quantidade mínima de entrevistados, foi calculada a amostra de acordo com as equações (1) e (2), apresentadas por Mairinque e Lima (2020), Antunes e Simões (2013) e Barcelos *et al.* (2017).

$$N_0 = 1 / \varepsilon^2 \quad (1)$$

$$n = (N_0 \cdot N) / (N_0 + N) \quad (2)$$

Em que:

N_0 : primeira aproximação do tamanho mínimo da amostra;

ε : erro amostral tolerável;

N : tamanho da população;

n : tamanho da amostra.

Para aplicar as equações (1) e (2), foi utilizada a população de Itajubá que, de acordo com o IBGE a

estimativa para 2021 é de 97.782 pessoas. Além disso, considerou-se um erro amostral de 6%.

Após a aplicação do questionário, verificou-se o tamanho da amostra, comparando-a com a quantidade mínima necessária. Finalizado esse processo, foi feita uma análise descritiva dos dados com base na frequência de cada variável. Além disso, o modo de execução da atividade principal, seja remoto, híbrido ou presencial, foi associado com as variáveis socioeconômicas, verificando as respectivas frequências.

Neste trabalho, foi aplicado o modelo de regressão logística multinomial utilizando o *software* R para explicar a relação entre o modo de execução da atividade principal e as variáveis socioeconômicas, considerando os efeitos globais e específicos das variáveis independentes.

Primeiramente, as variáveis dependentes (remoto, híbrido ou presencial) e independentes (características socioeconômicas) foram organizadas em uma planilha eletrônica, a qual foi inserida no banco de dados do software. O primeiro passo consistiu na análise de multicolinearidade das variáveis independentes, indicando se as mesmas possuem alta correlação ou não. Uma das maneiras de analisar o grau de multicolinearidade é por intermédio do fator de inflação da variância (cuja sigla, em inglês, é VIF), em que é indicado que uma determinada variável independente é explicada pelas demais variáveis independentes, a fim de classificar se o grau de multicolinearidade é severo ou não. Alguns autores citados por Pereira et al. (2014) consideram que VIF acima de 10 afeta os resultados da regressão, enquanto outros estipulam um limite entre 4 ou 5 unidades. Após a obtenção do VIF, verificou-se os efeitos globais e específicos das variáveis independentes e a acurácia do modelo.

Resultados e discussão

Após a coleta de dados, foram consideradas 319 respostas do questionário para análise. Na tabela 1, é possível comparar a quantidade de respondentes com o valor mínimo calculado utilizando as equações (1) e (2).

Tabela 1 - Tamanho mínimo e real da amostra

População	97782
ϵ	6%
Primeira aproximação	277,78
Tamanho mínimo da amostra	277
Tamanho da amostra	319

Após a coleta, os dados foram organizados de acordo com a frequência de cada categoria, com seus valores brutos e porcentagens. Os resultados são apresentados na tabela 2.

Tabela 2 - Distribuição dos dados amostrais.

Variável Independente		Frequência	
Gênero	Feminino	159	49,8%
	Masculino	160	50,2%
Idade	≤20	48	15,0%
	21-40	179	56,1%
	41-60	81	25,4%
	>60	11	3,4%
Escolaridade	Ensino Fundamental	1	0,3%
	Ensino Médio	132	41,4%
	Ensino Superior	87	27,3%
	Mestrado/Doutorado	99	31,0%
Renda	Até 1 salário mínimo	28	8,8%
	Entre 1 e 2	63	19,7%
	Entre 2 e 5	92	28,8%
	Entre 5 e 8	60	18,8%
	Mais de 8	76	23,8%
Atividade principal	Trabalho	149	46,7%
	Estudo	165	51,7%
	Cuidador de pessoas	2	0,6%
	Aposentado	1	0,3%
	Outra atividade	2	0,6%
Posse de carro	Sim	223	69,9%
	Não	96	30,1%
Posse de moto	Sim	61	19,1%
	Não	258	80,9%
Posse de bicicleta	Sim	211	66,1%
	Não	108	33,9%

Referente à atividade principal, foi verificado que, dos 319 entrevistados, 9 não realizam (2,8%), 27 executam de maneira remota (8,5%), 90 estão no modelo híbrido (28,2%) e 193 fora de casa (60,5%). Aproximadamente 90% dos respondentes realizam sua atividade principal presencialmente ou no modelo híbrido (Figura 1).

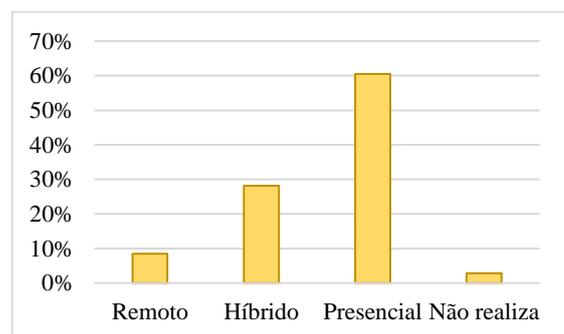


Figura 1 - Distribuição do modo de realização da atividade principal.

Em seguida, alguns dados foram analisados e agrupados de acordo com sua frequência em relação ao modo de realização da atividade principal dos respondentes, como pode ser visto nas figuras 2, 3, 4 e 5.

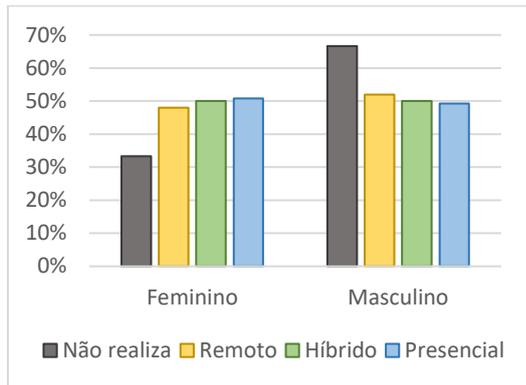


Figura 2 - Distribuição do gênero x atividade principal.

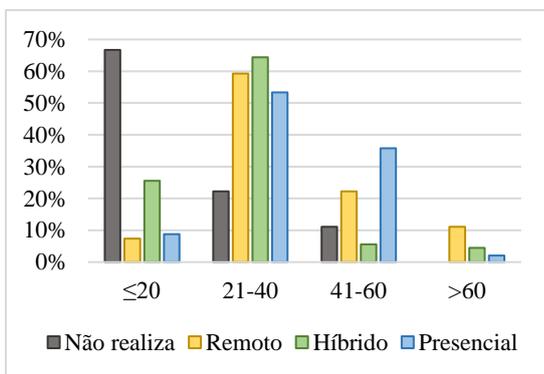


Figura 3 - Distribuição da idade x atividade principal

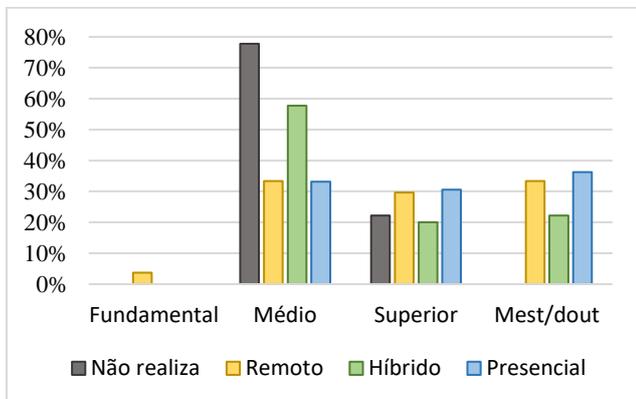


Figura 4 - Distribuição da escolaridade x atividade principal.

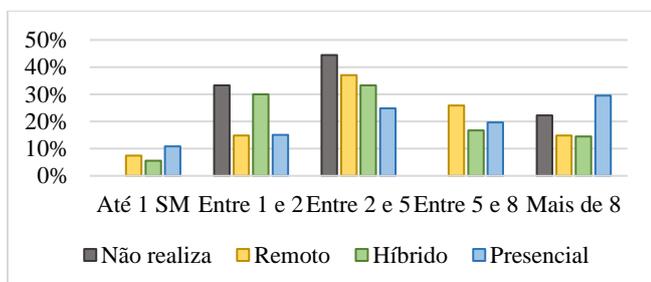


Figura 5 - Distribuição da renda x atividade principal.

Deste modo, seguiu-se para a elaboração do modelo de regressão logística multinomial. Inicialmente foi realizada uma análise de correlação entre as variáveis independentes. Nesta etapa é desejado que as variáveis não apresentem forte correlação. Em geral, são considerados ajustes fortes entre as variáveis valores maiores que 0,8 (COMBER *et al.*, 2020). Na figura 6 é possível perceber que nenhum par de variáveis apresentou forte correlação, sendo que o maior valor foi 0,69.

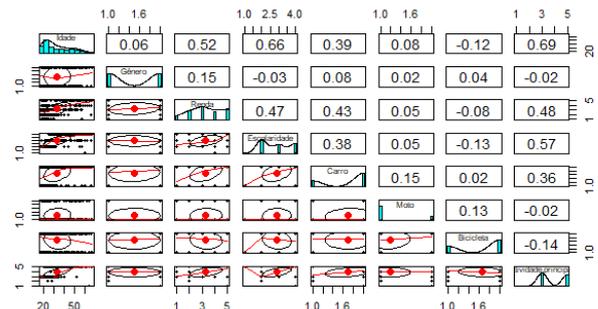


Figura 6 - Correlação entre as variáveis independentes

Em seguida foi avaliado o valor do VIF (*Variance Inflation Factor*) para cada variável independente selecionada, como forma de garantir a inexistência de multicolinearidade no modelo. A partir da análise da tabela 3 é possível perceber que os valores de VIF para as variáveis do modelo estão bem abaixo do valor máximo aceito (entre 4 e 5), confirmando que não existe multicolinearidade entre as variáveis.

Variável Independente	VIF
Idade	3,03
Gênero	1,12
Renda	2,27
Escolaridade	2,78
Posse de carro	1,41
Posse de moto	1,10
Posse de bicicleta	1,06
Atividade Principal	2,79

Pela análise global dos efeitos das variáveis independentes no modelo (Tabela 4), apenas as variáveis “Moto” e “Atividades Principal” foram significantes, apresentando p-valor < 0,05, o que indica que tais variáveis são preditoras significativas da variável dependente. Porém optou-se por manter todas as variáveis independentes no modelo uma vez que a teoria sobre o assunto apresenta diferentes colocações, logo elas podem ser importantes para o entendimento das atividades remotas.

Tabela 4 - Efeitos globais das variáveis independentes

Variável Independente	LR	Chisq	Df	Pr(>Chisq)
Idade	1,629		3	0,653
Gênero	0,939		3	0,816
Renda	18,18		12	0,110
Escolaridade	8,497		9	0,485
Posse de carro	0,828		3	0,843
Posse de moto	7,979		3	0,046*
Posse de bicicleta	1,123		3	0,772
Atividade Principal	58,21		12	4,774e-08***

Códigos de significância: '***' 0,001 '**' 0,01 '*' 0,05

Foi avaliada também a acurácia do modelo gerado, estimando a probabilidade de acerto das suas previsões, o valor obtido foi de 0,674. Além disso, foram analisados os efeitos específicos das variáveis independentes no modelo. Verificou-se que a probabilidade de realização de cada um dos modos da atividade principal ao longo das diferentes faixas de idade não apresenta grande diferença entre os gêneros, porém nota-se que as pessoas mais novas têm realizado mais as atividades no modelo híbrido, o que pode ser justificado pelas aulas online. As pessoas que não possuem carro e estão na faixa etária de 30 a 50 anos têm maior probabilidade de realizar atividades híbridas, já as pessoas acima dos 30 anos e com carro apresentam maior probabilidade de realizar presencialmente. Percebeu-se, também, que as pessoas com menos de 55 anos e que possuem moto apresentam maior probabilidade de realizar atividade no modo híbrido do que aquelas que não possuem.

Considerando a renda dos respondentes, é possível perceber que as atividades realizadas no modo híbrido têm maior probabilidade de serem realizadas por pessoas com renda entre 1 e 2 salários mínimos em todas as idades, e no geral o trabalho remoto tem maior probabilidade de ser realizado por pessoas com renda entre 2 e 8 salários mínimos. Por fim, a atividade principal realizada no modelo híbrido acontece com maior probabilidade para as pessoas entre 35 e 55 anos que possuem até o ensino médio, as atividades realizadas de maneira remota apresentam muito pouca diferença para os diferentes níveis de escolaridade.

Conclusões

Ao longo deste estudo foi avaliada a relação entre a forma de realização das atividades diárias com variáveis socioeconômicas no município de Itajubá, Minas Gerais,

por intermédio de um modelo de regressão logística multinomial. O modelo apresenta uma probabilidade de acerto em suas previsões de 0,674, e as variáveis independentes não apresentaram multicolinearidade.

Pela análise dos efeitos globais das variáveis independentes, foi identificado que apenas as variáveis posse de moto e tipo da atividade principal influenciaram significativamente na forma de realização da atividade. Porém, optou-se por manter todas as variáveis no modelo devido a sua importância para o tema. Ademais, o estudo também avaliou os efeitos específicos das variáveis independentes.

Por fim, a importância da pesquisa está atrelada com o entendimento dos fatores socioeconômicos que podem interferir na execução das atividades remotas no município, colaborando com o planejamento de transporte e uso do solo e aplicação de políticas de incentivo à realização das atividades remotas.

Agradecimento

Agradeço a FAPEMIG pelos incentivos financeiros que viabilizaram a realização desta pesquisa.

Referências

- ANTUNES, E. M.; SIMÕES, F. A. Engenharia urbana aplicada: um estudo sobre a qualidade do transporte público em cidades médias. **Revista Brasileira de Gestão Urbana**, v. 5, n. 2, p. 51-62, 2013.
- BARCELOS, M. et al. Inferindo a importância dos atributos do transporte coletivo a partir da satisfação dos usuários. **Transportes**, v. 25, n. 3, p. 36-48, 2017.
- COMBER, A. et al. The GWR route map: a guide to the informed application of Geographically Weighted Regression. **arXiv preprint arXiv:2004.06070**, 2020.
- MAIRINQUE, L. A.; LIMA, J. P. Impactos da pandemia por Covid-19 nos deslocamentos e atividades principais em cidades brasileiras de médio porte. **Anais do XXXIV Congresso de Pesquisa e Ensino em Transportes, ANPET**, p. 442-453, 2020.
- MOURATIDIS, K.; PAPAGIANNAKIS, A. COVID-19, internet, and mobility: The rise of telework, telehealth, e-learning, and e-shopping. **Sustainable Cities and Society**, v. 74, 1 nov. 2021.
- PEREIRA, G. A.; LIMA MILANI, L.; CIRILLO, M. Â. Uso de alguns estimadores *ridge* na análise estatística de experimentos em entomologia. **CERES**, v. 61, n. 3, p. 338-342, maio/jun. 2014.
- RASCHE, B. et al. Forschung für die Mobilitätswende: COVID-19-Pandemie als Treiber? **GAIA - Ecological Perspectives for Science and Society**, v. 30, n. 4, p. 276-277, 16 dez. 2021.